Response to Review Team’s Comments

Manuscript #: RA-JAIS-19-0085

Title: Examining the Impacts of Airbnb’s Review Policy Change on Listing Reviews

**Responses to the Senior Editor**

*In the context of sharing economy, the authors investigate the effects of the simultaneous review system on review generation. The simultaneous review system allows a host and a guest to view each other’s review only after both reviews are posted. The authors argue that compared to the old asynchronous review system, the simultaneous review system helps generate more reviews, reduce reviews’ sentiment scores, and increase the dispersion of sentiment scores. Data were collected from Airbnb before and after the company rolled out the simultaneous system. Using multiple analytical methods, the authors find considerable support to their proposed hypotheses.*

*The paper was sent to two highly qualified researchers for review. Both reviewers have substantial expertise in related domains and have published in top journals. Their reviews are largely consistent. On one hand, they think the paper is well written and the research topic is relevant, timely, and interesting. On the other hand, they have pointed out several issues on theorization, empirical designs, and data analysis. While R1 recommends major revision, R2 recommends reject and resubmit. My assessment of the paper is similar to the reviewers’. I applaud the authors for studying this practically important phenomenon. Yet, I agree that the theorization and analysis need to be strengthened. I am willing to offer the authors an opportunity to revise the paper.*

*I am not going to repeat all the concerns raised by the reviewers. Instead, I will focus on a few issues that the authors need to pay particular attention to.*

Thank you very much for organizing an excellent review panel for our paper. Both reviewers as well as you have provided very detailed and constructive comments to revise and improve our paper. We really appreciate the opportunity to strengthen the paper’s theorization and analyses and address all concerns raised by the review panel. We are pleased to submit a revised version of the paper. For ease of reference, the text of your original comments is repeated here in italics in the textbox, followed by our response.

In this revision…

All comments and suggestions raised by you and the reviewers are indeed necessary in order to ensure that our manuscript can meet the publication standard of *the Journal of the Association for Information Systems*. We sincerely hope that the revised paper is satisfactory and can make a nice contribution to the journal. Thank you!

1. *R1 has identified the weak connection between motivation and analysis. The authors argue that the simultaneous review system can mitigate review bias, but they do not present data to precisely measure bias. For example, although their data analysis shows the simultaneous review system increases topic variety and reduces sentiment scores, these findings do not necessarily mean that the previous review system was biased. It is possible that the previous review system generated fair reviews, while the current system leads to biased reviews.*

Mitigating review bias 🡪 review informativeness…

1. *The theoretical development of this study needs to be strengthened. First, the logic behind the hypotheses is relatively intuitive and lacks theoretical depth. Second, I am not convinced why the simultaneous review system can increase the amount of information content (H1). It is possible that in the asynchronous review system, reviewers want to please each other so write more. Third, do the authors expect the same pattern for topics, depth, and objectivity (H1)? After all, these aspects are different and should be justified separately. Fourth, as information amount and sentiment scores are very much different, what are the fundamental difference in the theoretical backing for H1 and H2?*

Rewrite and strengthen all hypotheses development…

1. *One major concern about data analysis is the long-time window (from 2010 to 2016). While rich data may give you more insights, but may unfortunately suffer from more potential confounding influences. How can you convince readers that the difference in reviews is only caused by the 2014 new system? Many things can have happened in the 7 years. It is difficult to attribute the change to only one system change. R2 has suggested that a shorter time window may yield more convincing results.*

Thank you for your comment. We repeated the analysis using quarterly data limited to 2013, 2014, and 2015. The results are reported in table xx. According to these results, the findings are all the same except for Depth. Furthermore, we created a monthly panel data (instead of quarterly panel data) and replicated the analysis using a 2-year period of time (from April 2013 to May 2015). We report the results in table xx. Our findings are still the same as reported previously, with the exception of Depth.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. The Results of RDD Analysis for Quarterly Data (2013, 2014, 2015)** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| After\_Policy | **0.630\*\***  **(0.231)** | 3.210  (2.748) | **0.032\***  **(0.015)** | **-0.107\*\***  **(0.040)** | **0.275\*\***  **(0.116)** | **0.140\*\*\***  **(0.049)** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. The Results of RDD Analysis for Monthly Data (2013, 2014, 2015)** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| After\_Policy | **1.273\*\*\***  **(0.117)** | 1.670  (1.327) | **0.031\*\*\***  **(0.007)** | **-0.058\*\***  **(0.019)** | **0.315\*\***  **(0.044)** | **0.165\*\*\***  **(0.030)** |

1. *In relation to the above point, the authors have admitted that there are some other changes in the review system (p. 11) and argue that those changes may not affect their results. However, the reviewers are not convinced.*

Will provide more evidence…

1. *R2 have raised several issues on the data analysis results. Some data patterns seem to imply that the change in reviews is not only simply caused by the introduction of the simultaneous review system. The authors must address these issues in their revision.*
2. *Both reviewers are concerned about your choice of IV. I also share the same concern. Please justify why the IV is legitimate.*
3. *The authors group textual features of reviews into four categories: topics, depth, objectivity/subjectivity, and sentiment/polarity/emotions. There are some potential overlap between objectivity/subjectivity and sentiment and overlap between depth and objectivity. The authors should make a strong case for this categorization and how the four textual features match informational content and personal opinion.*
4. *The hypotheses only present main effects, but data analysis includes interaction effects and temporal effects. This makes your findings look random and not theoretically-driven. Please see R1’s comments. Can you add corresponding hypotheses and test the hypotheses in the data analysis?*

*In sum, the reviewers have provided extremely useful feedback. The issues they have pointed out are sharp and pertinent. However, as the dataset is rich and the topic is interesting and relevant, I see considerable potential in the paper. I am hopeful that the authors can take advantage of reviews to improve the paper. Good luck!*

We would like to thank you again for your encouraging words and the time and effort you devoted to the manuscript. Your insightful comments have helped us to produce a significantly better manuscript compared to the original submission. We look forward to your feedback on the revised submission.

