The Effect of Rain on Yelp Reviews

Jonathan Monreal

Introduction

In this paper, I will explore the question of whether or not rain has an effect on Yelp reviews of restaurants. This topic is important for several reasons. Firstly, it fits into a larger context of the effect of weather on human behavior. While user reviews are perhaps not as important an area as the stock market, they are nonetheless an area where weather may have a real economic impact. Secondly, if precipitation is having an effect on user reviews, businesses may want to attempt to address the issue. With the meteoric rise of review sites such as Yelp, it seems likely that consumers are turning to such sites when making decisions about where to eat if they are unfamiliar with the business, and a higher review score may influence people to choose a particular business over another. If a business were able to change the impact of rain on their review scores, they could potentially gain an advantage over other businesses in the area. Additionally, it could potentially point to a situation where a business simply might want to try to put in additional effort to keep customers satisfied if rain is having a negative effect on customers' perception of the business.

While there are a number of factors that can be investigated as far as weather, precipitation seems the most appealing for initial study due to its common association with mood. Rain is associated with other changes in the weather as well such as greater cloud cover and change in barometric pressure, and thus precipitation serves as a useful metric by which to look at the collective effects of these changes in weather. For these reasons, while other factors such as temperature, wind, and pressure could also potentially affect review scores, this paper will focus on rain specifically as an initial explanatory variable.

Background

The idea that weather has a tangible effect on mood is probably familiar to most people. Conventional wisdom tells us that a cloudy, rainy day is often referred to as "a gloomy day," and that a sunny, temperate day can be characterized as "beautiful." Despite this, there does not seem to be a great deal of research in the field of psychology into the effect of weather on mood (Denissen et. al 2008).

Keller et. al (2005) found that pleasant weather, meaning higher temperature or barometric pressure, was associated with "higher mood, better memory, and 'broadened' cognitive style during the spring as time spent outside increased" (Keller et. al 2005). At the same time, however, they did not find that the effect continued into the summer, and that at that time, hotter weather was associated with lower mood. The authors point to time spent outside and Seasonal Affective Disorder (SAD) as factors rather than temperature alone. Fedderson, Metcalf, and Wooden (2012) found that precipitation and temperature did not have a statistically significant effect on self-reported life satisfaction, but that "day-to-day weather variation impacts life satisfaction by a similar magnitude to acquiring a mild disability" (Fedderson, Metcalf, and Wooden 2012). Furthermore, they suggest that past studies that found that precipitation did have an effect on mood may have found this relationship because precipitation serves as a proxy variable for air pressure. Denissen et. al (2008) looked into the effect of temperature, wind power, sunlight, precipitation, air pressure, and photoperiod (sunlight) on mood in three categories: positive affect, negative affect, and tiredness. They found that sunlight had a main effect on tiredness, and further that it mediated the effect of both pressure and precipitation on tiredness. They also found that the average effect of weather on mood tended to be small overall but that there was significant variance across individuals. This suggests that precipitation may serve as a proxy for sunlight as the actual variable effecting mood.

One study involved having "[a] server in a midscale restaurant" write on the back of their customers' checks either nothing at all, that the weather would be unfavorable the next day, or that it would be good the next day (Rind and Strohmetz 2001). They found that a favorable forecast was correlated with significantly higher tip percentages, suggesting that mere belief about the weather might have an effect on mood.

Most research on the economic impact of weather seems to focus on the stock market, for which there is extensive literature. There seems to be considerable debate, however, as to whether or not weather has any effect on the stock market. Chang et. al (2005) found a relationship between temperature, cloud cover, and stock returns in Taiwan. Symeonidis et. al (2010) suggest that cloudiness and nighttime length are both inversely related to stock market volatility. Goetzmann and Zhu (2002) found "virtually no difference" between cloudy days and sunny days with regards to buying and selling of equities, and suggests that if there is indeed an effect, it is related to "market-makers, news providers, or other agents physically located in the city hosting the exchange" rather than individual investors themselves (Goetzmann and Zhu 2002). Jacobsen and Marquering (2005) found no evidence to support that temperature or SAD and found that a winter/summer dummy best fit the effects on the market, and that "the conclusion that weather affects stock returns through mood changes of investors is premature" (Jacobsen and Marquering 2005). Based on the great variety in interpretation of these studies, there seems to be no consensus on whether or not weather has any effect on the stock market.

There seems to be very little research into the exact topic of the effect of weather on user reviews. One study by researchers from Georgia Tech and Yahoo Labs showed that the rating of user reviews of restaurants was affected negatively on days where precipitation was greater than zero for data collected using the CityGrid API between 2002 and 2011 in 32,402 U.S. cities (Bakhshi et. al 2014). This seems promising with regards to the prospects of the effect of precipitation on Yelp reviews.

Data

Data was collected from two different sources. Firstly, the Yelp Academic Challenge dataset was obtained from the Yelp website. The dataset as acquired included data from Yelp about reviews (pooled cross section), users (cross-section), businesses (cross-section), and check-in (pooled cross-section) data about when the user visited a business and checked-in. All entries were in a JSON format, which was converted to a comma-separated values (CSV) format using a Python program and then imported into Stata. The information regarding users and businesses was merged with the review data based on the user IDs and business IDs, respectively. All reviews that did not have any user or business information associated with them were dropped from the dataset, as were reviews for businesses other than restaurants. The cities of Charlotte, NC; Pittsburgh, PA; Madison, WI; and Urbana-Champaign, IL were selected for the easy accessibility of weather data and relatively rainy climates. Reviews from the other cities were dropped from the dataset.

