Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior

Nils Rudi, David Drake
INSEAD, Boulevard de Constance, 77305 Fontainebleau, France, nils.rudi@insead.edu, david.drake@insead.edu

In an experimental newsvendor setting where 310 subjects make 50 repeated newsvendor decisions with the same known ex-ante parameters, we investigate three phenomena: Level behavior – the decision-maker’s average ordering tendency; adjustment behavior – the tendency to adjust period-to-period order quantities; and observation bias – the tendency to let the degree of demand feedback influence order quantities. We measure ordering behavior in terms of decisions (quantities) and performance (expected mismatch cost). We find that the portion of mismatch cost due to adjustment behavior exceeds the portion of mismatch cost due to level behavior in three out of four conditions, highlighting the importance of considering order adjustment in addition to level behavior, which has thus far received more research attention. Observation bias is studied through censored demand feedback, a situation which arguably represents the majority of newsvendor settings. When demands are uncensored, subjects tend to order below the normative quantity when facing high margin and above the normative quantity when facing low margin, but in neither case beyond mean demand (a.k.a. the pull-to-center effect). Censoring in general leads to lower quantities, magnifying the below-normative level behavior when facing high margin but partially counterbalancing the above-normative level behavior when facing low margin, violating the pull-to-center effect in both cases.

May 2009, revised August 2010

1. Introduction and literature review

The newsvendor model is the fundamental model for managing inventory under demand uncertainty and has received significant attention by researchers since its introduction by Arrow et al. (1951). Most of this research has taken a normative approach, while behavioral issues – the focus of this paper – have only recently received attention.

We study two aspects of ordering behavior in a repeated newsvendor model, and a bias that affects them both. We define level behavior as a decision-maker’s average order quantity. Using statistical terminology, level behavior can be thought of as the first moment of ordering behavior.
If, for example, the normative order quantity is 730, and a newsvendor orders 810 on average, then there is a deviation of 80 between the normative quantity and the level behavior. A prevalent level behavior result in the existing literature is that subjects on average tend to order quantities between the expected demand and the normative order quantity, which has been coined the pull-to-center effect. We define adjustment behavior as the tendency to adjust period-to-period order quantities. Again, in statistical terminology, adjustment behavior can be thought of as the second moment of ordering behavior. If, for example, one newsvendor alternates between orders of 809 and 811, and another newsvendor alternates between orders of 780 and 840, we say that the second newsvendor exhibits a greater degree of adjustment behavior than the first. Cachon and Terwiesch (2009) give an example of such quantity adjustments with respect to influenza vaccines, where one would expect the same ex-ante season-to-season parameters and, hence, no strong reason to adjust the quantity between two consecutive seasons:

There were 95 million doses of the flu vaccine produced for the 2002-2003 flu season in the United States. Unfortunately, 12 million doses were not used and had to be destroyed (a vaccine is good only for one flu season). Only 83 million doses of the flu vaccine were produced for the next season, 2003-2004. (Not coincidentally, 95-12=83.) Unfortunately, in that season there were widespread shortages, leading to flu related deaths, especially in Colorado.

This example is consistent with demand chasing, which is a bias that drives adjustment behavior and has been observed frequently within the experimental newsvendor literature.

Arguably, in most practical settings, demand feedback tend to be in the form of demand realizations censored by the ordered quantity, i.e., the newsvendors observe sales realizations, not demand realizations. Furthermore, even when firms have data that could be used to develop a rough estimate of lost sales, in our experience, they do not tend to do so. Revisiting the example of the influenza vaccine given above, note that it reports a specific left-over quantity for the 2002-2003 season, but only concludes that the shortages in the 2003-2004 season were “widespread.” We define observation bias as the tendency that the degree of feedback available – here, whether
full demand realizations are observed or if demands are censored – influences level and adjustment behavior. In this study, we manipulate demand feedback (censored and uncensored) as an experimental treatment to induce and test for observation bias.