**Responses to Review 1**

*This paper studies the impacts of Airbnb’s review policy change (asynchronous vs. simultaneous review system) on listing reviews. Overall, the paper investigates a relevant and interesting question, and is generally well written. However, there are some important issues in the paper:*

We greatly appreciate your time and effort devoted to reading our manuscript. Thank you very much encouraging our efforts and offering us the opportunity to further revise our paper. Your insightful comments and constructive suggestions have helped us to improve the quality of our paper significantly. Based on your detailed referee report, we have substantially revised the entire paper in order to address your concerns. Please see our detailed responses below. We sincerely hope that this revision would be satisfactory. For ease of reference, the text of your original comments is repeated here in italics in the textbox, followed by our response.

*1. My first concern is about the paper’s weak connection between motivation and analysis. The central argument of the paper is that simultaneous review system can lower reciprocity and mitigate rating bias. But how can the proposed empirical model and analysis derive such a conclusion? For instance, does the significance of certain variables show such findings? This thus needs much more discussion.*

Mitigating review bias 🡪 review informativeness…

*2. The literature review is not comprehensive. First, the key aspect of the paper is on rating bias. There are plenty of papers studying rating bias, but the authors did not provide a comprehensive review of them. Particularly, the authors may refer to a recent ISR paper (Lin et al. "Free Product Sampling and Rating Bias", Information Systems Research, 2019, 30(1), pp. 260–275) which also talks about reciprocity and bias, and reviews many possible sources of biases in ratings. Second, since the paper is positioned in the sharing economy domain, it should also provide a review on recent sharing economy papers published in IS, marketing, etc.*

Done…

*3. As to the empirical section, the authors seemed to provide a rich set of analysis (RDD, single-group interrupted time-series, fixed effects, IV, matched sample, etc.) while, unfortunately, did not discuss the purpose of such usage (e.g., what issue each approach can address). At the same time, I would like to highlight that, having a large set of analysis does not guarantee the rigorousness of estimates. Because approach A may address issue a, it may not be able to address issue b. Thus the single isolated use of many approaches may still be subject to estimation biases. The authors may need to discuss the limitation of each approach used.*

Thank you for your comment. Our study design is informed by (Cavusoglu et al. 2016). However, as you noted, we did not clearly explain what types of issues each method would address. Therefore, we revised the manuscript to explain why we used each method.

*4. Related to the above point, though the authors proposed a rich set of analysis, a consistent issue that remained was the lack of a control group. After all, that was a platform-level policy change that applied to everyone. Thus, the authors only had differences across time (before vs. after), but not across reviews/sellers/buyers, to conduct a difference-in-differences estimate. I understand that this cannot be completely solved given such a setting, I still suggest the authors to make some attempts to lower the bias and acknowledge this limitation.*

Thank you for your comment. As you noted, all of the listings in our dataset were treated exactly at the same time. This prevents us from running a difference-in-difference analysis. However, we argue that the regression discontinuity design (RDD) can be useful in our case to obtain unbiased estimates for the coefficients. Primarily, we note that the platform-level change is in fact exogenous as the policy change was enforced by Airbnb and the reviewers did not play a role in this change. Secondly, we argue that regression discontinuity in time (RDiT) is capable of producing unbiased estimates because it compares the reviewing behaviors in a relatively short window of time right after the policy change with the reviewing behaviors in a short window of time right before the policy change. RDiT assumes that the listings and reviewers do not change in a short window of time (or at least any changes are uncorrelated with the policy change). According to (Cavusoglu et al. 2016; Glass 1997), interrupted time series design (which includes RDiT) is a viable alternative to true experiments due to the fact that it has many pre- and postintervention observations (in our case many listings/ reviews), and this permits it to distinguish true intervention effect from a time trend or seasonality. Furthermore, (Glass 1997) points out that this design has become the standard method of causal analysis in applied behavioral research.

We note that the same model setup has been used in (Cavusoglu et al. 2016), which was also about a platform-wide policy change regarding Facebook’s privacy policy. Although this methodology is regarded suitable for such research design, we certainly agree that we need to test the robustness of our analysis and acknowledge the limitations of this methodology in our study. To address the limitations of the study and study the robustness of our method, we followed these steps:

1. We limited our analysis to listings that existed before and after the policy change. This step would remove the listings that were removed because of the policy change or the new listings that were encouraged by the new policy change.
2. We shortened the time window before and the time window after the policy change in the RDiT analysis. This would make our assumptions about no listing-level and policy-level change during the study period more reasonable as we are now assuming that there were no changes in a shorter period of time.
3. We used robust standard errors and standard errors clustered at the listing level in our estimations in section 5 in the manuscript.
4. Our main analyses were done over listings that existed before and after the policy change. We repeated our analysis using a matched sample of listings that existed only before and only after the policy change. It could be argued that some hosts did not like the new policy change and therefore decided to exit the platform (and remove their listings) after they learned about the policy change. In interrupted time-series literature (Forastiere et al. 2016), this is known as mortality (i.e. some listings were removed after the intervention), and is considered a threat to the internal validity of the study (Cavusoglu et al. 2016). Given that we are interested in offering generalizable insights, we decided to test if our results would have been the same if those listings were not removed from the platform. To conduct this analysis, first we identified 65 listings that were only listed before the policy change. Then we found the most similar listings that only existed after the policy change. We then compared the values of the variables of interest for the matched groups. We found that our findings are still valid.
5. Another threat to the internal validity of our study would be related to the presence of time trends and seasonal effects. One may argue that the use of Airbnb is seasonal and has a time trend. For instance, people travel more during the Holiday season and therefore more reviews would be generated during this time of year. Also, Airbnb has been expanding its market in the U.S. during the time of our study (time trend). Although interrupted time-series analysis is claimed to be robust in presence of seasonality and time trend, we further study the robustness of our results by using structure-free analysis. In particular, we used time-series analysis based on data generated before the policy change to establish a baseline for the variables of interest. Given that these time-series models are based on data generated prior to the policy change, the forecasts made by these time-series models would mimic the patterns as if the policy change was not rolled out throughout the platform.

