Peer Effect and Online Ratings in Hotel Consumption: Evidence from China

Xiangzhe Xu[1]

School of Economics and Management, Tsinghua University

February 2026

[1]School of Economics and Management, Tsinghua University. Email: xuxz23@mails.tsinghua.edu.cn; xiangzhexu096@gmail.com. I thank advisors for helpful comments. All errors are my own. This research received no specific grant from any funding agency. During preparation I used generative AI tools to assist with language and readability. I reviewed and edited all content and take full responsibility for this work.

Abstract

I use check-in records from a large star-rated hotel in China to provide causal evidence on peer effects in hotel consumption. Exploiting institutionally arranged stays by government delegations and enterprise training cohorts as exogenous shocks, I find that the monthly first-visit hazard among residential compound neighbors doubles in months 7 to 12 after the shock and is statistically zero in the first six months, consistent with peer information arriving before the travel opportunity it eventually triggers. The effect is geographically contained within compound boundaries and drops discontinuously to zero at compound walls. The peak-window peer effect is equivalent in demand magnitude to a 0.11-point improvement in the hotel's online rating. Interacting online rating signals with the event-study windows reveals that a higher own-hotel rating amplifies peak-window peer conversion while a higher competitor rating suppresses it, with both effects concentrated precisely in the months when peer information is most potent.

Keywords: Peer effects, word-of-mouth, online reputation, hazard model, experience goods

JEL Classification: L83, D83, C41, M31

1 Introduction

Consumers learning about experience goods face a fundamental information problem: quality cannot be verified before purchase. Two channels have emerged to help resolve this uncertainty. The first is offline peer information, the firsthand accounts of neighbors and colleagues who have already consumed the good (Cai et al., 2009; Dahl et al., 2014; Bailey et al., 2022). The second is online reputation, the aggregated ratings that platforms broadcast to anonymous audiences (Anderson and Magruder, 2012; Luca, 2016; Cho et al., 2024; Seiler et al., 2017). Each channel independently shapes demand, and each has attracted a substantial causal literature. Yet the interaction between them has not been studied. Whether a neighbor's private recommendation and a hotel's public rating reinforce or undercut each other, and whether a rival's strong online reputation can neutralize the demand a hotel generates through offline word-of-mouth, are questions that prior work has not answered. This paper provides the first causal evidence on both. The core finding is that the social multiplier of offline peer effects is endogenous to the online reputation environment: private social signals and public platform ratings are structural complements in consumer quality learning.

Establishing this requires first measuring the peer effect itself. In high-frequency markets, peer information translates into behavior within days or weeks and standard event-study windows capture the effect (Moretti, 2011; Gilchrist and Sands, 2016; Jin et al., 2025; Lalive and Cattaneo, 2009). Hotel stays are different. Consumption opportunities are constrained by travel schedules rather than immediate availability, so peer information arrives well before the purchase occasion it might eventually trigger. Standard frameworks will record a statistically zero effect for the first several months and miss the bulk of the cumulative impact.[2]

Identifying the peer effect, and then asking how the rating environment moderates it, requires a design that can trace the full dynamic path of social information transmission over years rather

[2]Longer-run peer effects in consumption and adoption decisions have been documented in solar panels (Bollinger and Gillingham, 2012), parental leave take-up (Dahl et al., 2014), electric vehicle adoption (Tebbe, 2025), and consumption network spillovers (De Giorgi et al., 2016). The present paper contributes by tracing the full dynamic path in a low-frequency experience goods market where consumption opportunities are externally constrained.

than weeks.

I use check-in records from a star-rated hotel in Changsha, Hunan Province, covering 86 months from December 2018 to January 2026 and comprising over 200,000 guests with permanent registered addresses. The identification strategy exploits institutionally arranged stays, specifically government delegations, enterprise training cohorts, and conference participants, whose timing is set by administrative calendars with no connection to the preferences or residential locations of local consumers. I define the earliest such exogenous stay among a focal individual's residential compound neighbors as an individual-level shock and combine it with a discrete-time hazard model to trace the causal dynamic path of peer effects over 86 months. The design directly addresses the three canonical identification threats, homophily, common shocks, and reverse causality (Manski, 1993; Bramöulle et al., 2009), and is validated by a series of geographic placebo tests.

The baseline results establish that peer effects in this market are delayed, non-monotonic, and geographically precise. The monthly first-visit hazard shows no significant response in the first six months after the exogenous neighbor stay, then doubles relative to the pre-event baseline in months 7 to 12, consistent with peer information arriving before the travel opportunity it eventually triggers. A second elevation emerges in year five, consistent with a second wave of within-compound transmission as initially converted neighbors become information sources themselves. The effect drops discontinuously to zero at compound walls. Assigning the shock to a different compound on the same administrative street produces a precisely zero post-period estimate, ruling out neighborhood-level common shocks and identifying the gated residential compound, rather than the postal code or administrative district, as the natural unit of word-of-mouth diffusion in Chinese cities.

With the peer effect established, I first quantify its magnitude relative to a familiar unit of account. The peak-window peer effect (AME $= 0.0061$ percentage points) divided by the direct demand effect of a one-point improvement in the hotel's own online rating (AME $= 0.057$ percentage points) yields a ratio of approximately 0.11: one exogenous neighbor stay generates a demand increment equivalent to a 0.11-point improvement in the hotel's Dianping score. This benchmark

anchors the economic significance of social information in this market.

I then ask whether the online reputation environment moderates the strength of this peer effect. I interact the event-study windows with the focal hotel's own Dianping rating and a composite competitor rating index. Two interpretations are possible. If ratings operate through a general demand-level channel, interaction terms should be uniformly significant across all post-shock windows. If ratings instead function as a verification mechanism for the private social signal, their moderating effect should concentrate precisely in the peak window where peer information is most potent, and be absent elsewhere. The data support the second interpretation. A higher own-hotel rating amplifies peak-window peer conversion, with the own-rating interactions jointly significant and the effect concentrated in months 7 to 12, roughly three times the baseline peer effect itself. A higher competitor rating suppresses peer conversion in exactly the same window and is absent in all others. Ratings do not uniformly shift the conversion hazard; they selectively alter the informational value of the private social signal at the moment it is most potent.

First, within the literature on peer effects in consumer markets (Kuhn et al., 2011; Bollinger and Gillingham, 2012; Gilchrist and Sands, 2016; Dahl et al., 2014; Bailey et al., 2022; Tebbe, 2025), this paper establishes the full temporal dynamics of peer effects in a low-frequency experiential goods market using clean causal identification. Because the residential compound contains no natural control group of guaranteed non-visitors, as any neighbor may eventually visit and the exact number of non-visitors is unknown, the relevant outcome is the timing of first visits rather than a binary adoption indicator, motivating a discrete-time hazard model. Standard short-window frameworks record a zero effect for the first six months and miss the majority of the cumulative impact. The combination of the earliest exogenous neighbor stay with a discrete-time hazard model offers a replicable design for studying peer effects wherever consumption opportunities are externally constrained.

Second, this paper advances the literature on quality signals and consumer demand (Chevalier and Mayzlin, 2006; Anderson and Magruder, 2012; Luca, 2016; Mayzlin et al., 2014; Cho et al., 2024; Ameri et al., 2019). Within a single information environment, word-of-mouth and observa-

tional learning provide distinct rather than redundant information (Ameri et al., 2019). This paper shows that across the offline-online divide, private social signals and public platform ratings are not merely distinct but structurally complementary. The social multiplier of word-of-mouth is endogenous to the online reputation environment: the marginal value of a reputation investment is not fixed but depends on the peer information environment in which consumers are embedded, and vice versa.

Third, the competitive suppression result contributes to the literature on competitive spillovers in demand (Bailey et al., 2022; Cho et al., 2024). Prior work has shown that competitor reviews reduce booking performance through direct demand diversion. This paper identifies an additional channel: a rival's higher rating attenuates the social multiplier through which the focal hotel's current guests generate future guests, but only at the moment when peer information is actively shaping decisions. This imposes a second-order competitive externality through signal substitution rather than quantity diversion—one that is invisible to analyses focusing on aggregate demand and detectable only by tracing the full dynamic path of peer-effect transmission. A firm's reputation asset thus serves two functions: a direct demand effect and a role in sustaining the word-of-mouth multiplier that compounds over successive transmission cycles.

The remainder of the paper is organized as follows. Section 2 describes the data and sample construction. Section 3 presents the identification strategy and estimate model. Section 4 establishes the baseline peer effect results and robustness checks. Section 5 maps the geographic boundaries of transmission. Section 6 estimates the moderating effects of online ratings. Section 7 concludes.

2 Data

2.1 Institutional Setting

The research site is a five-star hotel in central Changsha, Hunan Province, affiliated with a provincial state-owned enterprise and established in the 1990s. The hotel operates nearly 400 rooms

alongside conference facilities, dining, and fitness amenities. Rack rates of 398–698 RMB per night place it in the mid-to-high segment, ruling out status-signaling as a demand driver. Business-related bookings account for 65.4% of stays, with corporate meetings alone comprising 37.6%, confirming that demand is largely occasion-triggered by external institutional calendars.

The identification strategy exploits three types of institutional stays whose timing is set entirely outside the local residential economy. *First*, inter-provincial government delegations attending policy briefings or regulatory inspections book through centralized government procurement, with scheduling governed by the administrative calendars of out-of-province agencies. *Second*, employees dispatched to mandatory enterprise training programs are assigned to the hotel by their employers' human-resources departments, typically months in advance and with no role for individual participants in venue selection. *Third*, national and international conference delegations arrive as a consequence of institutional affiliation and organizing-committee contracts, not personal preference. Across all three types, the defining feature is that scheduling rests with external institutions whose logic is orthogonal to the preferences and residential locations of local consumers.

2.2 Hotel Records and Sample Construction

Data. The hotel's property management system (PMS) contains complete guest records for all arrivals between December 2018 and January 2026. The full database comprises 222,838 stay records from 206,515 individual guests. Each record includes check-in date, a group-booking flag distinguishing institutionally arranged stays from voluntary bookings, national identification number and registered address. Merging stay-level and guest-profile records via fuzzy matching on name and identification number yields the linked panel used throughout the analysis. Externally arranged stays account for 15,452 records (6.9% of the total); the remaining 207,386 are voluntary individual bookings.

Residential compounds and neighbor groups. The permanent registered address links each guest to a specific residential compound (*xiaoqu*). Chinese urban residential compounds are

gated communities enclosed by walls, typically comprising multiple buildings sharing common entrances, gardens, and amenities; daily interactions among residents are largely confined within these boundaries. This physical structure constitutes the natural unit within which word-of-mouth transmission operates, a claim confirmed by the spatial discontinuity tests in Section 5. Chinese hotel regulations require guests to present their national identity card, which carries the household registration address. I geocode all addresses to the compound level and define each individual's neighbor group as all other guests sharing the same compound, excluding historical co-travelers.

Sample construction. Starting from 206,515 individuals in the raw PMS database, I first apply a preprocessing pipeline that retains only individuals for whom a valid exogenous peer-group shock can be defined: those with at least one identifiable compound neighbor in the hotel database whose earliest recorded stay is an exogenous institutional visit (arranged by out-of-province government delegations, large enterprise training cohorts, or national/international conference committees). This yields a first-stage candidate pool of 4,882 individuals across 2,321 residential compounds (Table 1).