The main contributions of this paper are two-fold. First, though the existing literature explores both level and adjustment behavior in newsvendor settings, we are, to our knowledge, the first to separate these behavioral patterns into each of their respective performance impacts. This allows us to disambiguate the overall behavioral cost (i.e., the increase of expected mismatch cost due to behavioral deviations from normative order quantity) by mapping level and adjustment behavior to their respective effects on expected mismatch cost, which we term level cost and adjustment cost, respectively. Strikingly, we find that adjustment cost exceeded level cost in three out of four of our experimental conditions, underscoring the importance that it be taken into account when studying newsvendor behavior. The second, and more significant, contribution of this paper is that it is, to our knowledge, the first experimentally controlled study of the impact of demand censoring on ordering behavior. We find that demand censoring induces a reduction in order quantities relative to uncensored settings, supporting the presence of an observation bias on level behavior. Hence, when margins are high, censoring magnifies the pull-to-center effect, but when margins are low, censoring partially counterbalances it – in both of our margin conditions this impact was strong enough that the pull-to-center effect typically found in experimental newsvendor studies was violated. We also investigate the impact of observation bias on adjustment behavior, proposing and testing mechanisms by which demand censoring may impact order adjustments. We find that the magnitude of adjustments are related to the degree of variability observed by subjects. Finally, we find that there tends to be a positive relationship between the degree of adjustment behavior and the distance between subjects’ level behavior and the normative order quantity, indicating that those who limit adjustment cost also tend to limit level cost.

1.1. Literature review

The seminal work of Kahneman and Tversky (see for example Tversky and Kahneman 1974) catalyzed a general interest in cognitive heuristics and the decision biases that these heuristics
expose decision-makers to. Bazerman and Moore (2009) provides a summary of these and other decision biases relevant to managerial settings.

The study of behavioral issues in operations management is far more nascent. Schweitzer and Cachon (2000) initiated the exploration of behavioral issues in newsvendor settings, primarily focusing on what we refer to as level behavior. They consider the anchoring and insufficient adjustment bias, as well as different preference functions: risk neutral, risk-averse and risk-seeking, prospect theory, loss-averse, waste-averse, stockout-averse, underestimated opportunity costs, and minimizing ex-post inventory error. From these ten preference structures, they find that only anchoring and insufficient adjustment and the minimization of ex-post inventory error explain the observed behavior – the tendency for subjects to order between the normative (risk neutral) solution and the mean demand – a result now referred to as the pull-to-center effect. This pattern of behavior has been ubiquitously reported within the experimental newsvendor literature (Bolton and Katok 2008, Lurie and Swaminathan 2009, Bostian et al. 2008, Benzion et al. 2008, etc.). Schweitzer and Cachon do not find support for learning in terms of order quantities trending toward the normative solution over time.

Extensions of Schweitzer and Cachon (2000) include Kremer et al. (2010), who manipulate task complexity by varying demand states and order options as well as investigate regret and operational versus context-neutral framing. Schultz et al. (2007) investigate cost versus profit framing, Corbett and Fransoo (2007) perform a survey of entrepreneurs to elicit risk profiles and newsvendor ordering behavior, and Gavirneni and Isen (2010) perform a protocol analysis on the questions participants asked to solicit information to aid them with a newsvendor decision.

In a comprehensive investigation of learning with respect to level behavior within newsvendor settings, Bolton and Katok (2008) find significant learning, with subjects trending toward the normative solution in eleven out of twelve of their experimental conditions. They attribute this to the inclusion of 100 periods per condition rather than the 15 periods per condition used by Schweitzer and Cachon (2000). Benzion et al. (2008) also find support for significant learning in terms of level behavior with a 100-period horizon, and so do Bostian et al. (2008) and Lurie and
Swaminathan (2009) despite a horizon of only 30 periods, while Schultz et al. (2007), and Kremer et al. (2010) do not find such learning with the number of periods per condition ranging from 20 to 30.

With respect to the treatment of adjustment behavior within the literature, Schweitzer and Cachon (2000) find weak support for demand chasing (i.e., the tendency to adjust the order quantity in the direction of the previous period’s imbalance between demand and quantity). Bolton and Katok (2008) provide perhaps the most thorough treatment of adjustment behavior to date, segmenting subjects by those whose order adjustment can be accounted for primarily by their trend in level behavior, those exhibiting a prevalence for adjustments toward the most recent demand realization (demand chasing), those tending to adjust away from the most recent demand realization (gambler’s fallacy), and those whose adjustments were not statistically different from random. However, Bolton and Katok conduct this study of adjustment behavior within a setting where subjects only had three possible order quantities to select from, which may have attenuated within subject order variability. Others have also reported demand chasing and trending in level behavior, which would both contribute to order adjustment (see for example Bostian et al. 2008, Benzion et al. 2008, and Lurie and Swaminathan 2009). Bolton et al. (2008) report adjustment behavior in terms of within subject standard deviation with respect to order quantities, which is the same measure that we use. Adjustment behavior “learning” has been reported in the form of a decreased frequency in order changes (Schweitzer and Cachon 2000), decreased magnitude of order changes (Benzion et al. 2008), and the attenuation of demand chasing over time (Bostian et al. 2008).

