Economics Bulletin

Volume 43, Issue 1

The effect of COVID-19 on Amazon MTurk

Jun Yeong Lee
Pusan National University

Elizabeth Hoffman
Iowa State University

Abstract

This paper examines the effect of COVID-19 on experiments run using Amazon MTurk. We run a Bertrand Oligopoly experiment with conversation before and after the pandemic outbreak. The outcome shows that after COVID-19, more participants do not fully engage in the experiment and "cherry-pick" the participation fee. There is a 36.51% decrease in the number of participants who actually participate in conversations and a 31.71% decrease in the number of groups discussing pricing behavior.

The authors would like to thank John P. Conley, Peter Orazem, and an anonymous referee for helpful comments on earlier drafts. The authors received no external funding for this work and are solely responsible for the contents.
Citation: Jun Yeong Lee and Elizabeth Hoffman, (2023) "The effect of COVID-19 on Amazon MTurk", *Economics Bulletin*, Volume 43, Issue 1, pages 583-588
Contact: Jun Yeong Lee - junyeonglee@pusan.ac.kr, Elizabeth Hoffman - bhoffman@iastate.edu.
Submitted: May 17, 2022. **Published:** March 30, 2023.

1. Introduction

Over the past two decades, experimental economists have come to rely on subjects recruited from the Amazon Mechanical Turk subject pool. Arechar et al. (2018) review experimental research using MTurk subjects going back to 2000 and compare student subjects with MTurk subjects playing a complicated game. Hoffman et al. (2020) compare student subjects in person, student subjects online, and MTurk subjects making lottery choice decisions. Both papers find small difference between students and MTurk but conclude that the results do not differ significantly as a result of using MTurk subjects instead of or in addition to student subjects.

These results, confirmed by the extensive review of articles in Arechar et al. (2018) have helped make the use of MTurk subjects increasingly common in economics experimental research. Besides working for far less money than student subjects, MTurk subjects bring the advantage of diversity more consistent with the U.S. population than student subjects who represent 18-25-year-old, largely white or Asian, and well-educated individuals.

The COVID-19 pandemic accelerated the need for online subjects as the NIH shut down approval of in-person experiments. However, as we found trying to replicate results generated in 2019, MTurk subjects recruited in the summer of 2020 behaved significantly differently than MTurk subjects recruited in 2019. The best way we can summarize the difference is that subjects recruited in the summer of 2020 were significantly less engaged in the experimental activity under study. The results of our 2020 experiments show such high levels of disengagement that we are in the process of redoing our replication with students online.

In this paper, we investigate the effect of COVID-19 on the quality of Amazon MTurk data. We run experiments before and after COVID-19 to examine the impact of the pandemic. We use a Bertrand oligopoly design that allows three-subject groups to text one another before independently setting prices. In our control set of experiments, subjects may not text one another. Thus, the control represents the pure Bertrand oligopoly model. The Nash equilibrium of this model predicts that subjects will compete with one another to set lower prices to win the right to sell and thus will converge to a zero-profit equilibrium. We consider as treatments allowing participants to text one another before setting prices. We vary how frequently participants may text. While game theory still predicts a zero-profit equilibrium with non-binding "cheap talk," previous experimental research found that cheap talk sometimes results in tacit collusion (Fonseca and Normann, 2012).

Our results using MTurk subjects collected in 2019 (Pre-COVID) showed a dramatic difference between subject pricing behavior when they could text before every round and when then could text less frequently. When we replicated the experiment in summer 2020 (post-COVID), the results were much less dramatic and more variable. The underlying difference appears to be a difference between how subjects Pre- and Post-COVID used the texting opportunities. Pre-COVID, 2.52 participants per group texted and 82% of the texts discuss pricing. Post-COVID, only 1.6 participants per group text anything and only 56% of the texts discuss price. These differences are statistically significant. These results indicate that MTurk subjects Post-COVID are less engaged in the studied activity than MTurk subjects Pre-COVID.

We outline the experimental design in Section 2, the results in Section 3, and conclude the paper in Section 4.

2. Bertrand Oligopoly Experiment with Conversation, Pre- and Post-COVID

As we outline in the introduction, we are studying the impact of the frequency of the opportunity to communicate on the extent of tacit collusion in a Bertrand oligopoly market with

inelastic demand and zero-marginal cost of production. Three participants play with one another for one experimental observation that runs for a given number of rounds. Each participant privately sets a price between 0 and 100. The lowest price in the group wins. If two or three participants in a group set the same lowest price, they divide the profits equally. The lowest price is disclosed to everyone at the end of each round, but the participant who has chosen it is not revealed.