The New Analysis:

1. T-test Results:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. T-test Results for 6 Months before and after Policy Change** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| Mean Before Change | 4.868 | 35.999 | 0.380 | 0.946 | 0.771 | 0.425 |
| Mean After Change | 5.707 | 37.028 | 0.390 | 0.903 | 1.002 | 0.489 |
| t-value | **4.068\*\*** | 1.138 | **2.899\*** | **-7.027\*\*\*** | **3.349\*** | **3.590\*\*** |

1. RDiT Analysis:

Regarding our RDiT analysis, we revised the entire section to better report our methodology and communicate our results. First, we decided to use a month as unit of time instead of a quarter. Second, we limited our RDiT analysis to 12 months before and 12 months after the policy change. This means that we used monthly data from April 2013 through June 2015 (excluding April, May, and June of 2014) for RDiT analysis. The data from April 2013 to March 2014 are generated before the policy change and the data from July 2014 to June 2015 are generated after the policy change. Although we are including 12 months of data before the policy change and 12 months of data after the policy change in RDiT analysis, we use Imbens-Kalyanaraman method (Imbens and Kalyanaraman 2009) to determine the optimal bandwidth for local linear regression in our RDiT analysis.

First, we visually inspect the impact of policy change on each one of our target variables:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. The Results of RDD Analysis for Monthly Data** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| After\_Policy | **2.843\*\***  **(0.607)** | 1.159  (3.034) | **0.050\*\***  **(0.008)** | **-0.090\* (0.034)** | **0.291 (0.139)** | **0.114\*\*\***  **(0.026)** |
|  | -1.098 (0.668) | 2.211 (3.336) | **-0.032\***  **(0.010)** | 0.042 (0.040) | 0.038  (0.057) | -0.015  (0.012) |
|  | -0.226 (0.213) | 0.471 (1.067) | **-0.009\***  **(0.003)** | 0.014 (0.012) | 0.011  (0.009) | -0.001  (0.001) |
|  | -0.014 0.020) | 0.016 (0.100) | <-0.001  (<0.001) | 0.001 (0.001) | <0.001  (<0.001) | <0.001  (<0.001) |
| After\_Policy | 1.088 (0.736) | -2.791 (3.677) | **0.036\***  **(0.011)** | -0.046 (0.044) | -0.069  (0.058) | 0.019  (0.014) |
| After\_Policy | 0.034 (0.263) | -0.429 (1.315) | 0.004  (0.004) | -0.007  (0.015) | -0.009  (0.009) | 0.003  (0.002) |
| After\_Policy | 0.048 (0.028) | -0.040 (0.142) | **<0.001\***  **(<0.001)** | -0.002 (0.001) | -0.001  (<0.001) | <0.001  (<0.001) |
| Adjusted R-Sq. | 0.914 | 0.738 | 0.916 | 0.840 |  |  |
| F-stat  [p-value] | **17.740\*\* [0.007]** | 5.435 [0.061] | **18.020\*\***  **[0.007]** | **9.249\* [0.024]** | 16.22  [<0.001] | 14.54  [<0.001] |

Thank you for your comment. [Reza’s Notes: One possible solution will be just to use fixed effects to determine if the effects are short-term or long-term.]

*5. IV validity test: I don’t quite understand the authors’ IV validity test. I think more common tests should be the over-identification test, among others.*

Thank you for your comment. The IV validity test was done to examine if the IV is significantly associated with the endogenous variable. The coefficient in the analysis was significant and positive, confirming that our choice of IV is at least associated with the variable it is instrumenting (the endogenous variable in the original model). As you noted, we should also test if the IV is uncorrelated with the error terms. The exogeneity of the IV is the null hypothesis in Sargan test. However, this test can only be used if there are more instruments than endogenous variables. In our case

Overidentification test: “One crucial property of instruments is that they ought to be uncorrelated to the regression error terms $u$. Instrument exogeneity is set as the null hypothesis of this following test with the alternative hypothesis being that the instruments are endogenous. This test can only be applied if you have more instruments than endogenous variables. It is therefore sometimes also called the test for overidentifying restrictions.”

*6. Interaction effects: The authors proposed only 3 hypotheses about the main effect of policy change, but suddenly provided the test of interaction effects. The authors may need to improve the paper structure to make it better organized and more coherent (theory-empirics).*

Will do…

**Responses to Reviewer 2**

*This study investigates the impact of the simultaneous review system on the informational content and the personal opinions of guest reviews. I enjoy reading this manuscript, and I appreciate the unique data access the authors created and the number of studies conducted for robustness checks. However, I have some considerable concerns with the theoretical contribution, theoretical development and the validness of empirical analysis. Below I provide a detailed list of my comments, suggestions, and concerns. I wish the authors luck with this research.*

We greatly appreciate your time and effort devoted to reading our manuscript. Thank you very much for your insightful comments and constructive suggestions, which have helped us to improve the quality of our paper significantly. Based on your detailed referee report, we have substantially revised the entire paper in order to address your concerns towards the theoretical contribution, theoretical development, and the validness of empirical analysis. Please see our detailed responses below. We sincerely hope that this revision would be satisfactory. For ease of reference, the text of your original comments is repeated here in italics in the textbox, followed by our response.

