

## **Direct Replication of a research claim from Zhang (2009) in Experimental Economics: Final Report**

Nathan Fong (Rutgers University, School of Business – Camden) and Sangsuk Yoon (University of Dayton, School of Business Administration)

### **Summary**

Following the preregistered replication protocol (Fong et al. 2022), data were collected to test the claim from Zhang (2009) that the average market price in the UNIFORM treatment (a common type of multi-unit auction) is significantly higher than that in the FIXED treatment (a fixed price of 1.94). Based on the planned analysis, the result did not replicate.

Data from a total of 32 intact (without dropouts or bankruptcies) groups of 4 participants were collected. For this full sample, the average price in the UNIFORM condition was 1.96, with 18 of the 32 groups attaining average prices over the benchmark of 1.94, resulting in a P-value of 0.71 in the proscribed Wilcoxon signed-rank test.

All materials are archived in a repository at <https://osf.io/3tczs/>.

### **Data Collection**

Data were collected in two rounds, with 15 intact groups in the first round and 17 in the second round. Participants were recruited using Prolific (<https://www.prolific.co/>), an online platform for matching participants to research studies. Participation in the study was restricted to US-based participants fluent in English. While the experiment was incentive compatible, there was a minimum wage of \$8/hour set by Prolific's rules, so a floor was set at \$8 for participants who lost money or went bankrupt during the experiment. The auction was implemented using the oTree platform for online experiments with multiple participants (Chen et al. 2016). A snapshot of the code is provided in the file IPO\_Study\_Github\_20220531.zip.

Note that the individual difference scales originally planned (but not needed as part of the replication) were not included, because in testing, completion of the study took longer than anticipated (taking a full hour without the survey, which would have been administered at the end of the study).

The data for the first round of 15 groups were collected on 5/31/2022 and 6/1/2022 and analyzed according to the preregistration. The target result from Zhang (2009) was not replicated, so following the plan, a second round was conducted from 6/6/2022 through 6/8/2022, with a stopping rule of collecting data from 16 intact groups, which resulted in 17 intact groups (since multiple groups were running simultaneously). The raw data for each round is provided in two separate files, round1\_raw.csv and round2\_raw.csv.

The preregistration did not precisely specify what constitutes a dropout, which can be less explicit in an online setting. We specified that 6 rounds missed would be considered a dropout, with time limits of 3 minutes for the first 5 rounds and 1.5 minutes for the remaining rounds. Participants were informed of these restrictions, and behaved accordingly: of those who did not drop out, one participant missed 5 rounds, and the remainder missed 2 or fewer. Of those participants who did drop out, they each missed

at least 18 rounds. No participants missed between 6 and 17 rounds, suggesting that once they decided to complete the experiment, they stuck with the decision.

## Analysis

The analysis was performed in two stages: the first round of data was analyzed on its own, and then the full sample with both rounds of data. The unit of observation is the group: while each group is made up of 4 participants, in each round of the experiment, one market price is produced for each group. This outcome was averaged across the 20 rounds for each group before the Wilcoxon signed-rank test was conducted. The Stata code for the analysis (matching the preregistration analysis plan sample code) is included in the file `replication_analysis.do` (output in `replication_analysis.log`). The analysis datasets are provided as `round1_analysis.csv` and `fullsample_analysis.csv`. An alternative analysis performed in R is also provided for accessibility (`replication_analysis.Rmd` and `replication_analysis.html`).

For the first round of data only, the average price exceeded the benchmark of 1.94 for 9 out of 15 groups, resulting in a P-value of 0.72 (this is the exact P-value as computed in Stata).

For the combined data, the average price exceeded the benchmark for 18 out of 32 groups, resulting in a P-value of 0.71 (including two groups with bankruptcies, the average price exceeded the benchmark for 20 out of 34 groups, resulting in a P-value of 0.48).

## Discussion

The main result from the target paper was not successfully replicated, and the observed market prices (roughly 1.96) were substantially lower than the market value (expected value of 2.5). The expectation is that competitive auctions would result in market prices closer to the market value for the fictional assets being sold. While a slight majority of groups had average prices exceeding the benchmark, many groups had surprisingly low average prices. Two groups even had average prices, across 20 rounds, of less than 1, which is notable because the market value of the fictional goods had a possible range of 1 to 4. Thus, on many occasions participants had strictly dominant profitable deviations where they could increase either the price or quantity of their bids.

While the reported study was intended as a direct replication of the target study, there were several differences that could potentially contribute to the difference in results. Firstly, while both the original and the replication were run as computer simulations, the original had participants using computers in person, in a lab, whereas the replication was conducted online, with remote participants. While the comprehensive instructions for the study hewed closely to the original and required passing an “attention check” that tested participants’ understanding of the auction mechanism, it was not practical for participants to ask for clarifications from the experimenters. Getting through the instructions and passing the test took roughly twice as long as anticipated (15-20 minutes), suggesting that some participants had difficulty comprehending the task. Remote participants may also be more prone to distractions during the experiment. Poor comprehension and inattention likely contributed to the low average prices in some groups.

In addition, the original study recruited from a pool of UK undergraduates, whereas the replication recruited from anonymous online participants in the US. University undergraduates, relative to the general population, might have higher capacity or inclination to learn the rules of the game, resulting in more “rational” outcomes in the auction.

One other potential explanation could be collusion, either explicit or tacit. Given the circumstances, we find explicit collusion to be unlikely. A search of forums where research participants share information (on Reddit and elsewhere) did not turn up any indication that our study was discussed in any way. Tacit collusion is more difficult to rule out, and quite possibly occurred to a limited extent: in a group with one or two sophisticated participants paired with one or two inattentive or miscomprehending participants, upon observing underbidding by other participants, may have decided not to bid aggressively so as to prolong the opportunity for excess profits (even if in any given round they could earn more by bidding aggressively).

Finally, we note that if the analysis were performed with the group-round as the unit of observation, we do observe a significant effect (with 193 out of 300 group-round observations exceeding the benchmark in the first round, for a P-value of 0.052, and 430 out of 640 exceeding the benchmark in the full sample, for a P-value<0.001). However, such an analysis does not account for correlated errors across rounds for a group, and in any case the observed effect size is substantially smaller than in the original study (with the replication plan, including the group-level analysis, designed to achieve 90% power for 50% of the original effect size).

## References

- 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.
- Fong, N., Field, S., Miske, O., Loomas, Z., & Yoon, S. (2022). Direct Replication of a research claim from Zhang (2009) in *Experimental Economics*. Retrieved from <https://osf.io/t3a6j>
- Zhang, P. (2009). Uniform price auctions and fixed price offerings in IPOs: an experimental comparison. *Experimental Economics*, 12(2), 202-219.