Replication Report for
"Designing Pricing Contracts for Boundedly Rational Customers: Does the Framing of the Fixed Fee Matter?"

By: Teck-Hua Ho and Juanjuan Zhang

Primary Team: Elena Katok and Kyle Hyndman
University of Texas at Dallas

Secondary Team: Jordan Tong and Xiaoyang Long
University of Wisconsin-Madison

Ho and Zhang (2008) investigate a supply chain contracting problem in which they test two-part tariffs as a way to increase supply chain efficiency. They vary the framing of the two-part tariff as either a fixed fee or a quantity discount. They find that supply chain efficiency is higher under the quantity discount framing.[1]

> **Hypothesis to replicate:**
>
> Supply chain efficiency is higher when a two-part tariff is framed as a quantity discount as opposed to a fixed fee.

Power Analysis

The original t-statistic, taken from Table 3(c) of the paper, is 1.99 ($p = 0.047$). This statistic is based on treating each observation – one observation per pair of subjects interacting per round – as independent. The relevant data are from Table 2 of their paper, which we replicate in Table 1 at the end of this document.

The original sample size is 48 subjects in the TPT treatment (4 sessions of 12 subjects per session) and 46 subjects in the QD treatment (4 sessions, two with 12 subjects and two with 11 subjects).[2] To achieve 90% power based on their method of statistical analysis, the required sample size is 1348 observations. Since each session of 12 subjects generates 66 observations, this implies that we need approximately 252 subjects, which translates into 11 sessions of one treatment and 10 sessions of the other treatment. Our target was to obtain data for 11 sessions of both treatments, which would allow us to slightly exceed the desired power.

---

[1] We phrase the hypothesis as directional in terms of the result that the original study authors observed in the paper. The original hypothesis was two-sided in nature, namely that supply chain efficiency is invariant to the framing of the fixed fee. We implement a two-sided $t-$test to be consistent with the original authors.

[2] This is a conjecture based on their reported methods.

Sample

Participants for the original study were "undergraduate students at a West Coast university". The target sample size for the primary replication was at least 252 University of Texas at Dallas students (11 sessions of 12 subjects per treatment would yield 264 subjects, thereby allowing some sessions to contain only 11 subjects, as was the case in the original paper). Given the COVID-19 restrictions on in-person studies and their spillover effects on subject recruitment, as well as the large sample required, we did not commit to conduct this study with a full replication sample at the secondary site. Indeed, as we discuss below, we did not conduct the experiment at all at the secondary site. Students were recruited from the general laboratory population at the University of Texas at Dallas.

Materials

The instructions for the TPT treatment were included in the supplementary materials of the published paper. The original experiment was conducted via paper and pencil. For logistical reasons, we decided to implement the experiment on a web-based software platform (SoPHIE). Every effort was made to ensure that the task, decision support, and interface was similar to the original experiment. For the TPT treatment, we used the original instructions provided by the authors modified slightly to update for the transition to a software-based experiment. For the QD treatment, the authors could not provide us with original materials. Therefore, we created these from scratch using the TPT instructions and the language in the paper as a guide. Discussions were held with the original authors in effort to ensure that the instructions and software corresponded closely to the original study materials.

Procedure

We follow the same protocols outlined in section "3.2 Experimental procedure" on pages 690–691 with some minor deviations, detailed in a later section. The primary dependent variable of interest is supply chain efficiency: the actual channel profit earned divided by the integrated channel profit benchmark. The pre-registration report for the experiment is available at https://aspredicted.org/ya2gu.pdf.

Analysis

The analysis of our data is identical to the original article: a two-sample t-test comparing the unconditional supply chain efficiency in the TPT and QD treatments, treating each observation as independent for the purpose of statistical analysis.

Differences from Original Study

The differences with respect to the original study are as follows:

1. The experiment was conducted at the University of Texas at Dallas, rather than "a West Coast university" (presumably, UC Berkeley).

2. The lab population at UT Dallas contains a mix of undergraduate and graduate students, rather than the undergraduate students reported in the original paper.

3. We used SoPHIE software rather than paper and pencil.

4. We made minor modifications to the instructions to accommodate the software platform implementation of the experiments. As mentioned above, the Quantity Discount treatment instructions were written entirely by us, using the Two-Part-Tariff treatment and the description in the published paper as a guide. We demonstrated the software and instructions to the original authors and tried to address any issues with instructions and software that they raised with us.

5. We provided a show-up fee for participation as is the standard procedure of the lab. This additional compensation also helps to account for inflation from the time the original experiments were conducted.

6. In the original study, subjects had to calculate the order quantity and profits manually by filling a table with this information on a sheet of paper. The Quantity Discount treatment also required an additional calculation of the wholesale price marked up by the fixed fee prorated by the order quantity. In the original experiment calculations were not checked. We implemented this process on the computer, with the exception that calculations entered into the computerized table were checked, and had to be entered correctly before the participant was allowed to submit the form. Subjects were given a calculator, scrap paper and a pen to make calculations.