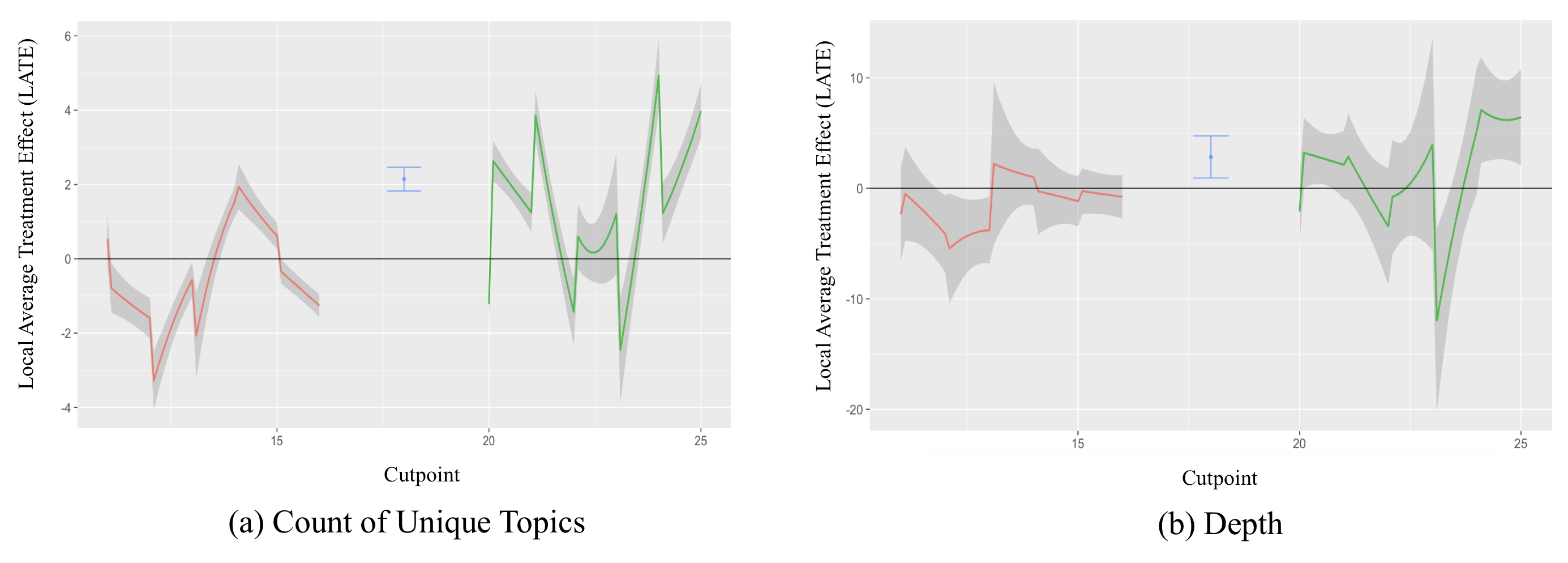
*First, I am not sure whether this work has sufficient contributions beyond previous studies mentioned in the paper (Bolton et al. 2013; Fradkin et al. 2018; Preserpio et al. 2018). The authors argued their contributions as testing the impacts of the simultaneous review system via large-scale observational data over an extended period of time, however, this time-series design may suffer from potential issues of a time trend or seasonality which were not clearly distinguished from the true effect.*

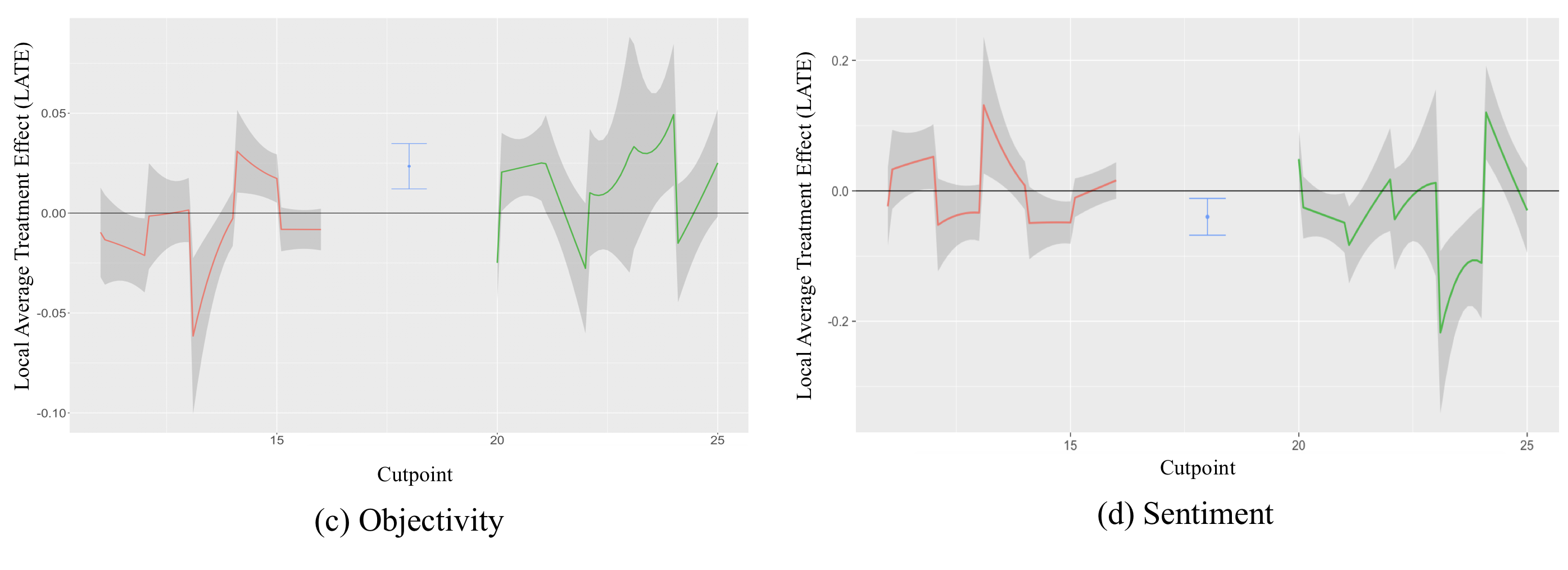
Strengthen theoretical contribution, and explain how a time trend or seasonality is controlled in empirical analyses…

