

Reply to Author Response 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

Tertiary Team: Samantha Keppler and Stephen Leider
 University of Michigan

On February 23, 2023, we received the final author response about our replication efforts for [Ho and Zhang \(2008\)](#). This followed a series of email exchanges between the replication team and the original authors. We provide further context below but copy one important extract from their response that we believe deserves further discussion and a formal reply. Their response stated (*italics added*):

Would the efficiency result be statistically significant had the Michigan site met the target number of observations? We tested this hypothesis by pooling the data collected in the second wave of replication (396 from UT Dallas and 629 from Michigan). Indeed, efficiency is higher in the QD condition than in the TPT condition at $p = 0.039$, replicating our original finding (see the table below).

Variable	Original Study			Replication Study (Second Wave)		
	TPT <i>N</i> = 264	QD <i>N</i> = 242	<i>p</i> -value	TPT <i>N</i> = 508	QD <i>N</i> = 517	<i>p</i> -value
Wholesale price	3.96 (1.17)	3.41 (1.25)	0.000	4.56 (1.42)	4.32 (1.48)	0.009
Fixed fee	5.24 (2.32)	6.95 (4.17)	0.000	4.17 (2.55)	5.18 (2.75)	0.000
Retail price (if accept)	6.86 (0.54)	6.71 (0.80)	0.029	7.16 (1.01)	6.98 (0.92)	0.007
Acceptance (%)	74.24 (43.81)	82.23 (38.30)	0.029	78.15 (41.36)	80.27 (39.83)	0.403
Efficiency (%)	69.51 (41.27)	76.37 (36.18)	0.047	66.33 (39.59)	71.25 (36.77)	0.039

Notes: Values in parentheses are standard deviations.

The replication team chose to pool data by site, combining the data obtained in both waves at UT Dallas. This is problematic since the data collected at UT Dallas in the first and second waves were not comparable (due to subject frustration and confusion in the first wave).

In the second wave of replication, once again the contract terms (i.e., wholesale prices, fixed fees, retail prices) were significantly more efficient in the QD condition than in the TPT condition. It appears that these results are very robust regardless of subject heterogeneity, frustration, and confusion during the study. These were the

main results of the original study, as highlighted in our original response (dated June 11, 2022, above).

Having seen an earlier version of their response, which included much the same language, on February 19, 2023 the replication team sent the original authors an email in which we sought clarification on how they defined the “second wave” in their response and also stressed to the authors that the sessions were conducted as described in Table 1. In so communicating, we wanted to emphasize that the majority of sessions conducted prior to the first submission were methodologically the same as all sessions conducted after the initial submission because they used the same experimental protocol. This is because their concerns center around the protocol changes introduced after the issues observed in the initial four sessions. We concluded our email by stating (*italics added*), “*If based on this clarification you would like to change how you wish to pool data across UTD and Michigan, feel free to edit your addendum.*”

On February 23, 2023, the authors responded that (*italics added*), “*Yes, the definition of waves is what we had in mind: the second wave refers to all data collection that happened since our initial response on June 11, 2022.*” The authors declined to change how they pool the data but they did edit their response in other dimensions, which is what we have included above. Therefore, in the “pooled analysis” in their response they do not include 12 sessions which employed the same experimental protocol as the 16 included in their analysis.

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.

We believe it is important to clarify a few points. First, while we did fall short of the target pre-registered sample size at one of the sites, we did not “choose to pool the data by site” in an ex post fashion. We analyzed data by site in all replications that we conducted, which also follows our pre-registration. Second, we would like to emphasize only two sessions for each treatment were conducted under the initial protocol in which subjects expressed some frustration.

In Table 2, we report the results including all sessions from the modified/informative protocol. In these results, the overall efficiency is directionally higher in the QD treatment, but the difference is not statistically significant ($p = 0.121$).

If the authors’ pooling analysis is intended to speak to whether further data collection to reach greater power would likely lead to replicating the original result, these results using all the data under the modified protocol answer that question more directly. When pooled in this manner, the overall effect is directionally consistent with the original result,

Table 2 Replication Results on Efficiency (Pooling All Sessions With Modified/Informative Protocol)

Parameter	TPT		QD		<i>p</i> -value
Efficiency	64.84% (40.92)		67.79% (39.93)		0.121
Wholesale Prices	4.59	(1.49)	4.329	(1.56)	0.000
Fixed Fees	4.33	(2.63)	5.38	(2.80)	0.000
Acceptance Rate	76.66		76.16		0.804
N	904		902		
Conditional Efficiency	84.58% (22.64)		89.00% (14.26)		0.000
Retail Prices	7.11	(1.11)	6.90	(0.98)	0.000
N	693		687		

although not statistically significant at the required level. However, we would also like to stress that the mechanism appears to be different than in the original paper. Ho and Zhang (2008) found that conditional efficiency was not significantly different across treatments but rejections were higher in the TPT treatment. We find that there is no difference in rejections across treatments, while the conditional efficiency is significantly higher in the QD treatment.

We recognize that our replication effort did not provide an exact replication because of differences in the study designs (original: paper and pencil, information entered not checked vs replication: computerized, information entered checked). We hope the reader and the authors understand our reasons for these deviations. This was not a frivolous decision. We wished to conduct a replication over the computer to make sure that our replication itself is easily replicable. Regarding checking calculations, if subjects were allowed to enter incorrect calculations, would they have been paid based on what they entered or what was correct? If we were to pay subjects based on what they entered, this would have created an obvious incentive compatibility problem. If we were to pay them based on correct calculations, this would have created a loss of control in the experiments.

In general, while the result for this paper (in the context of the replication project) is a failure to replicate the original efficiency result, we believe there remain several open questions around what aspects of the original results and mechanisms are robust (and under what conditions), and we would encourage further study to gain greater clarity.

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.