Fragile Intangible Capital: Measuring the Real Cost of Exposed Disinformation

John (Jianqiu) Bai, Yi Cao, Chi Wan, and Sean Wang[†]

June 2026

Abstract

What are the economic consequences when disinformation is exposed? We investigate staggered events where a market information intermediary posts a public alert that a firm's prior reputation has been built around fabricated information—a credibility shock that impairs the firm's belief-based reputational capital without conveying information about products, prices, or physical operations. Exposure of disinformation occurs via two distinct channels—documented evidence of the firm offering compensation in exchange for business reviews and algorithmic detection of suspicious review patterns—each triggering a salient alert to the intermediaries' users. Using the intermediary's algorithm to construct demand-side peer controls based on revealed preferences, and linking each triggering event to GPS foot-traffic and transaction data, we find that exposed firms lose about 8% of foot traffic and 5–6% of transaction volume relative to peers. These losses emerge within a month and persist for at least eighteen months—beyond the removal of public warnings and the end of associated advertising bans—with no sign of recovery. As predicted by our framework, losses scale with reliance on intermediated beliefs—penalties are largest when the exposure of disinformation is backed by clear evidence, during high-demand periods, in areas with deeper intermediary penetration, and for firms more tightly integrated with the intermediary. Exposure also degrades the firm's information environment: ratings decline, review activity falls, and further manipulation is deterred. Fabricated signals boost demand while hidden; their revelation causes lasting demand destruction. Our estimates capture only demand-side effects and exclude any impact on financing costs.

Keywords: Disinformation; reputational capital; credibility shocks; intangible assets; real effects; belief revision; information asymmetry; market discipline

JEL codes: D83; M41; O34; G14; G30; G40.

[†]John (Jianqiu) Bai is from the School of Accounting and Finance at Hong Kong Polytechnic University (PolyU), M715, 7/F, Li Ka Shing Tower, Hung Hom, Kowloon, Hong Kong, China (john.baijianqiu@gmail.com); Yi Cao is at the Costello College of Business, George Mason University, 4400 University Dr, Fairfax, VA 22030 (ycao25@gmu.edu); Chi Wan is from the Fowler College of Business, San Diego State University, San Diego, CA 92182 (cwan@sdsu.edu); Sean Wang is the corresponding author (sean@profsean.wang).

1 Introduction

Disinformation—the deliberate fabrication of reputational signals—is a central problem in intermediated markets. From used-car "lemons" (Akerlof, 1970) to gamed cardiac surgery report cards (Dranove et al., 2003) and inflated credit ratings (Bolton et al., 2012), intermediaries are trusted to certify signals that market participants cannot verify. Theory shows reputational capital can alleviate information asymmetries by shaping beliefs about firm type (Klein and Leffler, 1981; Shapiro, 1983). Such intangible capital is also fragile: its value depends on factors outside the firm's control (Eisfeldt and Papanikolaou, 2013). Such assets are also economically large yet absent from financial statements: accounting standards bar internally generated intangibles from the balance sheet, so their value—and any impairment to it—goes unrecorded (Ewens et al., 2025). When the intermediary's signals are manipulated, the firm's reputational capital becomes inflated. Fabricators attract unearned demand, and market participants allocate capital under false pretenses. Measuring the economic cost of disinformation is difficult. Exposure events usually coincide with news about fundamentals, e.g. earnings restatements, corporate misconduct, product recalls, regulatory actions—so the fabricated signal's cost cannot be separated from that of the problem it concealed. Thus, while we know disinformation distorts beliefs and misallocates resources while hidden, we do not know the cost of its exposure.

We measure that cost using staggered, plausibly exogenous shocks to reputational capital that leave firm fundamentals unchanged. Yelp, an information intermediary, aggregates reputational signals to reduce information asymmetry between firms and market participants. Its Consumer Alerts program publicly flags firms that manipulate their reputation by soliciting or purchasing fraudulent reviews. Each alert is a discrete public correction. It reveals that the firm's prior standing was artificially constructed, without adding information about quality, pricing, or operations. This setting offers an empirical laboratory for a central question in the real effects of information literature: what is the economic cost when a belief-based reputational capital is impaired—and does that cost subside once the impairment is no longer enforced, or is the damage persistent over time?

Our identification strategy is a difference-in-differences design that compares flagged manipulators to control firms identified through revealed demand substitution. The intermediary's algorithm links firms that share overlapping visitor flows, matching on location, category, price segment, and browsing behavior. We connect the intermediary's revelation of fabricated reputation to GPS-based foot traffic and credit card transactions, measuring changes in real economic activity at high frequency. The intermediary exposes disinformation through two independently timed channels—outside whistleblower reports and internal algorithmic detection—and neither is preceded by differential demand declines. The endogeneity concern is sharpest for the algorithmic channel, yet that subsample shows the cleanest pre-trends, ruling out a deteriorating-fundamentals story.

Flagged firms experience an approximately 8% decline in foot traffic and a 6% reduction in transaction volume relative to matched peers. Strikingly, the loss outlives every institutional support that might otherwise explain it: the warning banner is removed after about 90 days and the advertising ban lifts after one year, yet demand shows no reversion at any horizon through at least eighteen months. A penalty that would mechanically fade if it ran through enforcement instead persists once enforcement ends—the signature of impairment to the belief-based asset itself rather than a temporary stigma or a withdrawn subsidy. This rapid-impairment, slow-recovery asymmetry is what theory predicts for reputational capital: damaged swiftly, rebuilt only through sustained credible performance (Shapiro, 1983).

These penalties are not uniform: they are two to three times larger for firms that depend more heavily on the intermediary—those in high-penetration markets and with deeper integration—and roughly twice as large when the detection signal is most precise, i.e. when the alert carries direct evidence of compensation for activity versus when it rests on algorithmic detection alone. Together, these cross-sectional patterns point to a single driver: the fragility of belief-based intangible capital rises with the firm's reliance on the intermediary to generate demand.

Exposure also reshapes the information environment around flagged firms. Average ratings fall, review activity drops, and suspicious review content declines sharply. These findings support both stakeholder disengagement as well as the deterrence of further manipulation. Demand does not simply dissipate. It reallocates to local substitutes rather than exiting, and the reallocation we can detect accrues to peers with deeper verifiable reputation. Because we measure only the demand channel, the 8% estimate is a lower bound on total value destruction: any impairment to borrowing terms or the cost of capital would push losses higher.

Our study contributes to two literatures. First, we extend research on the real effects of information (Biddle et al., 2009; Kanodia and Sapra, 2016; Roychowdhury et al., 2019; Huang, 2025) to the integrity of intermediary signals. Whether reputational signals are genuine or fabricated has first-order real consequences through the demand channel. Fabricated signals sustain real economic activity while undetected, and their exposure destroys belief-based intangible capital in ways that are hard to reverse. This complements research on the real effects of formal financial disclosure (Leuz and Wysocki, 2016). We show that informal intermediary signals—outside the mandatory reporting system—generate first-order real consequences, and that disinformation within these systems imposes costs formal disclosure regulation alone cannot address. Where recent work recovers the value of intangible capital as it accumulates (Ewens et al., 2025), we trace the other end of its lifecycle: how quickly, and how durably, a belief-based intangible can be destroyed once the market learns it was inflated.

Second, we isolate the standalone real cost of exposing fabricated reputational signals: the asset-impairment event itself, separated from concurrent changes in firm fundamentals. Prior work documents the prevalence and determinants of review manipulation (Mayzlin et al., 2014; Luca and Zervas, 2016) and the value of favorable intermediary ratings for demand and financing (Huang, 2025). A large literature on auditors, credit rating agencies, and analysts further shows that compromised gatekeeper signals matter. In each case, however, exposure is bundled with news about fundamentals, so the fabricated signal's cost cannot be separated from that of the problem it concealed. Our setting strips away this confound. The alert reveals only that reputation was manufactured, holding fundamentals fixed, which lets us measure what disinformation costs along a dimension prior designs cannot isolate. This establishes reputation as a priced intangible asset—one whose fabrication generates real activity, and whose exposure destroys it.

The paper proceeds as follows. Section 2 presents our framework and hypotheses. Section 3 describes the data, institutional setting, and empirical methodology. Section 4 reports our main results on the economic impact of credibility shocks—persistence, heterogeneity, information environment effects, and competitive reallocation. Section 5 considers implications for the broader literature on intangible capital. Section 6 concludes.

2 Stylized Framework and Hypothesis Development

This section develops a stylized example to organize the empirical analysis. The goal is not to characterize equilibrium manipulation incentives or the intermediary's detection technology, but to derive testable predictions that discipline the identification strategy and provide a calibration anchor for interpreting the magnitudes we estimate. Appendix A provides additional details regarding the framework and empirical estimation.

2.1 Setup

A firm's perceived quality has two components:

$$\text{Total Perceived Quality} = \theta + \delta \tag{1}$$

where $\theta > 0$ is *verifiable true quality*—the track record from observable operations and legitimate, verified reviews—and $\delta \geq 0$ is the *unverifiable reputational premium*: belief-dependent reputation beyond what θ supports. Crucially, δ need not be fabricated. Much of it is legitimate—genuine quality not yet reflected in the verifiable record—and rational patrons price it accordingly, just as a firm can trade above book value on credible expectations of future performance. Manipulation inflates δ beyond its legitimate level by fabricating belief-dependent signals. Exposure reveals that some of this premium was manufactured; what it destroys is the market's willingness to price δ at all. The destroyed premium therefore bounds the fabricated component from above—a distinction we make precise in Sections 4.4 and 5.3.

This decomposition parallels the distinction between assets in place and belief-dependent intangible value, with θ playing the role of book value and δ the belief-driven premium above it. We define the firm's *reputational market-to-book ratio* as:

$$(M/B)_{rep} = \frac{\theta + \delta}{\theta} = 1 + \frac{\delta}{\theta} \tag{2}$$

A ratio of one means reputation rests entirely on verifiable quality. A ratio above one means a larger share of perceived quality is belief-dependent—whether legitimately or through fabrication—and thus exposed to credibility shocks. Manipulation raises $(M/B)_{rep}$ by padding δ with fabricated signals; exposure collapses the belief-dependent portion the market is willing to support.

2.2 Demand and Informational Intermediation

Demand depends on verifiable quality and a multiplicative reputation premium:

$$D = D_0(\theta)\left(1 + \frac{\delta}{\theta}\right)^{\lambda_i} \qquad \iff \qquad \log D = \log D_0(\theta) + \lambda_i \log\Big((M/B)_{rep}\Big). \tag{3}$$

An inflated reputational signal raises demand by $(1 + \delta/\theta)^{\lambda_i}$ above the baseline $D_0(\theta)$ from verifiable quality alone. In logs, the premium is additively separable. The parameter $\lambda_i > 0$ measures firm i's *intermediation intensity*—the rate at which the intermediated signal maps into demand, so that a one-unit rise in the log reputational premium scales log demand by

λ_i.[1] Only the intermediated premium is exposed to the credibility shock; we assume $D_0(\theta)$ is unaffected.[2] We need not specify $D_0(\theta)$: firm fixed effects absorb it, and the treatment effect isolates the premium $\lambda_i \log\Big((M/B)_{rep}\Big)$ (Appendix A).

The demand premium depends on the ratio (δ/θ) because fabricated signals matter more when a firm has a shorter track record.[3] The concavity of $\log\Big((M/B)_{rep}\Big)$ in δ/θ reflects diminishing marginal returns to ratings (Anderson and Magruder, 2012) and matches diminishing returns along the extensive margin: at low quality, extra reputation attracts new visitors; at high quality, most potential visitors are already aware, so further gains operate mainly on the intensive margin.[4]

2.3 The Credibility Shock and Its Consequences

At a random time t^*, the intermediary detects manipulation and flags the violating firm. While the firm's initial decision to manipulate ($\delta > 0$) may be endogenous, the detection timing is plausibly exogenous to changes in firm quality.[5]

When the intermediary reveals that the firm's reputation was manipulated, the market stops pricing the belief-dependent premium: $(1 + \delta/\theta)^{\lambda_i} \to 1$, so the log premium $\lambda_i \log\Big((M/B)_{rep}\Big) \to 0$. The collapse reflects lost belief in the premium as a whole, not the deletion of any particular signal; because exposure may also dispel legitimate belief-

[1]The parameter λ_i is conceptually parallel to the share of trusting investors in Bolton et al. (2012), who model naive investors that take ratings at face value and sophisticated investors who see through inflation, and for whom ratings distortion increases with reliance on the intermediated signal. The objects are not of the same type: their parameter is a population share bounded in $[0, 1]$, whereas λ_i is a firm-specific pass-through elasticity that need not lie below one. We do not observe λ_i directly, and because it enters demand only through the product $\lambda_i \log\big((M/B)_{rep}\big)$, its level is not separately identified from the premium. Section 4.7 identifies three empirical dimensions along which λ_i varies, plus a separate dimension—signal precision—that determines how fully exposure destroys the premium.

[2]In Section 5.3 we discuss the implications of relaxing this assumption.

[3]For example, adding 50 fabricated reviews (δ) greatly shifts perceived quality for a new firm with only 10 authentic reviews (θ), but barely affects an established firm with 1,000 authentic reviews. The ratio δ/θ captures this proportional vulnerability.

[4]The qualitative predictions below hold for any increasing transform of δ/θ; only the calibration from the demand drop to a particular δ/θ uses the $(M/B)_{rep}$ form, consistent with the log-approximating outcome.

[5]Detection is frequently triggered by random third-party reports—for example, a patron reporting compensation or a discount offered as a quid pro quo for a review—rather than by changes in firm operations. We formalize this identification argument and test for pre-trends in Section 3.

based reputation alongside the fabricated portion, the demand loss bounds the fabricated premium from above. Because the intermediary conveys no information about true quality θ, baseline demand $D_0(\theta)$ is unchanged. We treat θ as verifiable because it reflects experiential attributes—food, service, ambiance—that patrons observe independently of online signals. This holds cleanly for patrons with prior direct experience. For new, intermediary-dependent visitors who have not observed θ, exposure may also discount otherwise-genuine signals—a possibility we bound as *credibility contagion* in Section 5.3.

The demand loss from exposure is the collapse of this log premium:

$$\Delta \log D = -\lambda_i \log\Big((M/B)_{rep}\Big), \tag{4}$$

which corresponds to the first-difference our difference-in-differences design estimates (Section 4.4).