*(a). Results in Figure 1 show that Topic\_Count and PD\_sentiment are not robust which suggest the effect of Topic\_Count and PD\_sentiment may not due to the policy change. Strictly speaking, the confidence intervals of other variables also overlap with zero which didn’t suggest robust results of other variables. Besides, this figure is not clearly presented. What is the X-axis in the figure? Which hypothetical time is used because the authors use “these hypothetical time stamps” on page 24?*

Thank you for your comment. We should have clarified the information provided in Figure 1. In regression discontinuity design, the assumption is that the treatment effect only occurred at the cutoff (July 2014 in our study). To test this assumption, we perform placebo tests by re-estimating the treatment effects for different cutoffs. Preferably, we would like to see that the estimates are close to zero for placebo tests. This would indicate that the treatment effects were only present if we use the real cutoff and would disappear if we change the cutoff. Figure 1 in the original manuscript displays the effect at the cutoff in blue, effects at lower cutoffs in red (as if the policy change was implemented some time before July 2014), and effects at higher cutoffs in green (as if the policy change was implemented after July 2014). All effects are bounded by a confidence interval (95% confidence interval). In these plots, the horizontal axis represents the time interval (quarter or month) and the vertical axis displays the estimate and the confidence interval for the treatment effect.

According to Figure 1 in the original manuscript (also copied below), except for “Count of Unique Topics” and “Percentage Difference of Sentiment”, the placebo tests result in treatment effects near zero (i.e. the confidence intervals for the red and green lines overlap with zero for the majority of the placebo tests) for most of the placebo cutoffs. Therefore, these plots suggest some evidence against potential violations in treatment assignment for “Depth”, “Objectivity”,   
“Sentiment”, and “Coefficient of Variation of Sentiment”. For “Count of Unique Topics” and “Percentage Difference of Sentiment”, however, placebo tests result in treatment effects estimates significantly different from zero for the majority of placebo cutoffs. In both cases, the treatment effects estimates also change their signs as the cutoff changes. Therefore, the RDDiT results for “Count of Unique Topics” and “Percentage Difference of Sentiment” could be unreliable due to the potential violation in treatment assignment.





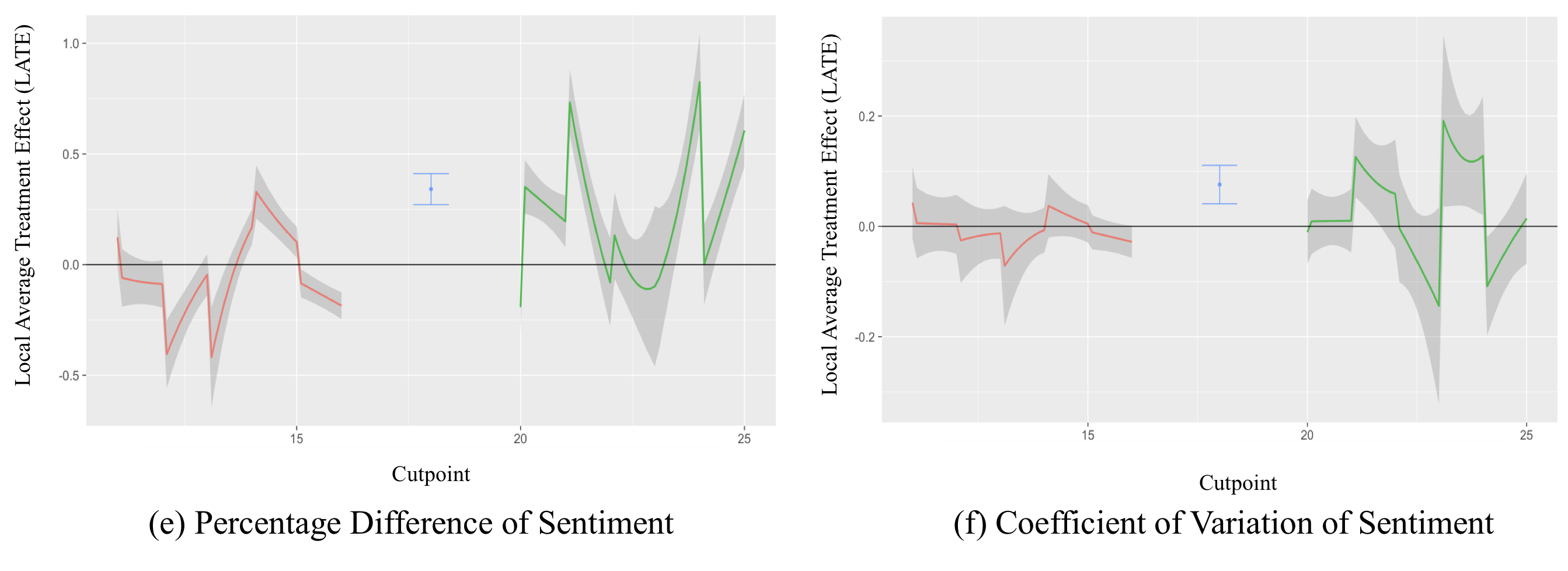
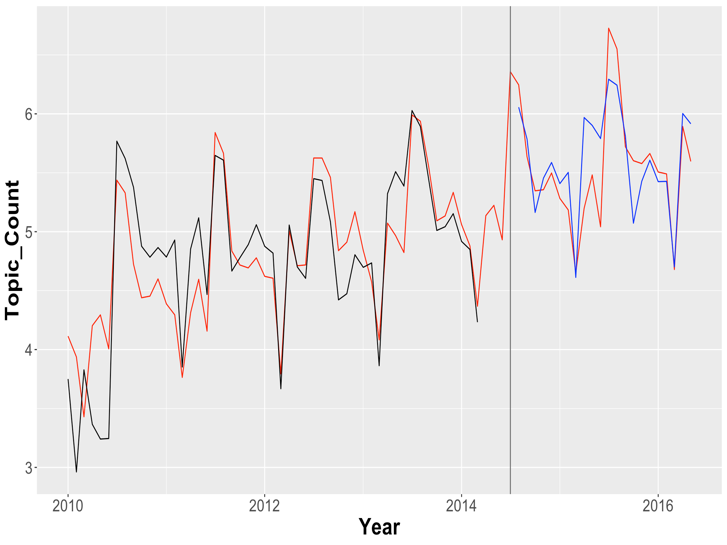