I then apply three additional restrictions motivated by identification concerns. I exclude 1,337 rural registered-address residents, for whom compound-level neighbor groups cannot be meaningfully defined; this subsample also exhibits significant pre-trends ($\hat{\beta}_{[-24,-13]} = 1.003$, $p = 0.062$). I further exclude 18 individuals whose address corresponds to a government administrative compound, whose hotel exposure may propagate through occupational rather than residential channels. Finally, I exclude 889 Hunan Province residents: the hotel's institutional visibility within Hunan, via state-owned-enterprise networks, generates pre-existing information exposure that violates the no-anticipation condition, as confirmed by a significant pre-trend in this subsample ($\hat{\beta}_{[-24,-13]} = 0.51$, $p = 0.02$). The resulting main analysis sample comprises 2,638 individuals.

Retained individuals fall into three mutually exclusive groups. Ω_A (*pre-shock converters*): 531 individuals (20.1%) who completed a voluntary stay before the compound's shock date S_c. Ω_B (*post-shock converters*): 1,057 individuals (40.1%) whose first voluntary stay occurs on or

after S_c; they are the primary source of identification. Ω_C (*right-censored non-converters*): 1,050 individuals (39.8%) who record no completed voluntary stay through January 2026, including individuals who held reservations but did not complete check-in (cancellations or no-shows) as well as those who never initiated a booking during the observation window. Summary statistics are reported in Table 2.

Table 1: Sample Construction Funnel

	Individuals	Cumulative %
Raw PMS database	206,515	100.0%
Less: Preprocessing pipeline	−201,633	97.6%
First-stage candidate pool	**4,882**	
Less: Rural household registration	−1,337	2.4%
Less: Government compound address	−18	1.7%
Less: Hunan Province residents	−889	1.7%
Main analysis sample	**2,638**	**1.3%**

Notes: The preprocessing pipeline processes all 206,515 raw PMS records and retains individuals for whom the earliest compound neighbor stay is an exogenous institutional visit. The combined pipeline exclusion is 201,633. Subsequent restrictions address identification threats within the first-stage candidate pool. Cumulative percentages are relative to the raw PMS count of 206,515.

Retained individuals fall into three mutually exclusive groups. Ω_A (*pre-shock converters*): 531 individuals (20.1%) who completed a voluntary stay before the compound's shock date S_c; they contribute to the pre-period baseline. Ω_B (*post-shock converters*): 1,057 individuals (40.1%) whose first voluntary stay occurs on or after S_c; they are the primary source of identification. Ω_C (*right-censored non-converters*): 1,050 individuals (39.8%) who record no voluntary stay through January 2026. Summary statistics are reported in Table 2.

The shock-to-conversion gap for Ω_B averages 31.64 months (median 27 months), foreshadowing the hazard model results: most individuals convert well outside a standard short event window, underscoring why a multi-year event-study framework is necessary. Province-of-origin statistics confirm geographic distance from Changsha (mean 959.82 km), with Beijing (20.9%), Guangdong (9.7%), and Sichuan (7.5%) as the three largest source provinces, ruling out independent prior local

Table 2: Summary Statistics

	Mean	SD	Min	Max
Panel A: Individual level ($N = 2{,}638$)				
Distance to hotel (km)	959.82	422.38	142.46	2,987.56
Compound peer group size	7.34	12.72	1	65
Male	58.6% (1,248 of 2,130 with non-missing gender)			
Panel B: Sample partition				
Ω_A: pre-shock converters	531 (20.1%)			
Ω_B: post-shock converters	1,057 (40.1%)			
Ω_C: right-censored	1,050 (39.8%)			
Panel C: Shock-to-conversion gap, Ω_B only ($N = 1{,}057$)				
Months from shock to first endogenous stay	31.64	22.93	0	82
Median	27 months			
Panel D: Individual-month level ($N = 171{,}215$)				
First endogenous visit (Y_{it})	0.0093	0.096	0	1
Relative time τ_{it} (months)	-1.61	35.39	-84	84
Panel E: Province of origin (top 10, individual level)				
Beijing (11)	550 (20.9%)			
Guangdong (44)	255 (9.7%)			
Sichuan (51)	196 (7.5%)			
Zhejiang (33)	140 (5.3%)			
Hubei (42)	138 (5.2%)			
Shaanxi (61)	118 (4.5%)			
Guangxi (45)	104 (4.0%)			
Jiangsu (32)	100 (3.8%)			
Henan (41)	98 (3.7%)			
Shandong (37)	94 (3.6%)			
Other (20 provinces)	437 (16.6%)			

Notes: Main analysis sample, defined as individuals satisfying all regression conditions: urban household registration, non-government-compound address, registered outside Hunan Province, and earliest compound neighbor stay being an exogenous institutional visit (see Section 2.2). Distance is the straight-line distance from the registered address compound centroid to the hotel. Ω_A: completed a voluntary stay before the shock month S_c. Ω_B: first voluntary stay occurs on or after S_c; primary source of identification. Ω_C: no voluntary stay recorded through January 2026; enter as right-censored observations. The shock-to-conversion gap in Panel C is computed as the number of months between S_c and the individual's first voluntary stay; the median of 27 months is consistent with the hazard model peak at months 7–12, as most individuals convert outside the peak window as travel opportunities arise. Gender is missing for 508 individuals (19.3%).

exposure.

Figure 1 maps the geographic origins of endogenous and exogenous stays. Institutional visitors (orange) originate from across the country, while endogenous residential visitors (blue) concentrate in central Changsha, confirming that institutional scheduling is orthogonal to local residential demand patterns.

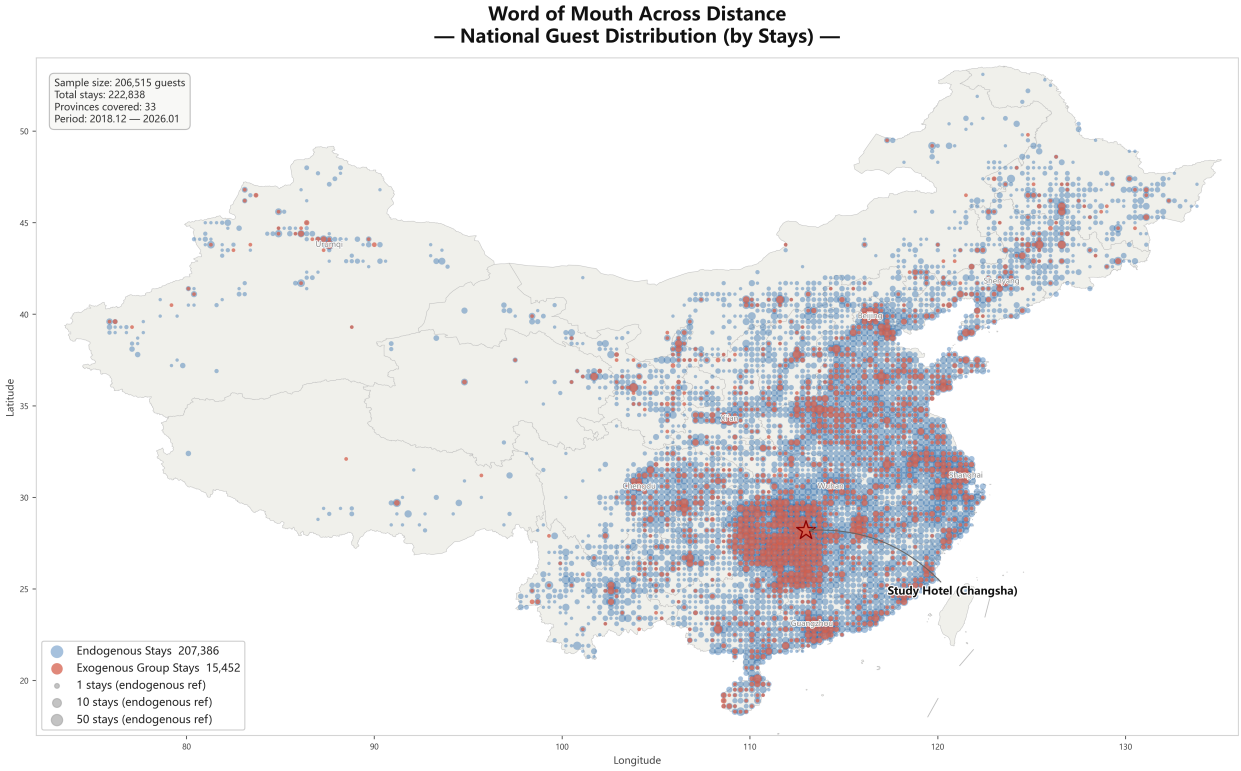


Figure 1: Geographic distribution of hotel guests, December 2018–January 2026. Each bubble represents a $0.3° \times 0.3°$ grid cell scaled by number of stays. Blue: endogenous voluntary visits; orange: exogenous institutional visits. Star marks the hotel location (Changsha).

2.3 Online Rating Data

Online rating data are drawn from Dianping, China's leading consumer review platform for hospitality and dining. I collect monthly scores for the focal hotel and five comparable five-star properties in Changsha that constitute its primary competitive set. Competitor ratings are aggregated into a single monthly index using a six-month backward rolling average, smoothing transitory fluctuations while retaining medium-run variation in competitive quality signals. Both the own-hotel

rating and the competitor index are matched to the individual-month panel by calendar month. The construction of rating-interaction terms and the moderating-effect specifications are presented in Section 6.

3 Empirical Strategy

3.1 Identification Challenges

Identifying peer effects in consumption requires confronting three canonical threats. The first is *correlated unobservables*: residents of the same compound share unobserved characteristics such as income trajectories, leisure preferences and employer affiliation that independently predict hotel consumption, so co-movement in visit timing may reflect taste similarity rather than information transmission (Manski, 1993). The second is the *reflection problem*: visit timing among neighbors is mutually determined, and without an exogenous shifter that breaks this simultaneity the causal direction cannot be recovered. The third is *contextual effects*: common shocks like a shared employer announcement or a corporate discount arrangement could simultaneously raise all residents' visit hazards without any peer-to-peer transmission occurring.

3.2 A Two-Part Identification Design

I address all three threats through a two-part design: an exogenous source of variation that breaks the reflection problem, and a sample construction rule that eliminates contamination from pre-existing endogenous information flows within the compound.

Part 1: Exogenous shock timing. For each residential compound c, the shock date S_c is the calendar month in which the first institutional visitor from that compound's neighbor group arrives at the hotel:

$$S_c = \min_{j \in \hat{N}(i)} \left\{ T_j^{(1)} : T_j^{(1)} \text{ is an exogenous visit} \right\}. \tag{1}$$

Because out-of-province government agencies, enterprise training departments, and conference organizing committees set schedules according to administrative calendars with no connection to the preferences or incomes of any particular urban compound, the timing of S_c is, conditional on calendar month fixed effects, orthogonal to residents' baseline conversion propensities. This exogeneity addresses all three identification threats at once: correlated unobservables cannot bias the estimates because external institutions schedule without reference to any compound's characteristics; the reflection problem is resolved because S_c is determined by out-of-province bodies rather than any local resident's decision, so causality runs unambiguously from the institutional visitor to the focal individual; and contextual effects are absorbed by calendar month fixed effects, which purge any aggregate demand fluctuation common to all compounds. The joint insignificance of pre-period hazard coefficients confirms the no-anticipation condition, ruling out the possibility that residents systematically adjust visit behavior before the shock arrives. Two residual concerns, that unobserved demand shifters specific to a neighborhood or to a compound might nonetheless correlate with S_c, are addressed by supplementary evidence reported later. The neighborhood-level concern is mitigated by the placebo test in Section 5, which assigns the shock to a different compound within the same residential community and finds no post-period effect. The compound-level concern is addressed in robustness checks that saturate the model with compound-by-year fixed effects, leaving the baseline estimates unchanged.

Part 2: Eliminating pre-existing endogenous contamination. The exogenous shock provides a valid first information source, but a second problem remains. Within a given compound, endogenous voluntary stays may have already occurred before the first institutional visit. If resident A stayed voluntarily before institutional visitor B arrived, any subsequent visit by A's neighbors cannot be attributed solely to B, since it may also reflect prior voluntary information transmission by A.