Empirical implication. Eq. (3) maps one-to-one into our estimating equation, which is why the difference-in-differences design recovers a structurally interpretable object rather than a reduced-form correlation. Taking logs of Eq. (3) separates demand additively into a verifiable-quality baseline, $\log D_0(\theta)$, and a belief-dependent premium, $\lambda_i \log\Big((M/B)_{rep}\Big)$. A Consumer Alert reveals the premium was inflated and collapses it to zero, so differencing log demand around the alert isolates $-\lambda_i \log\Big((M/B)_{rep}\Big)$—exactly the object the specification in Section 3.3 estimates: firm fixed effects absorb $\log D_0(\theta)$, time fixed effects absorb aggregate shocks, and the coefficient on *Flagged*$\times$ *Post* recovers the destroyed premium. The functional form supplies this interpretation; it carries no identifying weight, which rests entirely on the parallel-trends assumption stated in Section 3.3.

As stated earlier, our baseline calibration assumes exposure fully destroys the belief-based premium—a reasonable approximation during the alert window, when the warning is prominently displayed to all intermediary visitors. In practice the *fraction* of the premium that collapses depends on how credibly the alert disconfirms it: an alert backed by direct evidence of compensated reviews discredits the inflated signal more completely than an opaque algorithmic warning. This *completeness of belief revision* is distinct from intermediation intensity λ_i—it governs how much of the premium collapses, not what share of demand the premium carries—and we treat it as a separate dimension of heterogeneity below. To the extent that some intermediated visitors never encounter the alert, the observed decline understates the true premium, reinforcing the lower-bound interpretation.

2.4 Empirical Predictions and Hypotheses

2.4.1 Baseline Effect

For any $\delta > 0$ and $\lambda_i > 0$, the loss $-\Delta \log D = \lambda_i \log\Big((M/B)_{rep}\Big) > 0$. The credibility shock collapses the belief-dependent premium while leaving verifiable quality unchanged—a belief-revision channel rather than a change in fundamentals.[6]

Hypothesis 1: *Firms flagged for reputation manipulation experience a significant decline in foot traffic and transactions relative to peer firms, despite no concurrent change in firm quality or physical operations.*

2.4.2 Intermediation Intensity Amplifies the Penalty

The magnitude of the demand loss, $-\Delta \log D = \lambda_i \log\Big((M/B)_{rep}\Big)$, increases in intermediation intensity λ_i:

$$\frac{\partial\left(-\Delta \log D\right)}{\partial \lambda_i} = \log\Big((M/B)_{rep}\Big) > 0. \tag{5}$$

Firms whose economic activity is more heavily intermediated suffer larger penalties because the credibility shock reaches a greater share of their demand base.

This implies that penalty will scale with several observable characteristics. First, *Demand composition*: when the composition of arriving market participants shifts toward uninformed

[6] The alert is accompanied by one platform-side action that is not informational: flagged firms are barred from purchasing intermediary advertising for one year. We therefore read the decline as impairment of belief-based reputational capital with this bundled action acknowledged, rather than as a purely informational shock, and we bound the advertising and enforcement channel in Section 4.6.

agents who lack private information about local firm quality—i.e., new rather than repeat patrons—more demand flows through the intermediary, temporarily raising λ_i. Second, *Market penetration*: when the intermediary impacts more information flow, a larger share of each firm's demand is intermediated. Third, the *signal precision*: flagged firms with direct evidence (compensated reviews) will more fully revise belief priors and thus have a more negative demand effect versus algorithmically flagged firms.

The last λ_i dimension is *firm age*. Firms with longer signal histories and deeper intermediary integration have higher λ_i than newer firms whose visitors arrive disproportionately through direct channels.[7] For demand composition, market penetration, and signal precision, the prediction is clear: more intermediation, or a more credible alert, produces larger losses. Firm age, however, creates competing forces, because the loss $-\Delta \log D = \lambda_i \log\Big((M/B)_{rep}\Big)$ depends on *both* λ_i and $(M/B)_{rep}$, which age moves in opposite directions. Older firms will have more verifiable quality and thus a lower $(M/B)_{rep}$, thereby predicting *smaller* losses. But they also have higher λ_i: deeper intermediary integration channels more demand through the signal, predicting *larger* losses. Whether younger or older firms suffer greater penalties is therefore an empirical question, and the answer allows us to observe which channel dominates—the fragility of inflated reputation or the depth of informational intermediation. We take this to the data in Section 4.

Hypothesis 2: *The demand decline following exposure is larger (i) the more credibly the alert discredits the premium, as proxied by the verifiability of the alert signal; and the greater the intermediation intensity, as proxied by (ii) the timing of the alert during high intermediary usage periods, and (iii) intermediary penetration in the local market.*

[7] Flagged firms are also prohibited from advertising on the intermediary for one year, which could amplify age heterogeneity if established firms advertise more. Section 4.6 addresses this enforcement channel directly and shows it is not the primary mechanism.

2.4.3 Other Empirical Implications

The belief-dependent premium δ accumulates gradually through reputational signals but collapses at exposure time t^*. Shapiro (1983) suggests that rebuilding reputational price premiums requires many verifiable signals, and thus recovery will be slow and incomplete. Thus, we expect to observe a rapid loss of reputational capital and long-lived impairment with event-study estimates of demand around exposure. We also examine three potential consequences of the credibility shock: weaker review activity, less suspicious or policy-violating content, and reallocation of demand toward substitute peer firms. Because these reflect broader information-environment implications rather than standalone predictions, we test them without formal hypotheses.

3 Data, Sample, and Empirical Methodology

3.1 Data Sources and Sample Construction

We combine three data sources: intermediary data on firm listings, reviews, and alerts; GPS-based foot traffic from anonymized SafeGraph mobile signals; and, for a subset of firms, Consumer Edge credit card transactions. Foot traffic is a high-frequency, establishment-level measure of real economic activity that complements accounting outcomes (Chetty et al., 2024; Goolsbee and Syverson, 2021; Gurun et al., 2023). Recent work validates this design for identifying real responses to information events: Noh et al. (2025) use the same GPS data to show that earnings announcements move visits. Our 2019–2024 sample contains 288,426 firm-month observations with active intermediary listings, a period with broad alert coverage and high-frequency foot traffic data.[8]The baseline estimation sample includes 16,837 firm–month observations in a symmetric 3-month window around the month the firm was flagged

[8]Our sample begins in 2019, when SafeGraph foot-traffic and spending data via Dewey become available. SafeGraph reports visits in dwell-time buckets: less than 5 minutes, 5–20, 21–60, 61–240, and more than 240 minutes. We exclude visits over 240 minutes, which likely reflect employees or other non-patron activity.

for inflated reputation.[9]

Credit card transaction data are available for 4,509 firm–month observations, with limited coverage driven by the provider's merchant panel rather than selection on firms or alerts. We use foot traffic as our main outcome because it covers the full sample and captures demand regardless of payment method or whether a visit converts to a recorded transaction; the credit card data then confirm that these visit declines correspond to actual spending losses.

The intermediary issues two alert types. A *Compensated Activity Alert* is triggered when it receives documentary evidence (e.g., screenshots of solicitations or payment offers) that a firm bought or solicited reviews. These alerts arise from third-party whistleblower reports, which the intermediary investigates, and, if confirmed, posts the alert.

A *Suspicious Review Activity Alert* is triggered by automated detection software that flags statistical patterns consistent with manipulation, such as many positive reviews from a single IP address or accounts tied to known review-exchange rings. The key difference is timing: Compensated Activity Alerts depend on outsider reports, whereas Suspicious Review Activity Alerts depend on the intermediary's internal detection process. We use this difference in our identification strategy (Section 4.5) and heterogeneity analysis (Section 4.7).

In our sample, 6,867 firm-month observations involve evidence-based Compensated Activity alerts and 9,970 involve algorithmic Suspicious Review Activity alerts. For each flagged firm we observe the issuance month, alert type, and duration. Alerts appear as a prominent warning on the firm's page and remain visible for about 90 days.[10]

To construct the control group, we identify demand-side peers via revealed substitution: the intermediary's recommendation algorithm links firms sharing overlapping visitor flows,

[9] To link a Yelp business to a SafeGraph POI, we first fuzzy-match the Yelp business name to SafeGraph's location_name using RapidFuzz, which returns a 0–100 name similarity score. For each Yelp business, we keep the top SafeGraph candidates by name similarity, extract the Yelp ZIP code from the business-location string, and drop candidates in other SafeGraph ZIP codes. We then compute an address similarity score by comparing the SafeGraph street address with the Yelp location string using normalized string-edit distance. We average the name and address scores, rank candidates by this average, and select the highest-scoring SafeGraph POI as the Yelp–SafeGraph match.

[10] The Consumer Alert also carries a one-year ban on purchasing advertising, which could amplify demand declines for older firms if they advertise more. We bound this enforcement channel in Section 4.6.

matching on a vector of attributes such as location, category, price segment, and browsing behavior.[11] Unlike standard industry-code controls that group firms by what they produce, our control group consists of firms considered by the same potential buyers—a demand-side match that captures substitution along dimensions too fine for coarser controls. Monthly foot traffic data report visits and unique visitors per firm. The distribution is highly right-skewed (mean = 726; median = 121) and includes zeros, so we use the inverse hyperbolic sine transformation (*Visits*) as the baseline dependent variable, which accommodates zeros and approximates a log for positive values. In the credit card transaction subsample, we observe spending volume for in-person and online transactions (*Transactions*, *Online Transactions*).

Table 2 reports summary statistics for the estimation sample. Three features stand out. First, the median cumulative stock of four- and five-star reviews (*Positive Review Stock*) is zero, so half the sample has a thin verifiable reputational base, consistent with the framework's prediction that firms with little accumulated stocks of verifiable quality will have the most to gain from inflating belief-based signals. Second, firm age (*Firm age*; log age in months) is highly dispersed (mean = 3.23, SD = 1.16 in logs), providing variation to test the competing-forces prediction in H2. Third, cross-industry local traffic (*Cross-Industry Traffic*: mean = 7.86) far exceeds same-industry traffic (*Same-Industry Traffic*: mean = 3.97), foreshadowing the Table 3 result that generalized local demand exposure, rather than narrow product-market rivalry, better predicts which firms inflate reputational signals.

Our identification strategy controls for time-varying determinants of foot traffic that could confound the treatment effect, grouped into four categories. First, we control for reputational capital: cumulative four- and five-star reviews (*Positive Review Stock*) and firm age (*Firm Age*). Both predict manipulation propensity (Section 3.2) and mechanically relate to demand; because they vary over time, firm fixed effects do not absorb them. Second, we include calendar controls—days (*Days in Month*), Saturdays (*Saturdays*), and Sundays (*Sundays*) per month—to capture mechanical variation in foot-traffic opportunities. Third,

[11] The intermediary labels these "People Also Viewed" (PAV) recommendations; we refer to them as demand-side peers.

we include weather controls: cooling degree days (*Cooling Degree Days*) and total monthly precipitation (*Precipitation*), standard determinants of pedestrian activity. Omitting them could leave seasonal demand variation in the residual that correlates with alert timing if detection rates vary seasonally.[12] Fourth, we control for the local competitive environment: foot traffic to other firms within a one-mile radius, separately for same-industry (*Same-Industry Traffic*) and cross-industry (*Cross-Industry Traffic*) establishments.[13] These variables capture time-varying, location-specific demand shocks—such as a new transit stop, local festival, or seasonal tourism—that affect all nearby firms and could otherwise bias the DiD estimate if they coincide with alert issuance.

3.2 Descriptive Evidence: Which Firms Are Flagged?

We first examine which firms are flagged:

$$Flagged_i = \alpha + X_i\beta + \gamma_c + \eta_j + \varepsilon_i \tag{6}$$

where $Flagged_i$ equals one if firm i receives an alert during the sample period. The sample comprises focal firms that receive an alert and their matched demand-side peers, each observed in the month the focal firm is flagged; because alerts arrive at different times, each focal firm and its peers enter only for the corresponding alert month. X_i includes *Positive Review Stock*, *Firm Age*, calendar controls, and local competition. γ_c and η_j are city and 3-digit NAICS industry fixed effects. Section 4.1 below discusses are findings, revealing that firms with unusually many positive reviews (*Positive Review Stock*), shorter operating histories (*Firm Age*), and greater *Cross-Industry Traffic* are significantly more likely to be flagged, consistent with strategic inflation by firms with the most to gain.

[12] Weather controls are derived from CustomWeather hosted by Dewey. Each firm is assigned to its nearest weather station, selected from 2,069 stations nationwide based on geographic distance.

[13] *Same-Industry Traffic* and *Cross-Industry Traffic* measure total foot traffic to nearby establishments—SafeGraph "points of interest" (POIs)—within a one-mile radius of the focal firm. The former aggregates traffic to POIs in the same industry as the focal firm, the latter aggregates traffic to POIs in all other industries.

3.3 Identification Strategy

We estimate the causal effect of exposure on demand using a difference-in-differences design that compares flagged firms to their demand-side peers before and after alert issuance:

$$Visits_{it} = \alpha + \beta(\mathit{Flagged}_i \times \mathit{Post}_{it}) + X_{it}'\Gamma + \gamma_i + \delta_t + \varepsilon_{it} \tag{7}$$

$Visits_{it}$ is the inverse hyperbolic sine (IHS) of monthly foot traffic (or transaction volume) for firm i in month t. The IHS handles zero visits and approximates the log for positive values, so β is the approximate percentage change in foot traffic due to exposure. $\mathit{Flagged}_i$ indicates treated firms; Post_{it} equals one in the alert month and the three months thereafter (see the event window below). Firm fixed effects γ_i absorb time-invariant heterogeneity, and year-month fixed effects δ_t absorb aggregate shocks common to all firms in a given calendar period, including macroeconomic fluctuations and pandemic-era disruptions. The vector X_{it} collects time-varying controls: the current level of reputational capital (*Positive Review Stock*, *Firm Age*), weather (*Cooling Degree Days*, *Precipitation*), and local same- and cross-industry traffic. The calendar-composition controls used in Equation (6)—days, Saturdays, and Sundays per month—are not included here because they are common to all firms within a calendar month and are absorbed by δ_t. Under parallel trends—that absent the alert, flagged and peer firms would have followed similar demand paths—β identifies the causal effect of exposure. The sample includes 16,837 firm–month observations in a symmetric 3-month window around each firm's alert month.

Two distinct detection channels. The primary causal threat to our investigation is that firms might be flagged just as their demand was about to fall for unrelated reasons. Because our two alert types have their timing set by entirely different processes, this concern must be evaluated separately for each—and each clears the bar on its own grounds. Evidence-based alerts start with an outsider—a patron, competitor, or journalist—who reports the firm

(Dyck et al., 2010), so their timing reflects an outsider's decision rather than the firm's own trajectory. Algorithmic alerts, by contrast, are triggered inside the platform when automated detection flags statistical patterns in the review database—IP clustering, reviewer-network links, and unusual review timing. While this process is ostensibly orthogonal to the firm's demand or operations; in theory its timing timing could be tied to the firm's own path, since it moves with within-platform review activity. This makes the algorithmic channel the more demanding test—yet it shows the flattest pre-trends in the paper (maximum $|t| = 0.53$; Section 4.5, Figure 4). The two channels thus rest on separate identifying arguments—plausibly exogenous timing for one, flat pre-trends for the other—so the result does not hinge on either alone.

