Yelp Ratings vs City Development

Final Research Presentation

Royce Yang

University of Chicago

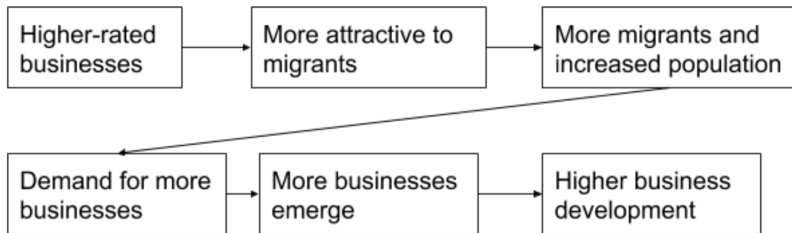21 May 2019

Research Question

What is the causal effect of Yelp Ratings on City Development?

Hypothesis

I hypothesize that cities with *higher* average Yelp Ratings (across businesses) experience *more* City Development (infrastructure, public transit system, schools, and population density, etc.).

Causal Theory

Key Literature: The Competitive Advantage of the Inner City (Porter 1995)

- Brief: Having more businesses in a city has a significant causal effect on overall city development.
- Therefore: Measuring growth of businesses is sufficient to answer my research question.

Key Literature: Law of Growth in Forecasting Demand (Prescott 2012)

Periods of Development:

1. Experimentation
2. Growth
3. Diminishing Growth
4. Stability

Remark

My hypothesis later turns out to be wrong because I did not consider (3).

Data Source

I use the official dataset from Yelp, provided as part of the Yelp Dataset Challenge. It is:

- Big (8 GB)
- Reliable (300+ citations on Google Scholars)
- Up-to-Date (up to December 2018)

Variables

- Independent Variable: Yelp Ratings[2] before 2014
    - Average Rating of All Businesses in a City
- Dependent Variable: Business Growth[3] from 2014 to 2018
    - Number of Newly Opened Businesses - Number of Closed Businesses in each city
- Control Variables: Average Price Range of Businesses, Review Count, Credit Card Acceptance Rate

---

[2]All independent variables (including control variables) are taken from before 2014 for precedence over the dependent variable.

[3]Sometimes referred to as *Business Development*. It is a good proxy for City Development due to our findings from slide 5.

Graphical Evidence
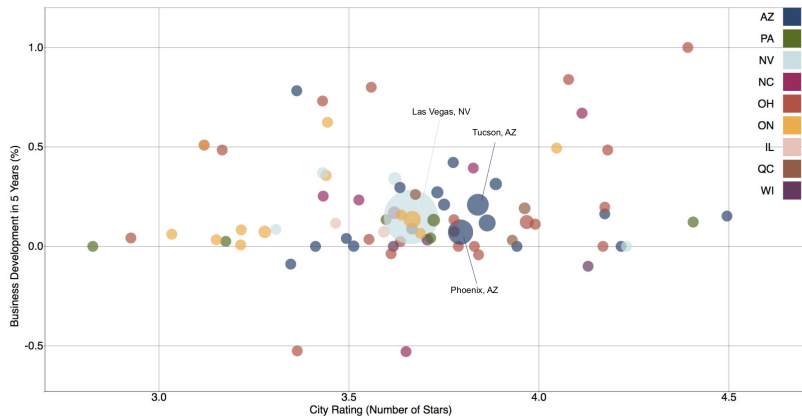
[4]

---

[4]Not the full dataset, because that would be too messy to show.

OLS Regression

Table: The Effect of Yelp Ratings on Business Growth, with Controls

	Dependent variable:
	development5yr
city_rating	−0.386* (0.216)
city_average_price	0.153 (0.294)
city_total_reviews_scaled	−0.828 (0.978)
city_credit_acceptance	0.506 (0.477)
Constant	0.965 (1.253)
Observations	46
R^2	0.399
Adjusted R^2	0.128
Residual Std. Error	0.366 (df = 31)
F Statistic	1.470 (df = 14; 31)

| *Note:*	*$p<0.1$; **$p<0.05$; ***$p<0.01$

Interpretation of OLS Regression Results

- $\hat{\beta}_{city_rating} = $ -0.386 (p<0.1) : For a one star increase in city_rating, our model predicts, in ceterus paribus, a 0.386 unit decrease in city development over the following 5 years.
- But our hypothesis suggests that $\hat{\beta}_{city\_rating} > 0$!
- No significance in controls, will not interpret for now.

What could be the problem?

- Reverse causality: crossover across our temporal boundary (2014).
- Omitted variables[5]: business' profit, viewer count of business on Yelp, category that business belongs to.
- Non-random measurement error: in pre-processing, filtered out very small businesses as a trade-off of variance for bias (which we can correct).

