

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: Xiaoyang Long and Jordan Tong
University of Wisconsin-Madison

Tertiary Team:* Samantha Keppler and Stephen Leider
University of Michigan

*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 2 at the end of this document.

The original sample size is 48 subjects in the TPT treatment (4 sessions of 12 sub-

jects 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 reasons explained, we needed to add a third location in order to attempt the second replication.

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.

for 11 sessions of both treatments, which would allow us to slightly exceed the desired power.

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. In the original submission of the overall replication project, we did not conduct the experiment at all at the original secondary site. However, during the revision stage, a new secondary site was added (University of Michigan) and a second replication was attempted. Students were recruited from the general laboratory population at the University of Texas at Dallas (primary site) and the University of Michigan (revised secondary site).

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 and the University of Michigan, rather than “a West Coast university” (presumably, UC Berkeley).
2. The lab populations at UT Dallas and the University of Michigan contain 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 and 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.

7. We conducted the experiment at the primary site (UT Dallas) in two waves, first in Spring 2022, conducting 8 sessions for each treatment and second in Fall 2022, collecting the remaining three sessions per treatment required to achieve the necessary sample size. Data from the revised secondary site (Michigan) was completed in a single wave between November 2022 and January 2023.

Replication Results

The findings regarding overall efficiency did not replicate. Given the differences between the original study and our replication (which we discuss further in the next section), there is ambiguity about whether

the non-replication stems from these differences or from theoretical weakness of the original result. The results obtained from 263 subjects from the UTD sample are given in Table 3 and the results from 120 subjects from the Michigan sample are given in Table 4. 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 Tables 3 and 4, the efficiency of the QD contract is never statistically significantly higher than the TPT contract ($p \geq 0.10$). Indeed, at UT Dallas, the two efficiencies are almost indistinguishable (63.18% and 63.61%, respectively), while at Michigan efficiency is directionally higher in the QD treatment (75.25% vs 70.73%).

In the original paper, Ho and Zhang (2008) show that the higher efficiency of the QD contract compared to the TPT contract 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 were observed in our replication samples. At UT Dallas, the rejection rate is marginally significantly higher in the Quantity Discount treatment ($p = 0.065$) – which is the opposite of the original result – while the conditional efficiency is significantly higher in the Quantity Discount treatment ($p < 0.01$), which attenuates the overall effect size. At Michigan, we do see that the rejection rate is lower in the QD treatment (consistent with Ho and Zhang (2008)) but the difference is not statistically significant ($p = 0.266$). Consistent with UT Dallas, and unlike the original study, conditional efficiency is higher in the QD treatment, but it does not reach statistical significance ($p = 0.148$).

Unplanned Protocol Deviations

At the primary location, 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 refer to this as the “initial protocol”. 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. We refer to this as the “modified/informative protocol”. 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 initial protocol and the modified/informative protocol described here. The efficiency results are qualitatively similar if the first two sessions of each treatment are dropped. All of the sessions at the University of Michigan were conducted with the modified/informative protocol and sessions there took approximately two hours as well.

Every effort was made to conduct 11 sessions per treatment at UT Dallas. In Spring 2022, 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. However, after the original paper submission, we conducted 3 additional sessions per treatment, for a total of 263 subjects and 1441 observations. We, therefore,

achieved the desired 90% power at UT Dallas.

In addition, given the difficulty to reliably recruit 12 subjects per sessions (and the large number of sessions required), it was originally determined that the experiment would not be replicated at the secondary location – University of Wisconsin at Madison. However, after the original paper submission, we attempted to conduct the replication at a new secondary site: the University of Michigan. While every effort was made to conduct 11 sessions per treatment, the decision was made to stop data collection after five sessions per treatment. Ten scheduled sessions had to be cancelled either due to insufficient sign-ups prior to the session or insufficient show-ups at the sessions. In Table 1 we provide details on the sessions, including timing, location and protocol.