Event window. Our baseline static regression specifications use a symmetric three-month window $(-3, +3)$: $Post_{it}$ equals one in the alert month and the three months after, and the pre-period comprises the three months before. We adopt this window for three reasons. First, it maps onto the institution: the Consumer Alert is displayed most prominently in the roughly ninety days following issuance, so a three-month post-period captures the interval during which the warning is actively shown. Second, the symmetric pre-period draws the counterfactual from the months most comparable to treatment. Third, a balanced window maximizes the matched panel: each event contributes the same seven-month window, whereas wider static windows would force unbalanced cohort composition into a single pooled coefficient.[14]

Across all specifications, standard errors are two-way clustered by industry and city (Cameron et al., 2011), capturing common demand shocks along both dimensions: firms in the same city share local economic conditions, and firms in the same industry share sector-wide shocks.

[14]We do not rely solely on this static 3-month window to establish pre-trends or persistence. We use Sun–Abraham event studies in Figures 2 and 3 to estimate the full dynamic path over horizons up to eighteen months and across a variety of alternative pre and post windows (Appendix Table A.1), confirming pre-trends are indistinguishable from zero and a persistent decline in demand.

Isolating credibility from fundamentals. The manipulation flag conveys no information about the firm's true quality θ. The firm's products, pricing, operations, and service quality remain unchanged on the issuance date and throughout the alert window. This separation is credible for several reasons. First, θ captures experiential attributes—quality, environment, service—that patrons evaluate directly. Prior visitors form beliefs about θ independently of online signals; when the flag appears, they revise only their assessment of the reputational premium δ. Second, the alert is informationally precise: it identifies manipulation of review signals, not a decline in underlying quality. A visitor who sees the alert learns that the firm's earlier reputation was artificially constructed, not that its products or service worsened—the program aims to correct informational distortion, and is orthogonal to firm operations. [15] Finally, the flat pre-alert trends we display in Figure 4 are inconsistent with any pre-existing quality deterioration driving both the flag and the demand decline.

Control group construction. Because peer status reflects existing browsing patterns rather than alert issuance, treatment does not affect control group membership, avoiding a reflection problem in peer selection. Peer firms also share the same local information environment and similar intermediation intensity, supporting parallel trends.[16]

Intermediation intensity. The framework parameter λ_i reflects accumulated intermediary tenure, review-history depth, and integration into the recommendation architecture —features that change over months and years, not week to week. Because the 90-day alert window is short relative to this evolution, we treat λ_i as time-invariant over the estimation horizon. One concern may be that treated firms were on a steeper pre-alert λ_i trajectory, creating spurious pre-treatment divergence—this would produce sloped pre-period foot-traffic

[15] One potential concern is that the alert could also weaken perceptions of genuine θ-based signals along with the fabricated δ, a phenomenon we call credibility contagion and discuss in detail in Section 5.3. Even so, this occurs only *after* the alert, making it a consequence of disclosure and unrelated to fundamentals.

[16] A possible concern is that review manipulation tainted the control group by drawing atypical visitors, distorting the peer set. This does not apply: the people also viewed signal used for peer assignment is based on within-session navigation—page content, location, price tier, and category—not rating-weighted rankings. Review inflation alters displayed ratings but not the session-level structure that defines peers.

coefficients for treated firms relative to peers, which we do not observe (Figure 2).[17]

Remaining threats. Firm fixed effects absorb any time-invariant selection into manipulation: characteristics correlated with both manipulation propensity and demand levels. What they cannot absorb is selection on time-varying unobservables—the residual form of the timing concern above, in which a firm is flagged just as some unobserved driver of demand turns down. The pre-trends evidence directly addresses this, and the observable imbalance strengthens the point—flagged firms start with *more* accumulated positive reputational signals than controls (Table 10, Panel A), a pattern inconsistent with firms on the verge of fundamental decline.

4 Empirical Results

4.1 Which Firms Are Flagged?

Table 3 reports regression estimates where the dependent variable is *Flagged*. Across all specifications, *Positive Review Stock*—the cumulative number of four- and five-star reviews—strongly predicts flagging. In the most saturated specification with industry and city fixed effects, a one-standard-deviation increase in log(Positive Review Stock) raises the flagging probability by about 13 percentage points (coeff. = 0.058, $t = 14.94$), large relative to the unconditional alert rate. Unusually positive reputational signals are a key marker of suspicious activity. *Firm Age* is negative and significant throughout. A one-standard-deviation increase in log(Firm Age) lowers the flagging probability by about 10 percentage points (coeff.= -0.088, $t = -3.35$ in the most saturated specification). Younger firms—with thinner verifiable reputational capital—are substantially more likely to manipulate, consistent with the framework's prediction that the marginal return to inflating reputation is greatest

[17] Within the alert subsamples, this concern is greatest for algorithmic alerts, whose detection timing is keyed to the same accumulating review activity that λ_i proxies; as discussed above, that subsample nonetheless exhibits the tightest pre-trends in the paper (Panel B).

when the verifiable base is small.

Calendar controls capture mechanical variation in foot traffic and review activity across months; *Sundays* is positively associated with flagging (coeff. ≈ 0.23–0.27, $t > 3.5$). Local competitive conditions also matter. *Cross-Industry Traffic*—foot traffic to nearby firms outside the focal firm's industry—is positively and significantly associated with flagging (coeff. ≈ 0.0076–0.0249), suggesting firms inflate reputation to attract visitors drawn to the area for other reasons. By contrast, *Same-Industry Traffic* is insignificant, implying manipulation responds to broad local demand rather than narrow product-market rivalry. Overall, firms are more likely to be flagged when perceived quality is especially valuable, prior reputation is weak, and public attention is high—consistent with theories of reputation as a belief-based intangible asset that can be strategically distorted.

4.2 The Demand Premium of Positive Reputation

The patterns in Table 3 suggest that manipulation is strategic, and Table 4 shows why: positive reputational signals yield demand gains, creating incentives for inflation. Estimated on the full sample, *Positive Review Stock* is strongly positively associated with both foot traffic (*Visits* coeff. = 2.9%, $t = 7.65$) and unique visitors (*Visitors* coeff. = 3.4%, $t = 10.77$). These relationships are robust to firm, month, city, and industry fixed effects, though we treat them as conditional correlations rather than causal. The magnitudes empirically anchor the reputational demand premium $\lambda_i \log\Big((M/B)_{rep}\Big)$ in the framework—the demand later destroyed when the manipulation flag is posted. For younger firms with limited verifiable track records, these gains can be large, making manipulation an equilibrium strategy despite the risk of exposure.

4.3 The Real Cost of Exposure

Table 5 reports difference-in-differences estimates of exposure on foot traffic. The *Flagged* $\times$ *Post* coefficient in our preferred specification (Column 3), which includes firm and calendar year-month fixed effects, is -0.076 ($t = -3.52$), implying an approximately 8% decline in *Visits* relative to peer controls. The coefficient attenuates as fixed effects are added—from -0.271 with month effects alone (Column 1), to -0.153 with city and industry effects (Column 2), to -0.076 once firm fixed effects absorb time-invariant heterogeneity (Column 3)—consistent with selection into manipulation on firm-level characteristics that firm fixed effects remove. Our preferred estimate uses within-firm, within-time variation in exposure.

Our findings support H1: exposure of artificially inflated reputation induces a discrete demand decline without concurrent changes in firm quality or physical operations. The roughly 8% drop in foot traffic reflects the estimated magnitude of loss of a belief-dependent reputation premium. The magnitude is consistent with evidence that reputational penalties are largest when misconduct harms a firm's own counterparties rather than unrelated third parties (Karpoff and Lott, 1993)—as here, where fabricated signals mislead the firm's own potential patrons.

4.4 Structural Interpretation: Recovering the Fabrication Share

The DiD coefficient in Table 5 has a direct structural interpretation under Section 2. Exposure removes the belief-dependent premium δ from priced demand while leaving the verifiable base θ unchanged. With IHS(visits) $\approx$ log(visits) at typical visit counts, demand is $\log(\text{visits}) = \log D_0(\theta) + \lambda_i \log(1 + \delta/\theta)$ before exposure and $\log(\text{visits}) = \log D_0(\theta)$ after. The first difference identifies

$$\Delta \log(\text{visits}) \;=\; -\lambda_i \log\left(\frac{M}{B_{rep}}\right), \tag{8}$$

where $M/B_{rep} \equiv (\theta + \delta)/\theta = 1 + \delta/\theta$ is the reputational market-to-book ratio: total reputational signal relative to its verifiable base.

Under the benchmark that the intermediated premium passes through one-for-one into demand ($\lambda_i \to 1$), the coefficient $\beta = -0.076$ in Column (3) implies $\log(M/B_{rep}) \approx 0.076$, so $M/B_{rep} \approx 1.08$ and $\delta/\theta \approx 0.08$: the belief-dependent premium the market withdrew on exposure was roughly 8% of flagged firms' verifiable base. If exposure also dispels legitimate belief-based reputation alongside the fabricated portion, this figure is an *upper bound* on the fabricated share; the manipulation itself accounts for at most 8% of perceived quality, and likely less.[18] Because demand depends only on the product $\lambda_i \log(M/B_{rep})$, the level of δ/θ is not separately identified; the $\lambda_i \to 1$ benchmark fixes the split so the implied premium coincides with the reduced-form coefficient by construction. Relaxing the benchmark moves the implied premium in a known direction: because demand identifies only the product $\lambda_i \log(M/B_{rep})$, the premium satisfies $\delta/\theta = e^{\beta/\lambda_i} - 1$, so weaker pass-through implies a larger premium and stronger pass-through a smaller one—roughly 0.04 at $\lambda_i \approx 2$ to 0.16 at $\lambda_i \approx 0.5$. Across this range the underlying inflation remains modest, yet its exposure triggers large and persistent demand losses—the central evidence of belief-based intangible capital's fragility.

Our calibration is a conditional illustration, not a separately identified estimate. The mapping from the reduced-form coefficient to a premium rests on assumptions about intermediation intensity λ_i, alert exposure, the completeness of belief revision, the absence of contagion to genuine signals, platform-visibility effects, and the conversion from foot traffic to firm value—several of which our design cannot sign. We therefore keep three objects distinct. First, the *reduced-form demand response*: the approximately 8% foot-traffic decline we identify directly, which persists through every horizon we observe. Second, the *belief-dependent premium* in the intermediated signal, δ/θ: the belief-priced reputation above verifiable quality that exposure destroys, of which the *fabricated* portion is a subset—so the destroyed premium bounds fabrication from above. Third, *total impairment of intangible capital*: the full value loss, which our demand-based estimate bounds from below because it

[18]This destroyed premium exceeds a naive ex-ante benchmark—flagged firms' 0.698 log-unit excess in pre-treatment *Positive Review Stock* (Table 10) maps, at the conditional elasticity in Table 4 (0.0291), to about 2% of demand—but that benchmark is a floor, not a contradiction: it counts only *surviving* reviews, while the fabricated signals are precisely those the intermediary filters (−43%) and removes (−10%), and the within-firm elasticity understates the cross-sectional return. A destroyed premium of the magnitude we estimate is therefore consistent with the ex-ante returns evidence.

omits the cost-of-capital channel.[19]

If the belief-dependent premium δ disappears permanently once exposed, what sustains the firm's remaining demand, and why doesn't the gap with peers ever close? The drop is limited by the firm's verifiable base: true quality θ is unchanged and still generates demand, and repeat patrons whose beliefs derive from experience keep visiting regardless of the intermediated signal. Persistence follows from the same logic. The alert permanently removes the artificial support—suspicious review activity collapses after exposure and remains low (Section 4.6)—so the post-alert level is demand at the firm's unassisted θ base, observed for the first time. Rebuilding the premium legitimately requires slowly reaccumulating verifiable signals in a degraded information environment, a process Figure 3 shows has not begun within eighteen months; the small short-run attenuation does not cumulate into reversion.[20] This is the rapid-impairment, slow-recovery asymmetry of reputational capital (Shapiro, 1983) in stark form: one disconfirming disclosure instantly destroys what accumulated signals built, and no offsetting signal has yet restored the lost premium.

The magnitude and persistence of the decline are best understood relative to the only comparable estimate of disclosure's effect on real foot traffic. Noh et al. (2025) use the same GPS data and find that earnings announcements increase visits by about 1% on average, peaking near 3% four to five days after the announcement within a two-week window. Their effect reflects public attention rather than belief updates about firm quality: it rises with media coverage and search intensity, disappears for low-attention announcements, and—for comparable service firms—responds to the salience of extreme news rather than its sign.

Our estimate builds on Noh et al. (2025), who show that disclosure affects visitor behavior, but we identify a different and more destructive channel. An earnings announcement is

[19] The recovered δ/θ reflects only the demand response; because our estimator ignores any effect of exposure on firms' cost of capital (Lambert et al., 2007; Biddle et al., 2009), the 8% demand decline is a lower bound on total value destruction. If credibility shocks also tighten reputation-dependent financing, total losses exceed those implied by foot traffic. We expound upon this further in Section 5.3.

[20] Persistence is notable because the warning banner is removed after about 90 days: the demand response outlasts the banner by more than a year, indicating that the operative shock is the revelation, not the banner.

a recurring investor signal that reaches visitors only indirectly and briefly raises a firm's salience; a credibility shock is the opposite—it arrives at the visit decision, directly challenges whether the firm's reputation is genuine, and destroys belief-based intangible capital rather than providing a temporary cue. The dynamics are qualitatively opposite: an attention shock generates a positive, mean-reverting bump that disappears within two weeks, whereas the credibility shock causes a negative decline that persists for at least eighteen months. The effects are not directly comparable—their coefficient averages over a 7-day window, ours over 90 days—since longer averaging mechanically dilutes transient effects but not persistent ones. Still, the size of the raw coefficients in the credibility shock is larger by nearly an order of magnitude (0.009 vs. -0.076), and the difference in sign and persistence is clear. This contrast rules out the idea that the alert merely increases attention—because an attention effect such as the one found in in Noh et al. (2025) would have vanished long before the persistence observed in Figure 2. Supporting evidence points the same way: the decline is about twice as large for evidence-based as for algorithmic alerts (Section 4.7), continues after the warning banner is removed, and coincides with measurable deterioration of the firm's information environment (Section 4.6)—none of which a transient attention shift would produce.

