Replication Report for "Trust in Forecast Information Sharing"

By: Özalp Özer, Yanchong Zheng, and Kay-Yut Chen

Primary Team: Steve Leider and Samantha Keppler
University of Michigan

Secondary Team: Andrew M. Davis and Blair Flicker
Cornell University and University of South Carolina

Özer et al. (2011) study a supply chain setting where a supplier must make a capacity decision, and a manufacturer has private forecast information about market demand. The manufacturer sends a cheap talk message about the private forecast to the supplier. Standard theory predicts that the manufacturer's message should be uninformative (i.e. uncorrelated with the true private forecast), and that the supplier should ignore the message (i.e. the capacity decision should be uncorrelated with the message). Özer et al. (2011) find in human-subject experiments that there is informative communication, with manufacturers sending higher messages when the private forecast is higher, and suppliers choosing higher capacity when receiving higher messages. This result is obtained in all four treatments (varying demand uncertainty and capacity cost). We focus here on the High Cost, High Uncertainty treatment $C_H U_H$ with the smallest effect in the original data, and hence the strongest test of the result.

Hypothesis to replicate:

In the $C_H U_H$ treatment, manufacturers' messages will be positively correlated with their private forecast, and suppliers' capacity decisions will be positively correlated with the messages received.

Power Analysis

In the original study, the p-value is reported as $p < 0.01$: *"To test Theorem 1 against Hypotheses 1–3, we first regress $\hat{\xi}$ on ξ for each treatment and present the resulting slopes in Table 2. The slopes are all significantly positive ($p < 0.01$), suggesting a strong positive correlation between ξ and $\hat{\xi}$ for all treatments.... A similar analysis shows that the slopes on $\hat{\xi}$ when we regress K on $\hat{\xi}$ are also significantly positive ($p < 0.01$). Therefore, $\hat{\xi}$ and K are positively correlated."* (p. 1118). This is based on two Ordinary Least Squares (OLS) regression models regressing manufacturer message $\hat{\xi}$ on private forecast ξ, and regressing supplier capacity K on message $\hat{\xi}$. For treatment $C_H U_H$ the reported coefficient is 0.65 for the messages regression, and 0.63 for the capacity regression (p. 1119, Table 2). We contacted the study authors to secure the original study data to find the exact p-value. For the messages regression the exact t-statistic is 21.98 (with an associated p-value of $p = 1.133 \times 10^{-70}$). For the capacity regression the exact t-statistic is 17.50 (with an associated p-value of $p = 2.920 \times 10^{-51}$). We also re-estimated the results using a regression with clustered standard errors by participants. For the messages regression with clusters the t-statistic is 8.69, with $p = 5.359 \times 10^{-5}$. For the capacity regression with clusters the t-statistic is 9.90, with $p = 3.421 \times 10^{-5}$.

In the replication, we only tested the first thirty periods (see below for more details). Note that there is no significant time trend in the original data for either decision variable. Rerunning the OLS regression on the original data for just the first 30 periods yields similar results. For the messages regression: $\beta = 0.643$ with a t-statistic of 12.04 and $p = 2.829 \times 10^{-22}$. For the capacity regression: $\beta = 0.662$ with a t-statistic of 12.49 and $p = 2.454 \times 10^{-23}$. We also re-estimated the results using a regression with clustered standard errors by participants. For the messages regression with clusters the t-statistic is 12.36, with $p = 5.209 \times 10^{-6}$. For the capacity regression with clusters the t-statistic is 9.42, with $p = 3.169 \times 10^{-5}$.

The original sample size for the $C_H U_H$ treatment is 8 participants. To achieve 90% power based on the results of the standard OLS analysis for the full original data set, the required sample size is 1 participant. To achieve 90% power based on the results of the OLS analysis for just the first 30 periods, the required sample size is 1 participant. The same holds if we consider the regression with clustered standard errors. Both are below the minimum threshold of $N = 40$ set by the replication project. Thus, our target sample size is $N = 40$.

Sample

Data was collected from 44 participants at Michigan, and 46 participants at Cornell. As in the original study, we permitted both graduate and undergraduate students. The replication experiment was conducted in person at the Michigan and Cornell experimental labs. Each session had at least 8 participants (the session size in the original study).

Materials

The original experiment was conducted using proprietary software from the Hewlett-Packard experimental lab, which is no longer available. With the approval of the original study's authors, we have re-implemented the experiment in z-Tree using as a basis drafts of the original instructions, as well as the z-Tree code (provided by the authors) from their follow-up paper Özer et al. (2014), which uses a very similar setup. Our re-implementation was approved by the original study's authors before the replication study began.