The retention rule is: *individual i enters the analysis sample if and only if the first recorded stay among all of individual i's compound neighbors is an exogenous institutional visit.* The restriction

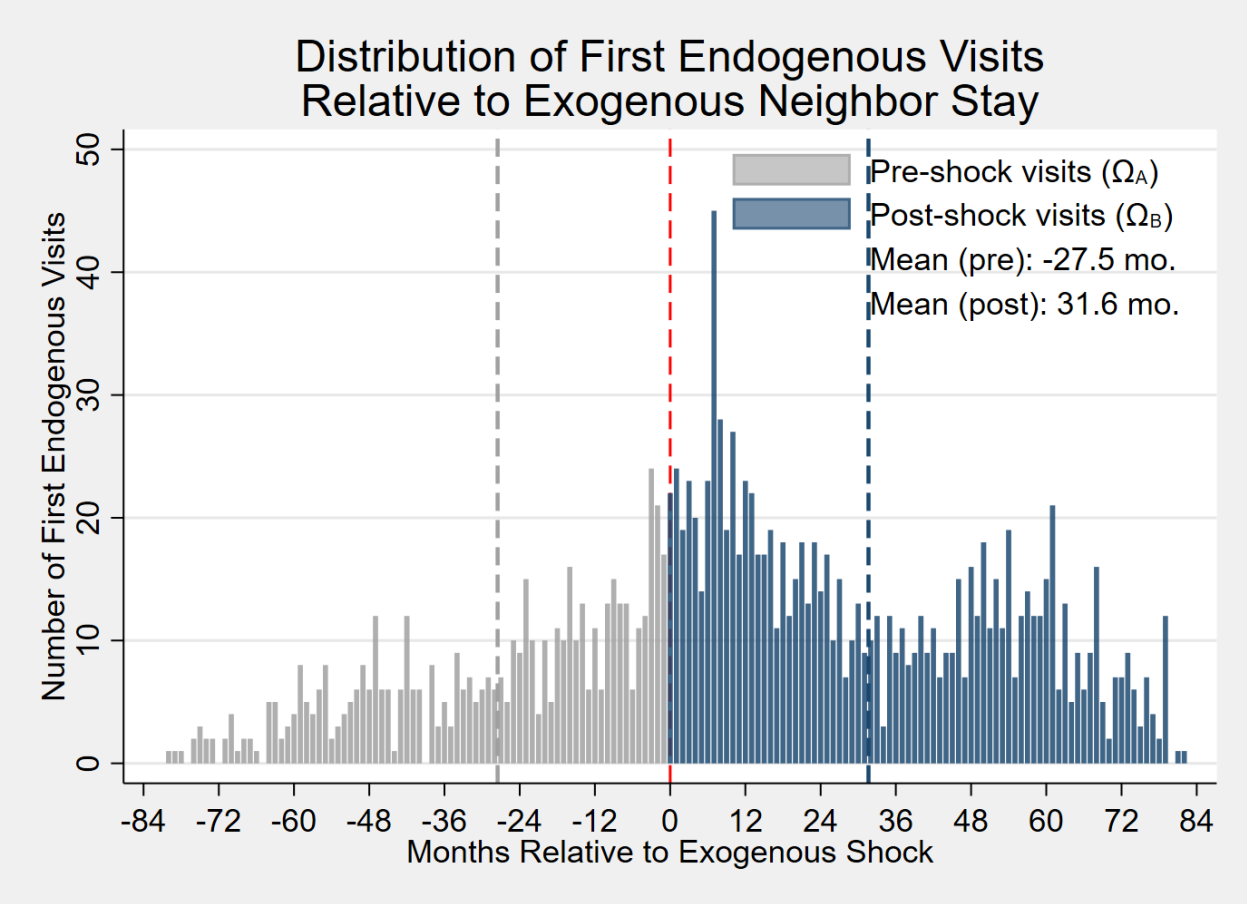


Figure 2: Distribution of first endogenous visits by relative time. Horizontal axis: months relative to the exogenous neighbor visit ($\tau = 0$). Vertical axis: number of first endogenous visits. Gray bars: pre-event stays (Ω_A); dark bars: post-event stays (Ω_B). Solid vertical lines indicate the mean relative time for each group. Red dashed line marks the event time ($\tau = 0$).

applies to the neighbor set $\hat{N}(i)$, not to the focal individual's own visit history; an individual who completed a voluntary stay before S_c may therefore still satisfy the retention rule, provided none of their neighbors did so earlier. Only under this condition can we be confident that the focal individual's first exposure to peer information about the hotel arrives through an institutionally determined, causally clean channel. To fix ideas, consider two compounds. In Compound 1, resident A completes a voluntary stay at time t_1, well before the exogenous institutional visit by D at time t_4 ($= S_{c1}$); resident C has a cancelled booking but no completed stay. Because A's voluntary visit is the earliest neighbor record for B, C, and D, all three fail the retention rule and are excluded. A herself is also excluded: she is the source of endogenous contamination, not a treated individual. In Compound 2, the first neighbor to visit is F, who arrives via an institutional channel at time t_2 ($= S_{c2}$). The earliest neighbor stay for E is F's exogenous visit; E satisfies the retention rule and is retained. For G, the earliest neighbor stay is E's subsequent voluntary visit at time t_5; G fails the rule and is excluded.

Retained individuals are partitioned into three mutually exclusive groups based on the timing of their own voluntary stay relative to S_c:

- Ω_A (*pre-shock converters*, $T_i^{\text{endog}} < S_c$): individuals who completed a voluntary stay before the shock. They are unaffected by peer information from the institutional visit and contribute to the pre-period baseline.

- Ω_B (*post-shock converters*, $T_i^{\text{endog}} \geq S_c$): individuals whose first voluntary stay occurs on or after the shock month. They are the primary source of identification.

- Ω_C (*right-censored non-converters*, $T_i^{\text{endog}} = \infty$): individuals who record no voluntary stay through January 2026. They remain in the risk set and enter the likelihood as right-censored observations.

Figure 2 displays the empirical distribution of first endogenous stays by month relative to the shock ($\tau = 0$). Pre-shock converters (Ω_A, gray bars) are roughly uniformly distributed across the pre-period, consistent with the no-anticipation assumption; post-shock converters (Ω_B, dark bars)

concentrate at positive relative times with a visible mass in the 7–15 month range, providing the first visual evidence of a delayed peer effect.

3.3 Identification Assumptions

Assumption 1 (No Anticipation)**.** *For all $t < S_c$, individual i's baseline conversion hazard is unaffected by the realized value of S_c.*

No-anticipation requires that residents do not adjust hotel visit behavior in anticipation of a future neighbor's institutional stay. This is guaranteed by design: S_c is determined by out-of-province administrative calendars that are unknowable to local residents in advance. The assumption has a directly testable implication that pre-period hazard coefficients should be jointly insignificant, which Section 4 confirms (joint $p = 0.82$).

Assumption 2 (Conditional Independence)**.** *Conditional on covariates X_i and calendar month fixed effects δ_t: $S_c \perp\!\!\!\perp T_i^* \mid X_i, \delta_t$.*

Conditional independence requires that, after conditioning on observables and calendar time, the shock month is unrelated to when the focal individual would have visited absent the shock. Because S_c is the minimum of visit dates each determined by out-of-province bodies independent of any resident's conversion propensity, the minimum inherits that independence (Appendix 7, Proposition A.1).

One subtlety deserves attention. Conditional independence holds for the full retained sample Ω but is mechanically violated within Ω_A and Ω_B individually, because subsample membership imposes restrictions on the joint distribution of S_c and T_i^*. Appendix 7 shows this does not introduce bias: Ω_A contributes zero to the score; Ω_B's score has zero expectation by a martingale argument; Ω_C exerts a conservative downward pull. Reported estimates are therefore lower bounds on the true complier effect.

3.4 Estimating Equation

Specification. Most individuals in the sample visit the hotel at most once over 86 months, and many never visit at all. The outcome of interest—the timing of the first voluntary stay is a discrete-time duration variable with substantial right-censoring. A standard difference-in-differences regression conflates "not yet visited" with "will never visit," producing biased estimates of the monthly conversion probability. A hazard model is the natural framework: it conditions on the risk set in each period, directly models the monthly conversion probability, and handles right-censored observations through partial likelihood.

I estimate a discrete-time complementary log-log (cloglog) hazard model:

$$\log\big(-\log(1-\lambda_{it})\big) = \alpha(t) + \boldsymbol{\beta}'\mathbf{D}_{it} + \gamma X_i + \delta_t, \tag{2}$$

where $\lambda_{it} = \Pr(T_i = t \mid T_i \geq t)$ is the monthly conditional conversion hazard; $\alpha(t)$ is a flexible nonparametric baseline hazard estimated via duration dummies; $\mathbf{D}_{it}$ is a vector of event-study bin indicators; X_i is straight-line distance from the guest's registered compound to the hotel; and δ_t are calendar month fixed effects absorbing aggregate demand fluctuations. Standard errors are clustered at the compound level. The cloglog link is preferred over logit as the natural discrete-time analog of the continuous-time proportional hazards model, ensuring consistent estimates as the time interval shrinks (Jenkins, 1995).

Event-study bins.

$$\begin{aligned}\mathbf{D}_{it} = \Big(&\mathbf{1}[\tau_{it} \leq -25],\ \mathbf{1}[\tau_{it} \in [-24,-13]],\ \mathbf{1}[\tau_{it} \in [-12,-7]],\ \mathbf{1}[\tau_{it} \in [0,2]],\ \mathbf{1}[\tau_{it} \in [3,6]],\\ &\mathbf{1}[\tau_{it} \in [7,12]],\ \mathbf{1}[\tau_{it} \in [13,24]],\ \mathbf{1}[\tau_{it} \in [25,48]],\ \mathbf{1}[\tau_{it} \in [49,60]],\ \mathbf{1}[\tau_{it} \geq 61]\Big)',\end{aligned} \tag{3}$$

with $\tau_{it} \in [-6,-1]$ as the omitted reference period. Pre-period bins ($\tau \leq -7$) serve as falsification tests for the no-anticipation assumption. Post-period bins trace the full dynamic path of peer effects

over five years following the shock, with bin widths chosen to balance statistical power against temporal granularity.

Average marginal effects. I report average marginal effects (AMEs) rather than raw coefficients to facilitate economic interpretation:

$$\widehat{\mathrm{AME}}_j = \frac{1}{N}\sum_{i=1}^{N}\big[\hat{\lambda}_{it}(\mathbf{D}_{it} = \mathbf{e}_j) - \hat{\lambda}_{it}(\mathbf{D}_{it} = \mathbf{0})\big], \tag{4}$$

where $\mathbf{e}_j$ is the j-th unit vector. AMEs are expressed in percentage points of monthly conversion probability and are directly comparable across specifications and subsamples.

3.5 Estimand

The identified parameter is a local average treatment effect (LATE): the causal effect of a neighbor's exogenous institutional visit on the monthly conversion hazard, for the subpopulation of motivated potential consumers in the risk set at the time of the shock, but have not yet completed a voluntary stay. The qualitative findings should generalize to other low-frequency experience goods that share persistent quality uncertainty, social signal dependence, and geographic concentration of peer interactions.

4 Baseline Results

4.1 Event Study

Figure 3 plots the raw empirical monthly hazard of a first endogenous stay before any covariate adjustment. The pre-event mean monthly hazard is 0.56% and the post-event mean is 1.44%. The post-event hazard exhibits a clear hump-shaped pattern, rising slowly in the first few months after the shock, peaking in months 7–12, and gradually declining thereafter. This descriptive pattern motivates the event-study bins used in the regression analysis.