4.5 Dynamic Effects and Validation of Parallel Trends

The event-study design tests the parallel-trends assumption and documents the dynamic demand response to exposure. It also provides the direct test of the detection-timing concern from Section 3.3: were firms flagged precisely when demand was about to decline, pre-treatment trajectories would diverge. We re-estimate Equation (7), replacing the single *Flagged* $\times$ *Post* indicator with a full set of event-time dummies. Let D_{it}^{k} equal one for treated firms in month k relative to the alert date, and zero otherwise. We estimate:

$$Visits_{it} \;=\; \alpha_i + \delta_t \;+\; \sum_{\substack{k=-6 \\ k\neq -1}}^{+6} \beta_k \, D_{it}^{k} \;+\; X_{it}'\Gamma \;+\; \varepsilon_{it} \tag{9}$$

where α_i and δ_t are firm and year-month fixed effects, X_{it} is the control vector from Equation (7), and the $k = -1$ dummy is omitted as the reference period. The coefficients $\beta_{-6}, \beta_{-5}, \ldots, \beta_{+6}$ capture foot-traffic dynamics before and after the alert.

Because firms are flagged in different months, we use the Sun and Abraham (2021) estimator rather than OLS. In staggered settings with heterogeneous treatment effects, OLS can yield event-time coefficients that are weighted averages with negative weights on some firms, distorting the dynamics (Goodman-Bacon, 2021; Baker et al., 2022). The Sun-Abraham estimator avoids this by estimating treatment effects separately for each treatment cohort, then averaging them by each cohort's share of treated firms. We use never-treated PAV peers as the control group, providing a clean comparison cohort and avoiding the contamination that arises when all units are eventually treated. The resulting $\hat{\beta}_k$ are plotted in Figure 2.

Pre-treatment coefficients are economically small (within ± 2 percentage points of zero) and statistically indistinguishable from zero, with a maximum absolute t-statistic of 1.24 at $t = -2$. The absence of differential pre-trends supports the parallel-trends assumption under a heterogeneity-robust estimator and indicates that our setting isolates a credibility shock rather than deteriorating fundamentals: the alert reveals that the firm's reputational signals were fabricated but conveys no new information about products, prices, service quality, or physical operations.[21] Figure 4 reports subsample event-study estimates using the Sun and Abraham (2021) interaction-weighted estimator, estimated separately for evidence-based (*Compensated Activity*) and algorithmic (*Suspicious Review Activity*) alerts. Both subsamples exhibit flat pre-treatment trajectories, supporting parallel trends for both de-

[21] Standard two-way fixed-effects event-study estimates are quantitatively similar to those in Figure 2: pre-treatment coefficients are statistically indistinguishable from zero in all pre-period bins, and post-treatment coefficients are negative and significant starting in the first post-alert month. Full TWFE estimates are available from the authors upon request.

tection mechanisms. For evidence-based alerts (Panel A), pre-period coefficients range from -0.019 to $+0.027$ and none is significant at the 5% level under our clustered inference (the largest, $|t| = 1.97$ at $t = -2$, has $p = 0.058$).[22] For algorithmic alerts (Panel B), pre-period coefficients are economically negligible, with a maximum $|t|$ of 0.53 and all p-values above 0.54. The algorithmic subsample—where detection timing is fully determined by the intermediary's automated process—thus displays cleaner pre-trends than the evidence-based subsample, undermining the concern that automated detection is endogenously triggered by deteriorating fundamentals.

Post-alert, both subsamples show sharp demand declines matching the pooled pattern in Figure 2, with no decline in the alert month itself (pooled: -0.6%, $t = -0.47$) and an abrupt onset thereafter that rules out anticipation. Evidence-based alerts trigger a larger immediate response, reaching -11.8% at $t = +2$ ($t = -4.18$), consistent with higher signal precision inducing stronger belief revision among intermediary-dependent visitors. Algorithmic alerts generate a smaller but similarly persistent response, reaching -8.1% at $t = +3$ ($t = -3.51$). The magnitude gap between alert types is analyzed formally in Section 4.7.

The post-alert decline persists over the full six-month window, with magnitudes aligning with the baseline DiD estimate of roughly 8% in Table 5. Foot traffic remains 8.2% below counterfactual at $t = +3$ ($t = -4.71$), 7.8% at $t = +4$ ($t = -3.08$), and 6.9% at $t = +6$ ($t = -2.53$), with confidence intervals excluding zero throughout. Beyond six months, Sun–Abraham estimates extended through $t = +18$ (Figure 3) show the decline does not reverse: point estimates remain negative throughout the eighteen months after the alert, including the six months after the one-year advertising ban expires, where they range from -8.1% to -11.4% with $|t|$-statistics above 2.8.

We vary two features of the event window to probe these conclusions separately: the lead length tests whether pre-treatment trends are flat at short and long horizons, and

[22] The $t = -2$ coefficient for evidence-based alerts is $+0.027$, indicating treated firms were slightly *above* peer controls in the month before the alert, contrary to a deteriorating-fundamentals story that would predict a negative pre-trend.

the post-horizon traces how long the decline lasts—through the alert window, after banner removal, and beyond the one-year advertising ban. Across all six combinations of six- and twelve-month leads and six-, twelve-, and eighteen-month horizons, the decline is stable in sign, magnitude, and significance, and the pre-period shows no systematic differential trend (Appendix Table A.1). The twelve-month lead, which more finely resolves the pre-period, shows treated firms drifting slightly upward to close the gap with their peers before the alert, not declining into it—evidence of a reputational premium from fabricated signals, not deteriorating fundamentals, which would imply the opposite pre-trend.[23] The persistence of negative coefficients well beyond the typical alert window indicates that the credibility shock erodes belief-based intangible capital that is hard to rebuild: once the intermediary publicly questions a firm's reputational signals, demand remains depressed for at least a year and a half. Consistent with the asymmetry in Section 2.4.3, rebuilding credibility requires slowly reaccumulating verifiable interactions, signal by signal. The degraded information environment in Section 4.6 slows this further: lower *Avg. Rating*, more *Missing Reviews*, and reduced manipulation weaken the infrastructure for new reputational signals.[24]

4.6 Effects on the Information Environment

Table 6 shows that exposure reshapes the information environment around flagged firms. After flagging, the average rating of newly received reviews (*Avg. Rating*) falls by about 0.19 stars ($t = -3.91$). The probability of receiving zero reviews in a month (*Missing Reviews*) rises by 5.5 percentage points ($t = 3.08$), indicating weaker engagement with the review system. *Filtered Reviews*, our proxy for suspicious reviews, fall by about 43%, and *Removed*

[23] Two of the eleven pre-period coefficients in the longest window are marginally significant (months $t = -7$ and $t = -6$); both trend positively toward zero, consistent with pre-exposure accumulation of the reputational premium documented in Table 4 rather than a demand decline predating the alert.

[24] The slow recovery also contrasts with capital-market reputation repair: Chakravarthy et al. (2014) show that firms actively rebuild reputation after restatements through targeted actions—press releases, CSR initiatives, governance reforms—that generate positive returns. Firms in our setting lack a comparable mechanism: the intermediary offers no channel to signal reform, and the deterrence effects in Section 4.6 suppress the review activity that would otherwise rebuild the signal stock, leaving recovery slow and largely passive.

Reviews fall by about 10% ($t = -7.12$). Because these categories capture suspicious or policy-violating content, the decline is consistent with deterrence: firms that might otherwise post low-credibility reviews self-select out under heightened scrutiny, potentially reinforced by stricter filtering for flagged firms.

Credibility shocks thus do more than reduce demand—they damage the firm's information infrastructure. Lower ratings, weaker review engagement, and reduced manipulation persist even after the alert is removed, slowing the signal-by-signal rebuilding of reputation and helping explain the persistent demand reduction in Figure 2.

Each alert includes an enforcement action: flagged firms cannot purchase intermediary advertising for one year. Four facts limit the role of this channel. First, both alert types impose the same ban, yet evidence-based alerts cut demand by about twice as much as algorithmic alerts (Table 7); a pure advertising channel cannot explain variation across alerts that differ only in information. Second, demand drops discretely within the first month, while lost advertising should reduce visibility gradually. Third, demand partially rebounds between the trough and month nine even though the ban is still active, the opposite of what a binding advertising constraint predicts. Fourth, when the ban expires after twelve months, demand does not recover. Instead, estimates in the six months after expiry are among the largest and most precise in the event study (Figure 3), from -8.1% to -11.4% with t-statistics above 2.8 in absolute value. We therefore attribute the demand response to belief revision rather than the bundled enforcement action.

4.7 Heterogeneity

In Section 2 we predict that the demand penalty grows with two features of the intermediated channel: its *intensity* λ_i and how *completely* the alert discredits the belief-based premium. Table 7 presents heterogeneity analyses along four dimensions spanning these two margins. Across these splits the destruction rate varies several-fold—from roughly 3% where intermediary reliance is lowest to upward of 20% where it is highest—so the penalty tracks the strength of the intermediated channel rather than applying uniformly. This dispersion is itself the empirical content of fragility: the same credibility shock is far more damaging precisely where belief-based demand is most exposed. First, *signal precision*. Evidence-based alerts—those backed by concrete proof of solicited or purchased reviews—produce much larger drops in foot traffic than algorithmic alerts (coeff. $= -0.1120$, $t = -5.82$ vs. coeff. $= -0.0545$, $t = -1.70$). The verifiable proof, rather than an opaque classification, spurs stronger belief revision, more fully discrediting the inflated premium—a completeness margin, not an intensity one.

Second, *demand composition*. Flags issued during holidays—when many users are uninformed and rely more on the intermediary—have much larger effects than flags outside holidays (coeff. $= -0.2273$, $t = -3.08$ vs. coeff. $= -0.0488$, $t = -1.91$). In these high-demand periods, more decisions run through the intermediary as visitors to unfamiliar areas lean on reputational signals, temporarily raising λ_i. The effect—about 4.5 times the non-holiday estimate—is large, though the smaller holiday subsample ($N = 2{,}973$ vs. 13,864) cautions against strong inference about the exact gap.

Third, *market penetration*. The penalty is much stronger where intermediary usage is higher (coeff. $= -0.1151$, $t = -3.05$ vs. coeff. $= -0.0333$, $t = -1.97$). Where the intermediary controls more local search and discovery, each firm's intermediated demand premium is larger and the credibility shock reaches more potential visitors, directly testing H2(iii).

Last, *firm age*. The framework gives opposing forces—younger firms carry a higher $(M/B)_{rep}$, older firms a higher λ_i—so the question is which dominates. The estimates favor intermediation: older firms face larger penalties (coeff. $= -0.1009$, $t = -3.17$) than younger firms (-0.0573, $t = -3.54$), both significant at 1%. The difference is not itself statistically significant, but the ordering is opposite the inflation channel: if inflated reputation were dominant, younger, higher-$(M/B)_{rep}$ firms would bear the larger penalty. We read this as directional evidence, consistent with the other dimensions, that intermediation intensity rather than inflation drives the penalty.

Across all four dimensions, the pattern consistently supports H2: the real economic cost of credibility shocks rises with the strength of the intermediated channel. Demand composition, market penetration, and firm age increase that cost by raising λ_i, while signal precision does so by making the alert more credible. In all cases, the decline works through the intermediated information channel, not through generalized stigma, operational disruption, or indiscriminate loss of trust.

4.8 Competitive Reallocation

Reputational shocks can affect both the focal firm and its close substitutes, with ambiguous direction ex ante. If stakeholders generalize mistrust, exposure creates negative market-level spillovers; if they reallocate toward firms with credible signals, demand shifts according to relative reputational strength. Table 8 examines where displaced demand goes, splitting peers by reputational strength. The reallocation we can estimate with precision accrues to peers with deeper verifiable reputation: foot traffic rises 5.7% for peers with a large stock of four- and five-star reviews (*Positive Review Stock* > 50; coeff. $= 0.057$, $t = 3.20$) but is insignificant for thin-reputation peers (*Positive Review Stock* ≤ 50;coeff. $= 0.040$, $t = 1.62$), and a *Firm Age* split at two years is significant for older peers (coeff. $= 0.073$, $t = 2.65$) but not younger ones (coeff. $= 0.033$, $t = 1.66$). Demand reallocates locally rather than exiting. We read this as suggestive, not definitive: within each split the high- and low-strength estimates are statistically indistinguishable, so the destination of displaced demand cannot be pinned to deeper-reputation or older peers with confidence.

4.9 Corroborating Evidence from Transaction Data

Table 9 analyzes credit card outcomes for firms with spending data (n = 4,509), showing that lower foot traffic becomes realized revenue losses.[25] The *Flagged* $\times$ *Post* coefficient is negative and significant for total (coeff. $= -0.053$, $t = -2.96$) and online transactions

[25]The sample reduction is due to the limited availability of spending-pattern data.

(coeff. = -0.067, $t = -1.95$), a 5–6% decline in volume. The close alignment of GPS-based foot traffic and credit card spending in both magnitude and significance indicates that the demand decline reflects real economic damage, not measurement error.

4.10 Entropy Balanced Covariates

Table 10 reports entropy balanced baseline regressions as a robustness check for selection on observables. Panel A compares flagged firms to PAV peer controls on baseline covariates. Flagged firms have more positive pre-treatment review stock (-0.70, $t = -7.48$) and are in slightly denser markets (-0.32 for same-industry traffic, $t = -2.14$; -0.32 for different-industry traffic, $t = -3.20$). Other covariates (firm age, calendar composition, weather) are similar or only slightly different. The review-stock difference reflects selection: firms with unusually favorable reputations are more likely to be flagged. Our identification relies on parallel trends, and Figure 2 shows these level differences do not imply differential pre-treatment trends. Panel B re-estimates the baseline DiD with entropy-balancing weights (Hainmueller, 2012) that match flagged firms' pre-treatment moments for the Panel A covariates. The headline coefficient is -0.077 with firm and month fixed effects ($t = -4.95$), essentially identical to the unweighted -0.076 in Column (3), Table 5. This similarity under entropy balancing suggests observable imbalances do not drive the treatment effect.

