Replication Report for "Inferring Quality from Wait Time"

By: Mirko Kremer and Laurens Debo

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

Secondary Team: Samantha Keppler and Stephen Leider
University of Michigan

Kremer and Debo (2016) investigate purchasing decisions as a function of consumer wait time in a setting with informed consumers, who know the quality of the product, and uninformed consumers, who do not. In a simulated queuing experiment, they find that uninformed consumers can infer product quality from observed wait times if there are enough informed consumers. For short waits, the purchase probability of uninformed consumers decreases in the presence of informed consumers ("empty restaurant syndrome"). As the wait time increases, uninformed consumers are more likely to purchase when there are some informed consumers in the population relative to when there are none.

Hypothesis to replicate:

Relative to the setting with no informed consumers ($q = 0$), the presence of informed consumers ($q = 0.50$) makes uninformed consumers (a) less likely to purchase upon observing a short wait ($w = 1$) and (b) less sensitive to the purchase probability reduction associated with each marginal unit of wait time. We test these two findings in the setting with a high prior of quality ($p_0 = 0.50$, treatments Q_{00} and Q_{50}).

Power Analysis

Kremer and Debo (2016) "study the impact of wait time on consumers' purchasing behavior when product quality is unknown to some consumers (the 'uninformed consumers') but known to others (the 'informed consumers')" (p. 3023). The predicted short-wait time purchase probability is 95% in the condition with no informed consumers (Q_{00} and $w = 1$). The corresponding probability is 68% when half of consumers are informed (Q_{50} and $w = 1$). According to Table 2, these are significantly different at the 5% level. The authors do not report the precise significance values, but they do describe their statistical methods on page 3032: "Table 2 includes the results from a series of two-sided Wald tests, based on the relevant coefficient estimates." The authors published their original experimental data, and our analysis of it reveals a highly significant difference between the regression coefficients of interest when including the two treatments that are part of our replication ($\chi^2(1) = 31.23$, $p = 2.29 \times 10^{-8}$). The original sample size is 100 (32 in Q_{00} and 68 in Q_{50}). To achieve 90% power, the required replication size is much smaller than this original sample size. The MS Replication Project team has adopted a policy of using the original sample size as a lower bound for replication. In this case, that lower bound of 100 is binding.

To demonstrate that the presence of "informed consumers in the population is sufficient to change the behavior of the uninformed consumers," (p. 3032) the authors estimate a probit regression model with cohort-level random effects. The purchasing decision is the dependent variable; dummy variables are used to represent each possible wait time for each condition (conditions differ on the proportion of informed consumers in the population, $q \in \{0, 0.05, 0.50\}$). The authors then compare corresponding dummy coefficients between conditions in a joint test of equality and reject the null hypothesis in all cases. "For the sharpest comparison, consider conditions without ($q = 0$) and with many ($q = 0.50$) informed consumers, for which the data shows a significant difference in both the high prior conditions Q_{00} and Q_{50} ($\chi^2(4) = 68.99, p < 0.01$), and the low prior conditions Q'_{00} and Q'_{50} ($\chi^2(4) = 107.56, p < 0.01$)" (p. 3032). For the purposes of replication, repeating this statistical approach does not seem sufficiently robust: We could theoretically find coefficient differences between the Q_{00} and Q_{50} conditions—thereby "confirming" the original result—even if the pattern of data we collect bears little resemblance to the original data. Therefore, in consultation with the original authors, we opted for a related test of replication that captures the directionality of the observed behavioral trends.

Like the original test, our test utilizes a probit regression model. We define a dummy variable for experimental condition:

$$D_{Q50} = \begin{cases} 0 & \text{if } q = 0 \quad \text{(zero informed)} \\ 1 & \text{if } q = 0.50 \text{ (many informed)} \end{cases}$$

We then estimate the following probit model with cohort-level random effects:

$$\begin{aligned} \Pr(Purchase) = \beta_1 WaitTime + \beta_2 D_{Q50} \\ + \beta_3 WaitTime \times D_{Q50} + Const + \epsilon \end{aligned}$$

In general, longer waits reduce the purchase probability (captured by $\beta_1 < 0$). If, however, longer waits also signal high quality (which is possible in the condition with informed consumers, Q_{50}), then the inverse relationship between wait time and purchase probability should be less strong in the presence of informed consumers, and we would observe a positive coefficient on the interaction term between wait time and the proportion of informed consumers (i.e., $\beta_3 > 0$). This test was not reported in the original paper, but estimating the model on the original data results in $\beta_3 = 0.85$, ($\chi^2(1) = 74.34, p < 0.0001$). A power analysis reveals that, once again, the sample size required to achieve 90% power is much smaller than the original sample size of 100 (our target sample size for replication). The experiment requires groups of exactly four participants. Groups were randomly assigned to the two experimental conditions.

Sample

Participants for the original study were recruited "from an experimental subject pool at Pennsylvania State University" (p. 3030). The sample for the primary replication consists of 100 students (12 groups in Q_{00} and 13 groups in Q_{50}) from the University of South Carolina. The sample for the secondary replication consists of 104 students (13 groups in both Q_{00} and Q_{50}) from the University of Michigan.

Materials

The study authors kindly provided the original instructions and zTree software, both of which we used for our two replication studies.

Procedure

The replication protocol follows that outlined in Section "4.2 Experimental Design and Implementation" on pages 3029–3031 with some minor deviations, detailed in a later section.