Table 3 reports the complementary log-log coefficients and average marginal effects (AMEs) from the baseline specification. The sample comprises 163,794 individual-month observations, 1,584 first-visit events, and 1,305 compound clusters.

Pre-period. The three pre-period bin coefficients are 0.057, 0.116, and 0.104, all individually insignificant. The joint F-test yields $p = 0.82$, supporting Assumption 1 and confirming that exogenous shock timing is uncorrelated with residents' baseline conversion trajectories.

Post-period. The immediate post-shock bins—$\tau \in [0, 2]$ (AME $= 0.23$ pp, $p = 0.233$) and $\tau \in [3, 6]$ (AME $= 0.26$ pp, $p = 0.118$)—are positive but statistically insignificant. The hazard peaks sharply at $\tau \in [7, 12]$: complementary log-log coefficient 0.630^{***}, AME $= 0.61$ percentage points ($p < 0.001$). The pre-shock baseline monthly hazard is 0.61%,[3] so the peer effect doubles the baseline monthly conversion rate during this window. The effect remains marginally significant at $\tau \in [13, 24]$ (AME $= 0.24$ pp, $p = 0.088$) before declining.

Figure 4 plots the full dynamic profile with 95% confidence intervals. This delayed peak reflects a fundamental constraint of low-frequency markets: peer information arrives first, but conversion must await the next travel opportunity, generating a natural lag between information receipt and observable behavior. Standard evaluation frameworks with observation windows of a few months will record a statistically zero effect and substantially understate the true demand impact of peer information in this setting.

[3] In the subsample with $\tau < 0$, the mean of Y is 0.0061, based on 87,038 individual-month observations.

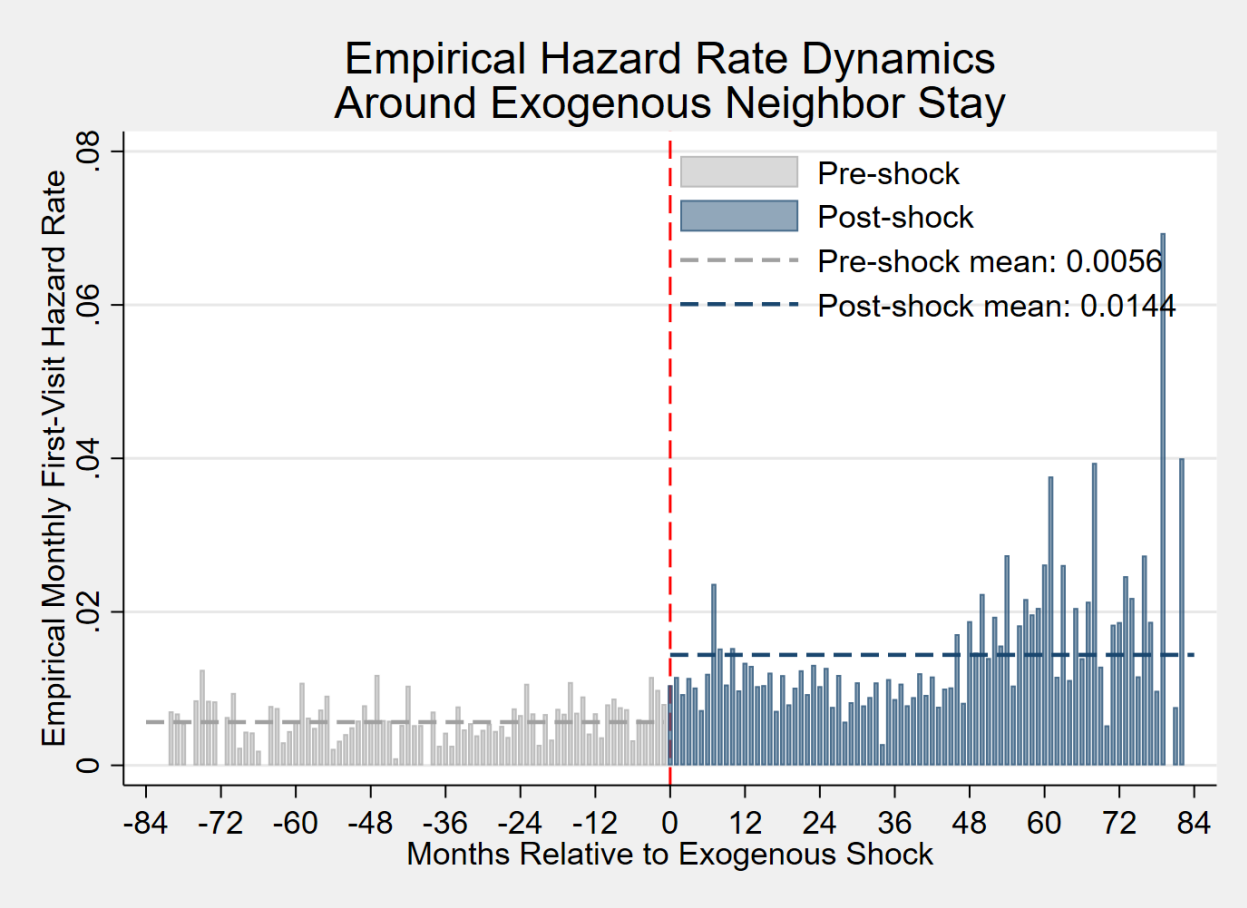


Figure 3: Empirical monthly hazard of a first endogenous stay. Horizontal axis: months relative to the exogenous neighbor visit ($\tau = 0$). Vertical axis: proportion of individuals still at risk who experience their first stay in that month. Gray bars: pre-event; dark bars: post-event. Horizontal dashed lines indicate the pre-event mean (0.56%) and post-event mean (1.44%). Red dashed line marks the event time ($\tau = 0$).

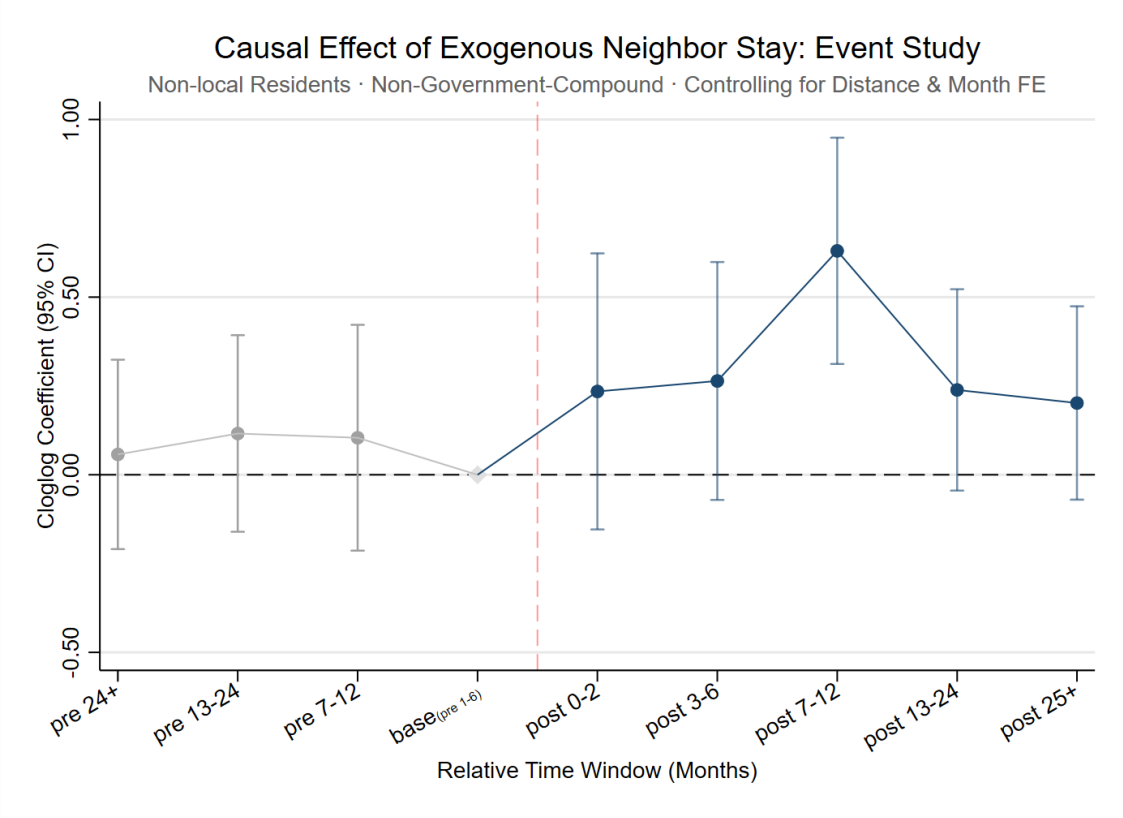


Figure 4: Event study: average marginal effects of a neighbor's exogenous visit on the monthly first-visit hazard. Horizontal axis: months relative to the exogenous shock ($\tau = 0$). Vertical bars: 95% confidence intervals (compound-clustered SE). Reference period: $\tau \in [-6, -1]$. The profile peaks at $\tau \in [7, 12]$ (AME $= 0.61$ pp, $p < 0.001$).

4.2 Long-Run Profile and Secondary Peak

Table 4 extends the profile to 84 months using 12-month bins for $\tau \geq 25$. The main peak at $\tau \in [7, 12]$ is confirmed (coefficient 0.632^{***}, AME $= 0.61$ pp). The intervening bins at $\tau \in [25, 36]$ (AME $= 0.13$ pp, $p = 0.369$) and $\tau \in [37, 48]$ (AME $= 0.10$ pp, $p = 0.527$) are small and insignificant, indicating that the elevated hazard from the first peak has substantially dissipated by year three. A secondary increase then emerges at $\tau \in [49, 60]$ (AME $= 0.33$ pp, $p = 0.050$), roughly half the magnitude of the primary peak and marginally significant. Figure 5 plots the full 84-month profile.

The non-monotonic pattern with a primary peak at months 7 to 12, a trough in years 3 to 4, and a

Table 3: Baseline Event Study: Complementary Log-Log Hazard Model

	Coefficient		AME (pp)	
Event Window	Coef.	SE	dy/dx	SE
Pre-period				
$\tau \leq -25$	0.057	0.136	0.047	0.130
$\tau \in [-24, -13]$	0.116	0.141	0.110	0.135
$\tau \in [-12, -7]$	0.104	0.162	0.100	0.156
Post-period				
$\tau \in [0, 2]$	0.235	0.198	0.228	0.191
$\tau \in [3, 6]$	0.264	0.171	0.257	0.165
$\tau \in [7, 12]$	0.630^{***}	0.162	0.610^{***}	0.159
$\tau \in [13, 24]$	0.239^{*}	0.145	0.238^{*}	0.139
$\tau \geq 25$	0.202	0.139	0.196	0.135
Observations		163,794		
Events		1,584		
Clusters (compound)		1,305		
Pre-period joint p		0.82		

Notes: Dependent variable: indicator for individual i's first endogenous hotel visit in month t. Complementary log-log hazard model with calendar month fixed effects and straight-line distance control. Reference period: $\tau \in [-6, -1]$. AMEs expressed in percentage points. Standard errors clustered at compound level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

secondary elevation in year 5 is consistent with a two-wave transmission process in which individuals converted by the original shock subsequently pass information to other neighbors, triggering a further round of conversions approximately four years later. This interpretation is corroborated by the summary statistics in Table 2: the median shock-to-conversion gap for Ω_B is 27 months, implying that converted individuals would themselves become information sources around years 4 to 5. The secondary peak should nonetheless be treated as suggestive given its marginal significance and the absence of direct evidence on second-wave transmission paths.