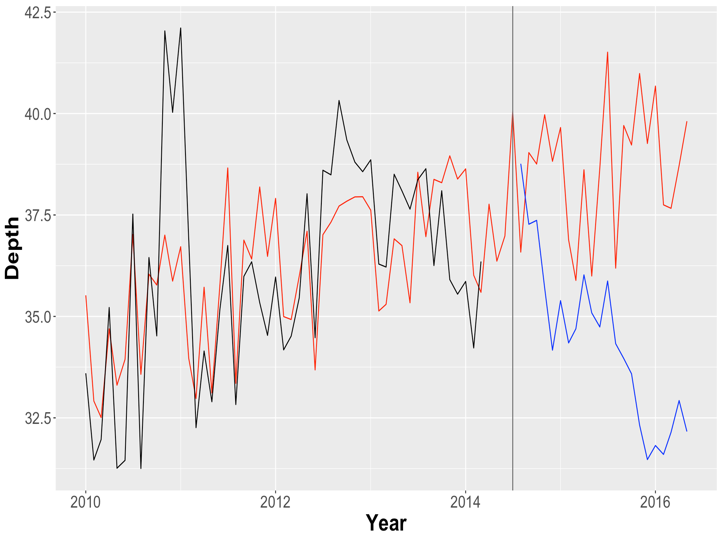
Figure 1. The Results of the Placebo Tests with Respect to Cutpoints (Hypothetical Timing of Policy Change)

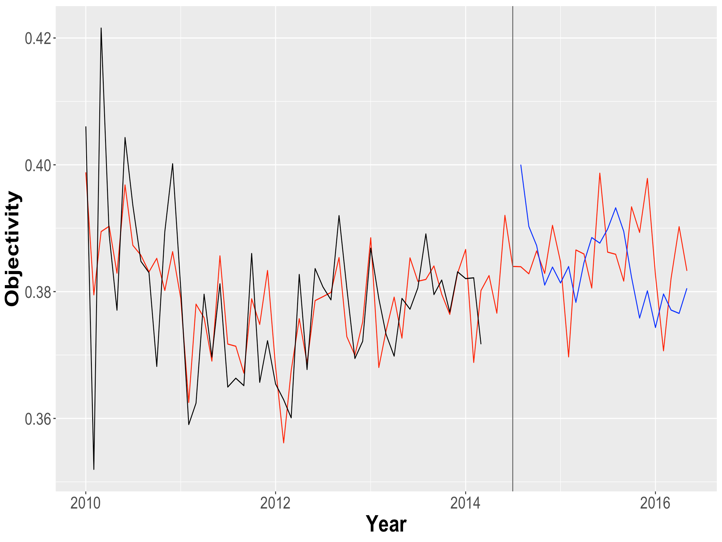
*(b) Figure 3 reinforces the existence of a time trend. If the impact were brought solely by the system policy change, we should expect that there is no trend of DVs before the policy change, and then after the policy was introduced, the effect started to be observed. However, from plot (a) we could see that there is a steady upward trend of the number of topics even before the policy change instead of a “sudden increase” according to the authors. It is not surprising why the authors didn’t get a consistent result for the number of topics in previous robustness check. This brings the concern that whether the increase of topic numbers was impacted by the simultaneous review system or other unobserved factors. Similarly, we could see the pattern for sentiment dispersion in plot (e) and (f) from 2012 to 2014. And for sentiment, we could observe a downward-trend from 2013.*