5 Discussion

5.1 Isolating the Cost of Reputational Capital Impairment

Identifying reputational capital impairment is difficult because credibility shocks usually coincide with fundamental news. Jarrell and Peltzman (1985) show product-recall penalties far exceed direct costs, and Karpoff et al. (2008) find penalties for financial misrepresentation dwarf legal sanctions; yet such events also involve SEC enforcement, management turnover, and operational disruption, obscuring the reputational component. Similar confounding arises in accounting restatements, which reveal misreporting while also triggering litigation risk and executive turnover (Palmrose et al., 2004; Hennes et al., 2008). Reputation repair likewise confounds identification, entailing operational changes that independently affect firm value (Chakravarthy et al., 2014). Armour et al. (2017) estimate reputational losses about nine times regulatory fines in UK enforcement actions, but those fines themselves change expected cash flows. Krüger (2015) finds sharp market reactions to ESG controversies that often coincide with value-relevant governance failures. Analogous issues arise in credit rating downgrades (Hand et al., 1992), which reflect updated default risk, and short-seller campaigns (Karpoff and Lou, 2010), which typically reveal fundamental problems alongside any credibility damage. Our setting strips away these confounds, providing a clean estimate of the standalone cost of impaired credibility.

Consistent with Eisfeldt and Papanikolaou (2013), who show organizational intangibles are risky because competitive firms can expropriate embedded human capital, we find belief-based reputational assets also lack a margin of safety and can be swiftly impaired. This distinguishes them sharply from misrepresented tangible assets. Inflating tangible assets does not increase productive capacity; the fabricated assets are inert, and profitability ratios such as ROA and ROE rise only because the denominator shrinks. Inflating belief-based reputational assets, by contrast, generates real demand: fabricated signals steer economic activity toward the firm while they remain credible. Exposure is therefore uniquely destructive, eliminating an asset that was generating real economic value rather than merely correcting an accounting fiction. The heterogeneity results (Section 4.7) are consistent with this fragility being governed by the depth of intermediation rather than the degree of inflation.

5.2 Measurement and Fragility of Intangible Capital

Our measurement approach also targets a separate broader challenge. Under current accounting standards, internally generated intangible assets—including reputational capital—do not appear on the balance sheet, and their impairment never enters reported earnings.

The economic losses we document are therefore invisible in financial statements for firms of any size; even if our sample firms were publicly traded, no impairment charge would arise because the destroyed asset was never capitalized. Ewens et al. (2025) address this gap from the accumulation side, using M&A due-diligence prices to estimate depreciation rates and recapitalize intangible stocks at market values. We address it from the impairment side, using real economic activity—GPS-based foot traffic and credit card transactions—to measure value destruction that accounting does not capture.

Dou et al. (2021) model customer capital as fragile to financial constraints; we show it is also fragile to informational shocks that undermine the signals sustaining it. Gourio and Rudanko (2014) treat the customer base as a sluggish state variable, costly to build and slow to adjust; our persistence results confirm this in reverse—once damaged, reputational capital rebuilds only gradually.

5.3 Omitted Channels and Bounding the Estimates

Information intermediaries can issue biased signals that mislead stakeholders, yet the real cost of exposing such manipulation has resisted measurement, because exposure usually coincides with fundamental news. Our setting provides it: the demand decline we estimate is the real-activity value propped up by belief-based reputation. A belief-based premium of roughly 8% above verifiable quality, destroyed on exposure, produces demand losses that persist for at least eighteen months without recovery—showing how strongly real activity depends on the integrity of intermediated information. We note two caveats below that follow from our reduced-form framework.

First, the decline we document is a direct first-order demand effect that impacts the firm's cash flow. If reputational damage also raises the firm's cost of capital—because lenders downgrade creditworthiness when public standing worsens—then total losses exceed what foot traffic alone captures, and the 8% demand decline understates total value destruction. To the extent that tighter or costlier financing constrains working capital and feeds back into operations, this second-order effect would compound the initial impairment over time, reinforcing the lower-bound reading; our design only captures the first-order demand channel and is silent on these downstream dynamics.[26]

Second, our framework treats the collapse of δ as withdrawal of the entire belief-based premium, with verifiable quality θ intact. This is plausible for repeat patrons, whose own experience dominates intermediated signals. But for visitors without direct experience—the intermediary-dependent group behind our main effect—the alert may devalue *all* accumulated signals, including those from genuinely satisfied patrons. If new visitors cannot distinguish genuine from fabricated reviews, the observed demand loss combines collateral damage to legitimate reputational capital with the removal of fabricated signals. Then the 8% estimate could *overstate* the impairment of the belief-based premium: less fabrication yields the same economic harm.[27]

Taken together, the two caveats bound the estimate from opposite sides: the omitted cost-of-capital channel makes 8% a lower bound on total value destruction, while contagion to genuine signals makes it an upper bound on the fabricated premium alone. The central finding is robust to both—a belief-based intangible whose exposure inflicts persistent real losses.

6 Conclusion

Manipulated signals in trusted intermediaries impose real economic costs, yet those costs have been almost impossible to isolate: the exposure of fabricated reputation typically arrives bundled with news about firm fundamentals. We exploit a setting that separates the two—an

[26] Recent research by Huang (2025) shows that intermediary-based reputational signals affect lending terms for small firms, suggesting exposed firms may face tighter credit. Information-risk models likewise predict that greater asymmetry raises the cost of capital (Easley and O'Hara, 2004; Lambert et al., 2007), and Graham et al. (2008) find that corporate misreporters pay higher borrowing costs.

[27] Under the one-for-one pass-through benchmark ($\lambda_i \to 1$), the full 8% maps to the belief-dependent premium $\delta/\theta \approx 0.08$, itself an upper bound on the fabricated share. If contagion implies 4 percentage points come from discounting genuine θ-based signals and only 4 from removing fabricated signals, the fabricated component falls to roughly 0.04.

information intermediary that publicly flags fabricated reputational signals without revealing anything about a firm's products, prices, or physical operations.

Comparing flagged firms to demand-side peers identified by revealed substitution, we find that exposure reduces foot traffic by approximately 8%, an effect that emerges sharply and persists well beyond the disclosure window. Because the alert signals fabricated reputational capital, the decline represents belief revision of the firm's reputation–stakeholders update their priors after learning that others' opinions of the firm were fabricated. We find that the magnitude of the impairment from the credibility shocks scales strongly with intermediation intensity—the rate at which the intermediated signal passes through into demand. The mechanism and framework we develop should translate to wherever firm value rests on belief-based intangible assets—a credit assessment, an ESG score, an analyst recommendation. While modest fabrication appears to generate real economic activity while undetected; as that share of firm value grows, intermediation-intensity risk becomes a first-order concern for firms who rely on demand generated by fabricated reputational capital—its exposure can permanently destroy real economic value.

References

Akerlof, G. A., 1970. The market for "lemons": Quality uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488–500.

Anderson, M., Magruder, J., 2012. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. Economic Journal 122, 957–989.

Armour, J., Mayer, C., Polo, A., 2017. Regulatory sanctions and reputational damage in financial markets. Journal of Financial and Quantitative Analysis 52, 1429–1448.

Baker, A. C., Larcker, D. F., Wang, C. C. Y., 2022. How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics 144, 370–395.

Biddle, G. C., Hilary, G., Verdi, R. S., 2009. How does financial reporting quality relate to investment efficiency? Journal of Accounting and Economics 48, 112–131.

Bolton, P., Freixas, X., Shapiro, J., 2012. The credit ratings game. Journal of Finance 67, 85–112.

Cameron, A. C., Gelbach, J. B., Miller, D. L., 2011. Robust inference with multiway clustering. Journal of Business & Economic Statistics 29, 238–249.

Chakravarthy, J., deHaan, E., Rajgopal, S., 2014. Reputation repair after a serious restatement. The Accounting Review 89, 1329–1363.

Chetty, R., Friedman, J. N., Stepner, M., 2024. The economic impacts of COVID-19: Evidence from a new public database built using private sector data. Quarterly Journal of Economics 139, 829–889.

Dou, W. W., Ji, Y., Reibstein, D., Wu, W., 2021. Inalienable customer capital, corporate liquidity, and stock returns. Journal of Finance 76, 211–265.

Dranove, D., Kessler, D., McClellan, M., Satterthwaite, M., 2003. Is more information better? the effects of "report cards" on health care providers. Journal of Political Economy 111, 555–588.

Dyck, A., Morse, A., Zingales, L., 2010. Who blows the whistle on corporate fraud? Journal of Finance 65, 2213–2253.

Easley, D., O'Hara, M., 2004. Information and the cost of capital. Journal of Finance 59, 1553–1583.

Eisfeldt, A. L., Papanikolaou, D., 2013. Organization capital and the cross-section of expected returns. Journal of Finance 68, 1365–1406.

Ewens, M., Peters, R. H., Wang, S., 2025. Measuring intangible capital with market prices. Management Science 71, 407–427.

Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. Journal of Econometrics 225, 254–277.

Goolsbee, A., Syverson, C., 2021. Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. Journal of Public Economics 193, 104311.

Gourio, F., Rudanko, L., 2014. Customer capital. Review of Economic Studies 81, 1102–1136.

Graham, J. R., Li, S., Qiu, J., 2008. Corporate misreporting and bank loan contracting. Journal of Financial Economics 89, 44–61.

Gurun, U. G., Nickerson, J., Solomon, D. H., 2023. Measuring and improving stakeholder welfare is easier said than done. Journal of Financial and Quantitative Analysis 58, 1473–1507.

Hainmueller, J., 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis 20, 25–46.

Hand, J. R. M., Holthausen, R. W., Leftwich, R. W., 1992. The effect of bond rating agency announcements on bond and stock prices. Journal of Finance 47, 733–752.

Hennes, K. M., Leone, A. J., Miller, B. P., 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. The Accounting Review 83, 1487–1519.

Huang, R., 2025. The financial consequences of online review aggregators: Evidence from Yelp ratings and SBA loans. Management Science 71, 59–82.

Jarrell, G., Peltzman, S., 1985. The impact of product recalls on the wealth of sellers. Journal of Political Economy 93, 512–536.

Kanodia, C., Sapra, H., 2016. A real effects perspective to accounting measurement and disclosure: Implications and insights for future research. Journal of Accounting Research 54, 623–676.

Karpoff, J. M., Lee, D. S., Martin, G. S., 2008. The cost to firms of cooking the books. Journal of Financial and Quantitative Analysis 43, 581–612.

Karpoff, J. M., Lott, Jr., J. R., 1993. The reputational penalty firms bear from committing criminal fraud. Journal of Law and Economics 36, 757–802.

Karpoff, J. M., Lou, X., 2010. Short sellers and financial misconduct. Journal of Finance 65, 1879–1913.

Klein, B., Leffler, K. B., 1981. The role of market forces in assuring contractual performance. Journal of Political Economy 89, 615–641.

Krüger, P., 2015. Corporate goodness and shareholder wealth. Journal of Financial Economics 115, 304–329.

Lambert, R., Leuz, C., Verrecchia, R. E., 2007. Accounting information, disclosure, and the cost of capital. Journal of Accounting Research 45, 385–420.

Leuz, C., Wysocki, P. D., 2016. The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. Journal of Accounting Research 54, 525–622.

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

Mayzlin, D., Dover, Y., Chevalier, J., 2014. Promotional reviews: An empirical investigation of online review manipulation. American Economic Review 104, 2421–2455.

Noh, S., So, E. C., Zhu, C., 2025. Financial reporting and consumer behavior. The Accounting Review 100, 407–435.

Palmrose, Z.-V., Richardson, V. J., Scholz, S., 2004. Determinants of market reactions to restatement announcements. Journal of Accounting and Economics 37, 59–89.

Roychowdhury, S., Shroff, N., Verdi, R. S., 2019. The effects of financial reporting and disclosure on corporate investment: A review. Journal of Accounting and Economics 68, 101263.

Shapiro, C., 1983. Premiums for high quality products as returns to reputations. Quarterly Journal of Economics 98, 659–679.

Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics 225, 175–199.

A From Demand to Empirical Estimation Strategy

This appendix shows how the demand specification in Section 2 yields the difference-in-differences estimator in Section 3.3. Identification is covered in Section 3.3, and our coefficient bounds in Section 5.3.

A.1 Two components of demand

The model decomposes a firm's demand into a part it earns from verifiable quality and a part it borrows from an inflated reputational signal:

$$D = D_0(\theta)\left(1 + \frac{\delta_i}{\theta_i}\right)^{\lambda_i}. \tag{10}$$

$D_0(\theta)$ is the demand generated by the firm's true quality θ alone. $(1 + \delta_i/\theta_i)^{\lambda_i}$ scales this by the reputational premium δ_i, where λ_i measures how much firm i relies on the intermediary to channel demand. If $\delta_i = 0$, the factor is one and demand depends only on verifiable quality; if $\delta_i > 0$, the firm attracts more traffic than its fundamentals alone would warrant.

Taking the natural log of Eq. (10) turns the product into a sum:

$$\log D = \underbrace{\log D_0(\theta)}_{\text{quality baseline}} + \underbrace{\lambda_i \log\left(1 + \frac{\delta_i}{\theta_i}\right)}_{\text{reputational premium}}. \tag{11}$$

The first piece depends on the firm's verifiable fundamentals; the second is the entire contribution of the intermediated signal.

A.2 Disclosure of fabricated reputation collapses the premium

Upon being flagged by the intermediary, the market stops pricing the reputational premium ($\delta_i = 0$), and demand reverts to the firm's verifiable quality θ alone.

$$\text{Before:} \quad \log D^{\text{pre}} = \log D_0(\theta) + \lambda_i \log\left(1 + \tfrac{\delta_i}{\theta_i}\right), \tag{12}$$

$$\text{After:} \quad \log D^{\text{post}} = \log D_0(\theta) + \lambda_i \log(1 + 0) = \log D_0(\theta). \tag{13}$$

Verifiable quality is untouched because the alert conveys nothing about θ, only that the prior stock of reputational capital δ_i was manufactured. Subtracting Eq. (12) from Eq. (13) gives the change in log demand caused by the alert:

$$\Delta \log D_i = \log D^{\text{post}} - \log D^{\text{pre}} = -\,\lambda_i \log\left(1 + \frac{\delta_i}{\theta_i}\right). \tag{14}$$

The alert destroys demand—the impact is larger when λ_i is higher (the firm relies more on the intermediary) and when δ_i/θ_i is higher (the premium was large relative to verifiable quality). These two parameters drive the cross-sectional predictions in Section 4.7.

A.3 From one firm to difference-in-differences

The single difference in Eq. (14) may be confounded by seasonal swings or local demand shocks. We address this by differencing against a demand-side peer firm j that was not flagged. The peer receives no credibility shock at t^*, so its reputational premium is unchanged and cancels out in its pre-post difference, leaving only the common aggregate shock.[28] Differencing the flagged firm's change from the peer's change yields

$$\hat{\beta} = \underbrace{\left(\log D_i^{\text{post}} - \log D_i^{\text{pre}}\right)}_{\text{flagged firm's change}} - \underbrace{\left(\log D_j^{\text{post}} - \log D_j^{\text{pre}}\right)}_{\text{peer firm's change}} = -\,\lambda_i \log\left(1 + \frac{\delta_i}{\theta_i}\right). \tag{15}$$

Because we observe many firms entering at staggered dates rather than a single treated–control pair, we implement this double difference through the panel regression of Eq. (7), reproduced here:

$$Visits_{it} = \alpha + \beta(Flagged_i \times Post_{it}) + X_{it}'\Gamma + \gamma_i + \delta_t + \varepsilon_{it}.$$

The interaction $Flagged_i \times Post_{it}$ then captures what is left only for flagged firms after exposure, so its coefficient β is the empirical counterpart of $-\lambda_i \log(1 + \delta_i/\theta_i)$: the reputational premium destroyed by exposure.

