Are Online Reviews Manipulated? Evidence from Amazon.com Dataset

Tiziano Rotesi

October 6, 2017

Abstract

In this paper we try to identify the presence of fake reviews by exploiting the way average ratings are presented to consumers on some popular websites. We use data from Amazon.com, where the average rating is rounded to the closest half star (the number of stars goes from 0 to 5). This creates discontinuities in the ability of the seller to manipulate the displayed number of stars. One single review can indeed increase the number of stars if the previous average is just below the threshold used to round the grade and/or if the total number of reviews is low. We find preliminary evidence that manipulation is at work.

In the past decade, online reviews have acquired more and more importance as a source of information for consumers. Websites that allow users to write reviews have become popular and it is easy now to find reviews about any kind of product or service available on the market. The possibility given to anyone to share opinions concerning a good has raised concerns regarding the possible influence that biased individuals could exert in order to distort the market. This topic has attracted interest not only in the academic community. Industry leaders such as Amazon.com have taken a number of actions to address the problem of fake reviews. For example suits were filed against owners of websites selling fake reviews[1] and new technologies have been developed to deal with this issue[2]. Still, identify the presence of fake

[1]http://www.forbes.com/sites/retailwire/2015/04/13/amazon-lawsuit-takes-on-fake-reviewers/

[2]http://www.theguardian.com/technology/2015/jun/22/amazon-ai-fake-reviews-star-ratings-astroturfing

reviews is a hard task as biased reviewers structure their reviews in such a way to mimic unbiased ones. The aim of this work is therefore to find conditions under which fake reviews are more likely to be written and show the effects on information available to consumers.

To do that we exploit the fact that on Amazon.com average ratings are rounded to the closest half star when first showed to the consumer (see Figure 1 and Figure 2). As consumers are likely to pay attention to this piece of information, sellers may want to manipulate the rating in order to increase sales. In principle sellers would always benefit from an extra positive review. The average rating would increase together with the number of consumers highlighting positive features of the product. Some sellers (or other people interested in increasing the success of the good) would therefore always consider fake reviews as part of their promotional activity. Still, not all reviews have the same impact. Depending on the average rating prior to the review and the number of reviews, there could be the possibility of increasing the number of stars. Therefore the benefit is likely to depend on factors that vary over time. Our hypothesis will be that instead , costs associated to the submission of promotional reviews are constant over time and only depend on firm characteristics. We will see that this idea will be the basis of our empirical strategy.

In order to link to the previous literature (cited below) we will start by considering a regression discontinuity setting and see whether there is evidence for selection. In particular the treatment that will be considered is the increase in half-star determined by the crossing of .25 and .75 thresholds. For example a book with average rating equal to 3.74 would get 3.5 stars while for another book with average rating 3.75 the number of star would equal 4. The first hypothesis that we will try to test is therefore whether we see evidence of selection, that is, the tendency of average ratings to lie on the right of the cutoff level. We will also analyze characteristics of reviewers, comparing the left and right side of the cutoff.

In order to strengthen the results, as a second step we will move from data at the item level (books) to data at the review level. This will allow us to overcome some limitations of the previous approach and get to some promising preliminary evidence. In particular we will study which are the factors that explain higher grades. In the presence of unbiased reviewers we would expect the rating to only depend on factors such as quality, opinion of previous reviewers and idiosyncratic noise. On the contrary, we will see that also the possibility of increasing the number of stars will explain some of the variation. In particular the regression shows that those reviews that can have an impact on the number of stars tend to be associated with a higher grade. This is compatible with the hypothesis described above of sellers actively involved in trying to manipulate the average rating and therefore the number of stars associated to the item. This allow us to partially solve the problem typically associated to this kind of analysis, that is the attempt of promotional reviewers to craft their reviews in such a way to appear unbiased. We will indeed avoid to focus on single characteristics of any particular review, but rather study the impact the review can have on the information presented to buyers.

This work proceeds as follows. In Section 1 we briefly discuss the literature on online reviews and their manipulation. Section 2 describes the data and present summary statistics. In Section 3 the first part of the analysis is done, with data on books. Section 4 includes results for single reviews. Section 5 we conclude and discuss possible extensions.

1. Prior Literature

The early literature on online reviews has focused on the effects that online reputation has on sales. Resnick and Zeckhauser (2002) and Resnick, Zeckhauser, Swanson and Lockwood (2006) were among the first contributions to show that on Ebay.com sellers characterize by higher ratings benefit higher sales and higher willingness to pay from consumers. Chintagunta, Gopinath and Venkataraman (2010) focus on the effect that online users review have on movie box office performances. The contributions by Chevalier and Mayzlin (2006), Luca (2011) and Anderson and Magruder (2012) are closer to our work. Chevalier and Mayzlin (2006) use data on books sales rankings from Amazon.com and BarnesandNobles.com and apply a difference in difference analysis to compare the evolution of sales in the two websites when ratings are different. They find that positive reviews are related to increase in relative sales. Luca (2011) and Anderson and Magruder (2012) focus instead on restaurants reviews on Yelp.com. This website displays average ratings in a way that is similar to what is done by Amazon.com: they round ratings to the closest half star and show the star rating in the main search pages. This fact allows the authors of both papers to use regression discontinuity design to measure the impact that an extra half-star has on sales. In Luca (2011) the author measures sales using data from the Washington State Department of Revenues, instead Anderson and Magruder (2012) rely on a database of restaurant reservation availability. In both papers it is found a sizable effect of star rating on sales.