Secondly, data was obtained about the weather in each city over the range of dates of the Yelp dataset from the National Oceanic and Atmospheric Administation (NOAA) National Centers for Environmental Information website, which includes information on precipitation, temperature, and type of weather for many weather stations in each city (and as such it was pooled cross section data as obtained). This data came in CSV format, which was imported into Stata. Entries with invalid values for precipitation and other key variables snowfall and snow depth were dropped from the dataset. The remaining entries were then collapsed and averaged on the basis of city and date such that there was one entry for each city and date, reflecting time series data. It was then merged into the reviews dataset on the basis of date and city. Review entries that did not have any weather data associated with them weUlre dropped, as were all entries with any reported snow, snow on the ground (by means of snow depth), fog, freezing rain, freezing drizzle, hail, high or damaging winds, smoke or haze, dust, volcanic ash, blowing dust, blowing sand, ice pellets, sleet, snow pellets, small hail, and thunder so as to control for any weather other than rain that might have an effect. This left a total n=59,260 reviews that had valid data for all key variables.

All told, the final dataset includes 59,260 observations reflecting individual reviews from December 19, 2004 until January 8, 2015. It seems impossible to determine whether or not this sample is random on multiple levels. Firstly, the language on Yelp's website makes no statements as to the randomness of the data included in the dataset. Secondly, the cities chosen may not represent a random sample of all Yelp users.

The key variables of interest here are *Stars* and *Precipitation*. Stars (see *Figures 2* and *3*), which is present from the Yelp reviews dataset, reflects the value on a scale of 1 to 5 stars that the user gave a particular business in a particular review. The scale of 1 to 5 stars is an arbitrary rating by the user of how highly they rate the business, with 1 star being the lowest possible score and 5 being the highest. As of this date, each star rating is associated with a different message on the Yelp website: "*Eek! Methinks not*" for 1 star, "*Meh. I've experienced better*" for 2, "*A-OK*" for 3, "*Yay! I'm a fan*" for 4, and "*Woohoo! As good as it gets!*" for 5 stars. Yelp's policies prohibit people from reviewing a business if they have a conflict of interest, such as if they received any incentive from the business[1].

The mean for stars is 3.7 while the standard deviation is 1.21, suggesting that reviews tend to fall more towards the higher end of the range. The most popular score in the dataset was 4 stars with 20455 reviews (34.5% of total), followed by 5 with 18082 reviews (30.5 %), 3 with 10212 reviews (17.2%), 2 with 6146 reviews (10.4%), and 1 star with 4365 reviews (7.37%).

[1] See http://www.yelp.com/guidelines.

This again suggests a bias towards more favorable reviews, and seems to point away from the popular wisdom of people using review sites mostly to rant about businesses they dislike for some reason.

Precipitation (see *Figures 4* and *5*) represents the amount of precipitation on average in a particular city on a particular day, measured in tenths of a millimeter. Given that types of precipitation other than rain have been dropped from the dataset, *Precipitation* in this case means rainfall. The mean is 13.07, while the standard deviation is 44.8. The lowest value is 0, corresponding with no precipitation at all, while the highest is 810, meaning 81 mm or about 3.19 inches of rainfall. For Charlotte, the mean was 8.10 and the standard deviation 28.24 with a max of 274. For Pittsburgh, the mean was 15.19 with a standard deviation of 41.49 and a max of 415. For Madison the mean was 23 and the standard deviation being 74.94, and it held the highest max at 810. Finally, in Urbana-Champaign, the mean was 13.75 and standard deviation 24.07, with a max of only 64. The majority of reviews, 41203 (69.53%), saw no rain on the days they were written. Days with greater and greater amounts of rainfall appeared more and more scarcely.

Other variables that are used as explanatory variables are *BusinessReviewCount*, which is the number of reviews that the company has on Yelp to date; *BusinessStars*, the average star rating of the business to date; *UserReviewCount*, the number of reviews a user has written on Yelp to date; *UserAverageStars*, the average star rating that a user gives; and monthly, representing monthly fixed effects in the dataset. *BusinessStars* (*Figure* 6) is a useful control because it reflects how the business in question is seen by many users on Yelp and thus serves as a useful basis for comparison as to how positive or negative the star rating is compared to the average. *UserAverageStars* (*Figure* 7) is also a useful control because it helps to account for how the user generally reviews businesses. Different users may tend to give higher or lower reviews, and so the individual review is likely to be correlated with their average star rating. The case with *UserReviewCount* (*Figure 8*) might seem less clear, but it seems likely that as users post more reviews, they become more experienced with reviewing businesses and have a clearer basis on which they assign a star rating to businesses. *BusinessReviewCount* (*Figure* 9) reflects the popularity of a business, which seems likely to have some influence on the star rating of businesses. A *Monthly* variable was also created based on the date of the review to account for monthly fixed effects in the data.

Population Model and Assumptions

The dependent variable used here is stars, specifically log(stars). Log form seems useful here due to the arbitrary nature of the review scale. For example, "one tenth of a star" seems like a far less useful metric than a percentage.

For this regression, I propose the following population regression model:

$$\log(stars) = \beta_0 + \beta_1 Precipitation + \beta_2 BusinessStars + \beta_3 BusinessReviewCount$$
$$+ \beta_4 UserAveageStars + \beta_5 UserReviewCount + \beta_6 Monthly + u$$

Quite a few unobservable factors undoubtedly go into the error term, such as the user's mood on that particular day, the quality of the service on that particular day, and the expectations of the user (among others), that seem likely to have an effect on the star rating. None of these, however, seem likely to be correlated with *Precipitation*, which is a natural event independent of the service itself.