A.4 A note on the inverse hyperbolic sine

Our primary dependent variable, $Visits$, is the inverse hyperbolic sine (IHS) of foot traffic rather than its logarithm. We use the IHS instead of the log because it is defined at zero, allowing us to keep firm-months with no foot traffic.[29] The IHS of x is $\operatorname{arcsinh}(x) = \log\left(x + \sqrt{x^2+1}\right)$. For x that is not close to zero, $\operatorname{arcsinh}(x) \approx \log(2x) = \log x + \log 2$, where the constant $\log 2$ is absorbed by firm fixed effects and removed by differencing. Thus, away from zero, the IHS and the logarithm move one-for-one, so the log-based derivation applies.

Because the β coefficient on $Flagged \times Post$ is the change in log foot traffic between post- and pre-alert periods,

$$\beta = \log(\text{visits}^{\text{post}}) - \log(\text{visits}^{\text{pre}}) = \log\left(\text{visits}^{\text{post}}/\text{visits}^{\text{pre}}\right).$$

For changes that are not too large, $\log(1 + r) \approx r$, so with r the fractional change in foot traffic, $\beta \approx r$, allowing the interpret of β as an approximate percentage change.[30]

[28] Explicitly, the peer's log demand is $\log D_0(\theta_j) + \lambda_j \log(1 + \delta_j/\theta_j)$ plus aggregate shocks in both periods. Because $\lambda_j \log(1+\delta_j/\theta_j)$ is constant at t^*, the peer's pre-post difference removes it and the fixed baseline, leaving only the common shock.

[29] Untabulated, our results are unchanged if we use $\log(1 + \text{Raw visits})$ instead of the IHS.

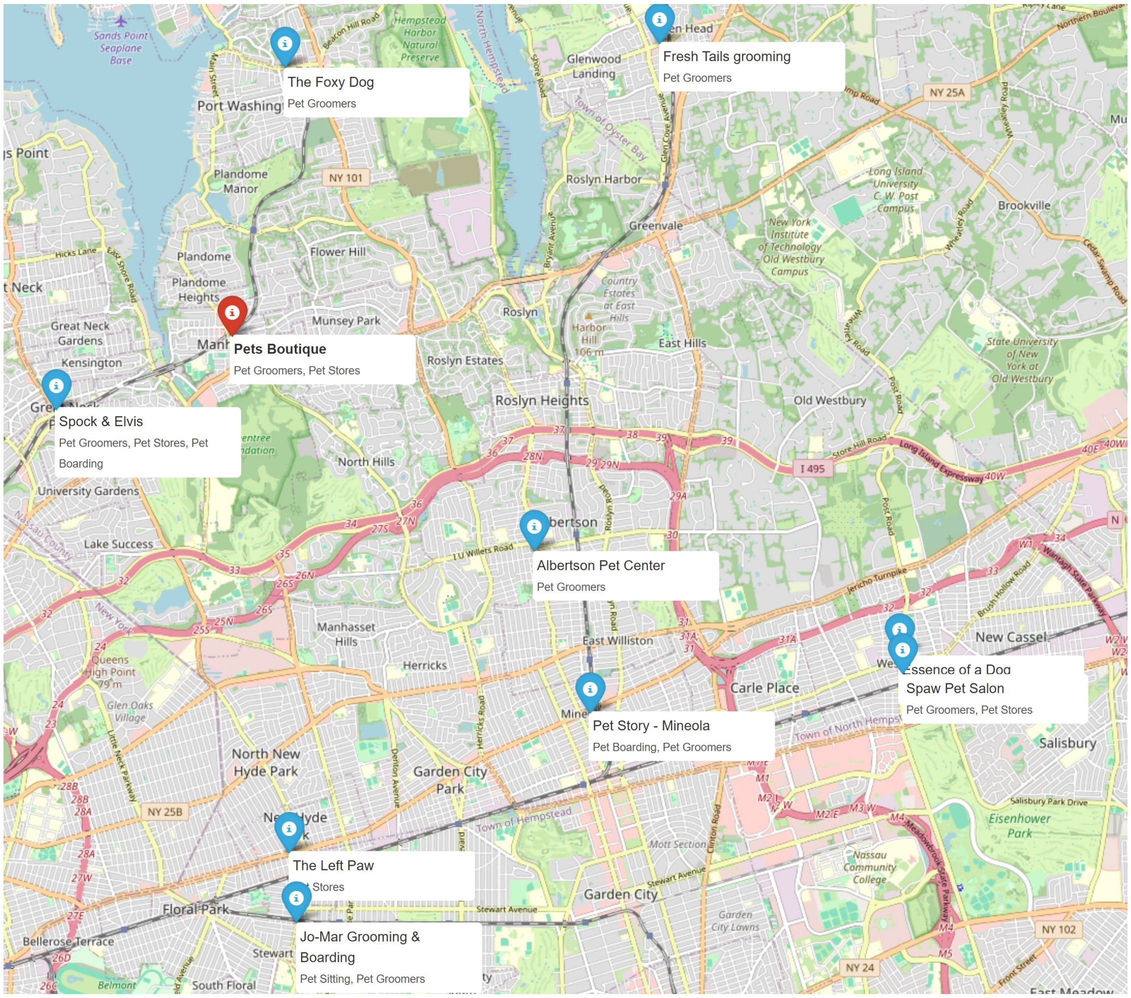
[30] The exact percentage change is $e^{\beta} - 1$.

B Variable Definitions

Variable	Definition
Firm Operating Outcomes	
Visits	The inverse hyperbolic sine (IHS) transformation of the number of visits to a firm in month t. (SafeGraph: Monthly Patterns - Foot Traffic)
Visitors	The IHS transformation of the number of distinct visitors to a firm in month t. Similarly, long-duration visitors are removed. (SafeGraph: Monthly Patterns - Foot Traffic)
Transactions	The IHS transformation of the number of total credit card transactions at a firm in month t (Consumer Edge, via Dewey)
Online Transactions	The IHS transformation of the number of online credit card transactions at a firm in month t (Consumer Edge, via Dewey)
Review Outcomes	
Avg. Rating	The average star rating of reviews received in month t, where 1 star indicates the lowest quality and 5 stars the highest
Missing Reviews	An indicator equal to one if the firm receives no reviews in month t
Filtered Reviews	The IHS transformation of the number of reviews classified as "Not Recommended" by the intermediary in month t
Removed Reviews	The IHS transformation of the number of reviews removed from the firm's page in month t
Key Identification Variables	
Flagged	An indicator equal to one if a Consumer Alert for "Compensated & Suspicious Review Activity" is issued to the firm in month t, and zero otherwise; in the cross-sectional determinants analysis, $Flagged_i$ is the firm-level analog, equal to one if the firm is ever flagged during the sample
Post	An indicator equal to one for the alert month and the three subsequent months following the issuance of a Consumer Alert to firm i, and zero otherwise (baseline three-month window)
Control Variables	
Positive Review Stock	Cumulative count (log(1 + count)) of four- and five-star reviews received by month $t-1$
Firm Age	Firm age (in logs), measured as the number of months from the firm's first intermediary review to month t
Days in Month	Logged calendar days in month t
Saturdays	Logged number of Saturdays in month t
Sundays	Logged number of Sundays in month t
Cooling Degree Days	Logged number of cooling degree days (cumulative degrees above 65°F) in month t (CustomWeather)
Precipitation	Logged total precipitation (inches) in month t (CustomWeather)
Same-Industry Traffic	Logged foot traffic to other firms in the same 3-digit NAICS industry within a 1-mile radius in month t

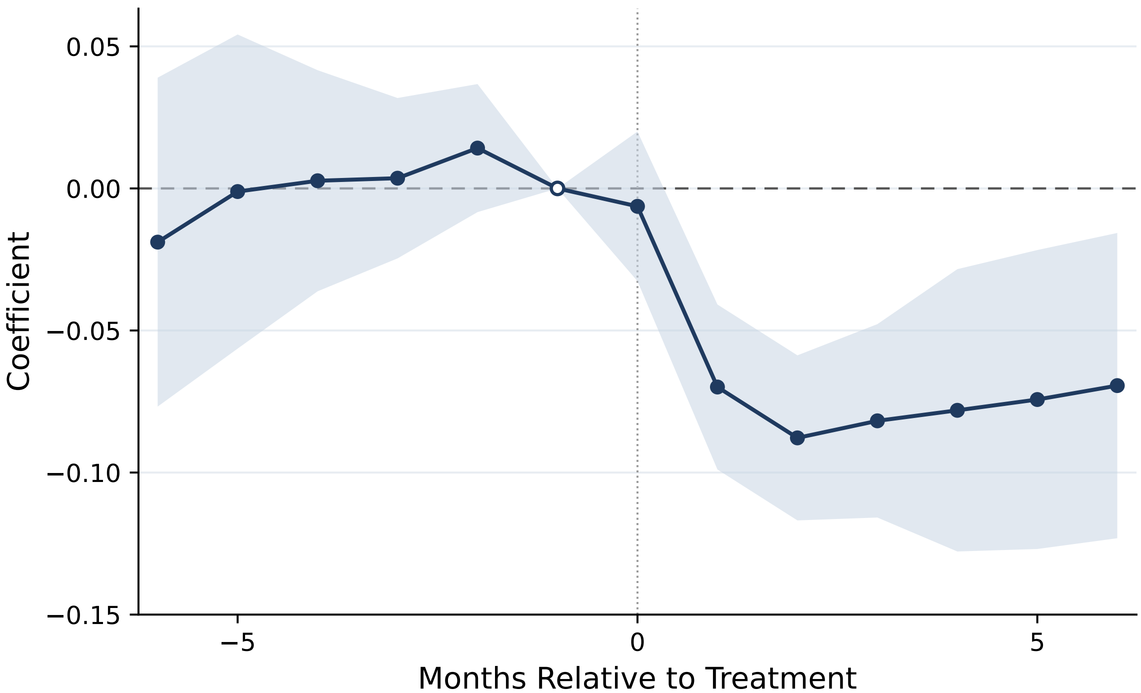
Variable	Definition
Cross-Industry Traffic	Logged foot traffic to firms in different 3-digit NAICS industries within a 1-mile radius in month t
App Usage	Firm-level proxy for intermediary penetration in the firm's local market, defined as the total foreground duration of the intermediary's mobile application aggregated across opt-in Android panelists in the firm's ZIP code, from the Global Wireless Solutions Magnify panel (via Dewey Data). Firms are split into "High App Usage" and "Low App Usage" subsamples at the sample median
Heterogeneity Split Variables	
Evidence vs. Algorithm Flag	Partition of alerts by the intermediary's own alert type: *Evidence Flagged* denotes *Compensated Activity Alerts*, issued on documentary evidence (e.g., screenshots of solicitations or payment offers) that a firm solicited or purchased reviews; *Algorithm Flagged* denotes *Suspicious Review Activity Alerts*, triggered by automated detection software flagging statistical manipulation patterns.
Holiday Season	Indicator equal to one for the months of November and December, and zero otherwise
Firm Age (split)	Firms (and, in Table 8, their demand-side peers) are classified by age as *Younger* (≤ 2 years) or *Older* (> 2 years)

Figure 1: Example of Flagged Firm (Treatment) and Demand-Side Peers (Control)



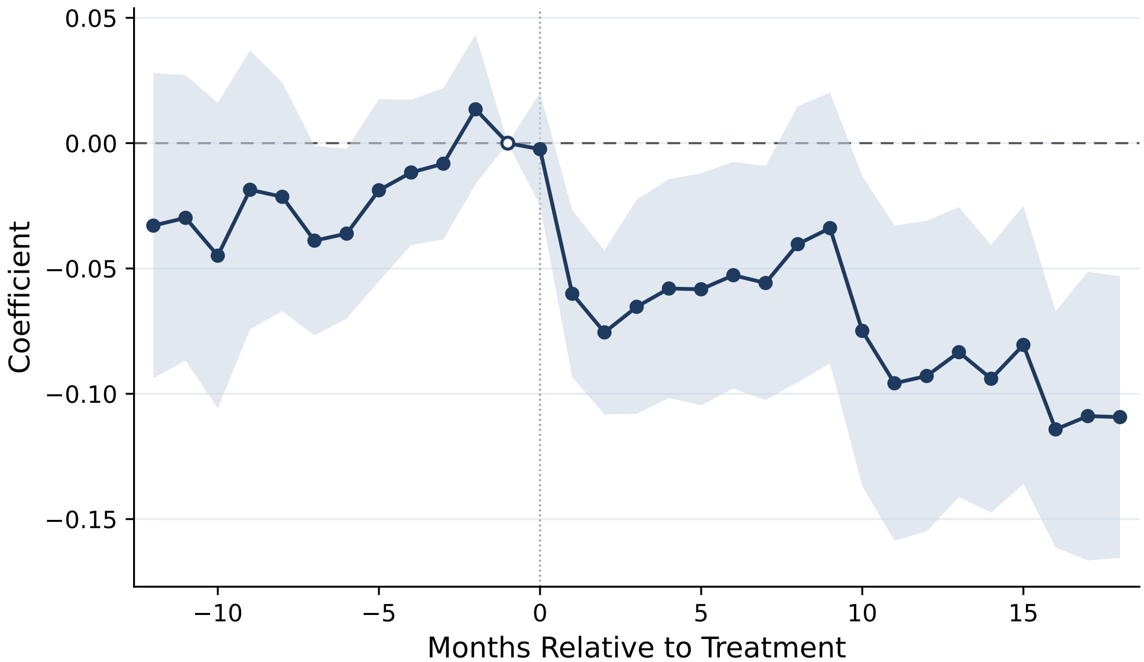
This figure illustrates the construction of the control group for a representative flagged firm in Nassau County, New York (Long Island). The red pin marks an flagged (treated) firm; blue pins mark the demand-side peer firms identified by the intermediary's "People Also Viewed" (PAV) recommendation algorithm, which selects peers from revealed co-viewing behavior rather than closest geographic proximity. The peers share the flagged firm's service category—here, pet grooming, boarding, and retail—yet are dispersed across the surrounding region, illustrating that demand-side comparability arises from substitutability in the eyes of visitors rather than from spatial adjacency. These peer firms constitute the comparison set in the difference-in-differences design. Map data © OpenStreetMap contributors.

