Online Reputation

Social Media and Web Analytics

Lachlan Deer
Tilburg University
Updated: 2022-04-14

Learning Goals for this Week

At the end of this lecture you will be able to:

1. Summarize the impact of negative reputation on sales
2. Diagnose situations where fake reviews are more or less prevalent
3. Explain how managerial responses to online reviews impact future reviews
4. Interpret regression estimates from existing studies

Online Reputation Matters

Online Reputation Matters

Image from WebpageFX

Online Reputation Effects Perceptions

Image from WebpageFX

Online Reputation Effects Sales

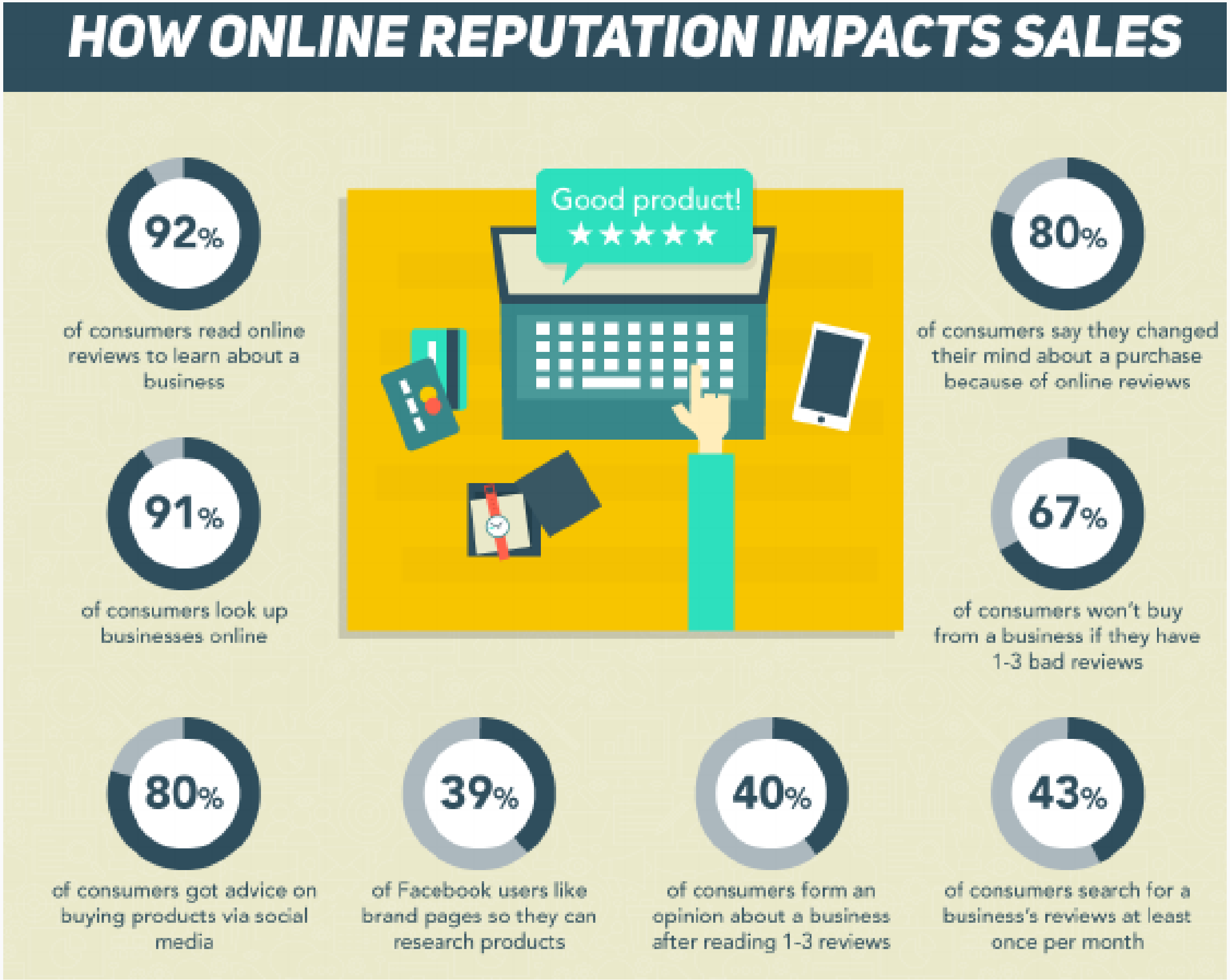

Image from WebpageFX

Why Online Reputation Matters

Buyers need to trust sellers

- Product descriptions
- Fulfilling transactions

Sellers need to trust buyers

- Ensure buyer will pay
- Abide by terms of service

Where does this trust come from?

$\implies$ **reputation systems**

… and design choices made by a marketplace

What do we want to know?

This class:

- How does seller reputation impact pricing and sales?
- Do fake reviews impact online reputation? How? When?
- Are managerial responses an effective way to manage online reputation?

Note: There's much more out there - let us know if you want further links in to the literature

Dynamics of Seller Reputation

Seller Reputation & Online

Motivation: Reputation mechanisms allow consumers to monitor firms

- How do consumers respond to changes in seller reputation?

Specific Questions: What is the effect of reputation on:

- Price / Willingness to Pay
- Sales Growth
- Subsequent reviews

Following discussion based on Cabral & Hortacsu, 2010

Empirical Approach

Data: eBay, follow sellers of five homogeneous products