Recent contributions have moved in the direction of studying fake reviews. Mayzlin, Dover and Chevalier (2014) study promotional reviews for hotels. They compare two platforms that follows different criteria in determining who is allowed to write reviews. On TripAdvisor.com anyone can submit a review, while on Expedia.com only those who have actually used the website to pay for a stay are allowed to. This generates a difference in the costs associated to the submission of a promotional review between the two platforms. The authors exploit this fact in conjunction to the hypothesis that independent hotel with single unit owners would benefit more from positive reviews than chain hotels that rely on better known brands. In this context, a difference in difference approach allow them to establish the presence of reviews manipulation done in particular by independent owners. Fake reviews are used to increase their average rating or to damage reputation of their closest competitors. Another relevant contribution in this literature is Luca and Zervas (2015). In this paper the authors study incentives to commit review fraud on Yelp.com. By looking at reviews that were filtered by Yelp.com (because considered fake) they find that fake reviews are more likely to appear when the number of reviews is low or after bad reviews. Moreover, big chains are less likely to manipulate reviews and increases in competition lead to higher number of fake reviews.

2. Data

We use data on reviews from Amazon.com. The dataset we use has been collected by Julian McAuley[3] and has been used by McAuley and coauthors for two publications: McAuley, Targett, Shi and van den Hengel (2015) and McAuley, Pandey and Leskovec (2015).

The dataset contains information on 83.06 million reviews written on Amazon.com between May 1996 and July 2014. For each review we know the item it refers to, who wrote it, whether it was voted as helpful by other users, the text, the day it was written and finally the rating. The dataset contains also products metadata. In this case we have information on the title, price, related items ("also bought", "also viewed"), the sales rank and the product category.

For the analysis presented here we only used metadata to select a sample a books [4]. In particular we selected a random sample of 476,342 books. For

[3]link: http://jmcauley.ucsd.edu/data/amazon/

[4]Price and sales rank vary significantly over time and since we don't know the exact day these data have been collected, it is unclear how to use them. For what concerns the "also viewed" and "also bought" graphs, they could in principle be used as a proxy for

Table 1: User Reviews - sample Amazon.com

	Mean	Standard deviation	Min	Max
Number of reviews - item	9.49	54	1	8280
Average rating - item	4.29	0.895	1	5
Number of reviews - reviewer	11.06	52.22	1	44557
Average rating - reviewer	4.27	0.85	1	5
Total number of items	476,342			
Total number of reviewers	2,598,773			
Total number of reviews	4,522,001			

each item then we collected the reviews, ordered them in time and calculated other variables such as the average before the review, the variance of ratings etc. Variables that refer to single reviewers are built in a similar manner, but in this case we take advantage of the availability of data and use the whole sample of reviews. This allows us to map the complete activity of each reviewer over time. Table 1 includes summary statistics for the three levels of observation: review, reviewer, product. Table 2 and Table 3 include informations on some of the others variables we considered.

One feature of this dataset that it is important to underline is the fact that has been collected in 2014. This could offer an advantage with respect to data downloaded now because Amazon.com seems to have changed strategy with respect to fake reviews. It was already filtering promotional reviews, when detected, but from June 2015 the effort in this direction has increased (see the link mentioned in the introduction). In Section 5 we discuss how improve on these data, as some information presented in the website have not been downloaded or tracked over time.

sales, but are limited to 100 items maximum.

Table 2: Reviews - variables - summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
n_rev	4522001	157.8587	511.488	0	8279
length	4522001	641.9334	890.9032	0	32729
avg_bef	4045659	4.344951	.5959879	1	5
var_bef	4045659	.8189537	.7576315	0	4
var_last4	3340851	.8062823	.8832875	0	3.84

Note:
- n_rev: number of reviews before
- length: length of the review (number of characters)
- avg_bef: average before the review
- var_bef: variance of ratings before the review
- var_last4: variance previous 4 ratings before the review.

Table 3: Reviewers - variables - summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
n_rev_sameID	2598773	11.06698	52.22035	1	44557
avg_sameID	2598773	4.279943	.8505047	1	5
var_sameID	2598773	.6917453	.9554295	0	4
dist_time_sameID	2598773	2.225856	3.125815	0	18.16438

Note:
- n_rev_sameID: number of reviews from the same reviewer
- avg_sameID: average rating from the same reviewer
- var_sameID: variance of rating, same reviewer
- dist_time_sameID: distance measured in years from the first to the last review.

3. Book level analysis

As a first step we study selection in a context that is comparable with what has been done in Luca (2011) and Anderson and Magruder(2012). We consider a regression discontinuity setting in which however the question is not the effect of stars on sales (as it was in the afore mentioned papers) but whether we can identify selection of individuals into the treatment. With treatment in this case we mean having an extra half star, while the exogenous rule that assigns the treatment is the criterion used to round the ratings. The variable that determines the treatment is therefore the average rating.