The positive AMEs across virtually all post-shock bins indicate that standard evaluation frameworks, which typically observe behavior over a few months, will substantially understate the total demand impact of peer information. The baseline monthly conversion probability is 0.61%; the peer effect doubles this rate during the primary peak window and maintains an elevated hazard, at varying intensity, for the full 84-month horizon.

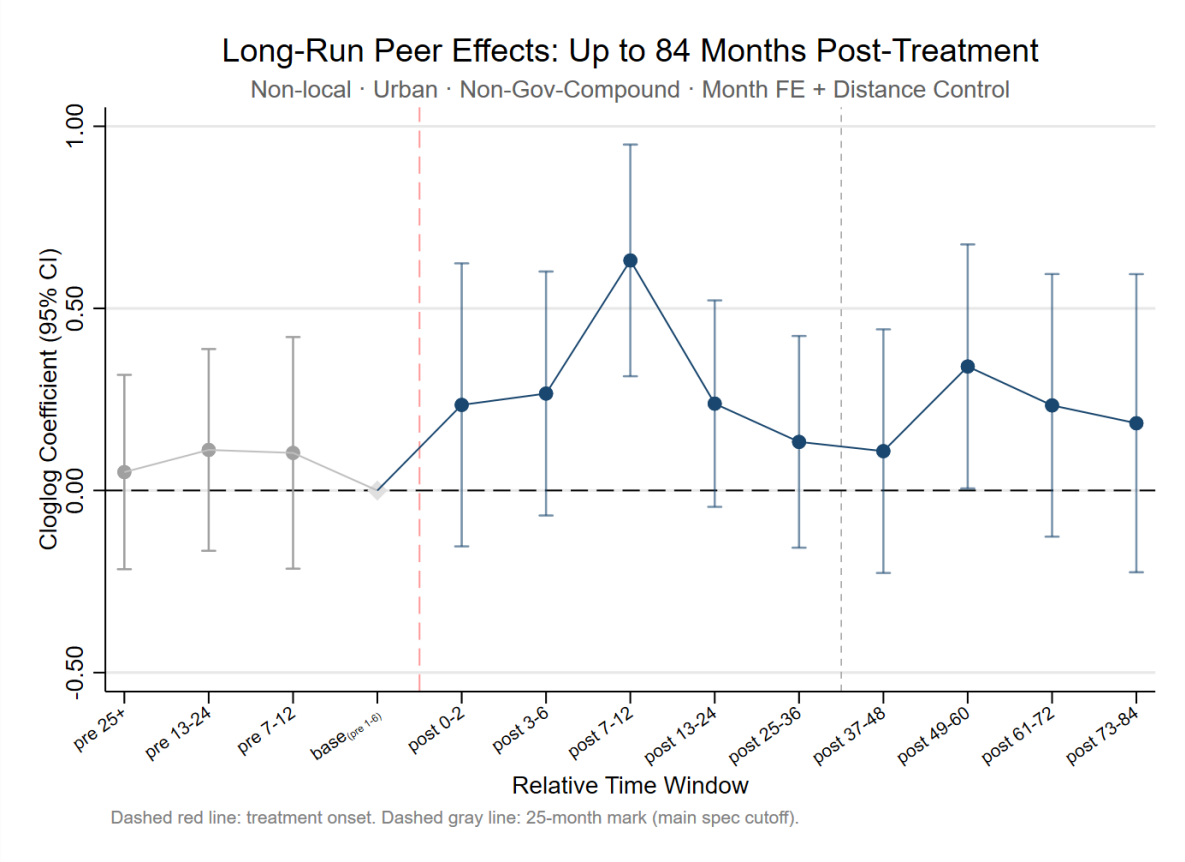


Figure 5: Long-run event study: complementary log-log coefficients extending to 84 months post-shock, with 12-month bins for $\tau \geq 25$. The primary peak at $\tau \in [7, 12]$ and the secondary increase at $\tau \in [49, 60]$ ($p = 0.050$) are both visible. Gray dashed line: $\tau = 25$ cutoff used in baseline binning.

4.3 Robustness

The peak-window estimate is robust across functional form, fixed effect structure, and unobserved heterogeneity specifications. Table 5 reports five specifications alongside the baseline.

Random effects. Estimating a random-effects complementary log-log model that allows for compound-level unobserved heterogeneity yields a peak-window coefficient of 0.631^{***}, virtually identical to the baseline 0.630^{***}. The near-zero estimated variance of the random effect confirms that compound-level unobservables do not materially confound the event-study profile.

Alternative link function. Replacing the complementary log-log link with a logit link leaves the $\tau \in [7, 12]$ coefficient unchanged at 0.635^{***}, and the full dynamic profile is identical to the

Table 4: Long-Run Event Study: Complementary Log-Log Hazard Model

	Coefficient		AME (pp)	
Event Window	Coef.	SE	dy/dx	SE
Pre-period				
$\tau \leq -25$	0.051	0.136	0.049	0.130
$\tau \in [-24, -13]$	0.111	0.141	0.107	0.135
$\tau \in [-12, -7]$	0.103	0.162	0.099	0.156
Post-period				
$\tau \in [0, 2]$	0.235	0.198	0.226	0.190
$\tau \in [3, 6]$	0.266	0.171	0.255	0.165
$\tau \in [7, 12]$	0.632^{***}	0.162	0.606^{***}	0.159
$\tau \in [13, 24]$	0.239	0.145	0.229	0.139
$\tau \in [25, 36]$	0.134	0.148	0.128	0.143
$\tau \in [37, 48]$	0.108	0.171	0.104	0.164
$\tau \in [49, 60]$	0.341^{**}	0.171	0.327^{**}	0.167
$\tau \in [61, 72]$	0.234	0.184	0.224	0.178
$\tau \in [73, 84]$	0.185	0.209	0.177	0.201
Observations	163,794			
Events	1,584			
Clusters (compound)	1,305			
Pre-period joint p	0.82			

Notes: Same specification as Table 3 with 12-month bins replacing the single $\tau \geq 25$ indicator. Standard errors clustered at compound level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

baseline.

Control variables. Table 6 sequentially adds cumulative neighbor visits, compound group size, Dianping rating variables, and the Baidu search index to the baseline specification. The $\tau \in [7, 12]$ coefficient ranges from 0.617 to 0.650 across all four specifications, within 5% of the baseline value in every case. The Baidu search index is individually insignificant and leaves the peak-window estimate unchanged, ruling out local search-interest trends as a confound. The positive and significant coefficient on cumulative neighbor visits is consistent with information accumulation within the compound but does not alter the event-study pattern.

County $\times$ month fixed effects. To absorb county-level demand shocks that may co-vary with institutional visit timing, I estimate a linear probability model with county-by-month fixed effects (logit is infeasible at this dimension). Only the peak-window coefficient remains significant (0.0067^{***}, $p = 0.004$); all other post-period bins are indistinguishable from zero. This confirms that the peak-window effect survives stringent geographic-time controls, while estimates for other windows should be interpreted with caution under this demanding specification.

Compound $\times$ year linear trend. Saturating the model with compound-specific linear time trends absorbs slow-moving compound-level confounders. Under this specification, all post-period bins are jointly and individually significant, with coefficients ranging from 0.007^{***} to 0.013^{***}. This pattern should be interpreted carefully: the compound-specific trend may also partial out genuine long-run peer effects that evolve smoothly over time, so the uniformity of significance across windows does not straightforwardly translate into a uniform causal effect in each period. What the result does establish is that the post-shock increase in conversion hazard is real and not driven by compound-level trends; the non-monotonic long-run profile documented in Section 4.2, and in particular the secondary peak at $\tau \in [49, 60]$, should nonetheless be treated as suggestive rather than definitive.

Ruling out coincidental business travel. A potential concern is that the peak-window effect reflects coincidental business trips to Changsha rather than peer information transmission: the shock month may simply correlate with periods of elevated institutional travel demand that

Table 5: Robustness: Alternative Specifications

	(1) Baseline Cloglog	(2) RE Cloglog	(3) Logit	(4) County×Ym LPM	(5) Cmpd×Yr LPM
Pre-period					
$\tau \leq -25$	0.057	0.057	0.058	-0.000	-0.002
	(0.136)	(0.121)	(0.137)	(0.001)	(0.001)
$\tau \in [-24, -13]$	0.116	0.116	0.117	0.001	-0.000
	(0.141)	(0.133)	(0.142)	(0.001)	(0.001)
$\tau \in [-12, -7]$	0.104	0.105	0.105	-0.000	0.001
	(0.162)	(0.157)	(0.163)	(0.002)	(0.001)
Post-period					
$\tau \in [0, 2]$	0.235	0.234	0.236	0.001	0.007^{***}
	(0.198)	(0.161)	(0.200)	(0.003)	(0.002)
$\tau \in [3, 6]$	0.264	0.264^{*}	0.267	0.000	0.008^{***}
	(0.171)	(0.152)	(0.172)	(0.002)	(0.001)
$\tau \in [7, 12]$	0.630^{***}	0.631^{***}	0.635^{***}	0.007^{***}	0.013^{***}
	(0.162)	(0.131)	(0.164)	(0.002)	(0.002)
$\tau \in [13, 24]$	0.239^{*}	0.241^{*}	0.240^{*}	-0.001	0.010^{***}
	(0.145)	(0.128)	(0.146)	(0.002)	(0.002)
$\tau \geq 25$	0.202	0.205^{*}	0.203	-0.001	0.010^{***}
	(0.139)	(0.116)	(0.140)	(0.002)	(0.002)
Absorbed FE	Cal.Ym	Cal.Ym	Cal.Ym	County×Ym	Cmpd, Ym, Cmpd×Yr
Observations	163,794	163,794	163,794	163,264	170,841
Clusters (compound)	1,305	—	1,305	1,198	1,305

Notes: Dependent variable: indicator for individual i's first endogenous hotel visit in month t. All specifications include straight-line distance control. Reference period: $\tau \in [-6, -1]$. Columns (1)–(3) report complementary log-log or logit coefficients; columns (4)–(5) report linear probability model coefficients, as logit is infeasible at these fixed-effect dimensions. Column (2) reports random-effects cloglog with compound-level random intercepts; standard errors are not clustered. Column (5) absorbs compound fixed effects, calendar month fixed effects, and a compound-specific linear time trend. Standard errors in parentheses, clustered at compound level except column (2). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Robustness to Control Variable Specification

	(A) +Neighbor Visits	(B) +Group Size	(C) +Rating Variables	(D) +Baidu Index
Pre-period (Coefficients)				
$\tau \leq -25$	0.061 (0.135)	-0.033 (0.132)	0.155 (0.141)	0.087 (0.136)
$\tau \in [-24, -13]$	0.111 (0.141)	0.087 (0.142)	0.175 (0.145)	0.127 (0.142)
$\tau \in [-12, -7]$	0.100 (0.162)	0.088 (0.162)	0.147 (0.168)	0.124 (0.163)
Post-period (Coefficients)				
$\tau \in [0, 2]$	0.236 (0.198)	0.260 (0.197)	0.274 (0.208)	0.250 (0.199)
$\tau \in [3, 6]$	0.261 (0.170)	0.299 (0.167)	0.321 (0.178)	0.277 (0.171)
$\tau \in [7, 12]$	0.617^{***} (0.160)	0.650^{***} (0.162)	0.645^{***} (0.168)	0.631^{***} (0.163)
$\tau \in [13, 24]$	0.212^{*} (0.142)	0.260^{*} (0.144)	0.301^{**} (0.148)	0.240^{*} (0.146)
$\tau \geq 25$	0.065 (0.134)	0.132 (0.131)	0.142 (0.138)	0.206 (0.139)
Cumulative neighbor visits	0.034^{***} (0.008)	0.073^{***} (0.018)	0.034^{***} (0.008)	—
Neighbor group size	—	-0.020^{***} (0.005)	—	—
Rolling avg. rating	—	—	2.708^{**} (1.326)	—
Baidu search index	—	—	—	-0.002 (0.002)
Observations	163,794	163,794	151,138	163,155
Events	1,584	1,584	1,498	1,577

Notes: All specifications are complementary log-log hazard models with calendar month fixed effects and distance control. Reference period: $\tau \in [-6, -1]$. Specification (C) has fewer observations due to Dianping data availability; (D) due to missing Baidu index values. Standard errors in parentheses, clustered at compound level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

also affects focal individuals. To address this, I restrict the sample to individuals who face no employer-imposed travel schedule—females aged 55 or above and males aged 60 or above (statutory retirement ages in China), with and without individuals aged 18 or below. Among retirees only (7,344 individual-month observations), the $\tau \in [7, 9]$ coefficient is 1.476^{***} (AME $= 2.72$ percentage points), roughly four times the full-sample baseline. The combined retiree-and-minor subsample (10,169 observations) yields a nearly identical estimate (1.419^{***}; AME $= 2.43$ pp). Pre-period coefficients remain jointly insignificant in both subsamples. Since this population has no employer-determined travel calendar, their conversion must reflect genuine peer-information transmission rather than coincident institutional demand.