Replication Results
The results obtained from 191 subjects from the UTD sample are given in Table 2. This is less than the target sample size, for reasons explained in the next section. Our main variable of interest is the overall efficiency of the supply chain in the two treatments. Ho and Zhang (2008) showed that efficiency was significantly higher in the Quantity Discount treatment (76.37%) than in the Two-Part Tariff treatment (69.51%). However, as can be seen in Table 2, the two efficiencies are closer (62.62% and 65.18%, respectively) and the difference is not statistically significant ($p = 0.328$). In the replication, the efficiency was directionally lower in the Quantity Discount treatment.

In the original paper, Ho and Zhang (2008) show that this result is driven by the significantly higher rejection rate in the Two-Part Tariff treatment ($p = 0.030$), while the efficiency conditional on an agreement was not statistically distinguishable between the two treatments ($p = 0.240$).

Neither of these results holds in our replication sample. The rejection rate is significantly higher in the Quantity Discount treatment ($p = 0.036$) – which is the opposite of the original result – while the conditional efficiency is significantly higher in the Quantity Discount treatment ($p = 0.021$), which attenuates the overall effect size.

Unplanned Protocol Deviations
We conducted two sessions of each treatment using the above process (in which subjects had to enter in correct calculations and would only proceed upon entering the requisite information correctly). We found that the the Two-Part Tariff sessions lasted approximately two hours, while the Quantity Discount sessions lasted substantially longer. Our perceived main driver for this long duration is that, like in the original paper, we used perfect stranger matching in groups of 12 participants, which means that each round took as long as the slowest participant in the session. Participants in both treatments expressed frustration by the long waits. Therefore, we modified the software so that if a participant entered an incorrect calculation, the error message provided the appropriate formula as well as the correct calculation. After this modification was made, most sessions were completed in either a little more or a little less than two hours. In our analysis, we include all data collected, both with the original protocol and the slightly modified protocol described here. The efficiency results are qualitatively similar if the first two sessions of each treatment are dropped.

Every effort was made to conduct 11 sessions per treatment at UT Dallas. However, the subject population that was willing to participate in in-person experiments was significantly reduced due to the COVID-19 pandemic. We were only able to conduct 8 sessions per treatment for a total of 191 subjects. Because of the smaller sample size, we do not achieve the desired 90% power.

Instead, given the actual sample, our power is approximately 80%.

In addition, given the difficulty to reliably recruit 12 subjects per sessions (and the large number of sessions required), it was determined that the experiment would not be replicated at the secondary location.

Discussion

Although we did not obtain the requisite number of subjects to achieve the desired 90% power, the fact that we do not replicate the results from Ho and Zhang (2008) does not appear to be due to the limited sample. Unlike in the original study in which rejections were significantly larger in the Two-part Tariff treatment, we observe that rejections are significantly *higher* in the Quantity Discount treatment. It is beyond the scope of this report to delve into why this may be the case.

We would like to conclude with an important aspect of the data in our replication study that does replicate the original finding. The qualitative differences in contract parameters are in line with the original findings. Consistent with Ho and Zhang (2008), wholesale prices are lower under Quantity Discount than under Two-Part Tariff ($p = 0.001$), the fixed fee is higher under Quantity Discount than under Two-Part Tariff ($p < 0.001$), and the average retail price conditional on acceptance is lower under Quantity Discount than under Two-Part Tariff ($p = 0.001$). Therefore, the lack of difference in overall efficiency in our data is due to the fact that the acceptance rate is lower under Quantity Discount, while in the Ho and Zhang (2008) study it was higher.

Table 1   **Original Results From Ho and Zhang**

Parameter	TPT		QD		$p-$value
Efficiency	69.51%	(41.27)	76.37%	(36.18)	0.047
Acceptance Rate	72.24		82.23		0.030*
Wholesale Prices	3.96	(1.17)	3.41	(1.25)	0.000
Fixed Fees	5.24	(2.32)	6.95	(4.17)	0.000
N	264		242		
Conditional Efficiency	93.62%	(5.29)	92.87%	(14.60)	0.240
Retail Prices	6.86	(0.54)	6.71	(0.80)	0.029
N	196		199		

* This $p-$value was not reported in the text but follows from a two-sided proportions test made possible by information report in the text.

Table 2   **Replication Results on Efficiency**

Parameter	TPT		QD		$p-$value
Efficiency	65.18%	(21.82)	62.62%	(17.56)	0.328
Wholesale Prices	4.54	(1.45)	4.23	(1.66)	0.001
Fixed Fees	4.34	(2.49)	6.02	(7.65)	0.000
Acceptance Rate	76.14		70.41		0.036
N	528		517		
Conditional Efficiency	85.60%	(21.82)	88.94%	(17.56)	0.021
Retail Prices	7.06	(1.09)	6.80	(1.06)	0.001
N	402		364		

References

Ho, Teck-Hua, Juanjuan Zhang. 2008. Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science* **54**(4) 686–700.