The average rating depends on reviews written by consumers and possibly on fake reviews. Unbiased reviews can be in principle considered to be a random process that mainly depends on the underlying quality of the product. If this is the case, unbiased reviews should generate a distribution of average ratings that is continuous around the .25 and .75 cutoffs. Similarly, if all sellers manipulate the average rating with the same ability, the final distribution should still be continuous. Instead, if we observe discontinuities around the cutoffs, this could be interpreted as evidence in favor of manipulation by a group of sellers.

We will be looking for two types of discontinuities. First we will apply McCrary test to see whether the density of observations jumps so that fewer books lie on the left of the threshold. Second, we will consider jumps in other variables, like characteristics of reviewers, that could suggest the presence of manipulation.

3.1 McCrary Test for manipulation of average rating

We want to test whether sellers are able to manipulate their average grade in order to gain extra stars. If fake reviews are written when the rating is just below the threshold in order to cross it, then we should observe more books with average rating just above the cutoff.

To understand whether this is the case it is useful to start from the observation of histograms. Figure 4 and Figure 5 display histograms for density of observations around the cutoffs. In both cases the variable we are considering takes value in the interval (-0.25, 0.25) that is the distance of the average rating from the closest cutoff. For example for two books with average rating 3.15 and 4.65 the variable would take value -0.1 while for a book with average rating 4.4 it would take value 0.15. The difference between the two figures is the fact that in Figure 4 we focus on books that have more than 20 reviews. The two histograms share the characteristics of

Table 4: McCrary test

	(1)	(2)	(3)
	More than 20 rev	More than 20 rev	More than 100 rev
Estimated Discontinuity	0.28	0.171	0
	.035	.028	.08
Bandwidth	.1	.15	.114
Observations	35,854	35,854	5,500

being almost symmetric around 0. It is hard to identify jumps apart those that are determined by the fact that these ratings are the average of integer numbers that go from 1 to 5 (for example .25, .75, .33, .66).

Figure 6 and Figure 7 present histograms for the average rating around cutoffs 3.25, 3.75, 4.25 and 4.75. Again, it is hard to spot jumps in densities.

In order to formally test the hypothesis described above we implement McCrary test as described in McCrary (2008). Results are presented in Table 4 and Figures 7, 8 and 9. We find that peaks in densities at the cutoff level play a crucial role. For samples that include books with fewer reviews the test rejects the null hypothesis of the absence of discontinuity (columns 1 and 2). However this discontinuity seems mostly determined by the density in .25 and .75 that can be explained with the fact that ratings are the average of integer numbers. We can underline that for the subsample of books with more than 100 reviews the test does not identify any jump in densities (column 3).

3.2 Regression discontinuity

The previous analysis suggests, in line with the previous literature, that manipulation is more likely to occur when the number of reviews is lower. Here we implement a regression discontinuity analysis on reviewers and review characteristics to understand whether there is evidence in favor of differences between the two subsamples.

We proceed as follows. For each book in our sample we take the average of reviewers' characteristics described in Table 2 and Table 3. For each book we calculate the average length of reviews, the average number of reviews written by reviewers etc. We then run the following regression:

$$Y_j = \alpha + \beta I(R_j > \hat{R}) + \gamma R_j \tag{1}$$

Where R_j measures the distance from the cutoff and $I(R_j > \hat{R})$ is a dummy variable that takes value 1 when the rating is above the cutoff.

Table 5 reports regression discontinuities estimates together with summary statistics for the variables considered. Figure 10 report scatter plots around the threshold. It is important to underline that although the coefficients are significantly different from zero, they can not still explain a significant fraction of the variance. We cannot therefore conclude that reviewers that wrote reviews for books that lie above the threshold are on average different from those who wrote reviews for books below the threshold.

4. Review Level Analysis

In order to gain better insights regarding the presence of review manipulation we now use data on single reviews. For each book we have information concerning the whole history of grades so that we can reconstruct the evolution in time of the statistics that are displayed to the user. When submitting a review, an unbiased user should base her decision on the perceived quality of the item. It is also possible that the decision of writing a review and the rating depend on the statistics presented in the website. A higher average, a higher number of stars or a higher number of previous reviews are, among others, elements that can potentially influence one person's opinion regarding a product. We therefore expect the grade to depend on some unobservable quantities like the true quality of the book and on the statistics that the user finds available on Amazon.com. Let us instead consider the case of a seller interested in manipulating the grade in order to sell more copies. When writing or buying fake reviews the seller incurs a monetary cost but also risks to be identified by customers or by Amazon as a dishonest seller. Let us admit that this cost does not vary over time. In particular, let us assume that, similarly as in Mayzlin et al. (2014) different sellers can be characterized by different costs, but that these costs are constant. If we instead consider the benefits from writing promotional reviews, it is clear that these benefits are a function of the history of grades that evolves over time. Reviews that arise over time alter the average rating and, with that, increase or decrease the influence that a new review can have on the statistics that buyers are likely to consider when making their decisions.