While the sample here might not be truly random as discussed earlier, it seems unlikely that this will have much if any effect on the outcome of the regression. Because the item of concern is how precipitation effects the number of stars and thus a behavior that would presumably be the same across populations, even if the cities chosen are not representative of Yelp reviews as a whole in some way, the effect should still be seen. Furthermore, it seems likely that even if the data posted by Yelp is not a random representative sample, the effect should still be seen. Thus, whether or not the sample is truly random does not seem to be an issue.

I anticipate that the sign on *Precipitation* will be negative, reflecting lower reviews on average as an effect of precipitation. For the control variables other than *Monthly*, I expect the coefficients to all be positive. As *BusinessStars* and *UserAverageStars* increase, the review score should also increase. For *BusinessReviewCount*, I anticipate that the more popular a business is, the higher their review score will tend to be if only for the reason that it seems likely that businesses with higher Yelp scores are probably more likely to be visited by Yelp users. Finally, I anticipate that *UserReviewCount* will be positively correlated with log(stars) because users with few reviews may have joined Yelp only to complain about a business.

For a first regression, I use only log(stars) and *Precipitation*. For the second, I add all of the control variables. For the third regression, I only use the data from restaurants that have a delivery service to see if the availability of such a service might have an effect on the model given that being able to order food by delivery would mean that consumers would not have to go out in the rain and therefore I would anticipate that the coefficient on *Precipitation* may actually be positive in this case.

Regression Models

For an initial regression including only log(stars) and *Precipitation*, I estimate the following regression model:

$$\underset{}{\log(stars)} = \underset{(.00190)}{1.23} + \underset{(.0000407)}{.0000488} Precipitation + u$$

For the second regression including the control variables, I estimate the following regression model:

$$\log(stars) = \underset{(.0495)}{-.536} + \underset{(.0000329)}{0.00000954} Precipitation + \underset{(.0030008)}{.2602} BusinessStars$$

$$+\underset{(.0000113)}{.0000626} BusinessReviewCount + \underset{(.00234)}{.319} UserAverageStars$$

$$+\underset{(0.00000485)}{.000043} UserReviewCount - \underset{(.0000764)}{.000629} Monthly + u$$

For the third regression restricted to only those restaurants that offered a delivery service, I estimate the following regression model:

$$\log(stars) = \underset{(.0495)}{-.254} - \underset{(.0000329)}{.000108} Precipitation + \underset{(.0030008)}{.245} BusinessStars$$

$$+\underset{(.0000113)}{.000235} BusinessReviewCount + \underset{(.00234)}{.331} UserAverageStars$$

$$+\underset{(0.00000485)}{.0000313} UserReviewCount - \underset{(.0000764)}{.00107} Monthly + u$$

Interpretation of Results

The initial regression of *Precipitation* on *log(stars)* failed to yield significant results (*Table 1.1*). With t=1.20, the result is not significant at even the 10% level. Furthermore, the coefficient of .0000488 means that a 1/10th of a millimeter increase in rainfall would increase the review score by an average of only 0.00488% per millimeter (or about 0.123952% per inch) of rainfall, which seems insignificant for any practical economic reason as well. The R^2 for this regression was 0, suggesting that at least precipitation alone is a very poor fit for explaining the review scores.

The second regression which added *BusinessStars*, *BusinessReviewCount*, *UserAverageStars*, *UserReviewCount*, and *Monthly* yielded similar results (*Table 2.2*). All of the variables other than *Precipitation* were significant at the 1% level, while at t=0.29 it remained statistically insignificant. The coefficient of 0.00000954 is considerably lower than in the first regression, meaning that a 1/10th of a millimeter increase in rainfall would yield an increase in review score on average of only 0.000954% per millimeter (or about 0.00242316% per inch) of rainfall, which is even less economically significant than the result of the first regression. With an R^2 of 0.3509, this model does seem to be a much better fit for the data, however.

The final regression used which restricted the regression to only reviews did show some interesting results (*Table 1.3*). All of the variables except for *Precipitation* continued to be significant at the 1% level except for *UserReviewCount*, which was significant only at the 10% level. This time, the value for *Precipitation* was t=-1.15, which is still not significant even at the 10% level, but much closer to the value in the first regression. The coefficient on *Precipitation* here of -.000108 means that for every additional 1/10th millimeter increase in rain, review scores would decrease by .00108% (or about 0.027432% per inch) which again is economically insignificant.

Based on the results of these three regressions, precipitation did not seem to have any significant effect on Yelp review scores. None of the coefficients were significant at even the 10% level, suggesting that for this particular dataset and model there was no correlation between rainfall and the number of stars users gave businesses on Yelp. In addition, even the largest percentage change as seen in the first regression would result in only a 0.123952% increase in the average review score, which seems so small and economically significant that even if it were statistically significant, addressing it would serve little purpose.

Limitations

There are some important limitations to note associated with the dataset. Since the weather data is based on collapsed and averaged data for the entire city associated with it, it is not as well localized as it could be. Considering the geographical size of some of the cities involved, it is entirely possible that the weather near the business varied from the average weather for the entire city. For example, if one small part of the city received a great deal of rain while the other part of the city had not, the average rainfall for the city is likely to be relatively low and not accurately reflect the precise rainfall in that location. This issue could be resolved by identifying which of the weather stations is closest to the business, which is entirely feasible due to the inclusion of latitude and longitude for each business in the Yelp dataset but would require more time to implement. A possible future direction might be to use a program to compare the latitude and longitude of each business with the location of each weather station and associate the review with the appropriate weather station. This is quite feasible, and could potentially make a difference in the results over using an average.

This, however, brings up another limitation: problems with the NOAA data. While it has undoubtedly been a great resource, there are some issues with it. Of the 214,442 observations included in the dataset, over half had invalid values for one important variable or another and had to be discarded. This is not merely a case of some of the stations missing data, but invalid data for each of the weather stations. While this seems unlikely to have had much effect on the regressions as the data was still averaged and if there was no valid data for a particular review it was dropped, it does seem important to note that this may have had some effect.