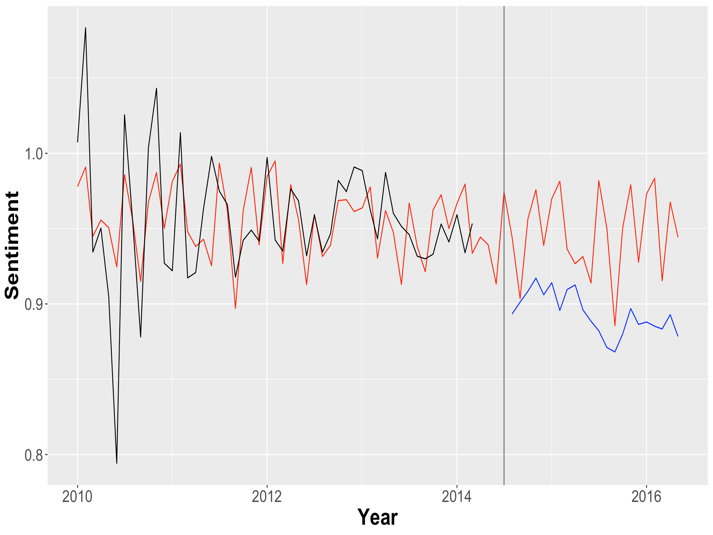
Thank you for your comment. We realized that our analysis that produced the plots in Figure 3 resulted in confusions about the results. In the manuscript, we reported that we used data after the policy change (data after July 2014) to build a time series model to forecast the values of the time series for the future months. The intention was to show if the policy change would result in a long-term effect or only in a temporary effect that would phase out after the first few months. According to the analysis reported in section 5.4 of the original manuscript, the impact of policy change on the informational content (Topic\_Count, Depth, and Objectivity) is somewhat unstable (the trend reverses after a few months after the policy change). However, we concluded that the impact of policy change on constructs related to personal opinions (Sentiment, PD\_Sentiment, and CV\_Sentiment) is stable over time.

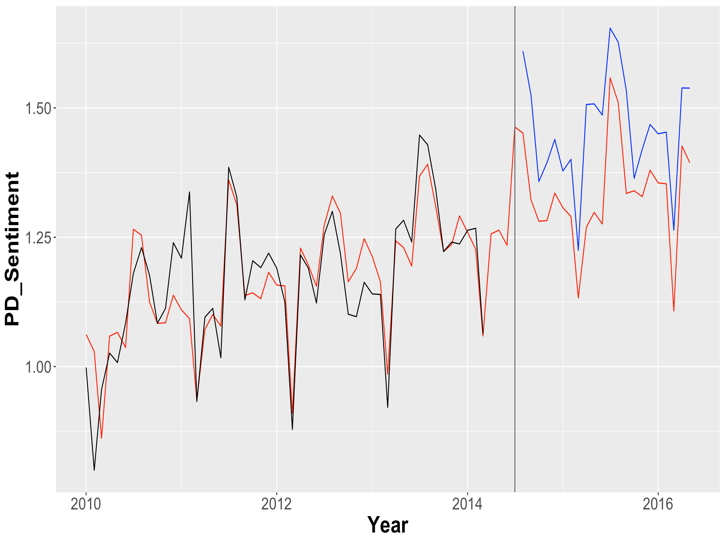
To show what the values would have been if there was not policy change, we repeated our time series analysis by using data from before the policy change to forecast the values of time series in the future months. We added the results of this analysis to Appendix xx (displayed the plots below).

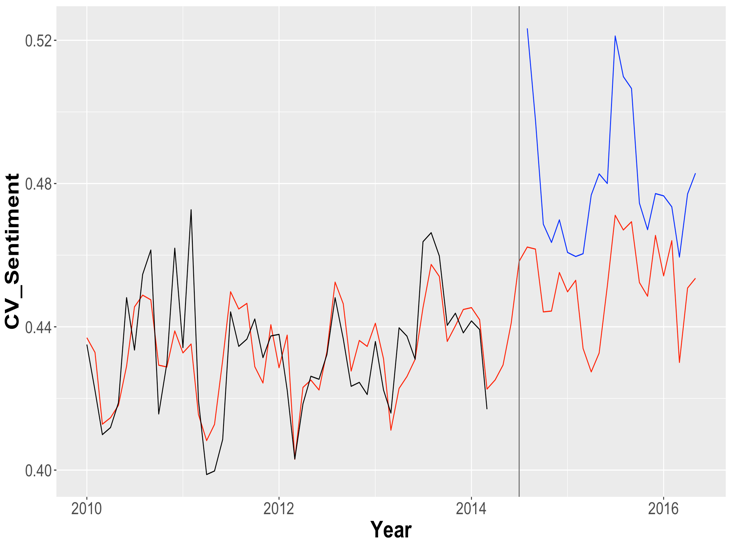












*Second, theory and hypotheses did not seem well organized and compelling. (a) The authors refer to reciprocity to explain impacts on informational contents and personal sentiments. However, the connections are not well established. Reciprocity implies that “reviewers can provide information more freely”, then they could either write longer or shorter. “Topics are less likely to be limited to those initiated by the first movers”. Any evidence from the pre-change period? What are the general topics covered by hosts and guests respectively? I am not sure that the two types of reviews would have many common topics. (b) The arguments of objectivity and sentiment are correlated with each other however not clearly distinguished. The authors argued that “both positive and negative feedbacks are disabled” in line 37 on page 9, however, in line 10 on page 10, they argued that “people are more likely to report negative feedback”.*

Rewrite the entire hypotheses section…

*Empirical setting and data:*

*4.1. The authors ignored the impact of the policy change of shortening review period from 21 to 14 days. Any details of the comparison of the DVs under the two review period from pre-change data?*

Using Fradkin’s paper to explain why shortening time window did not affect review behavior much…