Table 5: Regression Discontinuity estimates

Variable	Coef.	Std. Err.	z	P > z	[95 % C.	Interval]
avg_sameID	-.03354	.00638	-5.25	0.000	-.04605	-.021
var_sameID	-.01188	.00488	-2.43	0.015	-.02145	-.0023
n_rev_sameID	72.798	14.308	5.09	0.000	44.753	100.84
dist_time_sameID	.29431	.03471	8.48	0.000	.22627	.3623
length	23.372	6.9672	3.35	0.001	9.716	37.02

Note: this estimates refer to average values by book for the following variables
- avg_sameID: average rating from the same reviewer
- var_sameID: variance of rating, same reviewer
- n_rev_sameID: number of reviews from the same reviewer
- dist_time_sameID: distance measured in years from the first to the last review
- length: average length

Bandwidth chosen using algorithm by Imbens and Kalyanaraman (2012).

Summary statistics:

Variable	Obs	Mean	Std. Dev.	Min	Max
avg_sameID	476,342	4.28	.532	1	5
var_sameID	476,342	.809	.677	0	4
n_rev_sameID	476,342	342.4	1692.2	1	44557
dist_time_sameID	476,342	4.177	3.378	0	18.16
length	476,342	686.87	788.145	0	32729

The clearest example is the case of stars. Depending on the average at the moment of writing, a high grade can increase the number of stars that will be assigned to the product. Our empirical strategy goes in this direction. We want to see whether the possibility of increasing the number of stars is a factor that helps explaining the rating assigned in a review. First, we have that it is easy for a seller to understand the extent that her review can alter the rounded rating. Second, this information is in principle available also to any unbiased reviewer, but in this case it is hard to imagine why they should find it relevant. Therefore we want to test the hypothesis that the seller, while keeping track of the evolution of the rating and the number of reviews, is ready to take advantage of those cases in which the number of stars can be easily manipulated. It is clear that we are considering only one of the strategies that sellers may be implementing in order to modify buyers' perception of their products. Still, we try to take advantage of the fact that in this context the ability to manipulate information varies over time and depends on factors that are not perfectly controlled by the seller, like in particular the grades of unbiased users. We therefore propose a regression model in which the overall grade is regressed on a set of observables at the time of the review and on a set of reviewer characteristics. Among the observables, we are particularly interested on a variable that measures the ability of the review to alter the number of displayed stars. The estimating equation will be as follows:

$$r_i = \alpha + \beta gain\_if5_i + \gamma B1_i + \delta B2_j + \theta B3_r \qquad (2)$$

Where r_i is the rating attributed in review i by reviewer r for the product j. The variable $gain\_if5_i$ will describe the number of stars that can be gained in case the overall grade in equal to 5. This variable depends on the average at the time of the review and on the number of reviews. The closer (to the left) the average is from the cutoff, the higher $gain\_if5_i$ will be. Similarly, a lower number of reviews will typically be associated with higher levels of $gain\_if5_i$. $B1_i$ contains controls that pertain to the single review, like the average grade and number of reviews before that review. $B2_j$ will include characteristics of the book, ideally we would include a measure of quality. For the moment we use the total number of reviews. Finally $B3_r$ refer to controls about the reviewer's characteristics like the total number of reviews written over time, the average rating assigned and the variance. We are interested in the correlation between the rating and the number of stars that can be gained. With this model we want to capture the kind of opportunistic calculations that the seller may do.

Table 6: Estimation Results for Equation (1)

VARIABLES	(1) Whole sample	(2) More than 50 rev	(3) More controls
gain_if5	0.282***	0.180***	0.101***
	(0.00793)	(0.00959)	(0.00665)
n_rev	2.22e-06	3.15e-06	-1.57e-05**
	(5.74e-06)	(6.12e-06)	(7.66e-06)
avg_bef	0.286***	0.474***	-0.162***
	(0.00748)	(0.0119)	(0.00929)
var_bef	-0.00656***	0.0108**	0.231***
	(0.00221)	(0.00497)	(0.00339)
stars_bef	0.0820***	0.0144*	0.0351***
	(0.00655)	(0.00784)	(0.00520)
n_rev_end	2.69e-05***	1.17e-05**	9.25e-06
	(5.31e-06)	(4.59e-06)	(7.22e-06)
n_rev_sameID	1.14e-06***	4.52e-06***	4.71e-06***
	(1.28e-07)	(3.63e-07)	(2.21e-07)
avg_sameID	0.949***	0.929***	0.836***
	(0.00172)	(0.00331)	(0.00157)
var_sameID	0.0650***	0.0664***	0.0596***
	(0.00300)	(0.00544)	(0.00315)
dist_time_sameID	-0.00506***	-0.00625***	-0.00274***
	(0.000246)	(0.000415)	(0.000252)
avg_end			0.751***
			(0.00713)
var_last4			-0.310***
			(0.00181)
Constant	-1.417***	-1.819***	-1.967***
	(0.0198)	(0.0407)	(0.0247)
Observations	4,045,659	2,138,477	3,340,851
R-squared	0.31	0.309	0.353
σ_{gain_if5}	0.213	0.094	0.114

Results

Table 6 presents the results of the estimation of the equation described above. Columns (1) and (2) refer to the same model, but consider different samples. In column (3) we included two more variables.