Yet another limitation is in the way reviews inherently work. There is no real way of telling how long after the user visited the restaurant before they wrote the review. The weather could be significantly different by the time they write the review, and any influence that the weather may have had on a person's perception of their experience may disappear given the fact that it is no longer recent. It is also possible that the current weather, rather the weather when the restaurant was visited, would affect the score. While the check-in data might seem like a promising lead for addressing this problem, as seeing if a user both checked in and wrote a review for the same business on the same date could point towards more accurate data, the information in this dataset is limited to only the number of check-ins for the business itself and is therefore useless in determining whether or not a user checked in at a particular business.

Conclusion

Given the dataset and model used here, precipitation does not seem to be a factor. This seems to be contrary to past research on the subject. There seem to be a number of possible reasons for this difference, including differences in the models used and peculiarities in the dataset.

There also seems to be plenty of room for further research on the topic of weather and reviews in this dataset. While there may be no correlation between rain and the review scores, it may prove interesting to look at other factors such as barometric pressure, daylight hours, temperature, and even different types of precipitation in order to determine whether there are other weather-related factors that are correlated with review scores. Looking at more localized weather data for each restaurant rather than for the city as a whole may also offer better results due to better precision.

Bibliography

Akhtari, M. (2011). Reassessment of the weather effect: Stock prices and wall street weather. *Undergraduate Economic Review*, *7*(1), 19.

Bakhshi, S., Kanuparthy, P., & Gilbert, E. (2014). Demographics, weather and online reviews: a study of restaurant recommendations (pp. 443–454). ACM Press. http://doi.org/10.1145/2566486.2568021

Chang, T., Nieh, C.-C., Yang, M. J., & Yang, T.-Y. (2006). Are stock market returns related to the weather effects? Empirical evidence from Taiwan. *Physica A: Statistical Mechanics and Its Applications*, *364*, 343–354. http://doi.org/10.1016/j.physa.2005.09.040

Denissen, J. J. A., Butalid, L., Penke, L., & van Aken, M. A. G. (2008). The effects of weather on daily mood: A multilevel approach. *Emotion*, *8*(5), 662–667. http://doi.org/10.1037/a0013497

Dowling, M., & Lucey, B. M. (2005). Weather, biorhythms, beliefs and stock returns—Some preliminary Irish evidence. *International Review of Financial Analysis*, *14*(3), 337–355. http://doi.org/10.1016/j.irfa.2004.10.003

Feddersen, J., Metcalfe, R., & Wooden, M. (2012). *Subjective well-being: weather matters; climate doesn't*. [Parkville,] Vic.

Goetzmann, W. N., & Zhu, N. (2005). Rain or shine: where is the weather effect? *European Financial Management*, *11*(5), 559–578.

Jacobsen, B., & Marquering, W. (2008). Is it the weather? *Journal of Banking & Finance*, *32*(4), 526–540. http://doi.org/10.1016/j.jbankfin.2007.08.004

Keller, M. C., Fredrickson, B. L., Ybarra, O., Côté, S., Johnson, K., Mikels, J., … Wager, T. (2005). A warm heart and a clear head the contingent effects of weather on mood and cognition. *Psychological Science*, *16*(9), 724–731.

Kelly, P. J., & Meschke, F. (2010). Sentiment and stock returns: The SAD anomaly revisited. *Journal of Banking & Finance*, *34*(6), 1308–1326. http://doi.org/10.1016/j.jbankfin.2009.11.027

Rind, B., & Strohmetz, D. (2001). Effect of beliefs about future weather conditions on restaurant tipping. *Journal of Applied Social Psychology*, *31*(10), 2160–2164.

Symeonidis, L., Daskalakis, G., & Markellos, R. N. (2010). Does the weather affect stock market volatility? *Finance Research Letters*, *7*(4), 214–223. http://doi.org/10.1016/j.frl.2010.05.004

Tables and Figures

Results

Dependent Variable: log(stars)

Regressor	[1]	[2]	[3]
Precipitation (*Precip*)	.0000488 (.0000407)	0.00000954 (.0000329)	-.000108 (.0000941)
Average business stars (*BusinessStars*)		.2602 (.0030008)***	.245 (.00853)***
Business review count (*BusinessReviewCount*)		.0000626 (.0000113)***	.000235 (.0000849)***
Average user stars (*UserAverageStars*)		.319 (.00234)***	.331 (.00666)***
User review count (*UserReviewCount*)		.000043 (0.00000485)***	.0000313 (.0000176)*
Monthly fixed effects (*Monthly*)		-.000629 (.0000764)***	-.00107 (.000239)***
Intercept	1.23 (.00190)***	-.536 (.0495)***	-.253 (.154)**
R-squared	0	0.351	0.378
n	59260	59260	7004