5 Spatial Boundaries of Transmission

The baseline results establish the existence and dynamic pattern of peer effects. This section addresses two related questions: whether the estimated effect could reflect common demand shocks rather than peer transmission, and what geographic unit constitutes the natural boundary of word-of-mouth diffusion in Chinese cities.

5.1 Estimation Strategy

I re-estimate the baseline complementary log-log specification at three spatial tiers by redefining the peer group $\hat{N}(i)$ and shock date S_c at each level.

Building. The smallest unit, typically a single staircase and elevator bank housing 20 to 60 households. Residents share the same entrance, elevators, and mailboxes; daily face-to-face contact is nearly unavoidable.

Residential compound. A gated community enclosed by walls, containing multiple buildings with controlled entry points, shared gardens, and a property management office. The baseline specification operates at this level.

Administrative neighborhood. The lowest tier of urban administration, spanning multiple res-

idential compounds. Residents share a 15-minute living circle but lack institutionalized physical contact spaces.

Two placebo tests further map the hard boundary of transmission. Placebo I assigns the shock date based on the first exogenous visit by a neighbor in the same compound but a different building, testing whether cross-building transmission within a compound is real. Placebo II assigns the shock date based on the first exogenous visit by a neighbor in the same administrative neighborhood but a different compound, directly testing whether the effect persists beyond compound walls.

5.2 Results

Table 7 reports coefficients and AMEs across all five specifications.

Spatial gradient. The compound-level estimate is largest (0.630^{***}, AME $= 0.61$ pp), capturing the full within-compound transmission network. The building-level estimate is attenuated (0.434^{**}, AME $= 0.39$ pp), but this attenuation should be interpreted with caution: the average building peer group (4.49 members) is substantially smaller than the compound-level group (7.13 members), so the smaller coefficient reflects in part the limited effective demand population within a single building rather than necessarily weaker transmission intensity per contact. The administrative neighborhood estimate (0.438^{***}) is diluted by the inclusion of cross-compound pairs but remains significant because within-compound pairs dominate the sample.

Placebo I: same compound, different building. The peak-window coefficient is 0.414^{**} ($p = 0.026$), confirming that cross-building within-compound transmission is real. However, a significant pre-trend appears at $\tilde{\tau} \in [-12, -7]$ (0.388^{**}, $p = 0.027$), and the pre-period joint test yields $p = 0.083$. This pre-trend reflects correlated consumption propensities among residents of the same compound, illustrating precisely why exogenous variation is necessary: an observational design relying on voluntary neighbor visits would confound taste similarity with genuine information transmission.

Placebo II: same neighborhood, different compound. All five post-period coefficients lie between -0.086 and 0.077, none individually significant; the joint test yields $p = 0.45$. Pre-

period coefficients are also jointly insignificant ($p = 0.712$). The peer effect drops to zero at the compound boundary.

The contrast between the two placebos is the central finding of this section. Cross-building within-compound effects are real but contaminated by correlated propensities; cross-compound effects are absent entirely. Word-of-mouth transmission occurs within compounds and ceases at compound walls.

Table 7: Peer Effects Across Spatial Tiers and Placebo Tests

	(1) Building		(2) Compound (Baseline)		(3) Neighborhood		Placebo I Cross-building		Placebo II Cross-compound
Event Window	Coef.	AME	Coef.	AME	Coef.	AME	Coef.	AME	Coef.
Pre-period									
$\tau \leq -25$	-0.329^{*}	—	0.134	—	0.077	—	0.211	—	0.030
$\tau \in [-24, -13]$	-0.013	—	0.146	—	0.146	—	0.162	—	-0.040
$\tau \in [-12, -7]$	-0.094	—	0.077	—	0.134	—	0.388^{**}	—	-0.080
Post-period									
$\tau \in [0, 2]$	0.068	0.0006	0.235	0.0023	0.130	0.0014	0.465^{**}	0.0054	0.077
$\tau \in [3, 6]$	-0.009	-0.0001	0.264	0.0026	0.159	0.0017	0.046	0.0005	-0.017
$\tau \in [7, 12]$	0.434^{**}	0.0039^{**}	0.630^{***}	0.0061^{***}	0.438^{***}	0.0047^{***}	0.414^{**}	0.0048^{**}	-0.086
$\tau \in [13, 24]$	0.067	0.0006	0.239^{*}	0.0023^{*}	0.270^{**}	0.0029^{**}	0.169	0.0020	0.029
$\tau \geq 25$	0.153	0.0014	0.202	0.0019	0.182^{*}	0.0020^{*}	-0.038	-0.0004	-0.082
Observations	99,121		163,794		234,754		93,213		522,477
Events	906		1,584		2,566		1,100		6,179
Mean peer group size	4.49		7.13		8.37		—		—
Pre-period joint p	—		0.82		—		0.083		0.712
Post-period joint p	—		—		—		0.021		0.45

Notes: Columns (1)–(3): baseline complementary log-log hazard model with peer group and S_c redefined at each spatial tier. Column (2) reproduces the main analysis. The attenuated building-level coefficient reflects in part the smaller effective demand population within a single building (mean group size 4.49 vs. 7.13 at compound level). Placebo I: pseudo-treatment based on same compound, different building; the pre-trend at $\tilde{\tau} \in [-12, -7]$ reflects correlated consumption propensities within compounds. Placebo II: same administrative neighborhood, different compound; the flat pre-period ($p = 0.712$) and null post-period ($p = 0.45$) establish the compound boundary as the geographic limit of peer effect transmission. All specifications include calendar month fixed effects and distance control. Standard errors clustered at compound level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.3 Interpretation

Two conclusions follow from the spatial analysis. First, the sharp discontinuity at compound boundaries directly addresses the identification concern raised in Section 3: if the estimated effect reflected a demand stimulus common to the administrative neighborhood, the same pattern would

appear in Placebo II. It does not, ruling out neighborhood-level common shocks as an alternative explanation.

Second, the results speak to the appropriate geographic unit for peer effect identification in Chinese cities. Prior work has often used postal codes or county-level units, implicitly assuming that peer effects decay continuously with distance. The compound-boundary discontinuity documented here suggests instead that institutionalized physical contact spaces, rather than distance per se, determine the reach of word-of-mouth transmission. In the context of Chinese urban residential organization, the residential compound constitutes the natural unit within which peer information circulates, and the compound wall marks a genuine social boundary.

6 Online Ratings and Peer Effects

6.1 Setup

The baseline results establish that a neighbor's exogenous stay raises the monthly first-visit hazard by 0.61 percentage points during the peak window of months 7–12. This section asks whether the strength of that conversion effect depends on the online rating environment at the time of conversion. I examine two rating signals: the focal hotel's own rolling six-month Dianping score and a composite competitor index constructed as the six-month rolling average across five comparable five-star properties in Changsha. Table 8 reports pairwise correlations. The own rating and competitor index are positively correlated ($\rho = 0.233$), reflecting a common quality trend; this moderate collinearity does not prevent separate identification once both enter the same specification. The own rating has a sample mean of 4.35 and standard deviation of 0.26; the competitor index has a mean of 4.28 and standard deviation of 0.11, reflecting its construction as an average across five properties.

I proceed in two steps. First, I estimate direct rating effects by adding own and competitor ratings as time-varying covariates without interacting them with the event-study windows, establishing each signal's independent demand effect. Second, I interact both rating variables with the

Table 8: Pairwise Correlations Among Rating Variables

	Own rating	Competitor index
Own rating (6-month rolling)	1.000	
Competitor index	0.233	1.000
Mean	4.351	4.276
SD	0.265	0.112

Notes: Own rating is the focal hotel's six-month backward rolling average Dianping score. Competitor index is the unweighted average of six-month rolling scores for five comparable five-star properties in Changsha. $N = 307{,}677$ individual-month observations.

post-period event-study bins. The identifying logic is transparent: if ratings merely shift the overall demand level, interaction terms should be uniformly significant across all windows; if ratings specifically moderate the peer effect, the interactions should concentrate in the peak window where peer effects are most active and be absent elsewhere. This temporal specificity is the core identification evidence for the moderating channel. Because month-level fixed effects become collinear with the continuous rating variables when both signals enter jointly, I replace calendar month dummies with quarter dummies in the interaction specification.[4]

6.2 Direct Rating Effects

Table 9 reports the direct effect specifications. Column (1) adds own rating and review count to the baseline model with calendar month fixed effects. The own-rating AME is 0.057^{**} percentage points per rating point ($p = 0.019$); review count is insignificant. Column (2) adds the competitor index alongside own rating, replacing month with quarter fixed effects. The own-rating AME compresses to 0.011^{**} ($p = 0.013$) once the competitor index and seasonal aggregation absorb part of the common quality variation; the competitor index itself is insignificant (AME $= 0.004$, $p = 0.651$). The event-study profile is unaffected across both columns: the peak-window coefficient remains at 0.006^{***} and pre-period coefficients remain jointly insignificant.

[4]Including both own rating and competitor index alongside calendar month fixed effects produces near-perfect collinearity, as both variables vary only at the monthly level and the competitor index is a smooth six-month rolling average. Replacing month with quarter fixed effects restores identification; the baseline event-study profile is unchanged under this substitution, as confirmed by comparing columns (1) and (2) of Table 9.

The column (1) estimate anchors the conversion ratio reported in the introduction. The peak-window peer effect (AME $= 0.0061$ pp) divided by the own-rating direct effect (AME $= 0.0567$ pp per rating point) yields $0.0061/0.0567 \approx 0.11$: one exogenous neighbor stay generates a demand increment equivalent to a 0.11-point improvement in the hotel's online rating.