Participants began by reading the instructions which detailed "a simple game in which wait times can be a signal of quality information." They were told that they were one of four human "consumers" that formed a "market" seeking to purchase a series of 26 independent products to be offered sequentially by a firm. Participants were informed of the payoffs from consuming high-quality and low-quality products ($3.50 and $0, respectively), the likelihood of a product being of high quality (50% in the replicated conditions, $p_0 = 0.50$), their waiting cost ($1 per "week"), and the likelihood of being an informed consumer (0% or 50% in the replicated conditions, $q = 0$ and $q = 0.50$). For each product, participants were randomly ordered (they were not informed of this ordering) and asked, one-by-one, whether they would like to purchase the product based on the wait time (i.e., the number of consumers who had purchased the product before them).

The pre-registration report is available at https://aspredicted.org/b43k3.pdf.

Analysis

The analysis for short wait times is identical to the original article: a two-sided Wald test (p. 3032) that is based on a regression with cohort-level random effects (see Table 5 in the online appendix of the original paper). As detailed in the Power Analysis section, we test the effect of longer lines using an alternative probit regression model that was not reported in the original paper.

Differences from Original Study

The differences with respect to the original study are as follows. First, we will use subject pools at the University of South Carolina and the University of Michigan, rather than Penn State. And second, we will use an alternative statistical test, detailed in the Power Analysis section, to establish the effect that uninformed consumers, in the presence of informed consumers, are less sensitive to the purchase probability reduction associated with each marginal unit of wait time.

Replication Results

We find strong evidence for the "empty restaurant syndrome" (that uninformed consumers may balk at empty restaurants, despite short wait times (i.e., when $w = 1$) due to an inferred signal of low quality). The original paper reported purchase probabilities of 95% vs. 68% ($\chi^2(1) = 31.23, p < 0.0001$) for Q_{00} vs. Q_{50} for short waits ($w = 1$). These probabilities were 95% vs. 80% ($\chi^2(1) = 22.72, p < 0.0001$) at the primary replication site (University of South Carolina)—directionally consistent with the original finding and highly significant. Purchase probabilities were 94% vs. 68% ($\chi^2(1) = 47.94, p < 0.0001$) for Q_{00} vs. Q_{50} at the secondary site (University of Michigan), replicating the original data almost perfectly.

We also find support for longer wait times signalling higher quality more generally. Again, we are interested in the interaction term, β_3, in the following probit regression:

$$\Pr(Purchase) = \beta_1 WaitTime + \beta_2 D_{Q50} + \beta_3 WaitTime \times D_{Q50} + Const + \epsilon.$$

Using data from the original paper, we find that $\beta_3 = 0.85$, ($\chi^2(1) = 74.34, p < 0.0001$). At the primary replication site, $\beta_3 = 0.22$, ($\chi^2(1) = 5.35, p = 0.021$). The effect is weaker than in the original data, but it is directionally consistent and statistically significant. At the secondary replication site, $\beta_3 = 0.69$, ($\chi^2(1) = 57.72, p < 0.0001$). Once again, data from the secondary site matches the original data closely.

In summary, we tested the relationship between wait time and inferred quality using two specifications at two sites and found support for the effect in all cases.

Unplanned Protocol Deviations

The replication experiments were conducted as described above with only one deviation from the original protocol. The University of Michigan participants come from two different subject pools (60 from the ORSEE pool and 44 from the Ross pool). Due to low sign-ups from the ORSEE pool, we recruited additionally in the Ross pool to achieve the desired power. The results are consistent and remain significant when the data from each subject pool are analyzed separately.

Discussion

A summary of predicted purchase probabilities when customers face short wait times is provided in Table 1. The purchase probability with no informed consumers (Q_{00}) is similar across the original study and both replications. In the presence of informed consumers (Q_{50}), this probability decreases in the original study and in both replications (highly significant). While the Q_{50} purchase probability at the secondary replication site mirrors the original data almost perfectly, the treatment effect is less pronounced at the primary site. Nevertheless, it is clear that the presence of informed consumers consistently alters the behavior of uninformed consumers who encounter an empty service system.

Table 2 presents the results of a regression (not reported in the original paper, see specification above) designed to roughly capture the overall pattern of purchase probabilities as a function of wait time. The coefficient of interest is β_3 (the coefficient on the interaction of $WaitTime \times D_{Q50}$). Estimated from the original data set, this coefficient has a value of 0.85 and can be interpreted as follows: In Q_{50}, long waits (while undesirable) also serve as a signal of quality, so the negative relationship between wait time and purchase probability is less strong than it is in Q_{00} (where wait time cannot signal quality). At both replication sites, this interaction coefficient is positive and significant. Its magnitude and significance are smaller at both replication sites, though this difference is particularly pronounced at the primary site. Even so, the general effect appears to be robust.

Table 1 Short-wait-time purchase probabilities

Study	Q_{00}	Q_{50}	χ^2	Significance
Kremer and Debo (2016)	95%	68%	31.23	$p < 0.0001$
Primary Site (South Carolina)	95%	80%	22.72	$p < 0.0001$
Secondary Site (Michigan)	94%	68%	47.94	$p < 0.0001$

Table 2 Interaction between wait time and presence of informed consumers

Study	β_3 $(WaitTime \times D_{Q50})$	χ^2	Significance
Kremer and Debo (2016)	0.85	74.34	$p < 0.0001$
Primary Site (South Carolina)	0.22	5.35	$p = 0.021$
Secondary Site (Michigan)	0.69	57.72	$p < 0.0001$

References

Kremer, Mirko, Laurens Debo. 2016. Inferring quality from wait time. *Management Science* **62**(10) 3023–3038.