The first specification is done using the whole sample of reviews that were preceded by at least another review (we cannot calculate the variable average_before for the first review). Controls include the observables that are likely to influence reviewer's opinion regarding the quality of the book. As described above, we are interested in the coefficient of the variable called gain_if5. When we include the whole sample our results show a non negligible and statistically significant effect of the variable on the overall rating. In particular, 1 standard deviation change in gain_if5 explains approximately 1.5% variation in the overall rating. This effect is positive also when we only consider books with more that 50 reviews. This result is in line with the findings by Mayzlin et al (2014): books with a higher number of reviews are likely to be have been published by bigger firms that could be less prone to manipulate reviews.

In the third specification we include two more explanatory variables. We use avg_end, the average rating after all recorder reviews, as a proxy for quality and we exploit var_last4 to look for evidence in favor or coordination among reviewers. The variable var_last4 indeed measures the variance of rating of the four reviews that preceded the one we consider. Importantly, this variable forces us to consider only reviews that are at least fifth in the sequence of all reviews for a book. There are three interesting elements to underline about Column(3). The coefficient for gain_if5 signals now the presence of weak correlation. The coefficient for average_before switches sign, likely as an effect of avg_end. Finally, the coefficient for var_last5 is negative. This last point suggests the possibility that sellers tend to submit more reviews in a short time period.

Conclusions

Preliminary results presented above suggest that on Amazon.com a fraction of sellers may be using fake reviews as an way to manipulate information available to buyers. In particular we showed that this activity may be aimed at increasing the number of stars associated to the products they sell. To show this we used data on reviews. First we aggregated reviews by book and ran tests aimed at identifying the presence of selection. These tests are McCrary test on the density of observations around cutoff levels and regression discontinuity estimates of jumps in average books characteristics around these thresholds. In both cases we found weak signals that point towards manipulation. We then moved to consider data on single reviews. In this case we exploited variation in the ability to manipulate the number of stars and measured the response of ratings. We found a positive correlation between the ability to manipulate the number of stars and the assigned rating. In order to claim causality we need our explanatory variable to be randomly determined by the stream of reviews written by unbiased users. This may not be the case in presence of sellers producing more reviews with the aim of improving the number of stars.

There are a number of strategies we could implement to improve on these results. First, we could follow Mayzlin et al (2014) and identify groups of sellers that are likely to respond to different incentives. One possibility is to compare books that are published by small vs. big firms. Amazon.com allows writers to self publish their books. These are books that entirely depend on Amazon.com website for their sales and are therefore those who would benefit the most from any sort of promotional activity done on the website. We could test the hypothesis that for these books the explanatory variable we defined above is a stronger predictor of the rating. Another possibility would be to study groups of products that are sold by the same seller. Amazon.com indeed allows independent shops to be created inside its website. We could look for correlation in ratings assigned to products sold by the same seller and see how it varies over time.

Finally, this dataset could be improved in order to study also different questions that concern the topic of review manipulation. First we could explore the effect of competition on ratings. We could identify product categories in which competition among sellers is likely to be strong and study the stream of reviews written for these items. Another interesting question concerns the response of buyers to manipulation. Do consumers take into account the possible presence of manipulation when taking their decisions? To address this question we could compare the reaction of demand to changes in stars rating for categories of products that are more or less likely to suffer the presence of fake reviews. There could be three possible dimensions to explore. The first one based on the number of reviews: when the number of reviews is higher, manipulation is harder. The second dimension refers to the type of product. Ratings for products that face weaker competition inside the market are less likely to be manipulated. The third dimension pertains the seller. Seller whose brand is more known and that rely on online sales for a smaller fraction of their revenues will likely produce less fake reviews. In all this cases buyers should react more to changes in the rating. To run this type of analysis we would need data on sales. Unfortunately this typology of data is not available. Two things are available though. First, Amazon.com provides sales ranks per category that are updated every hour. Second, in some cases (usually for less popular goods) Amazon can run out of stocks. As these data are available online, we could explore the possibility of using these variables as proxy of sales.

References

Anderson M. and J. Magruder (2012) *Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database* The Economic Journal

Chevalier J. and D. Mayzlin (2006) *The Effect of Word of Mouth on Sales: Online* Journal of Marketing Research

Chintagunta P., S. Gopinath and S. Venkataraman (2010) *The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets* Marketing Science

Imbens G., K. Kalyanaraman (2012) *Optimal Bandwidth Choice for the Regression Discontinuity Estimator* The Review of Economic Studies

Luca M. (2011) *Reviews, Reputation, and Revenue: The Case of Yelp.com* Working Paper

Luca M. and G. Zervas (2015) *Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud* Working Paper

McAuley J., R. Pandey, J. Leskovec (2015) *Inferring networks of substitutable and complementary products* Knowledge Discovery and Data Mining

McAuley J., C. Targett, J. Shi, A. van den Hengel (2015) *Image-based recommendations on styles and substitutes* SIGIR

McCrary J. (2008) *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test* Journal of Econometric

Resnick, P. and R. Zeckhauser (2002) *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. The Economics of the Internet and E-Commerce.* Advances in Applied Microeconomics

Resnick P., R. Zeckhauser, J. Swanson, K. Lockwood (2006) *The value of reputation on eBay: A controlled experiment* Experimental Economics