---

[5]Variables we would like to include but do not have data for.

Solution: Instrumental Variable Regression
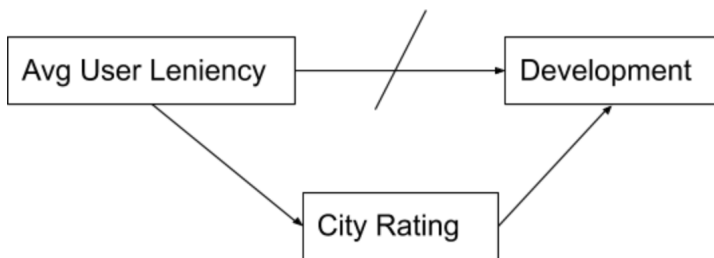
- Relevance: It makes sense that more lenient reviewers, in ceteris paribus, will give overall higher city ratings.
- Exclusion Restriction: Avg User Leniency is a metric specific to Yelp, and it would be difficult for such a specific metric to impact city development in a framework other than through aggregated rating scores.

Instrumental Variable Regression

Table: The Effect of Yelp Ratings on Business Development, with Instrument and Controls

	Dependent variable:
	development5yr
city_rating	-0.450^{***} (0.001)
city_average_price	-0.290^{***} (0.001)
city_total_reviews_scaled	-0.117^{***} (0.001)
city_credit_acceptance	0.244^{***} (0.002)
Constant	2.207^{***} (0.006)
Observations	665,527
R^2	0.297
Adjusted R^2	0.297
Residual Std. Error	0.083 (df = 665522)

Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Interpretation of Instrumental Variable Regression Results

- $\hat{\beta}_{city\_rating}$ = -0.450 (p<0.01) : (Still contrary to my hypothesis...) For a one star increase in city_rating, our model predicts, in ceterus paribus, a 0.450 unit decrease in city development over the following 5 years.

- $\hat{\beta}_{city\_credit_acceptance} > 0$: Contributes to development through increased economic activity.

- Other controls all have negative coefficients because it turns out that they are all indicative of how far we are in the development cycle (Prescott 2012).

- Prescott says that the 4 periods of development are Experimentation, Growth, Diminishing Growth, and Stability.

New Causal Theory
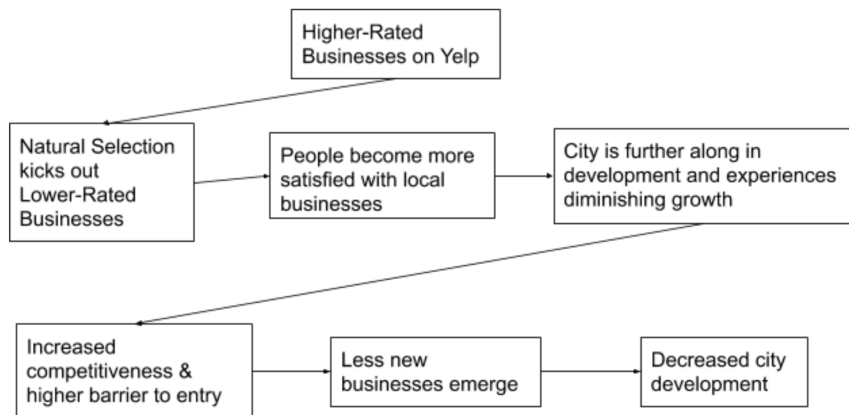
Findings Summary

- Higher ratings on Yelp help cities move along stages of development.
- Cities far along the stages of development experience diminishing growth and stability (Prescott 2012).
- Therefore, higher Yelp ratings have a negative effect on city development because of diminishing growth.

Extensions

- Further verification of my new theory
- How much power does Yelp have over a city's development?

Thank you!

References

1 T. Buch, S. Hamann, et al. "What Makes Cities Attractive? The Determinants of Urban Labour Migration in Germany." Urban Studies. July 2014.

2 M. Clemens. "Migration is a Form of Development: The Need for Innovation to Regulate Migration for Mutual Benefit." United Nations, Department of Economic and Social Affairs. 2018.

3 G. Dahl, A. Kostol and M. Mogstad. "Family Welfare Cultures." The Quarterly Journal of Economics, Volume 129, Issue 4, November 2014, Pages 1711–1752, 21 August 2014.

4 E. Glaeser, H. Kim and M. Luca. "Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity." Harvard Business School Working Paper, No. 18-022, October 2017.

5 M. Luca. "Reviews, Reputation, and Revenue: The Case of Yelp.com." Harvard Business School Working Paper, No. 12-016, September 2011. (Revised March 2016. Revise and resubmit at the American Economic Journal - Applied Economics.)

6 M. Porter. "The Competitive Advantage of the Inner City." Harvard Business Review, May 1995.