Table 9: Direct Effects of Own and Competitor Ratings

	(1) Own rating		(2) Own + Competitor	
	Coef.	AME	Coef.	AME
Pre-period				
$\tau \leq -25$	0.049 (0.136)	—	0.042 (0.136)	—
$\tau \in [-24, -13]$	0.115 (0.141)	—	0.113 (0.141)	—
$\tau \in [-12, -7]$	0.104 (0.162)	—	0.088 (0.162)	—
Post-period				
$\tau \in [0, 2]$	0.237 (0.199)	0.0023	0.217 (0.199)	0.0020
$\tau \in [3, 6]$	0.268 (0.171)	0.0026	0.250 (0.171)	0.0023
$\tau \in [7, 12]$	0.636^{***} (0.163)	0.0061^{***}	0.617^{***} (0.164)	0.0058^{***}
$\tau \in [13, 24]$	0.248^{*} (0.145)	0.0024^{*}	0.247^{*} (0.145)	0.0023^{*}
$\tau \geq 25$	0.205 (0.139)	0.0020	0.191 (0.139)	0.0018
Own rating (AME)	0.057^{**} (0.024)		0.011^{**} (0.004)	
Competitor index (AME)	—		0.004 (0.008)	
Review count (AME)	0.000 (0.001)		-0.000^{***} (0.000)	
Time FE	Month		Quarter	
Observations	163,794		168,209	
Clusters	1,305		1,305	

Notes: Complementary log-log hazard model. All specifications include age, age squared, and distance controls. Reference period: $\tau \in [-6, -1]$. Column (1) uses calendar month fixed effects; column (2) uses quarter fixed effects. AMEs expressed in percentage points. Standard errors in parentheses, clustered at compound level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

6.3 Moderating Effects

Table 10 reports the full interaction specification. The early post-period bins $\tau \in [0, 2]$ and $\tau \in [3, 6]$ are merged into a single $\tau \in [0, 6]$ bin, as neither carries significant peer-effect content in the baseline and the theory predicts a uniform zero in this window due to travel-opportunity constraints. Each rating variable is interacted with the four resulting post-period bins; pre-period bins and the reference period are left uninteracted.

Own rating interactions. The four own-rating interaction terms are jointly significant ($\chi^2(4) = 13.23$, $p = 0.010$). Individually, two windows are significant: the peak window $\tau \in [7, 12]$ (AME $= 0.018^{**}$, $p = 0.011$) and the long-run window $\tau \geq 25$ (AME $= 0.009^{**}$, $p = 0.018$). The early post-period window $\tau \in [0, 6]$ and the medium-run window $\tau \in [13, 24]$ are individually insignificant. A one-point increase in the own rating amplifies peak-window peer conversion by 0.018 pp, roughly three times the baseline peer effect of 0.0061 pp. Evaluated at the observed standard deviation of the own rating (0.265 points), moving from the 25th to the 75th percentile of the rating distribution raises peak-window peer conversion by approximately 0.010 pp—a 16% increase relative to the baseline.

Competitor rating interactions. The peak-window interaction is negative and significant (AME $= -0.023^{*}$, $p = 0.050$), while all three remaining post-period windows are individually insignificant. The joint test across all four post-period windows is insignificant ($\chi^2(4) = 6.18$, $p = 0.186$), which mechanically reflects the three null windows rather than an absence of the peak-window effect: a joint test that pools significant and null coefficients will understate the precision of the active term. The competitor rating's narrow observed range, with a standard deviation of only 0.112 points reflecting its construction as an average across five properties, further limits statistical power. Given this constraint, the economic magnitude of the peak-window estimate is notable: a one-standard-deviation increase in the competitor index suppresses peak-window peer conversion by 0.003 pp, offsetting roughly half the amplification produced by a one-standard-deviation increase in the own rating. The robustness specification with city fixed effects yields a tighter estimate (AME $= -0.028^{**}$, $p = 0.025$), confirming that the suppression effect is not an artifact of cross-city demand heterogeneity.

The concentration of both moderating effects in the peak window, and their absence in all other windows, is the central empirical result of this section. Online ratings do not uniformly shift conversion hazards across the full post-shock horizon; they exert their influence precisely when peer information is most actively translating into demand. The competitor rating's absence of a direct main effect, combined with its significant negative peak-window interaction, sharpens the

mechanism: competitor reputation does not divert demand broadly but specifically attenuates the social multiplier through which the focal hotel's current guests generate future voluntary visitors.

Robustness. A potential concern is that the moderating effects reflect unobserved heterogeneity in city-level demand rather than the rating environment per se: guests from cities with systematically higher consumption propensities may happen to arrive during periods of favorable ratings. To address this, I re-estimate the interaction specification as a logit model adding city fixed effects alongside quarter fixed effects, absorbing city-level average demand differences while retaining quarterly variation in ratings.[5] The peak-window interactions are robust under this demanding specification: the own-rating AME is 0.017^{**} ($p = 0.021$) and the competitor-rating AME is -0.028^{**} ($p = 0.025$), both tightening relative to the main specification. Pre-period coefficients remain jointly insignificant. The strengthening of the competitor-rating result under city controls suggests that the baseline estimate, if anything, understates the suppression effect.

6.4 Economic Interpretation

Three results taken together establish that the social multiplier of word-of-mouth is endogenous to the online reputation environment. First, the baseline peer effect at the peak window is equivalent to a 0.11-point improvement in the hotel's own rating—a concrete unit of account for the economic value of one neighbor's information. Second, the own-rating interaction is jointly significant across the post-shock horizon and individually concentrated in the peak and long-run windows, showing that reputation quality amplifies peer conversion not just at the moment of peak transmission but also in the slower diffusion that follows. Third, the competitor interaction is concentrated precisely in the peak window and absent in all others, consistent with a signal-substitution mechanism that operates only when peer information is potent enough to constitute a meaningful decision input. Both interactions strengthen under city fixed effects, confirming that the moderating effects are not driven by cross-city demand heterogeneity.

[5] City-by-quarter fixed effects would absorb all identifying variation in the rating variables, since both the own rating and the competitor index vary only at the monthly level and are common across all cities. City and quarter fixed effects entered separately are therefore the tightest feasible specification.

Table 10: Interaction of Rating Signals with Event-Study Windows

	(1) Baseline		(2) City FE	
	Coef.	AME	Coef.	AME
Pre-period (main effects)				
$\tau \leq -25$	0.058 (0.141)	—	0.050 (0.147)	—
$\tau \in [-24, -13]$	0.095 (0.142)	—	0.111 (0.144)	—
$\tau \in [-12, -7]$	0.074 (0.162)	—	0.074 (0.162)	—
Own rating × event window				
$\times\, \tau \in [0, 6]$	−0.409 (0.553)	−0.004	−0.547 (0.560)	−0.005
$\times\, \tau \in [7, 12]$	1.964** (0.765)	0.018**	1.836** (0.789)	0.017**
$\times\, \tau \in [13, 24]$	0.055 (0.484)	0.001	−0.217 (0.497)	−0.002
$\times\, \tau \geq 25$	0.920** (0.389)	0.009**	0.799** (0.396)	0.008**
Joint test (χ^2, p-value)	13.23 ($p = 0.010$)		—	
Competitor rating × event window				
$\times\, \tau \in [0, 6]$	0.476 (1.087)	0.004	0.196 (1.122)	0.002
$\times\, \tau \in [7, 12]$	−2.515* (1.272)	−0.023*	−2.955** (1.312)	−0.028**
$\times\, \tau \in [13, 24]$	0.126 (1.008)	0.001	0.042 (1.028)	0.000
$\times\, \tau \geq 25$	−0.946 (0.808)	−0.009	−1.017 (0.844)	−0.010
Joint test (χ^2, p-value)	6.18 ($p = 0.186$)		—	
Own rating (main AME)	0.007 (0.005)		0.008 (0.005)	
Competitor index (main AME)	0.008 (0.009)		0.009 (0.010)	
Model	Cloglog		Logit	
Time FE	Quarter		Quarter	
City FE	No		Yes	
Observations	168,209		163,619	
Clusters	1,305		1,277	

Notes: Column (1): complementary log-log hazard model with quarter fixed effects, age, age squared, review count, and distance controls. Column (2): logit with additional city fixed effects; city-by-quarter fixed effects are infeasible as rating variables vary only at the monthly level and are common across cities. Reference period: $\tau \in [-6, -1]$. The early post-period windows $\tau \in [0, 2]$ and $\tau \in [3, 6]$ are merged into $\tau \in [0, 6]$, as neither carries significant peer-effect content in the baseline and both are predicted to be zero by the travel-opportunity mechanism. Post-period main effects lose precision due to multicollinearity with the interaction terms; economic content is carried by the interaction AMEs. Joint tests in column (1) are Wald χ^2 tests on the four post-period interaction terms for each rating variable. Standard errors in parentheses, clustered at compound level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The pattern is inconsistent with ratings operating through a general demand-level channel.In that case, interactions would be uniform across windows. It is consistent with ratings functioning as a verification mechanism for the private social signal: consumers who have received a neighbor's recommendation are more likely to act on it when a high own-hotel rating corroborates the account and less likely when strong competitor ratings suggest alternative options of demonstrated quality. Mayzlin et al. (2014) document that consumers discount online reviews of uncertain provenance; it is precisely this credibility gap that makes the neighbor's firsthand experience valuable, and a high own-hotel rating amplifies that value by resolving residual uncertainty about the public signal's authenticity. The result extends Ameri et al. (2019)'s finding that word-of-mouth and observational learning provide distinct information within a single platform, by showing that private offline signals and public online ratings are not merely distinct but structurally complementary across the offline-online divide.

7 Conclusion

This paper provides causal evidence on peer effects in hotel consumption and their interaction with the online reputation environment. The peak-window peer effect is equivalent in demand magnitude to a 0.11-point improvement in the hotel's online rating, providing an intuitive benchmark for the economic significance of social information in this market.

These findings carry direct implications for firms operating in markets where consumption is constrained by travel opportunities. The demand value generated by an institutional visitor extends well beyond the direct booking revenue: through downstream word-of-mouth, each such guest generates future voluntary visits among residential neighbors, with a multiplier that persists for years. Crucially, this social multiplier is not fixed. It is larger when the hotel's own online rating corroborates the neighbor's private recommendation, and smaller when competitors simultaneously post strong ratings that reduce the relative informational value of the focal hotel's social signal. Reputation investment and peer-effect cultivation are therefore complements: the marginal

return to attracting word-of-mouth-generating guests is higher when a firm's public quality signal is strong, and lower in periods of intense competitive reputation pressure.

The identification framework proposed here, combining the earliest exogenous neighbor stay with a discrete-time hazard model, offers a replicable design for studying peer effects in other low-frequency experience goods markets, including tourism destinations, performing arts venues, and private healthcare. The key requirement is an institutional source of exogenous visits whose timing is orthogonal to local residential demand. Whether the temporal structure documented here, a delayed primary peak driven by travel-opportunity constraints and a secondary peak driven by within-network cascade, generalizes across such markets is an open empirical question. Future work drawing on richer data linking residential networks to consumption records across multiple markets and firms will further illuminate how social information shapes demand in settings where consumption opportunities are externally constrained.

Funding

This research did not receive any specific grant from funding agencies.

Declaration of Generative AI

During preparation I used generative AI tools to improve readability and language. I reviewed and edited the content and take full responsibility for the publication.

References

Ameri, M., Honka, E., & Xie, Y. (2019). Word of mouth, observed adoptions, and anime-watching decisions: The role of the personal vs. the community network. *Marketing Science*, 38(4), 567–583.

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

Bailey, M., Johnston, D., Kuchler, T., Stroebel, J., & Wong, A. (2022). Peer effects in product adoption. *American Economic Journal: Applied Economics*, 14(3), 488–526.

Bollinger, B., & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, 31(6), 900–912.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1), 41–55.

Cai, H., Chen, Y., & Fang, H. (2009). Observational learning: Evidence from a randomized natural field experiment. *American Economic Review*, 99(3), 864–882.

Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354.

Cho, M., Pekgün, P., Janakiraman, R., & Wang, Y. (2024). The effect of competitor reviews on hotel booking performance. *Journal of Marketing*, 88(1), 1–20.

Dahl, G. B., Løken, K. V., & Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7), 2049–2074.

De Giorgi, G., Frederiksen, A., & Pistaferri, L. (2016). Consumption network effects. *Review of Economic Studies*, 87(1), 130–163.