- Transaction level data
- Seller's sequence of reviews

Empirical Approach:

- Descriptive Regressions
- Differences in Means

None of the effects we discuss here are **causal**

- Think of the results as descriptive associations between two variables

Reputation & Price

Estimating Equation:

$$price = \beta(\text{reputation\_measure}) + \gamma(\text{other demand factors}) + \text{error}$$

Excerpt from Table 2:

TABLE II
CROSS SECTIONAL REGRESSIONS

Model #	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	log(p)	log(p)	log(p)	log(p)	completed sale	log(# bids)
% negative comments	– 7.54	– 7.54	0.68	5.16	– 1.96	– 5.35
	(2.51)*	(9.88)	(6.81)	(7.75)	(1.09)*	(3.31)
Total # of feedbacks	0.05	0.05	0.00	0.00	– 0.003	– 0.011
	(0.04)	(0.03)*	(0.00)	(0.00)	(0.001)**	(0.004)**
% negative comments after format change				– 15.80	0.01	1.90
				(7.83)**	(1.92)	(3.65)
Total # of feedbacks after format change				0.00	– 0.001	– 0.002
				(0.01)	(0.001)	(0.003)

Negative Feedback & Sales Growth

Metric: Difference in Sales Growth before / after first negative feedback

TABLE III
IMPACT OF NEGATIVES ON SALES GROWTH (%)

Avg. Week.		Object			
	Growth R.	Thinkpad	Eagle	Silver	Teddy
First Negat.	Before	5.17	6.88	5.07	12.06
	After	− 7.56	− 4.67	− 8.25	− 5.28
	Difference	− 12.74 ***	− 11.56 ***	− 13.32 ***	− 17.34 ***
	Std. Error	4.89	3.56	3.44	3.69
	N	66	95	130	136
Second Negat.	Before	2.57	− 1.67	3.41	6.41
	After	9.53	9.00	7.61	7.51
	Difference	+ 6.96	+ 10.67 ***	+ 4.20	+ 1.10
	Std. Error	5.03	4.82	5.96	6.12
	N	37	70	78	83
Third Negat.	Before	8.14	2.75	2.81	1.00
	After	4.91	− 2.53	2.13	9.70
	Difference	− 3.23	− 5.28	− 0.68	+ 8.70
	Std. Error	6.14	7.47	3.21	6.22
	N	28	52	57	64

