

Replication Report for “Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence”

By: Maurice E. Schweitzer and Gérard P. Cachon

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

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

Schweitzer and Cachon (2000) investigate newsvendor order quantity decisions for both high-profit products (critical fractile of 75%) and low-profit products (critical fractile of 25%). They find that average order quantities are set too low for high-profit products and too high for low-profit products.

Hypothesis to replicate:

Newsvendor order quantities are set too low for high-profit products and too high for low-profit products.

Power Analysis

[Schweitzer and Cachon \(2000\)](#) state that the “average high-profit order, 176.68, was significantly lower than the expected profit-maximizing order quantity of 225 ($t(32) = 6.58$, $p < 0.001$), and the average low-profit order, 134.06, was significantly higher than the expected profit-maximizing order quantity of 75 ($t(32) = 12.15$, $p < 0.001$)” (p. 412). These t -statistics correspond to p -values of 2.05×10^{-7} and 1.58×10^{-13} .

The original sample size is 34, but one participant’s data was unusable. This left 33 participants overall. To achieve 90% power, the required replication size is smaller than this original sample size (this is true for both the high-profit and low-profit conditions). The MS Replication Project team

adopted a policy of using the original sample size as a lower bound for replication. Further, the team agreed that no replication shall have fewer than 40 participants. In this case, the floor of 40 is binding.

Previous Replications

The pull-to-center effect was first demonstrated by [Schweitzer and Cachon \(2000\)](#) and has been replicated—in spirit, if not in precise methodology—many times in other published papers. [Bolton and Katok \(2008\)](#) demonstrate that the effect persists under a wide range of conditions: when the number of experimental rounds is increased to one hundred; when the number of order options is reduced to three; when realized profit or future expected profit feedback information is given about all ordering options;

and when participants are forced to place ten-period standing orders allowing them to better sample the profit distribution. Further, the too high/too low ordering pattern has been shown to hold across multiple subject pools. The original paper includes only MBA students. Bolton et al. (2012) demonstrate that undergraduate freshmen, graduate students, and professional purchasing managers all suffer from pull-to-center. And Lee et al. (2018) show that the effect is observed among participants recruited on Amazon's Mechanical Turk platform.

Because the general phenomenon has already been well-established, our replication effort focuses on testing the original methodology, specifically. As such, the experimental interface, protocol, subject pools, and payment mechanisms were kept as similar to the original paper as possible.

Sample

The sample at each location consisted of 40 participants. Cornell University was the primary replication site, and the University of Texas at Dallas was the secondary site. Participants for the original study were recruited "from a Duke University MBA operations management course." For consistency with the original study, each site began by recruiting MBA students who were either currently taking or had completed courses related to operations management. The primary site (Cornell) was able to collect 40 such MBA students for its data set. The secondary site (UT-Dallas) was able to collect 38 such MBA students. After exhausting the relevant MBA pools, the secondary site recruited two additional participants from a master's program to complete its data set of 40 participants.

Materials

The authors kindly provided the original software, which consisted of a Microsoft

Excel file with VBA macros, and instructions. Due to the online nature of the replication and technical issues with the original software, we converted the program to SoPHIE software while ensuring that the task, decision support, and interface was similar to the original software. We used the original instructions provided by the authors modified slightly to update any Excel-specific experimental mechanics (e.g., we deleted the sentence beginning "Once you finish the exercise please e-mail this workbook back to ...").

Procedure

We followed the same protocols outlined in section "3.1.1 Methods" on pages 409–410 with some minor deviations, detailed in a later section.

Each participant began by reading the instructions which detailed an inventory task. They were informed that the demand would be uniformly distributed between 1 and 300 and that the selling price would be 12 francs in each round. They were not aware of the number of rounds (30) or the production cost in future rounds (3 francs in "high-profit" rounds and 9 francs in "low-profit" rounds). Profits were calculated in francs and converted at a rate of 300 to \$1 U.S. dollar in the original study. In our replication, we used a more favorable exchange rate of 200 francs to \$1 U.S. dollar to account for inflation between the year of the original publication (2000) and the year of replication (2021). In the original study, one participant was "selected at random and paid in dollars." We employed a similar payment scheme: paying a large sum of money (optimal ordering would earn about \$85) based on performance to two randomly chosen participants (we chose to pay two participants rather than just one due to the increased sample size) per "wave." Data collection at Cornell was completed in a single wave (i.e., a single recruitment effort in

the month of April, 2021, yielded 40 participants). At UT Dallas, however, the first wave (April, 2021) only yielded 23 participants, so a second wave was run with a fresh cohort of MBA students in August, 2021. Because of this, a total of four participants were paid real money at UT Dallas while only two were paid at Cornell.

As in the original study, participants were provided with the ability to test order quantities and could observe the following information: the profit distribution as a table and graph, the probability that sales will exceed the order quantity, the probability that sales will be lower than the order quantity, the break-even sales level, and the probability that sales will be no lower than this break-even sales level. Participants could view results from previous rounds in a table that included price, cost, demand realization, order quantity, sales, round profit, and cumulative profit.