Figure 2: Dynamic Effects of Reputation Manipulation Alerts on Foot Traffic



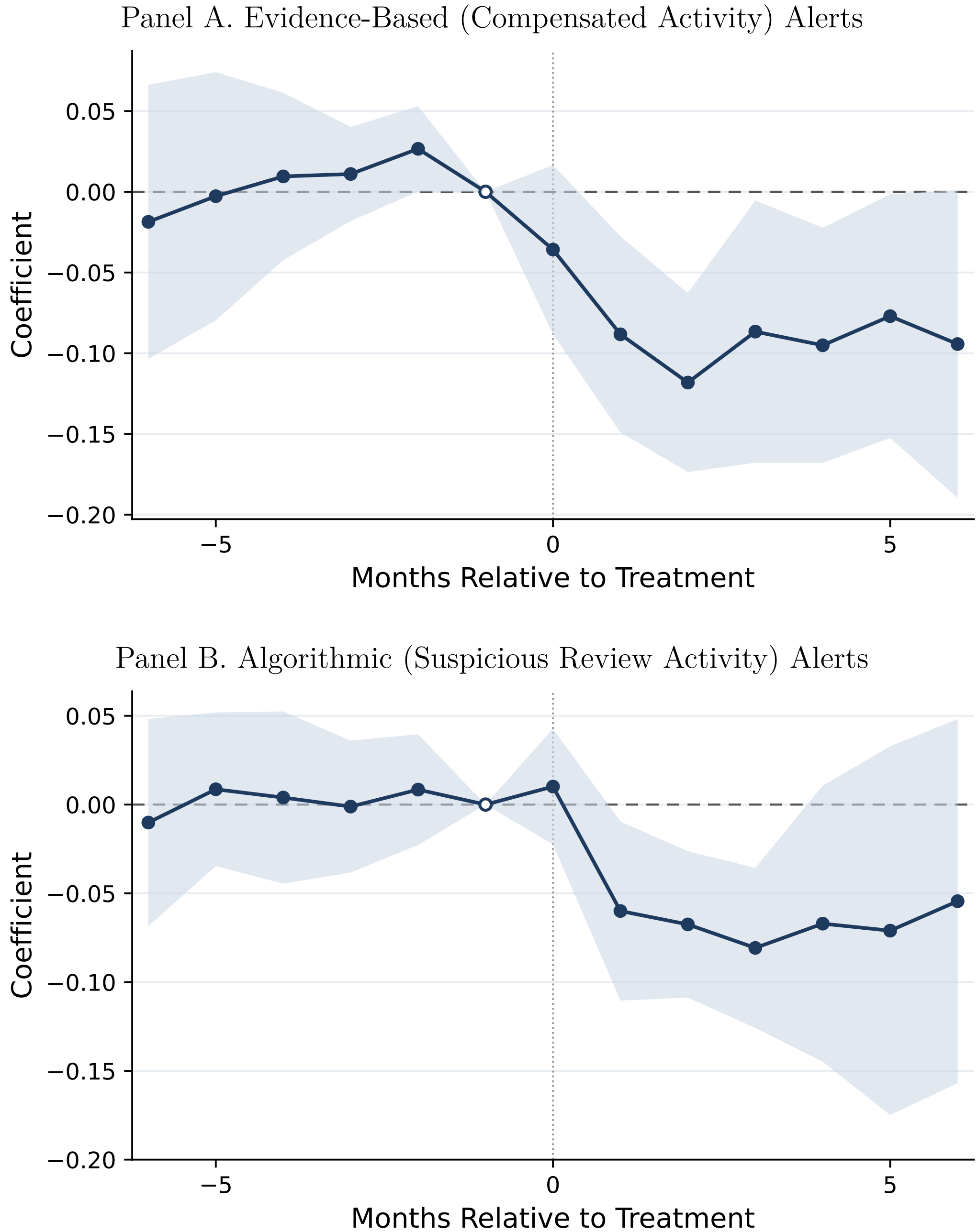
This figure plots 6-month pre and post coefficient estimates of Equation (9) using the Sun and Abraham (2021) interaction-weighted estimator, with 95% confidence intervals. The comparison group is never-treated PAV demand-side peer firms. Each estimate represents the cohort-share-weighted average treatment effect on the inverse hyperbolic sine of monthly foot traffic in the indicated event-month relative to the alert date. The open circle at $t = -1$ marks the omitted reference period. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Coefficients and t-statistics are reported in Appendix Table A.1.

Figure 3: Dynamic Effects of Reputation Manipulation Alerts on Foot Traffic — Extended Horizon



This figure plots estimates of Equation (9) using the Sun and Abraham (2021) interaction-weighted estimator over an extended $[-12, +18]$ event window, with 95% confidence intervals. The comparison group is never-treated PAV demand-side peer firms. Each estimate represents the cohort-share-weighted average treatment effect on the inverse hyperbolic sine of monthly foot traffic in the indicated event-month relative to the alert date. The open circle at $t = -1$ marks the omitted reference period. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Coefficients and t-statistics are reported in Appendix Table A.1.

Figure 4: Dynamic Effects of Credibility Shocks — Evidence-Based vs. Algorithmic Alerts



This figure plots estimates of Equation (9) using the Sun and Abraham (2021) interaction-weighted estimator over an extended $[-12, +18]$ event window, with 95% confidence intervals. Panel A restricts to firms receiving evidence-based (*Compensated Activity*) alerts; Panel B restricts to firms receiving algorithmic (*Suspicious Review Activity*) alerts. In both panels the comparison group is never-treated PAV demand-side peer firms, and the two panels share a common vertical scale. Each estimate represents the cohort-share-weighted average treatment effect on the inverse hyperbolic sine of monthly foot traffic in the indicated event-month relative to the alert date. The open circle at $t = -1$ marks the omitted reference period. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Coefficients and t-statistics are reported in Appendix Table A.2.

Table 2: Summary Statistics

Variable	N	Mean	SD	Min	p25	p50	p75	Max
Raw Visits	16,837	726.05	6,995.75	0	36	121	352	419,975
Raw Visitors	16,837	399.01	2,087.04	0	20	80	230	89,732
Visits	16,837	5.205	2.178	0	4.277	5.489	6.557	9.845
Visitors	16,837	4.675	2.250	0	3.690	5.075	6.131	9.349
Positive Review Stock	16,837	1.602	2.295	0	0	0	3.932	6.725
Firm Age	16,837	3.233	1.155	0	2.485	3.258	4.025	5.231
Days in Month	16,837	3.428	0.024	3.332	3.434	3.434	3.434	3.434
Saturdays	16,837	1.467	0.107	1.386	1.386	1.386	1.609	1.609
Sundays	16,837	1.462	0.106	1.386	1.386	1.386	1.609	1.609
Cooling Degree Days	16,837	0.911	0.910	0	0	0.606	1.763	2.792
Precipitation	16,837	4.174	2.177	0	3.021	5.068	5.779	6.972
Same-Industry Traffic	16,837	3.970	3.579	0	0	4.635	6.806	11.524
Cross-Industry Traffic	16,837	7.859	2.440	0	6.777	7.948	9.184	13.587

This table reports summary statistics for the estimation sample. The sample consists of firm–month observations for flagged firms and their demand-side peers. Demand outcomes are measured using GPS-based foot traffic data; raw visit and raw visitor counts are reported here in raw form for ease of interpretation, while along with the inverse hyperbolic sine transformation used in the analyses. Variable definitions are in Appendix B.

Table 3: Determinants of Review Manipulation

	Dependent Variable: *Flagged*		
VARIABLES	(1)	(2)	(3)
Positive Review Stock	0.0403***	0.0498***	0.0583***
	(9.050)	(15.951)	(14.937)
Firm Age	−0.0824**	−0.0858**	−0.0877***
	(−2.457)	(−2.637)	(−3.346)
Days in Month	−0.4112**	−0.6462***	−1.7119***
	(−2.227)	(−4.647)	(−3.280)
Saturdays	−0.0541	0.0031	0.1271*
	(−1.255)	(0.076)	(2.004)
Sundays	0.2668***	0.2326***	0.2600***
	(4.198)	(3.531)	(3.816)
Cooling Degree Days	−0.0001	−0.0016	−0.0090
	(−0.023)	(−0.271)	(−0.903)
Precipitation	0.0053***	0.0071***	−0.0055
	(5.398)	(5.247)	(−1.521)
Same-Industry Traffic	0.0001	0.0014	−0.0016
	(0.061)	(0.480)	(−0.236)
Cross-Industry Traffic	0.0092***	0.0076*	0.0249***
	(2.795)	(2.001)	(3.603)
Industry FE	No	Yes	Yes
City FE	No	No	Yes
Observations	2,391	2,391	2,391
Adj. R^2	0.0405	0.0463	0.0467

This table reports estimates from Equation (6), examining the determinants of reputational signal manipulation, *Flagged*. The dependent variable equals one if a firm receives an alert in a given month. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 4: The Demand Premium of Positive Reputation

VARIABLES	(1) *Visits*	(2) *Visitors*
Positive Review Stock	**0.0291*****	**0.0338*****
	(7.649)	**(10.772)**
Firm Age	0.0012	0.0083
	(0.044)	(0.361)
Cooling Degree Days	−0.0188*	−0.0116
	(−1.750)	(−1.298)
Precipitation	0.0002	0.0016
	(0.057)	(0.459)
Same-Industry Traffic	0.0406***	0.0403***
	(4.074)	(4.047)
Cross-Industry Traffic	0.1110***	0.1071***
	(5.125)	(4.847)
Firm, Year-Month FE	Yes	Yes
Observations	288,426	288,426
Adj. R^2	0.8169	0.7722

This table reports estimates examining the association between reputational capital and firm demand. The dependent variables are the inverse hyperbolic sine of monthly foot traffic (*Visits*, Column 1) and the inverse hyperbolic sine of monthly distinct visitors (*Visitors*, Column 2). The key explanatory variable is the cumulative stock of four- and five-star reviews (*Positive Review Stock*). All specifications include firm and month fixed effects and control for local competitive conditions and weather. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 5: Cost of Review Manipulation — DiD Estimation

	Dependent Variable: *Visits*		
VARIABLES	(1)	(2)	(3)
Flagged* × *Post	**−0.2706*****	**−0.1532****	**−0.0763*****
	(−4.314)	**(−2.002)**	**(−3.524)**
Post	0.0339	0.0004	0.0149
	(0.761)	(0.009)	(0.466)
Positive Review Stock	0.0452**	0.0711***	0.0191**
	(2.556)	(3.789)	(2.189)
Firm Age	0.1085***	0.0885**	−0.0173
	(2.873)	(2.540)	(−1.289)
Cooling Degree Days	0.2534***	−0.0144	−0.0043
	(4.199)	(−0.271)	(−0.181)
Precipitation	0.0512***	0.0153	0.0038
	(3.100)	(1.351)	(0.970)
Same-Industry Traffic	0.0460***	0.0508***	0.0087***
	(4.863)	(6.660)	(7.007)
Cross-Industry Traffic	0.1005***	0.0959***	0.0221***
	(6.025)	(5.650)	(3.788)
Firm FE	No	No	Yes
Year-Month FE	Yes	Yes	Yes
City FE	No	Yes	absorbed
Industry FE	No	Yes	absorbed
Observations	16,837	16,837	16,837
Adj. R^2	0.0683	0.4115	0.9382

This table reports difference-in-differences estimates from Equation (7) examining how exposure of fabricated reputational capital affects firm demand in the post-alert period. The treatment group consists of flagged firms; the control group consists of their demand-side peers. The dependent variable is the inverse hyperbolic sine of foot traffic (*Visits*). The key independent variable is *Flagged* × *Post*. All specifications include firm and month fixed effects, with additional city and industry fixed effects as indicated. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 6: Effects on the Information Environment

VARIABLES	(1) Avg. Rating	(2) Missing Rev.	(3) Filtered Rev.	(4) Removed Rev.
Flagged* × *Post	**−0.1900*****	**0.0546*****	**−0.5694*****	**−0.1001*****
	(−3.911)	**(3.078)**	**(−9.009)**	**(−7.117)**
Post	0.0137	0.0039	0.0098	−0.0059
	(0.253)	(0.653)	(0.877)	(−0.617)
Positive Review Stock	−0.0016	−0.1045***	0.0297***	0.0012
	(−0.161)	(−13.214)	(3.501)	(0.320)
Firm Age	−0.1374*	−0.2733***	0.0059	0.0000
	(−1.798)	(−7.742)	(0.502)	(0.009)
Cooling Degree Days	−0.0104	0.0017	0.0210*	0.0007
	(−0.319)	(0.113)	(1.757)	(0.092)
Precipitation	−0.0033	0.0045***	0.0022	−0.0012
	(−0.379)	(7.200)	(0.568)	(−1.339)
Same-Industry Traffic	0.0015	−0.0002	−0.0008	−0.0008
	(0.307)	(−0.351)	(−0.551)	(−1.030)
Cross-Industry Traffic	−0.0075	−0.0007	0.0014	0.0010
	(−0.761)	(−0.161)	(0.525)	(0.582)
Firm, Year-Month FE	Yes	Yes	Yes	Yes
Observations	7,873	16,837	16,837	16,837
Adj. R^2	0.2544	0.7059	0.5915	0.4000

This table reports estimates from Equation (7) examining how exposure of fabricated reputational capital affects the post-alert information environment of flagged firms. The dependent variables are the firm's monthly average star rating (Column 1, in raw stars). Column 2 is an indicator for no reviews in month t; and Columns 3–4 are the IHS counts of filtered and removed reviews. The key independent variable is *Flagged* × *Post*. All specifications include firm and month fixed effects. The sample size in Column (1) is smaller because *Avg. Rating* is undefined for firm-month observations with no reviews; Columns (2)–(4) use the full estimation sample. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 7: Cost of Review Manipulation — Cross-Sectional Heterogeneity Analysis

VARIABLES	(1) *Visits*	(2) *Visits*
Panel A: Signal Precision	**Evidence Flagged**	Algorithm Flagged
Flagged × *Post*	**−0.1120***** **(−5.823)**	−0.0545* (−1.695)
Observations	6,867	9,970
Adj. R^2	0.9479	0.9222
Panel B: Demand Composition	**Holiday Season**	Non-Holiday Season
Flagged × *Post*	**−0.2273***** **(−3.078)**	−0.0488* (−1.910)
Observations	2,973	13,864
Adj. R^2	0.8875	0.9421
Panel C: Market Penetration	**High App Usage**	Low App Usage
Flagged × *Post*	**−0.1151***** **(−3.053)**	−0.0333* (−1.965)
Observations	8,652	8,185
Adj. R^2	0.9286	0.9393
Panel D: Firm Age	Younger Firm	**Older Firm**
Flagged × *Post*	−0.0573*** (−3.541)	**−0.1009***** **(−3.166)**
Observations	8,149	8,688
Adj. R^2	0.9508	0.9075
Other Control Variables	Yes	Yes
Firm, Year-Month FE	Yes	Yes