*4.2. What’s the time window of listing active before and after change, column 5 in table 2? There is also a huge number of reviews difference before and after change. I would like to see some robustness tests of a shorter time period like a one or two-year time frame.*

Thank you for your comment. We repeated the analysis using quarterly data limited to 2013, 2014, and 2015. The results are reported in table xx. According to these results, the findings are all the same except for Depth. Furthermore, we created a monthly panel data (instead of quarterly panel data) and replicated the analysis using a 2-year period of time (from April 2013 to May 2015). We report the results in table xx. Our findings are still the same as reported previously, with the exception of Depth.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. The Results of RDD Analysis for Quarterly Data (2013, 2014, 2015)** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| After\_Policy | **0.630\*\***  **(0.231)** | 3.210  (2.748) | **0.032\***  **(0.015)** | **-0.107\*\***  **(0.040)** | **0.275\*\***  **(0.116)** | **0.140\*\*\***  **(0.049)** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. The Results of RDD Analysis for Monthly Data (2-year period)** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| After\_Policy | **1.273\*\*\***  **(0.117)** | 1.670  (1.327) | **0.031\*\*\***  **(0.007)** | **-0.058\*\***  **(0.019)** | **0.315\*\***  **(0.044)** | **0.165\*\*\***  **(0.030)** |

*4.3. Any specific example of how to measure the number of topics? The authors discussed the probability of a specific topic represented in a specific document, and then they use the top most represented topic. Is the probability of the top topic in the document the measurement of the number of topics?*

*4.4. Is there a specific reason why use quarter-year as the unit of time? Any robustness test of using month-year?*

**Authors’ Response:**

As you suggested, we repeated the analysis but using month as the unit of time. We also limited the data to 2013, 2014, 2015, and 2016. Table xx reports the results of RDD when we use monthly data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table xx. The Results of RDD Analysis for Monthly Data** | | | | | | |
|  | Topic\_  Count | Depth | Objectivity | Sentiment | PD\_  Sentiment | CV\_  Sentiment |
| After\_Policy | **1.183\*\*\***  **(0.109)** | **2.170\***  **(1.034)** | **0.030\*\*\***  **(0.006)** | **-0.056\*\*\***  **(0.012)** | **0.297\*\***  **(0.109)** | **0.114\*\*\***  **(0.026)** |
|  | **-0.147\* (0.051)** | **0.389**  **(0.535)** | **-0.009\*\***  **(0.003)** | 0.008  (0.008) | 0.038  (0.057) | -0.015  (0.012) |
|  | -0.001 (0.008) | **0.117**  **(0.082)** | **-0.001\***  **(<0.001)** | 0.001  (0.001) | 0.011  (0.009) | -0.001  (0.001) |
|  | **0.001\***  **(<0.001)** | 0.006  (0.004) | **<-0.001\***  **(<0.001)** | <0.001  (<0.001) | <0.001  (<0.001) | <0.001  (<0.001) |
| After\_Policy | 0.100  (0.053) | -1.074  (0.552) | **0.007\***  **(0.003)** | -0.012  (0.009) | -0.069  (0.058) | 0.019  (0.014) |
| After\_Policy | 0.005  (0.009) | **-0.067**  **(0.082)** | **0.001\*\***  **(<0.001)** | -0.001  (0.001) | -0.009  (0.009) | 0.003  (0.002) |
| After\_Policy | -0.001  (0.003) | -0.007  (0.004) | **<0.001\***  **(<0.001)** | <0.001  (<0.001) | -0.001  (<0.001) | <0.001  (<0.001) |
| F-stat  [p-value] | 74.55  [<0.001] | 94.71  [<0.001] | 6.853  [<0.001] | 93.39  [<0.001] | 16.22  [<0.001] | 14.54  [<0.001] |

*5.1. Any control variables used in specification 2?*

**Authors’ Response:**

Thank you for your comment.

*5.2. Any control variables used in specification 3?*

**Authors’ Response:**

Thank you for your comment.

*5.3.1. IV should be correlated with After\_policy which is the focal variable but not with the error terms. However, in the analysis, IV is correlated with the review count instead of the focal variable after\_policy. I am not persuaded that this is a valid IV. Besides, how is the neighborhood defined and how is the IV measured?*

**Authors’ Response:**

Thank you for your comment.

*5.3.2. Again please try different time window. Besides, any explanations of why there is a decreasing trend for depth?*

**Authors’ Response:**

Thank you for your comment.

*5.3.3. How do you find the neighbors? How many neighbors for each node? Which factors are used for matching? Any guesses these listings do not exist anymore? Any summary characteristics of these listings. It is possible that these listings are systematically different from other listings. If that is the case, it is difficult to argue the matching is valid.*

**Authors’ Response:**

Thank you for your comment.

*5.4. Again we could clearly observe the trends from Figure 2 and Figure 3. For all the short-term and long-term analysis, the authors only describe the results, however, provide no explanation. The lack of theory development disables the authors to explain the results nor distinguish the results from time trend and other unobserved factors.*

**Authors’ Response:**

Thank you for your comment.

*5.5. Mit should also be included in the equation other than policy and the interaction term.*

*These three variables, rating, price and number of listings, are time-invariant. Why use Mit in the regression? what does t represent? Besides, these three variables only reflect the status at the time of data collection, I doubt it is proper to use them in the analysis, especially in such a long time window. It is possible to manually calculate review rating and the number of host listings based on data and see how are the results change.*

*Line 5 on page 37, review rating didn’t diminish the impact on “sentiment” whose coefficient is positive. Again, I didn’t see meaningful discussions of results in this part.*