Frequency of Negative Feedback

Metric: Frequency of Arrival of Negative Feedback

TABLE V
FREQUENCY OF NEGATIVE FEEDBACK

	All Cat.	Thinkpad	Eagle	Silver	Teddy
T1	240.88	93.24	339.66	267.71	226.99
T2	188.76	58.59	199.24	261.26	199.86
ET	162.39	50.8	216.1	189.61	163.5
T1 – T2	52.12	34.66	140.41	6.45	27.13
T1 > T2 : p-val	0.021	0.036	0.017	0.452	0.27
T1–ET	78.48	42.44	123.56	78.09	63.49
T1 > ET: p-val	0.0002	0.0083	0.02	0.025	0.044
T2–ET	26.36	7.79	– 16.86	71.64	36.36
T2 > ET: p-val	0.032	0.176	0.73	0.027	0.089
N	311	58	79	78	96

T1: Sale-Related Feedbacks to First Negative.
T2: Sale-Related Feedbacks Between 1st and 2nd Negative.
ET: Average Number of Sale-Related Feedbacks Between Negatives.

Managerial Implications

1. **Price / Willingness to Pay** is **not the main metric** through which decreases in reputation operates
2. **Quantity sold** is an **important** metric
    - Sales decrease with negative feedback
3. **Negative Feedback can generate more negative feedback**
    - Though authors argue this is moral hazard - less effort by sellers

Online Review Manipulation

Reputation Manipulation

Motivation: Reputation is most useful when it's not tainted by "fake reviews"

- Fake reviews lead to:
  - Lower consumer welfare through sub-optimal choices
  - Mistrust in online reviews and reputation

Question: When does review manipulation occur?

- Are there more fake reviews when competition is close by?
- Do smaller hotels try to boost their reputation?
  - More positive fake reviews for small hotels?
  - More negative reviews for competitor nearby a small hotel?

Following discussion based on Mayzlin, Dover and Chevalier, 2014

Empirical Approach

Data: Travel sites in the US - TripAdvisor & Expedia

- Star Ratings of all reviews for all hotels in subset of cities in the US
- Supplement with hotel industry data from Smith Travel Research

Empirical Approach: Linear Regression

- Authors argue its some kind of difference in difference regression
- This paper is not DiD in a 'standard sense'

What makes all this work:

- TripAdvisor: Anyone can post at anytime
- Expedia: Can only post if booked on Expedia and stayed one night in last 6 months
- $\implies$ fake reviews are harder to post on Expedia
- **Assumption**: Users on each platform value hotel characteristics equally

Regression Equation

Estimate the following equation:

$$y_{ij} = X_{ij}B_1 + \mathbf{OwnAf}_{ij}B_2 + \mathbf{Nei}_{ij}B_3 + \mathbf{NeiOwnAf}_{ij}B_4 + \sum \gamma_j + \varepsilon_{ij}$$

Notation:

- i hotels, j city
- y_{ij} Difference in share of N star reviews between TripAdvisor and Expedia
- X_{ij} are hotel characteristics
- $\mathbf{Nei}_{ij}$ = 1 if competitor hotel within 0.5 km, else zero
  - We care about these coefficients, B_3
- $\mathbf{OwnAf}_{ij}$ are hotel ownership characteristics
  - We care about these coefficients, B_2
- $\mathbf{NeiOwnAf}_{ij}$ are competitor hotels ownership characteristics
  - We care about these coefficients, B_4

Why this approach will work ...

Authors don't observe review manipulation directly $\implies$ infer it from data patterns

- It's easier to manipulate reviews on TripAdvisor...

The story goes something like this:

- If the fraction of low (high) reviews on TripAdvisor is larger than on Expedia
- And consumers value the hotel equally between platforms
- Then differences are likely due to review manipulation on TripAdvisor

So let's check out the results...