Discussion

As noted earlier, there is ambiguity about whether the non-replication of results from [Ho and Zhang \(2008\)](#) stem from methodological issues with the replication or from theoretical weakness. The evidence we have suggests that, although we were only able to achieve 90% power at one of the two locations, the non-replication does not appear to be due to the under-powered sample. Consider first the UT Dallas sample, where we do achieve 90% power. 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. Moreover, also unlike [Ho and Zhang \(2008\)](#), we found that conditional efficiency was significantly higher in the QD treatment.

At University of Michigan, where we achieve only 60% power, the data are in closer alignment with the original study. In particular, overall efficiency is higher in the

QD treatment but the result was not significant ($p = 0.111$). It is certainly possible that with more data significance could have been achieved. Yet, even at University of Michigan, there are results theoretically inconsistent with the original study. Like the original study, there are fewer rejections in the QD treatment, which partially drives overall efficiency higher. However, unlike the original paper, but consistent with UT Dallas the conditional efficiency was higher in the QD treatment. Therefore—and importantly—even if the result *might have* replicated with a larger Michigan sample, it would have been due to a different mechanism.

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 ($p < 0.05$), the fixed fee is higher under Quantity Discount than under Two-Part Tariff ($p < 0.01$). For the UT Dallas sample, the average retail price conditional on acceptance is lower under Quantity Discount than under Two-Part Tariff ($p < 0.01$), while retail prices are virtually indistinguishable in the Michigan sample. Therefore, while some of the comparative statistics do appear robust, the main result on efficiency is not.

Table 1 Details on Sessions

Location	Timing	Protocol	Sessions	Note
UT Dallas	S2022	Initial	4	Prior to Original Submission
UT Dallas	S2022	Modified/Informative	12	
UT Dallas	F2022	Modified/Informative	6	After Original Submission
Michigan	F2022	Modified/Informative	8	
Michigan	S2023	Modified/Informative	2	

Notes: (1) S2022 denotes Spring 2022, F2022 denotes Fall 2022 and S2023 denotes Spring 2023. (2) Recall that the “initial” protocol required subjects to correctly input relevant numbers without any support before proceeding, while the “modified/informative” protocol still required subjects to correctly enter relevant numbers but gave subjects informative support to assist them in the event of errors.

Table 2 Original Results From Ho and Zhang

Parameter	TPT		QD		p -value
Efficiency	69.51%	(41.27)	76.37%	(36.18)	0.047
Wholesale Prices	3.96	(1.17)	3.41	(1.25)	0.000
Fixed Fees	5.24	(2.32)	6.95	(4.17)	0.000
Acceptance Rate	74.24		82.23		0.029
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)	0.030
N	196		199		

Notes: Standard deviations reported in parentheses.

Table 3 Replication Results on Efficiency (UT Dallas)

Parameter	TPT		QD		<i>p</i> -value
Efficiency	63.61%	(40.95)	63.18%	(42.35)	0.845
Wholesale Prices	4.66	(1.44)	4.38	(1.69)	0.001
Fixed Fees	4.17	(2.43)	5.68	(6.71)	0.000
Acceptance Rate	75.62		71.33		0.065
N	726		715		
Conditional Efficiency	84.12%	(22.14)	88.58%	(16.16)	0.000
Retail Prices	7.17	(1.09)	6.87	(1.04)	0.000
N	549		510		

Notes: Standard deviations reported in parentheses.

Table 4 Replication Results on Efficiency (Michigan)

Parameter	TPT		QD		<i>p</i> -value
Efficiency	70.73%	(38.68)	75.35%	(34.03)	0.111
Wholesale Prices	4.27	(1.38)	4.04	(1.27)	0.030
Fixed Fees	4.47	(2.71)	5.42	(2.73)	0.000
Acceptance Rate	80.97		84.33		0.266
N	310		319		
Conditional Efficiency	87.35%	(19.78)	89.36%	(10.84)	0.148
Retail Prices	6.96	(1.06)	6.95	(0.89)	0.958
N	251		269		

Notes: Standard deviations reported in parentheses.

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.