Figure 1

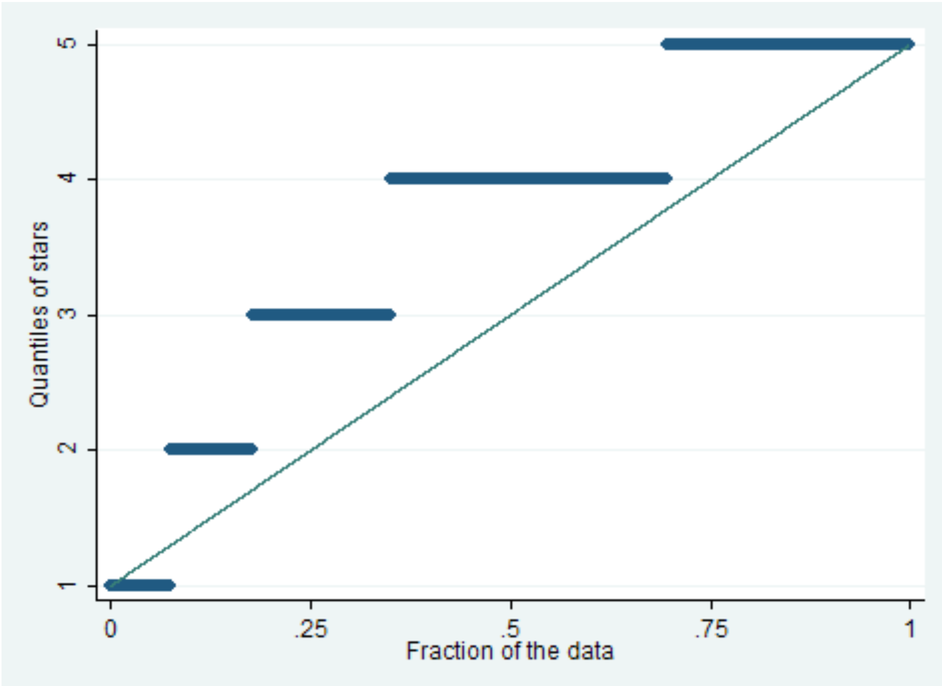


Figure 2

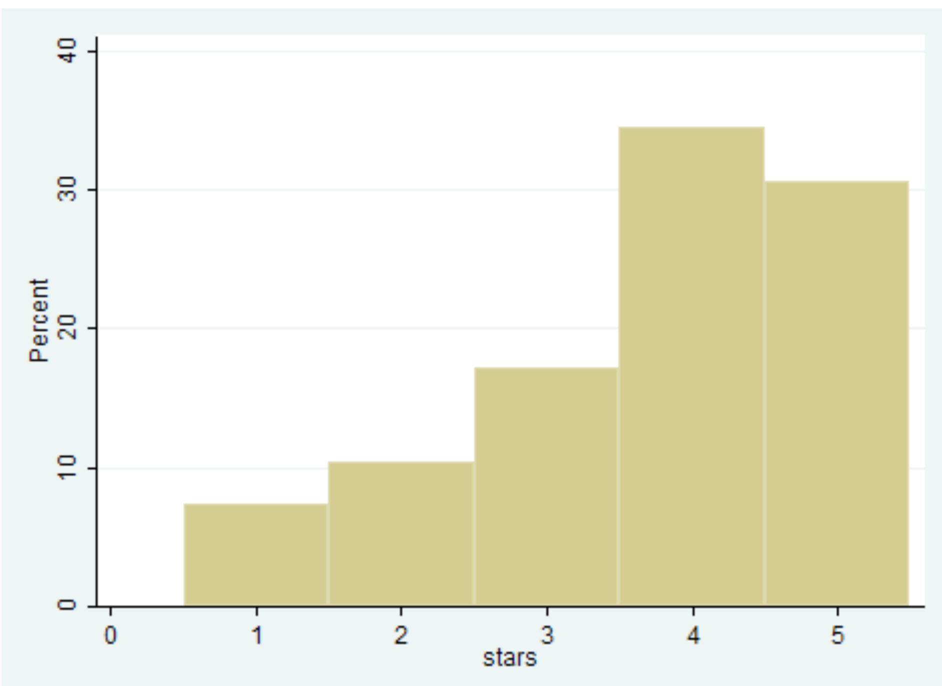
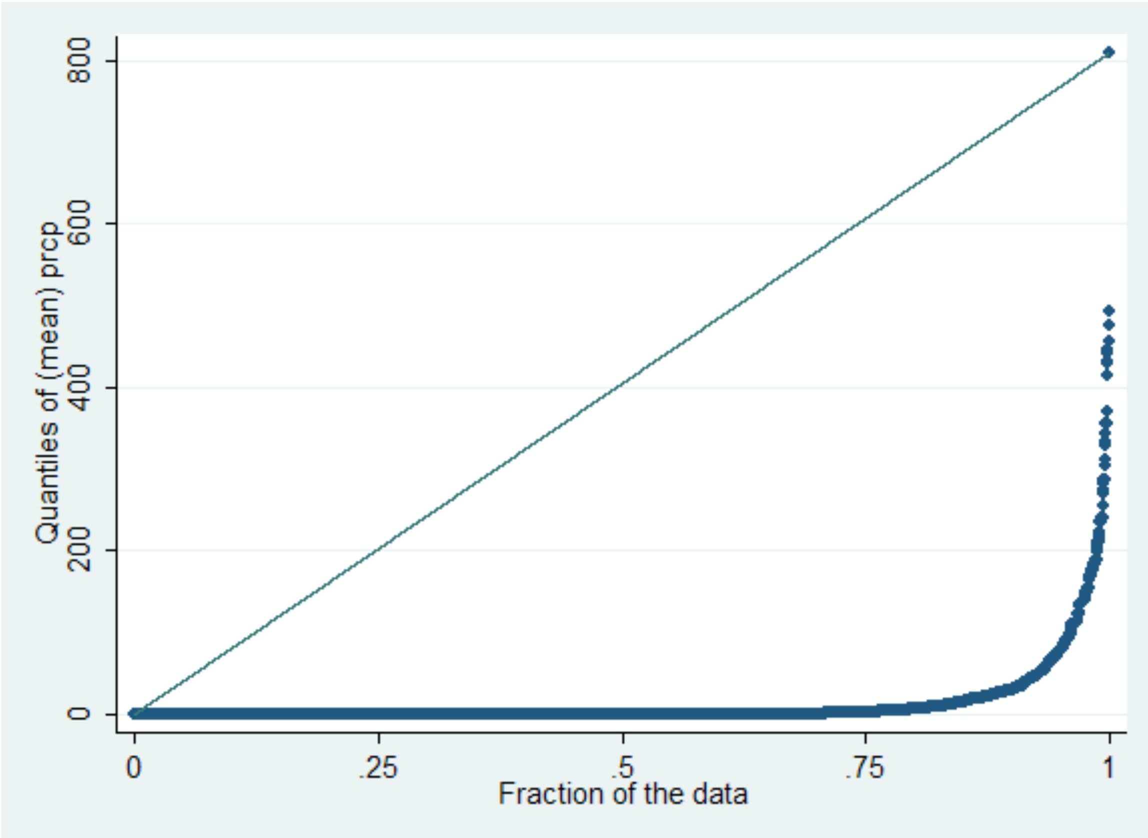