Main Results

Table 3—Estimation Results of Equation (1)

		Difference in share of one- and two-star reviews	Difference in share of one- and two-star reviews	Difference in share of five-star reviews
$\mathbf{X}_{ij}$	Site rating	−0.0067	−0.0052	−0.0205**
		(0.0099)	(0.0099)	(0.0089)
	Hotel age	0.0004***	0.0003*	0.0002
		(0.0002)	(0.0002)	(0.0002)
	All suites	0.0146	0.0162*	0.0111
		(0.0092)	(0.0092)	(0.0111)
	Convention center	0.0125	0.0159*	−0.0385***
		(0.0086)	(0.0091)	(0.0113)
	Restaurant	0.0126	0.0114	0.0318***
		(0.0093)	(0.0092)	(0.0099)
	Hotel tier controls?	Yes	Yes	Yes
	Hotel location controls?	Yes	Yes	Yes
$\mathbf{OwnAf}_{ij}$	Hotel is independent		0.0139	0.0240**
			(0.0110)	(0.0103)
	Multiunit owner		−0.0011	−0.0312***
			(0.0063)	(0.0083)
Nei_{ij}	Has a neighbor	0.0192**	0.0296**	−0.0124
		(0.0096)	(0.0118)	(0.0119)
$\mathbf{NeiOwnAf}_{ij}$	Has independent neighbor		0.0173*	−0.0051
			(0.0094)	(0.0100)
	Has multiunit owner neighbor		−0.0252***	−0.0040
			(0.0087)	(0.0097)
γ_j	City-level fixed effects?	Yes	Yes	Yes
	Observations	2,931	2,931	2,931
	R^2	0.05	0.06	0.12

Notes: Regression estimates of equation (1). The dependent variable in all specifications is the share of reviews that are N star for a given hotel at TripAdvisor minus the share of reviews for that hotel that are N star at Expedia. Heteroskedasticity robust standard errors in parentheses. All neighbor effects calculated for neighbors within a 0.5 km radius.

Interpreting Results

Column 1:

- B_3: 0.0192 $\implies$ **hotels with a neighbouring competitor have a 1.9 percentage point increase in share of bad reviews**
  - approx. 7.5 percent increase compared to the baseline of 25 percent bad reviews

Column 2:

- $B_3 + B_4$: $\implies$ **hotels with an independent hotel as a neighbouring competitor have a 4.7 percentage point increase in share of bad reviews**
  - approx. 20 percent increase compared to the baseline of 25 percent bad reviews

Column 3:

- B_2: $\implies$ **independent hotels have a 2.4 percentage point increase in share of positive reviews**
  - approx. 7.5 percent increase compared to the baseline of 31 percent five star reviews

Takeaways

1. **Hotels with neighbors have more negative reviews**
    - Suggestive of competitors giving each other negative fake reviews
2. If **neighbor is an independent** hotel, (1) is even **more likely**
3. Independent hotels have higher reviews
    - Suggestive of positive review manipulation
    - But there are competing stories

Punchline: Evidence for fake reviews and manipulating online reputation

- Either by competitors (negative) or by the firm itself (positive)

Managerial Implications?

- More for platform owners ...
- There's a need to try and monitor / control reviews

Managerial Response to Online Reviews

Managerial Reviews & Reputation

Motivation: Business increasingly proactive to managing reputation

One Approach: Managerial Responses

Question: What is the effectiveness of Managerial Responses on future reviews?

- Are there more or less? $\implies$ volume effects
- Are the more or less positive $\implies$ valence effects

Following discussion based on Chen et al, 2019

Empirical Approach

Data: Travel Agencies in China (two): Ctrip & eLong

Empirical Approach: Linear Regression

- They implement what is known as a "difference in the difference of differences" regression
  - Think DiD with y being a difference between the outcomes between Ctrip and eLong
  - We can interpret the results through what we already!

Why this works …

Ctrip introduces managerial response, eLong does not

- Intuitively: comparing differences in reviews between platforms before and after managerial responses are introduced
    - If there's a change - its due to the introduction of MR
    - That's the "difference in difference" part
- Extra layer of concern: hotels choose whether to adopt managerial response
    - So it's not "random"
    - Trying to control for that is where the "extra difference" comes in
    - (Though I am slightly skeptical…)

Regression Equation

How do the authors do that as a regression?

$$\Delta Y_{it} = \gamma MR_i + \beta MR_i \times After_{it} + \delta' X_{it} + \alpha_i + \theta_t + \varepsilon_{it}$$

- β is the treatment effect $\implies$ this is the (only) number we care about in this regression

Some notation:

- i is a hotel, t is time
- ΔY_{it} difference in review volume (valence) between Ctrip and eLong
- MR_i has hotel done any managerial response on Ctrip
    - Binary variable -- 0 or 1
- $After_{it}$ tells us whether managerial response feature has been "turned on" in the Ctrip Platform
    - Binary variable -- 0 or 1

$\implies$ β is the average effect of managerial responses on the difference in review volume (valence) between platforms

Main Results - Volume, not Valence

Table 3: Impact of MR on Subsequent Reviews

DV	$\Delta logVol_{it}$			$\Delta logVal_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)
MR_i	0.069***	0.102**		0.030	0.070	
$MR_i \times After_{it}$	0.142***	0.147***	0.123***	0.071	0.072	0.092
$\Delta CumVol_{i,t-1}$	0.005***	0.005***	-0.008***	0.003***	0.003***	-0.004***
$\Delta CumVal_{i,t-1}$	-0.030	0.007	0.145***	1.882***	1.893***	1.462***
$Month_t$		-0.007***			-0.002	
$MR_i \times Month_t$		-0.002			-0.002	
Hotel Dummies	No	No	Yes	No	No	Yes
Month Dummies	Yes	No	Yes	Yes	No	Yes
Observations	23082	23082	23082	23082	23082	23082
Adjusted R^2	0.169	0.066	0.147	0.138	0.135	0.029
Model	OLS	OLS	FE	OLS	OLS	FE

* p<0.10 ** p<0.05 *** p<0.01

Interpretation:

- Specification (3) and (6) are the richest
- On average, 12.3% increase in monthly volume after adopting managerial responses

Target & Style of Managerial Response

Table 8: Impact of MRs Target and Style

DV	$\Delta logVol$			$\Delta logVal$		
	(1)	(2)	(3)	(4)	(5)	(6)
$MR_i \times After_{it}$	0.095**	0.094**	0.096**	0.059	0.060	0.060
$PosMR_{i,t-1}$	0.071**	0.368***		0.020	0.104	
$NegMR_{i,t-1}$	-0.007	-0.528**		-0.098	-0.403	
$LenMR_{i,t-1}$	0.004	0.004	0.027	0.039	0.035	0.040
$PosMR \times LenMR_{i,t-1}$		-0.074***			-0.020	
$NegMR \times LenMR_{i,t-1}$		0.127**			0.075	
$PosRatio_{i,t-1}$			0.700***			0.200
$NegRatio_{i,t-1}$			-0.517*			-0.606
$PosRatio \times LenMR_{i,t-1}$			-0.176***			-0.020
$NegRatio \times LenMR_{i,t-1}$			0.129*			0.077
$\Delta CumVol_{i,t-1}$	-0.008***	-0.008***	-0.008***	-0.004***	-0.004***	-0.004***
$\Delta CumVal_{i,t-1}$	0.143***	0.143***	0.145***	1.462***	1.462***	1.462***
month dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23082	23082	23082	23082	23082	23082
Adjusted R-squared	0.148	0.148	0.148	0.030	0.029	0.030

* p<0.10 ** p<0.05 *** p<0.01

Main Takeaways:

- Short Responses to Positive Reviews to not distract consumers
- Longer Responses to Negative Reviews to mitigate concerns

Recap

Summary

- Online reputation matters --- suggestive evidence that decreasing reputation is associated with decreases in sales
- Competitors seem to use online platforms to post negative fake reviews about each other
    - And they might provide positive fake reviews about themselves
- Managing reputation via responses to comments on large platforms stimulates more volume
    - Does it effect reputation though?
    - Group Assignment 1 will try and tackle this question!

License & Citation

Suggested Citation:

```
@misc{smwa-lecture-03,
      title={"Social Media and Web Analytics: Online Reputation"},
      author={Lachlan Deer},
      year={2022},
      url = "https://github.com/tisem-digital-marketing/smwa-lecture-03"
}
```

This course adheres to the principles of the Open Science Community of Tilburg University. This initiative advocates for transparency and accessibility in research and teaching to all levels of society and thus creating more accountability and impact.