Figure 1: Best Sellers books on Amazon.com

Figure 2: How informations are presented to the customer

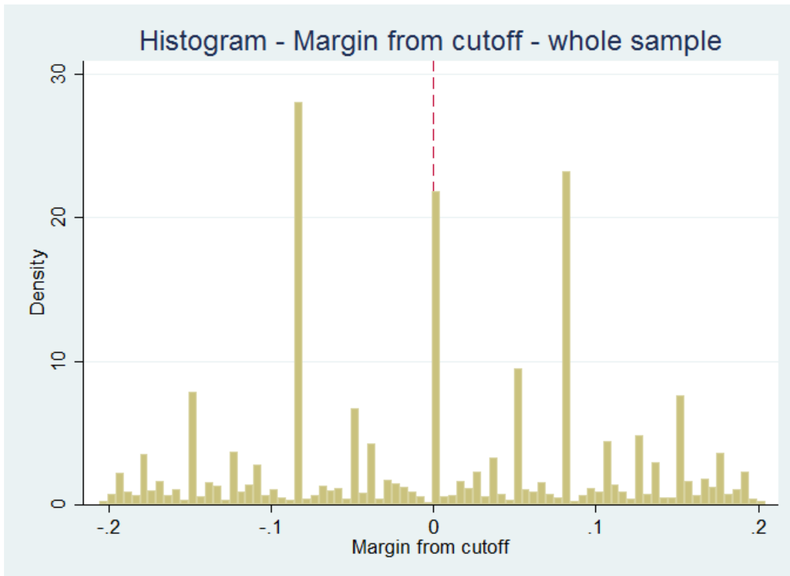

Figure 3: Density of observations around cutoff levels.

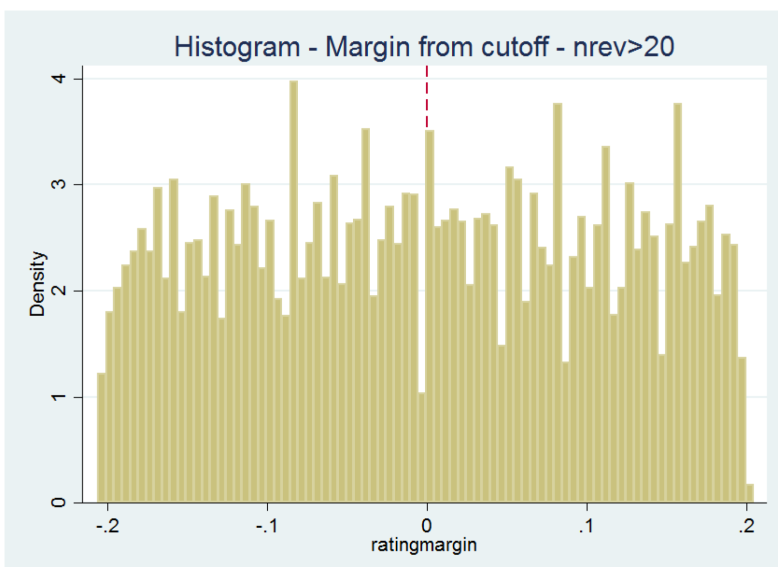

Figure 4: Density of observations around cutoff levels.

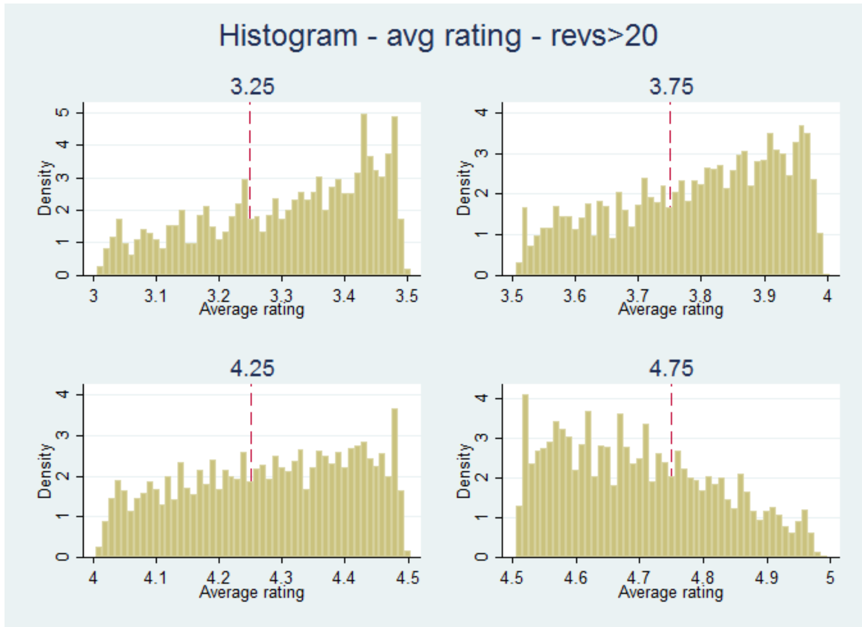

Figure 5: Density of observations around cutoff levels.