Gilchrist, D. S., & Sands, E. G. (2016). Something to talk about: Social spillovers in movie consumption. *Journal of Political Economy*, 124(5), 1339–1382.

Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1), 129–136.

Jin, M., Liu, Y.-J., Meng, J., & Zhang, Y. (2025). Peer effects in the hierarchy: Evidence from the workplace. *Journal of Economic Behavior & Organization*, 231, 106932.

Kuhn, P., Kooreman, P., Soetevent, A., & Kapteyn, A. (2011). The effects of lottery prizes on winners and their neighbors: Evidence from the Dutch postcode lottery. *American Economic Review*, 101(5), 2226–2247.

Lalive, R., & Cattaneo, M. (2009). Social interactions and schooling decisions. *Review of Economics and Statistics*, 91(3), 457–477.

Luca, M. (2016). Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School Working Paper*, 12–016.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3), 531–542.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421–2455.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *Review of Economic Studies*, 78(1), 356–393.

Seiler, S., Yao, S., & Wang, W. (2017). Does online word of mouth increase demand? (And how?) Evidence from a natural experiment. *Marketing Science*, 36(6), 838–861.

Tebbe, A. (2025). Peer effects in electric vehicle adoption. *Review of Economics and Statistics*, Revise and Resubmit.

Online Appendix

Appendix A Identification: Formal Proofs

This appendix formalizes the identification arguments sketched in Section 3. Section A.1 establishes the basic setup and notation. Section A.2 states and proves the key propositions supporting the identification assumptions. Section A.3 shows that subsample-level violations of conditional independence do not propagate to bias in $\hat{\beta}$. Section A.4 interprets the estimand.

A.1 Setup and Notation

Time and individual structure. The observation period is $t \in \{1, 2, \ldots, T\}$ with $T = 86$. Each individual i belongs to a unique residential compound c and building $b \subset c$. The neighbor set is $\hat{N}(i) = N(i) \setminus \{i\} \setminus R(i)$, where $N(i)$ denotes all compound residents in the guest database and $R(i)$ denotes historical co-travelers.

Potential outcomes. Let $T_i^* \in \{1, \ldots, T, \infty\}$ denote individual i's latent first endogenous visit time under no treatment. The baseline monthly hazard is $\lambda_{it}^{\circ} = \Pr(T_i^* = t \mid T_i^* \geq t,\, X_i,\, \varepsilon_i)$, where X_i collects observable covariates and ε_i captures unobserved heterogeneity.

Treatment timing. A visit is *exogenous* if it is arranged by an out-of-province government delegation, state-owned enterprise training cohort, or national conference organizing committee. The treatment date is $S_c = \min_{j \in \hat{N}(i)} \{T_j^{(1)} : T_j^{(1)} \text{ is exogenous}\}$. The relative time index is $\tau_{it} = t - S_c$ and the treatment indicator is $D_{it} = \mathbf{1}(t \geq S_c)$.

A.2 Key Propositions

Proposition 1 (Independence of the Minimum)**.** *Let $\{S_c^{(1)}, \ldots, S_c^{(n_c)}\}$ be the calendar months of all exogenous institutional visits within compound c. If each $S_c^{(k)}$ is independent of the individual latent conversion time T_i^* conditional on X_i, then $S_c = \min_k S_c^{(k)}$ is also independent of T_i^* conditional on X_i.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be arbitrary Borel sets. By the assumed independence of each $S_c^{(k)}$ from T_i^* given X_i, the σ-algebra generated by $\{S_c^{(1)}, \ldots, S_c^{(n_c)}\}$ is independent of the σ-algebra generated by T_i^*, conditional on X_i. Since $S_c = \min_k S_c^{(k)}$ is a measurable function of $\{S_c^{(k)}\}$, the event $\{S_c \in \mathcal{A}\}$ lies in the former σ-algebra and is therefore independent of $\{T_i^* \in \mathcal{B}\}$ conditional on X_i. Hence $\Pr(S_c \in \mathcal{A},\, T_i^* \in \mathcal{B} \mid X_i) = \Pr(S_c \in \mathcal{A} \mid X_i)\Pr(T_i^* \in \mathcal{B} \mid X_i)$. □

The key premise—that each institutional visit is independent of local residents' conversion propensities—is guaranteed by institutional design: out-of-province ministries and training departments set their schedules without reference to the preferences, incomes, or residential locations of any local community.

Proposition 2 (Conditional Independence on the Full Sample)**.** *Under the institutional orthogonality of $S_c^{(k)}$ and the sample-inclusion conditions of Section 3.3, Conditional Independence (Assumption 2) holds on the full analysis sample $\Omega = \Omega_A \cup \Omega_B \cup \Omega_C$: $S_c \perp\!\!\!\perp T_i^* \mid X_i,\, \delta_t$.*

Proof. The analysis sample Ω consists of individuals satisfying: (i) $|\hat{N}(i)| \geq 1$; (ii) the first recorded neighbor visit is exogenous. Condition (i) depends on who appears in the guest database, which reflects realized visits but is uncorrelated with the scheduling calendars of out-of-province institutions. Condition (ii) is a property of the observed sequence of neighbor visits; it does not select on the individual's own latent conversion time T_i^* or on compound-level demand characteristics correlated with S_c, because the institutional visits that define S_c are determined by external calendars. Formally, let E denote the event "the first neighbor visit is exogenous." Since E is determined by the institutional calendars and the ordering of visit times within $\hat{N}(i)$, and since institutional calendars are independent of T_i^* by Proposition 1 and the ordering of neighbor visit types does not depend on the focal individual's propensity, $\Pr(i \in \Omega \mid S_c = s, T_i^* = t, X_i) = \Pr(i \in \Omega \mid T_i^* = t, X_i)$ for all s and t. Applying Bayes' rule, $S_c \perp\!\!\!\perp T_i^* \mid X_i, i \in \Omega$. Calendar month fixed effects δ_t absorb city-wide seasonal demand variation; conditional on δ_t, the independence continues to hold. □

A.3 Score Function Analysis and Absence of Bias

Conditional independence holds on Ω but is mechanically violated on Ω_A (pre-treatment converters, requiring $T_i^* < S_c$) and Ω_B (post-treatment converters, requiring $T_i^* > S_c$), because subsample membership imposes restrictions on the joint distribution of S_c and T_i^*. This section shows that the violation does not propagate to bias in the cloglog MLE $\hat{\beta}$.

The cloglog log-likelihood contribution of individual i is $\ell_i(\beta) = \sum_t \big[Y_{it} \log \lambda_{it} + (1-Y_{it}) \log(1-\lambda_{it})\big]$, where $\lambda_{it} = 1 - \exp(-\exp(\eta_{it}))$ and $\eta_{it} = \alpha_t + \beta D_{it} + \gamma X_i$. The score with respect to β is $\partial \ell_i / \partial \beta = \sum_t D_{it}\, w_{it}\, (Y_{it} - \lambda_{it})$, where $w_{it} = \exp(\eta_{it}) > 0$. Unbiasedness requires $\mathbb{E}[\partial \ell / \partial \beta \mid \beta = 0] = 0$.

Proposition 3 (Ω_A Contributes Zero to the Score)**.** *For all $i \in \Omega_A$, $\partial \ell_i / \partial \beta = 0$ identically, regardless of the value of β.*

Proof. By definition, $i \in \Omega_A$ implies $T_i^{\text{endog}} < S_c$, so $t < S_c$ for all t in the risk set. Hence $D_{it} = \mathbf{1}(t \geq S_c) = 0$ for all t in the contribution to ℓ_i. Substituting $D_{it} = 0$ gives $\partial \ell_i / \partial \beta = \sum_t 0 \cdot w_{it} \cdot (Y_{it} - \lambda_{it}) = 0$. □

Proposition 4 (Ω_B Score has Zero Expectation)**.** *Under Assumptions 1 and 2 on the full sample Ω, $\mathbb{E}[\partial \ell_i^B / \partial \beta \mid \beta = 0, i \in \Omega_B] = 0$.*

Proof. The post-treatment score of $i \in \Omega_B$ is $\partial \ell_i^B / \partial \beta \big|_{\beta=0} = \sum_{t=S_c}^{T_i^*} w_{it}^\circ\, (Y_{it} - \lambda_{it}^\circ)$. Let $\mathcal{F}_t$ be the natural filtration. By the definition of the hazard rate, $\mathbb{E}[Y_{it} \mid \mathcal{F}_{t-1}, T_i^* \geq t] = \lambda_{it}^\circ$, so $M_{it} \equiv Y_{it} - \lambda_{it}^\circ$ is a martingale difference sequence with $\mathbb{E}[M_{it} \mid \mathcal{F}_{t-1}, T_i^* \geq t] = 0$.

The membership condition $i \in \Omega_B$ (i.e., $T_i^* > S_c$) imposes a lower bound on T_i^* relative to S_c, creating a truncated joint distribution. However, Proposition 2 establishes that $S_c \perp\!\!\!\perp T_i^* \mid X_i$ on the full sample Ω, which implies that S_c is independent of the individual-specific heterogeneity ε_i driving λ_{it}° conditional on X_i. The truncation condition $T_i^* > S_c$ can therefore be written as a measurable event in the filtration $\mathcal{F}_{S_c}$ that does not shift the conditional mean of M_{it} for $t \geq S_c$.

Formally, conditioning on $S_c = s$ and applying the optional stopping theorem to the bounded martingale $\{M_{it} : t \geq s\}$ with stopping time $\tau_i = T_i^*$: $\mathbb{E}\big[\sum_{t=s}^{\tau_i} w_{it}^\circ M_{it} \mid \mathcal{F}_{s-1},\, S_c = s\big] = 0$.

Taking the expectation over the distribution of S_c yields $\mathbb{E}[\partial \ell_i^B / \partial \beta \mid \beta = 0, i \in \Omega_B] = 0$. □

Proposition 5 (Ω_C Contributes a Conservative Downward Pull)**.** *Under Assumption 3 (strictly positive baseline hazard for Ω_C individuals), the Ω_C score contribution is strictly negative:* $\partial \ell_i^C / \partial \beta \big|_{\beta=0} = -\sum_{t=S_c}^{T} w_{it}^{\circ} \lambda_{it}^{\circ} < 0$.

Proof. Ω_C individuals are never observed to convert, so $Y_{it} = 0$ for all t. Substituting into the score: $\partial \ell_i^C / \partial \beta \big|_{\beta=0} = \sum_{t=S_c}^{T} D_{it}\, w_{it}^{\circ}\, (0 - \lambda_{it}^{\circ}) = -\sum_{t=S_c}^{T} w_{it}^{\circ} \lambda_{it}^{\circ}$. Since $w_{it}^{\circ} > 0$ and $\lambda_{it}^{\circ} > 0$ by Assumption 3, every term is strictly negative. □

Interpretation. Proposition 3 establishes that the pre-treatment converter group Ω_A plays no role in identifying β, which is appropriate: these individuals experienced their first visit before the information shock and carry no information about its effect. Proposition 4 establishes that the post-treatment converter group Ω_B identifies β consistently, despite the subsample-level violation of Conditional Independence. Proposition 5 establishes that the right-censored non-converter group Ω_C—individuals who remain in the risk set throughout the observation window—pulls $\hat{\beta}$ downward, so the reported estimates are conservative lower bounds on the true average treatment effect for the complier subpopulation.

A.4 The Estimand

The estimand identified by this design is a local average treatment effect (LATE): the average causal effect of a neighbor's exogenous visit on the monthly conversion hazard, for the subpopulation of motivated potential consumers who are in the risk set at the time of the shock. This is the population of individuals who have had prior contact with the hotel (through a booking, a cancellation, or an inquiry) but have not yet completed a first voluntary stay. The Ω_C group introduces a downward conservative bias, so the reported estimates are lower bounds on the true effect for the complier population.