This table reports cross-sectional heterogeneity analyses based on Equation (7). The dependent variable is the inverse hyperbolic sine of foot traffic (*Visits*). Each panel splits the sample along one of the four dimensions of intermediation intensity identified in H2. Panel A (signal precision) contrasts *Evidence Flagged* alerts (the intermediary's Compensated Activity Alerts, issued on documentary evidence that a firm solicited or purchased reviews) against Algorithm Flagged alerts (the intermediary's Suspicious Review Activity Alerts, triggered by automated detection of manipulation patterns). Panel B (demand composition) defines *Holiday Season* as the months of November and December and *Non-Holiday Season* as all other months. Panel C (market penetration) splits firms at the sample median of *App Usage*—the foreground duration of the intermediary's mobile application aggregated across opt-in panelists in the firm's ZIP code—into *High* and *Low* subsamples. Panel D (firm age) splits firms at two years from the first intermediary review into *Younger* (≤2 years) and *Older* (>2 years). All specifications include firm and month fixed effects. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 8: Demand Reallocation to Peer Firms

	Peer Positive Review Stock		Peer Firm Age	
	(1)	(2)	(3)	(4)
VARIABLES	Deep	Thin	Older	Younger
Flagged × *Post*	**0.0568*****	0.0403	**0.0730****	0.0331
	(3.197)	(1.624)	**(2.645)**	(1.659)
Post	0.0040	0.0162***	0.0101	0.0174***
	(0.210)	(2.744)	(0.688)	(3.937)
Positive Review Stock	−0.4794***	0.0146	0.0005	0.0527***
	(−4.007)	(1.143)	(0.090)	(3.700)
Firm Age	0.3511***	−0.0072	0.0298	−0.0361*
	(5.134)	(−0.770)	(1.299)	(−1.713)
Cooling Degree Days	−0.1178***	−0.0643***	−0.1216***	−0.0005
	(−3.664)	(−5.074)	(−6.043)	(−0.029)
Precipitation	0.0147***	0.0024	0.0058	0.0042**
	(4.681)	(1.091)	(1.240)	(2.063)
Same-Industry Traffic	0.0206***	0.0192***	0.0268***	0.0086***
	(9.195)	(6.030)	(5.211)	(3.029)
Cross-Industry Traffic	0.0972***	0.0703***	0.1059***	0.0286***
	(4.189)	(5.146)	(5.051)	(4.221)
Firm, Year-Month FE	Yes	Yes	Yes	Yes
Observations	54,120	207,665	143,385	118,400
Adj. R^2	0.8634	0.9101	0.8718	0.9518

This table reports estimates from Equation (7) examining demand reallocation from flagged firms to their demand-side ("People Also Viewed") peers following exposure. The dependent variable is the inverse hyperbolic sine of foot traffic (*Visits*) to peer firms; the key independent variable is *Flagged* × *Post* for the focal firm. Columns (1)–(2) split peers by *Positive Review Stock*—the peer's cumulative count of four- and five-star reviews—into Deep (>50) and Thin (≤50) reputation. Columns (3)–(4) split peers by *Firm Age* into Older (>2 years) and Younger (≤2 years), using the same two-year threshold as the firm-age panel of Table 7. All specifications include firm and month fixed effects. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 9: Corroborating Evidence from Transaction Data

VARIABLES	(1) Transactions	(2) Online Transactions
Flagged × Post	**−0.0526***** **(−2.959)**	**−0.0667*** **(−1.947)**
Post	0.0650 (1.446)	0.0076 (0.307)
Positive Review Stock	0.0286*** (6.599)	0.0156*** (9.767)
Firm Age	−0.0242** (−2.093)	−0.0239*** (−2.842)
Cooling Degree Days	0.0801** (2.237)	0.0246 (1.204)
Precipitation	0.0047 (1.348)	−0.0015 (−0.625)
Same-Industry Traffic	0.0007 (0.225)	−0.0001 (−0.098)
Cross-Industry Traffic	−0.0051* (−1.955)	−0.0048* (−1.809)
Firm FE	Yes	Yes
Year-Month FE	Yes	Yes
Observations	4,509	4,509
Adj. R^2	0.8741	0.8821

This table reports difference-in-differences estimates from Equation (7) using credit card transaction outcomes. The sample is restricted to 4,509 firm-month observations with available transaction data from Consumer Edge. The dependent variables are the inverse hyperbolic sine of the monthly count of total transactions (Column 1) and online transactions (Column 2). The key independent variable is *Flagged* × *Post*. All specifications include firm and month fixed effects. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B.

Table 10: Robustness: Covariate Balance and Entropy-Balanced Re-estimation

Panel A: Pre-treatment Covariate Balance

Variable	Control (Treat = 0)	Flagged (Treat = 1)	Difference (Control−Flagged)	t-stat	p-value
Positive Review Stock	1.332	2.030	−0.698***	−7.482	0.000
Firm Age	3.255	3.112	0.143***	2.882	0.004
Days in Month	3.428	3.428	0.000	−0.013	0.990
Saturdays	1.463	1.461	0.002	0.493	0.622
Sundays	1.456	1.453	0.003	0.680	0.496
Cooling Degree Days	0.926	0.924	0.002	0.055	0.956
Precipitation	3.957	4.039	−0.082	−0.870	0.384
Same-Industry Traffic	3.841	4.160	−0.319**	−2.142	0.032
Cross-Industry Traffic	7.722	8.046	−0.324***	−3.198	0.001

Panel B: Entropy-Balanced DiD Estimation

	Dependent Variable: *Visits*
VARIABLES	(3)
Flagged × Post	**−0.0772*** (−4.947)**
Controls	Yes
Firm FE	Yes
Year-Month FE	Yes
Entropy Weights	Yes
Observations	16,837
Adj. R^2	0.9358

This table reports two robustness exhibits addressing potential selection-on-observables concerns. Panel A reports pre-treatment means for the control group (PAV peers, *Treat = 0*) and treatment group (flagged firms, *Treat = 1*), the difference in means, and the t-statistic and p-value from a two-sample t-test. All continuous variables are reported in their estimation-form transformation (logged or IHS-transformed; see Appendix B). Panel B reports the baseline difference-in-differences specification from Equation (7) re-estimated using entropy-balanced weights (Hainmueller, 2012) that match the pre-treatment moments of flagged firms across the covariates documented in Panel A. The dependent variable is the inverse hyperbolic sine of foot traffic (*Visits*). The key independent variable is *Flagged × Post*. Controls include *Post*, *Positive Review Stock*, *Firm Age*, *Cooling Degree Days*, *Precipitation*, *Same-Industry Traffic*, and *Cross-Industry Traffic*. Firm and month fixed effects are also included. Standard errors are clustered two-way by industry and city (Cameron et al., 2011). Variable definitions are in Appendix B. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table A.1: Sun–Abraham Event-Study Estimates: Alternative Lead and Horizon Windows

Window	(1) [−6, +6]	(2) [−6, +12]	(3) [−6, +18]	(4) [−12, +6]	(5) [−12, +12]	(6) [−12, +18]
$t = -12$				−0.0289	−0.0381	−0.0329
				(−0.64)	(−1.08)	(−1.06)
$t = -11$				−0.0287	−0.0389	−0.0298
				(−0.71)	(−1.20)	(−1.03)
$t = -10$				−0.0364	−0.0487	−0.0449
				(−0.84)	(−1.36)	(−1.44)
$t = -9$				−0.0135	−0.0213	−0.0186
				(−0.41)	(−0.71)	(−0.66)
$t = -8$				−0.0162	−0.0279	−0.0214
				(−0.54)	(−1.06)	(−0.92)
$t = -7$				−0.0247	−0.0421*	−0.0389**
				(−0.91)	(−1.93)	(−2.02)
$t = -6$	−0.0189	−0.0384*	−0.0361*	−0.0242	−0.0388*	−0.0361**
	(−0.64)	(−1.72)	(−1.93)	(−0.89)	(−1.84)	(−2.09)
$t = -5$	−0.0011	−0.0225	−0.0190	−0.0040	−0.0221	−0.0188
	(−0.04)	(−1.04)	(−0.95)	(−0.16)	(−1.09)	(−1.01)
$t = -4$	0.0027	−0.0142	−0.0117	0.0012	−0.0137	−0.0117
	(0.14)	(−0.91)	(−0.81)	(0.07)	(−0.88)	(−0.79)
$t = -3$	0.0036	−0.0096	−0.0086	0.0011	−0.0099	−0.0082
	(0.25)	(−0.63)	(−0.54)	(0.09)	(−0.69)	(−0.53)
$t = -2$	0.0142	0.0105	0.0145	0.0112	0.0090	0.0135
	(1.24)	(0.78)	(0.97)	(0.93)	(0.66)	(0.89)
$t = +0$	−0.0063	−0.0018	−0.0013	−0.0080	−0.0034	−0.0024
	(−0.47)	(−0.16)	(−0.12)	(−0.60)	(−0.29)	(−0.21)
$t = +1$	−0.0699***	−0.0662***	−0.0591***	−0.0717***	−0.0680***	−0.0601***
	(−4.72)	(−4.08)	(−3.51)	(−4.50)	(−3.92)	(−3.53)
$t = +2$	−0.0878***	−0.0793***	−0.0734***	−0.0893***	−0.0824***	−0.0755***
	(−5.92)	(−4.80)	(−4.17)	(−6.03)	(−5.27)	(−4.53)
$t = +3$	−0.0818***	−0.0688***	−0.0640***	−0.0818***	−0.0715***	−0.0653***
	(−4.71)	(−3.76)	(−2.83)	(−4.49)	(−3.91)	(−2.99)
$t = +4$	−0.0781***	−0.0668***	−0.0569**	−0.0788***	−0.0704***	−0.0580***
	(−3.08)	(−3.04)	(−2.29)	(−3.67)	(−3.58)	(−2.60)
$t = +5$	−0.0743***	−0.0634***	−0.0564**	−0.0744***	−0.0676***	−0.0583**
	(−2.77)	(−2.66)	(−2.15)	(−3.16)	(−3.22)	(−2.47)
$t = +6$	−0.0694**	−0.0567**	−0.0505**	−0.0714***	−0.0624***	−0.0527**
	(−2.53)	(−2.34)	(−1.97)	(−2.87)	(−2.80)	(−2.29)
$t = +7$		−0.0655**	−0.0539*		−0.0706***	−0.0558**
		(−2.38)	(−1.95)		(−2.92)	(−2.34)
$t = +8$		−0.0481	−0.0379		−0.0532*	−0.0403
		(−1.56)	(−1.21)		(−1.93)	(−1.44)
$t = +9$		−0.0409	−0.0313		−0.0475*	−0.0339
		(−1.36)	(−0.99)		(−1.79)	(−1.23)
$t = +10$		−0.0852**	−0.0719**		−0.0922***	−0.0749**
		(−2.57)	(−2.07)		(−2.92)	(−2.37)
$t = +11$		−0.1070***	−0.0941***		−0.1124***	−0.0958***
		(−3.68)	(−2.89)		(−3.65)	(−2.99)
$t = +12$		−0.1005***	−0.0899***		−0.1086***	−0.0929***
		(−3.66)	(−2.79)		(−3.67)	(−2.94)
$t = +13$			−0.0801**			−0.0834***
			(−2.44)			(−2.83)
$t = +14$			−0.0895***			−0.0940***
			(−3.04)			(−3.45)
$t = +15$			−0.0764**			−0.0805***
			(−2.37)			(−2.84)
$t = +16$			−0.1077***			−0.1142***
			(−3.82)			(−4.76)
$t = +17$			−0.1034***			−0.1089***
			(−3.00)			(−3.71)
$t = +18$			−0.1014***			−0.1093***
			(−2.92)			(−3.81)
Observations	31,198	48,487	66,943	42,079	59,368	77,824

This table reports Sun and Abraham (2021) interaction-weighted event-study estimates of Equation (9) for all combinations of six- and twelve-month lead windows and six-, twelve-, and eighteen-month event horizons. The dependent variable is the inverse hyperbolic sine of monthly foot traffic; the comparison group is never-treated PAV demand-side peer firms; $t = -1$ is the omitted reference period. t-statistics, in parentheses, are based on standard errors clustered two-way by NAICS3 industry and city (Cameron et al., 2011). Blank cells indicate event-months outside the estimation window. Column (1) corresponds to Figure 2 and column (6) to Figure 3. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (two-tailed).

Table A.2: Subsample Event-Study Estimates: Evidence-Based vs. Algorithmic Alerts

	Evidence-Based		Algorithmic	
Event Time	Coefficient	(t)	Coefficient	(t)
$t = -6$	-0.0186	(-0.43)	-0.0101	(-0.34)
$t = -5$	-0.0027	(-0.07)	0.0086	(0.39)
$t = -4$	0.0095	(0.36)	0.0040	(0.16)
$t = -3$	0.0110	(0.74)	-0.0011	(-0.06)
$t = -2$	0.0266^{**}	(1.97)	0.0084	(0.53)
$t = -1$	—	—	—	—
$t = 0$	-0.0358	(-1.34)	0.0101	(0.61)
$t = +1$	-0.0883^{***}	(-2.85)	-0.0600^{**}	(-2.33)
$t = +2$	-0.1182^{***}	(-4.18)	-0.0675^{***}	(-3.21)
$t = +3$	-0.0866^{**}	(-2.09)	-0.0807^{***}	(-3.51)
$t = +4$	-0.0951^{**}	(-2.56)	-0.0671^{*}	(-1.69)
$t = +5$	-0.0770^{**}	(-2.00)	-0.0710	(-1.34)
$t = +6$	-0.0943^{*}	(-1.94)	-0.0544	(-1.04)

This table reports the Sun and Abraham (2021) interaction-weighted event-study coefficients plotted in Figure 4, estimated separately for evidence-based (*Compensated Activity*, $N =$ 12,763) and algorithmic (*Suspicious Review Activity*, $N =$ 18,435) alerts. The dependent variable is the inverse hyperbolic sine of monthly foot traffic; the comparison group is never-treated PAV demand-side peer firms; $t = -1$ is the omitted reference period. t-statistics, in parentheses, are based on standard errors clustered two-way by NAICS3 industry and city (Cameron et al., 2011). Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (two-tailed).