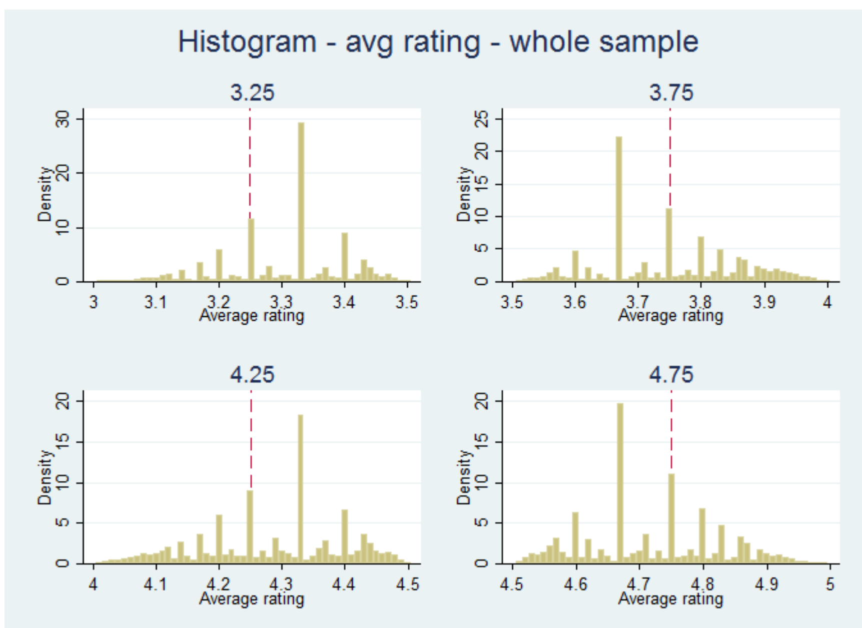

Figure 6: Density of observations around cutoff levels.

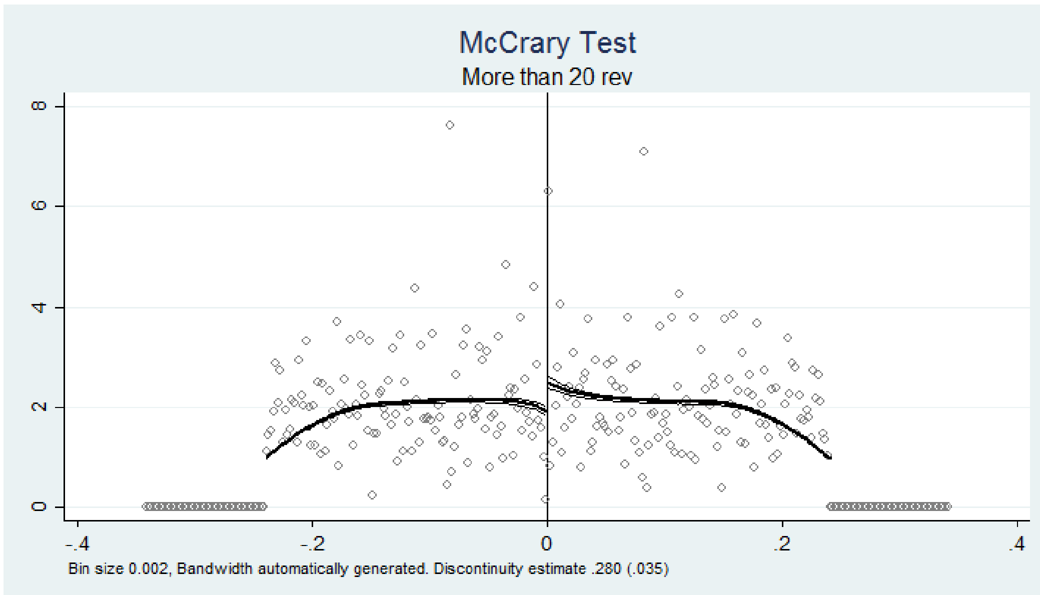

Figure 7: McCrary test.