Participants cannot communicate with each other for the first 10 rounds; however, they may text each other, beginning with the 11^{th} round. The frequency of communication varies according to the treatment level to which a participant is assigned. There are five treatment groups: texting each round (Every), every other round (Every2), every three rounds (Every3), every four rounds (Every4), and every five rounds (Every5). Each conversation lasts for 30 seconds, and messages are sent through the Instant Messenger built into the game. There are no restrictions on the topic of conversation, but we advise participants to protect their own privacy and keep the discussion respectful. Participants also have the right to skip the conversation or not participate at all.

The experiment was done on Amazon MTurk via oTree (Chen et al., 2016). We prevent retaking and set regional restrictions to the US so that participants can text each other smoothly. From July 2019 to November 2019, we ran 20-round experiments with 180 participants, including 10 groups per treatment and control groups (IRB ID 19-189, Iowa State University). Subjects were paid based on 20 rounds of profit, and 2000 points were worth $2. Each participant earned aggregated profit and a $2 completion fee at the end of the experiments.

From August 2020 to September 2020, we ran 30-round experiments with 360 participants, including 20 groups per treatment and control groups. We increased the compensation and set more qualifications to participate in the experiment. The new experiment required participants to have completed more than 10 approved Human Intelligence Tasks (HIT), with an approval of more than 95%. They were paid a $2 completion fee, and 3000 points were worth $7.5. We increased the payments and the subject requirements because the first few Post-COVID runs resulted in 0 engagement by the MTurk subjects.

3. Results

Table 1 presents texting results for experiments conducted Pre- and Post-COVID. The variable *Texted participant in the group* measures the number of people in each group who actually texted English sentences during the experiments. Pre-COVID, 72% of subjects are in groups in which all three participants text English sentences and only 8% are in groups in which no participants text English sentences. Post-COVID, the percentage of groups in which all three text English sentences drops to 36%. The percentage who are in groups in which no participants post English sentences increases to 30%. Moreover, 82% of pre-COVID texts involved price, while only 56% post-COVID involved price. A Kruskal-Wallis equality-of-populations rank test confirms that the difference of distributions is statistically significant ($p<0.001$).

Table 1: Distribution of texting behavior

Texted participant in the group	Pre-COVID	Post-COVID
0	8%	30%
1	4%	16%
2	16%	18%
3	72%	36%
Was it about the price?	82%	56%

NOTE: We exclude the control group for this analysis

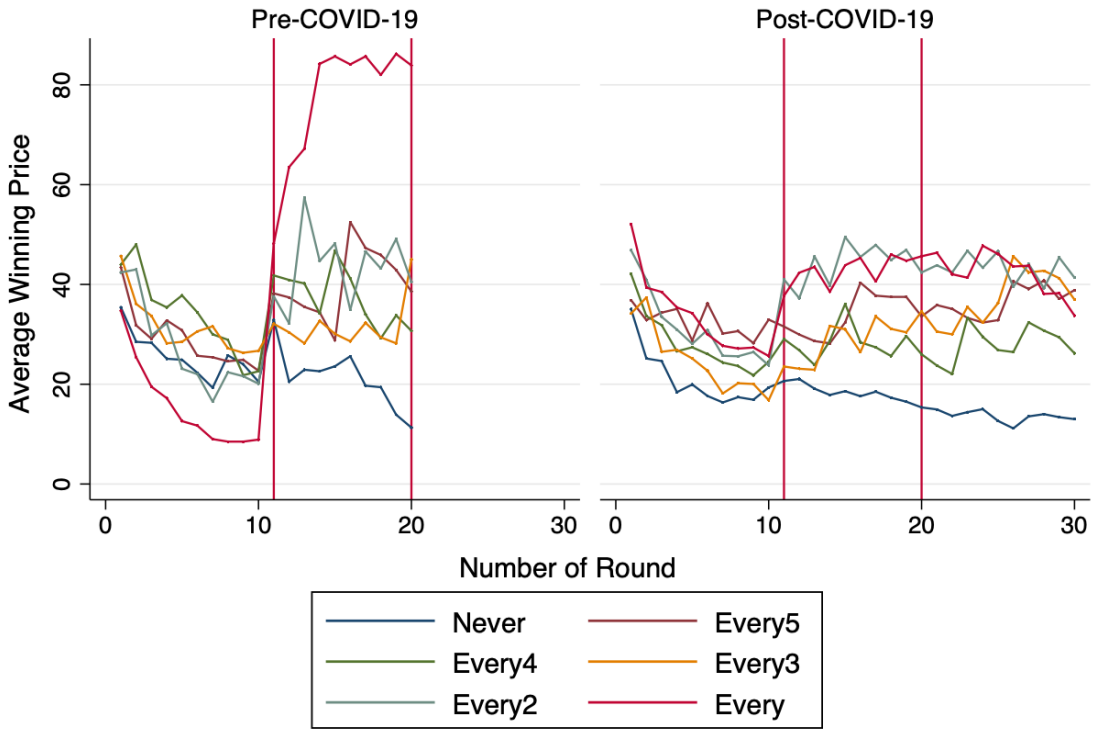
Table 2 reports participants' contributions in each of the five treatments. Comparing Pre- and Post-COVID, the average number of people in each group who actually text decreases. For Every, Every4, and Every5, average participation post-COVID is considerably lower than pre-COVID. In addition, lower percentages of all treatment groups (except Every 4) discuss price post-COVID.