After setting an order quantity, participants observed realized demand for the round and their profit. We used the same approximate demand draws from the original experiment (see Figures 1 and 2 on page 411, and the approximated values in Table 2 in the Appendix below).

Each participant made 15 decisions under the high-profit condition and 15 under low-profit. In the original experiment, 20 out of 34 participants played the high-profit condition first and 14 out of 34 participants played the low-profit condition first. For our replication, we randomly assigned the ordering of the profit conditions for each participant. At each location, 21 played the high-profit condition first and 19 played the low-profit condition first.

The pre-registration report is available at <https://aspredicted.org/759k8.pdf>.

Analysis

The analysis is identical to the original article: a one-sample t -test for each of the two conditions, high-profit and low-profit (top

of page 412). A successful replication of the “too low/too high” pattern would require that mean order quantities differ from the optimal levels *in the predicted directions*: the average order quantity should be significantly below the optimal level in the high-profit condition and significantly above the optimal level in the low-profit condition.

Differences from Original Study

The differences with respect to the original study are as follows. First, we conducted the study using SoPHIE, rather than Excel with VBA, at Cornell University and the University of Texas at Dallas. Second, the participant pool did not solely consist of MBA students at the secondary site (38 MBA and 2 MS students). Third, we added a page of instructions to the experiment detailing the random payment mechanism, which was not in the original instructions (as noted earlier, we adjusted payments for inflation and paid two participants).

Finally, we made a small edit to the experimental interface. The original authors calculated several probabilities related to *sales* on the main “info” page of the experiment, the first of which was the “Probability sales will be greater than your order quantity.” The statistics that the original authors actually reported were related to *demand*, rather than *sales*. Thus, we changed the word “sales” to “demand” for the four statistics reported on the info page (e.g., we reported the “Probability demand will be greater than your order quantity”). The numbers that participants saw remained unchanged; this was only a change of one word in the descriptive label.

Replication Results

For both the primary and secondary sites, the replication hypothesis is supported. For the primary site, the average high-profit order quantity was 196.29, significantly lower than the normative prediction of 225 ($t(39) = 3.95$, $p < 0.001$). The average

low-profit order quantity was 131.31, significantly higher than the normative prediction of 75 ($t(39) = 11.21$, $p < 0.001$).

For the secondary site, the average high-profit order quantity was 177.51, significantly lower than the normative prediction of 225 ($t(39) = 6.12$, $p < 0.001$). The average low-profit order quantity was 134.50, significantly higher than the normative prediction of 75 ($t(39) = 14.54$, $p < 0.001$).

In summary, at both replication sites, newsvendor order quantities were set too low for high-profit products and too high for low-profit products.

Unplanned Protocol Deviations

The only unplanned event in running this study was the need to recruit a second wave of participants at UT Dallas resulting in four participants earning real money. Otherwise, the replication experiments at both sites were conducted as described as planned without any protocol deviations.

Discussion

A summary of the average order quantities in the original study and each replication site, in the high-and-low profit treatments, is provided in Table 1. Comparing the average order-quantity results of Schweitzer and Cachon to the primary site, there appears to be a marginal difference in the high-profit treatment (176.68 versus 196.29). Despite this, in both data sets, there remains clear evidence of average-order quantities being set too low in the high-profit treatment. In the low-profit treatment, between the original data and the primary site, the average order quantities are quite close to each other (135.06 versus 131.31). Turning to the secondary site, average order quantities are virtually identical between Schweitzer and Cachon’s original data and the secondary site (176.68 versus 177.51 in high-profit and 135.06 versus 135.50 in low-profit).

Table 1 Order Quantity Results

Study	High Profit	Low Profit
Schweitzer and Cachon (2000)	176.68 (7.34)	135.06 (4.86)
Primary Site (Cornell)	196.29 (7.27)	131.31 (5.02)
Secondary Site (UT-Dallas)	177.51 (7.76)	134.50 (4.09)

Note: Standard errors reported in parentheses.

References

- Bolton, Gary E, Elena Katok. 2008. Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management* **10**(3) 519–538.
- Bolton, Gary E, Axel Ockenfels, Ulrich W Thonemann. 2012. Managers and students as newsvendors. *Management Science* **58**(12) 2225–2233.
- Lee, Yun Shin, Yong Won Seo, Enno Siemsen. 2018. Running behavioral operations experiments using amazon’s mechanical turk. *Production and Operations Management* **27**(5) 973–989.
- Schweitzer, Maurice E., Gérard P. Cachon. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* **46**(3) 404–420.

Appendix

The demand realizations, which are approximated from Figures 1 and 2 of the original study, are provided below in Table 2:

Table 2 Demand Realizations

Round	High Profit	Low Profit
1	281	210
2	260	157
3	13	172
4	275	80
5	106	82
6	155	236
7	233	2
8	12	233
9	167	244
10	146	208
11	87	9
12	180	122
13	188	252
14	77	240
15	80	107