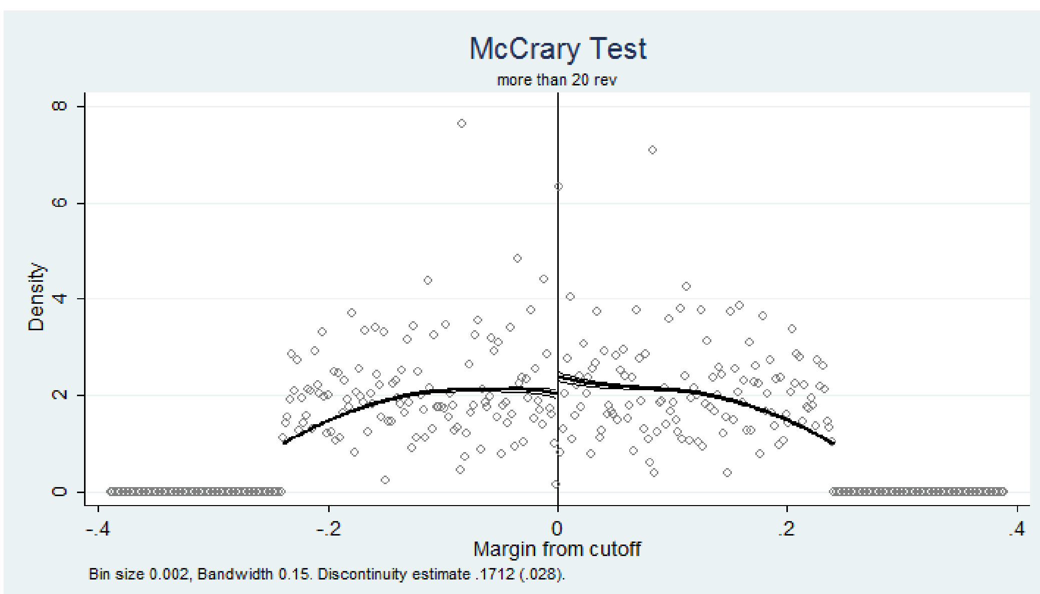

Figure 8: McCrary Test, specification 2.

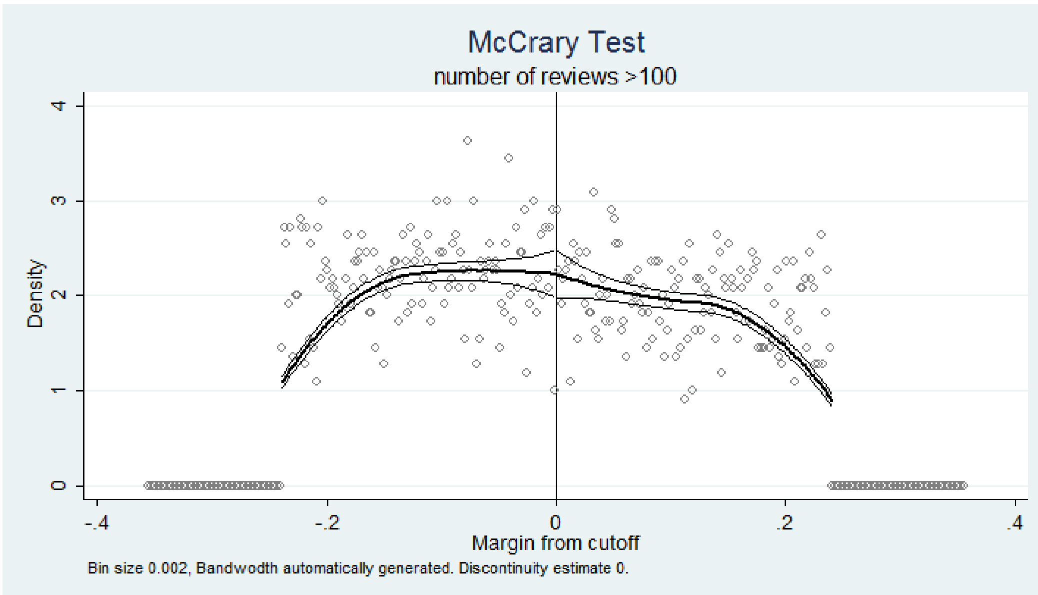

Figure 9: McCrary Test, specification 3.

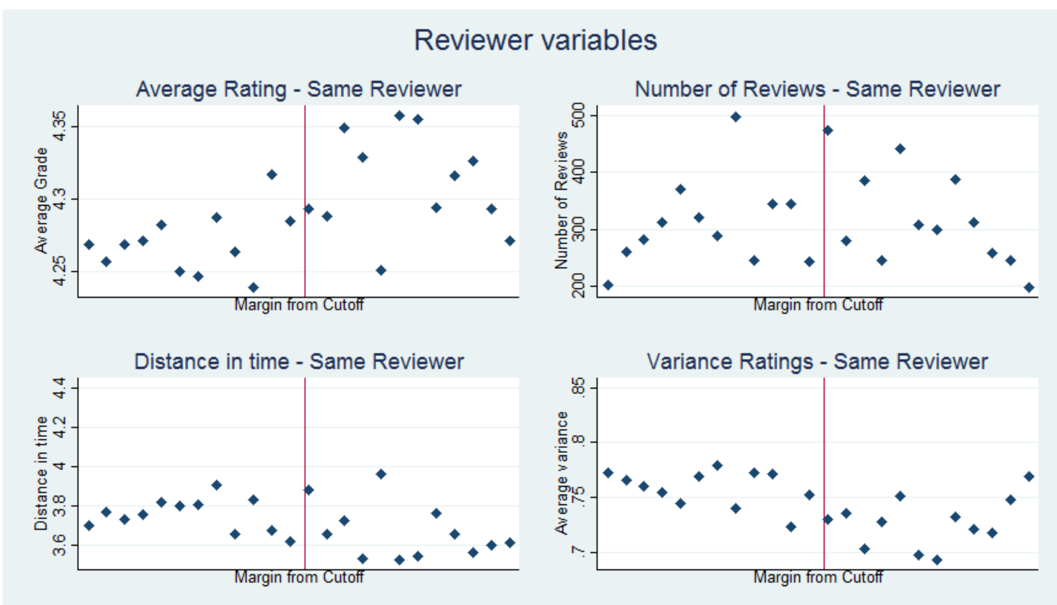

Figure 10: Scatter plot around cutoff