Figure 3



Quantiles of (mean) prcp
0
200
400
600
800
0
.25
.5
.75
1
Fraction of the data

Figure 4

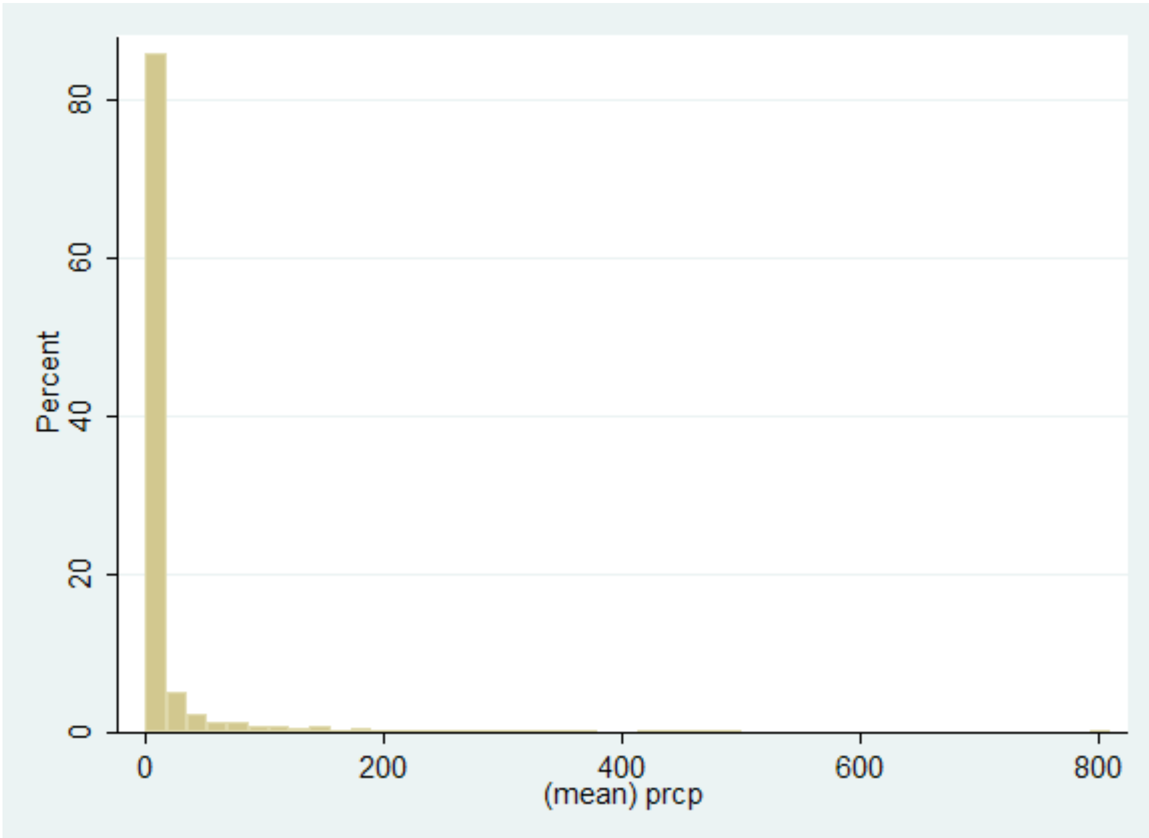


Figure 5

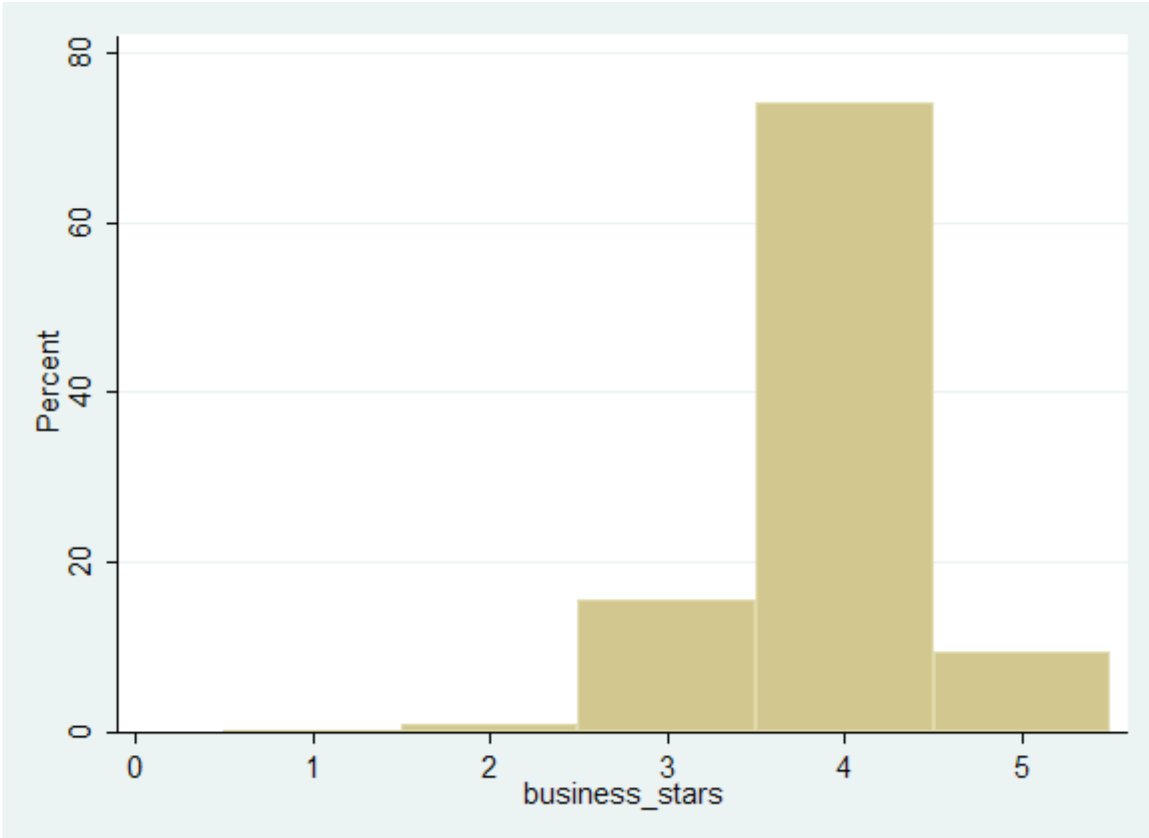