Table 2: Changes in participant's communication

	Pre-COVID		Post-COVID	
	Texted Participants	Was it about the price?	Texted Participants	Was it about the price?
Every5	2.1	80%	0.95	35%
Every4	2.4	70%	1.2	70%
Every3	2.4	70%	2.05	65%
Every2	2.7	90%	2.15	60%
Every	3	100%	1.65	50%

Figure 1 shows the average winning price of each round by treatment level, pre- and post-COVID. As shown in Table2, big differences are found in Every, Every4, and Every5, which show large decreases in texted participants and discussing price. The most dramatic difference is between Every groups, Pre- and post-COVID. When Every groups actually texted in English sentences and about price, they were able to coordinate and maintain their coordination to rise to and sustain a price close to the profit maximizing price of 100. No other treatment groups could achieve that degree of coordination.

Figure 1: Average winning price by treatment



We can only guess why the post-COVID MTurk subjects did not take advantage of the opportunities to communicate afforded to them by the experimental design. Several papers have studied the effect of the COVID pandemic. Ettman et al. (2020) document emotional effects of long hours working from home, and closed parks, bars, and amusement parks. Arechar and Rand (2021) point out that there are many new participants in MTurk as they lose their regular jobs and need to make money from home. Their personal and demographic characteristics and demographics could be different from the previous participants. This difference may lead them to "cherry-pick" the completion bonus and to not care about the content of HIT. We should have eliminated those subjects by requiring that our subjects have experience and high marks from other researchers.

The third possible explanation is that there are more participants who use Virtual Private Servers (VPS) to join the US only HIT requirement. Dennis et al. (2020) caution that there are some MTurk workers who use the VPS to circumvent screening rules. Since COVID-19 is a worldwide problem, unemployment is also higher worldwide. While the use of virtual private servers has been a problem for some time, we find evidence that some of our participants try to text in their own languages.

4. Conclusion

The purpose of this paper is to provide updated information on the quality of data from MTurk workers during the COVID-19 pandemic. We find a significant decrease in the number of people who actively engage in the experiments compared to the pre-COVID era. As Alsharawy et al (2021) suggest that the generalizing experimental work during pandemic could

be cautious. We do not think that all MTurk studies will be affected by the pandemic, but experiments requiring active engagement among participants may result in results that do not replicate at other times or with other groups of subjects.

References

Alsharawy, A., Ball, S., Smith, A., & Spoon, R. (2021). Fear of COVID-19 changes economic preferences: evidence from a repeated cross-sectional Mturk survey. *Journal of the Economic Science Association*, *7*(2), 103-119.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. Experimental economics, 21, 99-131.

Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. Behavior research methods, 53(6), 2591-2595.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88-97.

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. Behavioral Research in Accounting, 32(1), 119-134.

Ettman, C. K., Abdalla, S. M., Cohen, G. H., Sampson, L., Vivier, P. M., & Galea, S. (2020). Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA network open, 3(9), e2019686-e2019686.

Fonseca, M. A., & Normann, H. T. (2012). Explicit vs. tacit collusion—The impact of communication in oligopoly experiments. European Economic Review, 56(8), 1759-1772.

Hoffman, E., Schwartz, D. L., Spitzer, M., & Talley, E. (2020). Patently Risky: Framing, Innovation, and Entrepreneurial Preferences. Harv. JL & Tech., 34, 191.