Procedure

We follow the procedure of the experiment outlined in Experimental Design and Procedures (Section 4, p. 1117) of the original study. First, instructions are given. The game is described in terms of "experimental dollars." Participants are informed the game has two roles, a "supplier" and a "manufacturer." The supplier secures capacity to make a "product" at a capacity cost of 60 experimental dollars per unit. The manufacturer purchases the product from the supplier at a wholesale price of 75 experimental dollars per unit, selling it in the retail market at a retail price of 100 experimental dollars per unit. The manufacturer's order is set after "final customer demand" is determined. Customer demand is equal to X + Z, where X is uniformly distributed between 100 and 400, and in the $C_H U_H$ treatment Z is uniformly distributed between -75 and +75. The manufacturer has better information about the final customer demand (i.e. knowing the value of X in each period), and sends a "report" to the supplier. After receiving the report, the supplier sets their "capacity."

The game is played for 30 periods with random anonymous rematching in each period and a random assignment of the supplier and manufacturer role. The original study ran for 100 periods (lasting approximately 3 hours) and had a show-up fee of $25 plus additional payments based on performance; average total compensation was $81.74 (and ranged from $68.19 to $102.03). Our replication uses fewer periods due to recruitment issues (described below), with the change approved by the original study authors. We estimated that our shorter sessions would last 1.5 hours and therefore set average compensation to be half as large (to match the shortened duration). We used a showup fee of $12.50 and a conversion ratio of 3300 experimental dollars equaling $1 (projected to yield an average task earnings of $29, for a total compensation of $41.50).

The pre-registration report for this experiment is available at https://aspredicted.org/id2sz.pdf.

Analysis

Our analysis is identical to that in the original study. That is, we use OLS regressions to regress manufacturers' messages on their private forecast, and suppliers' capacity on received message. For robustness, we additionally conduct the regression analysis with clustered standard errors by participants.

Differences from the Original Study

There are two notable differences between the original study and the replication: (i) we conduct the experiment using a z-Tree re-implementation of the original experimental software, and (ii) we shortened the experiment from 100 periods to 30 periods (with an associated re-scaling of the total compensation). The first change was necessary due to the original software no longer being available. The second change was necessary due to recruitment issues caused by Covid-19. We originally attempted to run the full sessions (100 periods / 3 hours), with an adjusted compensation rate in line with Özer et al. (2014). However, Covid-19 protocols limited the capacity of the lab, and the show up rates were low. As a result, we were unable to run sessions successfully. We then changed to the current protocols of 30 period sessions with re-scaled compensation proportional to the original, in consultation with the original study authors.

Replication Results

Following the protocol in the pre-registration (and matching the original paper), all participants ($N = 44$ at Michigan; $N = 46$ at Cornell) are used in the analysis. Table 1 reports the regression analysis for the original data and the two replications. For reference, the first two columns report the analysis (run on the first 30 periods) for the original data from Özer et al. (2011), while the next two columns report the results from the Cornell sample, and the final two columns report the results from the Michigan sample. Following the original analysis, OLS standard errors are reported in parentheses. Additionally, standard errors clustered at the participant level are also reported in square brackets. The results at both sites are consistent with the original results: manufacturer reports are positively and significantly correlated with their private forecast, and supplier capacity decisions are positively and significantly correlated with the report they receive from the manufacturer. All coefficients are significant with $p < 0.001$ (both with OLS and clustered standard errors).

Unplanned Protocol Deviations

There were no unplanned protocol deviations.

Discussion

In summary, we verify the results from Özer et al. (2011) that there is informative communication: manufacturers send higher messages when their private forecast is higher, and suppliers choose higher capacity levels when they receive higher messages. Both effects are statistically significant at both replication sites with $p < 0.001$.

Table 1 Regression results for replications and original experiment

Study:	Özer et al. (2011)		Cornell Replication		Michigan Replication	
Dep Var:	$Report(\hat{\xi})$	$Capacity(K)$	$Report(\hat{\xi})$	$Capacity(K)$	$Report(\hat{\xi})$	$Capacity(K)$
$Constant$	62.671*	162.724*	108.46*	46.500*	88.825*	50.153*
	(4.683)	(5.312)	(6.964)	(8.967)	(7.639)	(8.355)
	[7.310]	[6.242]	[15.778]	[12.451]	[14.050]	[11.144]
$Forecast(\xi)$	0.643*		0.716*		0.749*	
	(0.053)		(0.025)		(0.029)	
	[0.052]		[0.046]		[0.046]	
$Report(\hat{\xi})$		0.662*		0.609*		0.587*
		(0.053)		(0.029)		(0.029)
		[0.070]		[0.051]		[0.051]

Note: $^{*}p < 0.01$. Standard errors reported in parentheses; Clustered Standard Errors (at the participant level) reported in square brackets. $Forecast(\xi)$ is the private forecast that the manufacturer observes. $Report(\hat{\xi})$ is the report that the manufacturer sends to the supplier. $Capacity(K)$ is the supplier's capacity decision.

References

Özer, Özalp, Yanchong Zheng, Kay-Yut Chen. 2011. Trust in forecast information sharing. *Management Science* **57**(6) 1111–1137.

Özer, Özalp, Yanchong Zheng, Yufei Ren. 2014. Trust, trustworthiness, and information sharing in supply chains bridging china and the united states. *Management Science* **60**(10) 2435–2460.