Figure 6

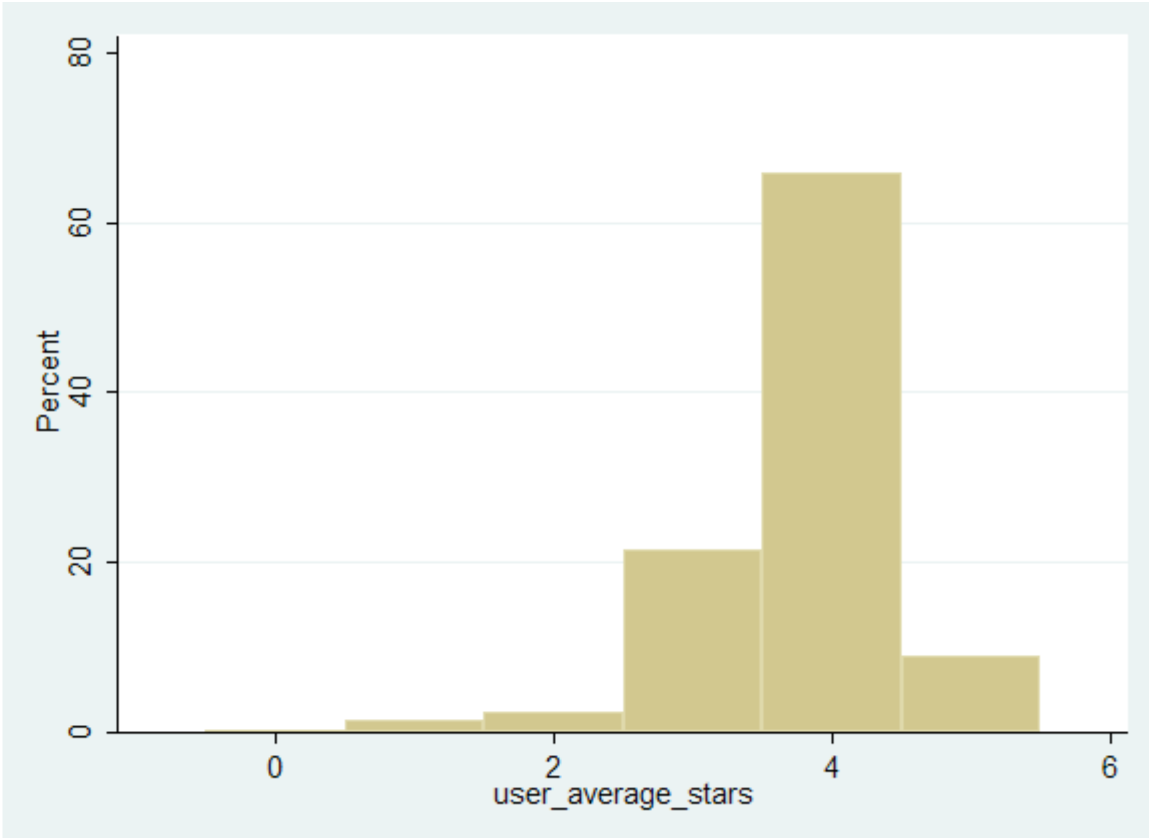


Figure 7

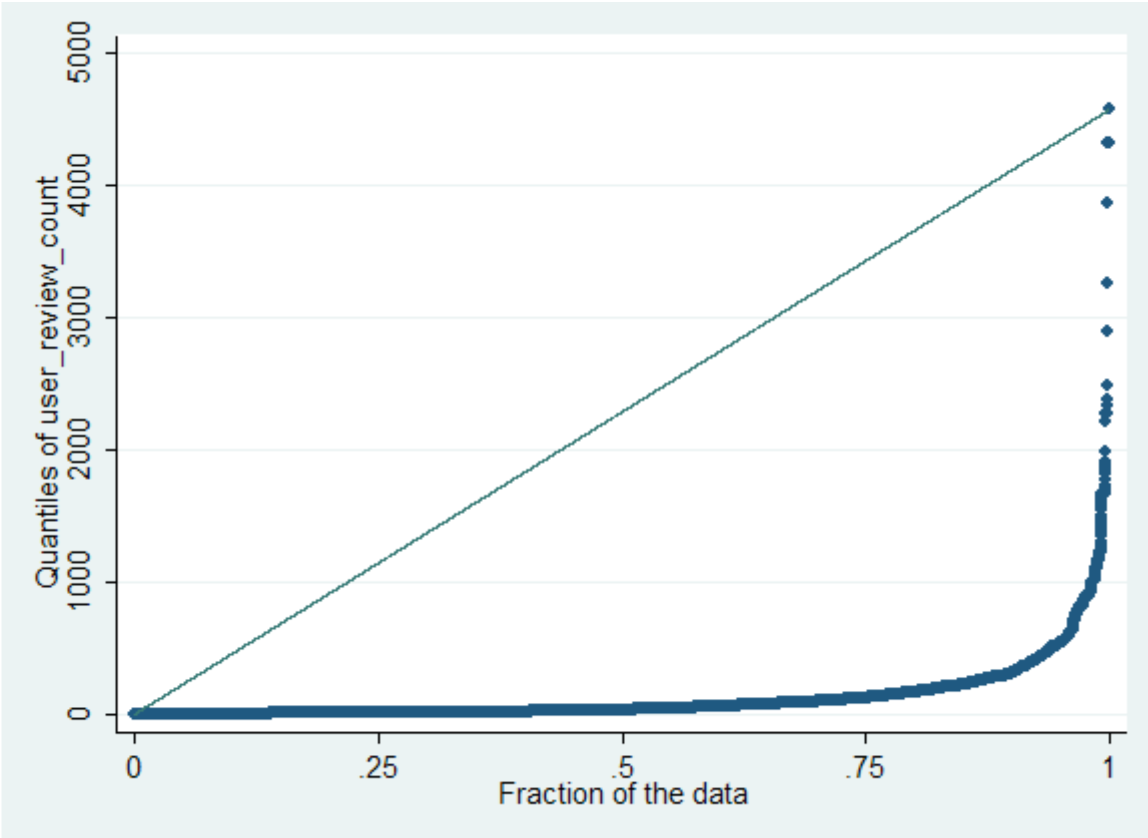

Quantiles of user_review_count
0
1000
2000
3000
4000
5000
0
.25
.5
.75
1
Fraction of the data

Figure 8

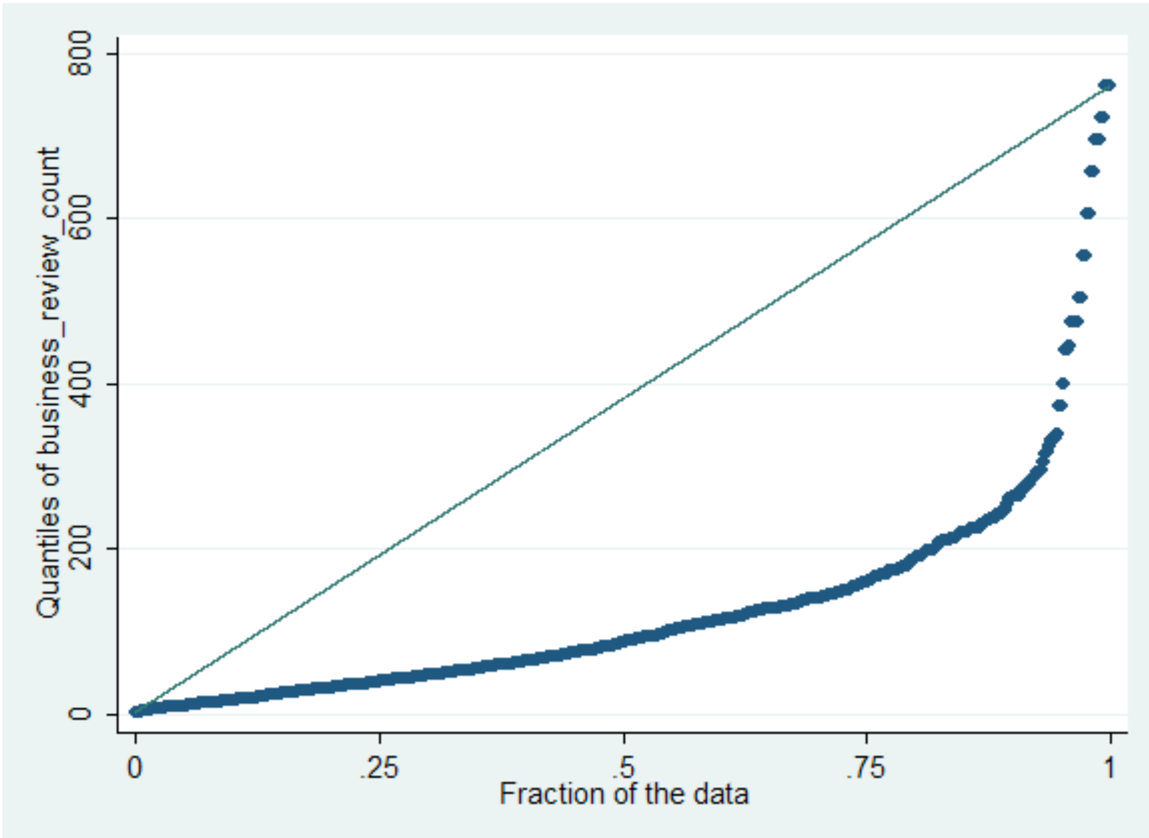

Quantiles of business_review_count
0
200
400
600
800
0
.25
.5
.75
1
Fraction of the data

Figure 9