While various aspects of level and adjustment behavior have been reported within the literature, measures of performance applied within the literature have only considered the impacts of these behaviors in an aggregate metric. Those who report on performance often do so in terms of profits (Benzion et al. 2008, Lurie and Swaminathan 2009), with others not reporting performance impacts at all (Schweitzer and Cachon 2000, Tong et al. 2009). Bolton and Katok (2008) and Bolton et al. (2008) report performance in terms of the proportion of the normative expected profit earned. Bolton et al. (2008) refer to this measure as “efficiency” and note that “it penalizes for order
variability; for example, ordering 75 every period [the normative solution in their setting] is a more efficient strategy than alternating between orders of 100 and 50.” This efficiency metric still combines the impact of level and adjustment behavior. We extend the literature by offering a means to map performance erosion into these two aspects of ordering behavior.

The literature studying normative issues related to demand censoring include its modeling (see for example Braden and Freimer 1991) and demand estimation using censored data (see for example Conlon and Mortimer 2010). For normative models prescribing ordering policies when the underlying demand distribution is unknown (in contrast to our setting) and demands are censored, a common result is that demand censoring leads to higher normative order levels due to the value of improved demand information through the ability to observe a larger range of demand realizations (Harpaz et al. 1982, Nahmias 1994, Lariviere and Porteus 1999 and Ding et al. 2002). In a recent paper, Tong et al. (2009) study Bayesian updating in an experiment with censored demand feedback and a newsvendor fractile of 1/2. They also find that demand censoring leads to lower quantities. However, in their study it is not clear what aspects of ordering behavior can be attributed to observation bias and what can be attributed to demand learning, an important distinction as each would be from different thought processes and require different improvement approaches.

Additional work in the field of behavioral operations management includes studies of behavioral impacts on the bullwhip effect (Croson and Donohue 2005, Croson et al. 2005 and Bloomfield et al. 2007), consumer estimation of household inventories (Chandon and Wansink 2006), revenue management (Bearden et al. 2008), and the effect of social preferences (Loch and Wu 2008a) and service-level agreements (Katok et al. 2008) on supply chain coordination. For an overview of behavioral research in operations, readers are referred to Bendoly et al. (2006), Gino and Pisano (2008) and Loch and Wu (2008b).

The next section outlines the normative newsvendor model. Section 3 describes the experimental design of the paper and the notation related to each experimental condition. In Section 4 we present the theory, hypotheses and results corresponding to level and adjustment behavior as well
as observation bias. Section 5 presents additional findings related to learning in terms of both level and adjustment behavior and the relationship between both behaviors. Finally, Section 6 provides a summary discussion of key results and managerial implications.

2. The Normative Newsvendor

In the normative model, the newsvendor decides order quantity $Q$ at unit cost $w$ while facing uncertain demand $D$ in a single-period environment. After demand is realized, the newsvendor sells at unit revenue $r$ a quantity limited by demand $D$ and order quantity $Q$. Here we consider the situation in which leftover inventory is disposed of (at no value). Define $a^+ = \max(0, a)$. The newsvendor’s decision then results in realized profit

$$\Pi(D, Q) = (r - w)D - G(D, Q),$$

where $G(D, Q) = c_u(D - Q)^+ + c_o(Q - D)^+$ is the mismatch cost when ordering $Q$, with parameters unit underage cost $c_u = r - w$ and unit overage cost $c_o = w$. Hence, the profit function (1) is separated into two parts: The profit if one could obtain exact demand information before committing to the quantity; and the cost of demand uncertainty (mismatch cost). The latter consists of the financial consequence per unit of unmet demand (unit underage cost) times the expected unmet demand, and the financial consequence per unit of leftover quantity (unit overage cost) times the expected leftover quantity. Let $\Pi(Q) = E\Pi(D, Q)$ denote expected profit and $G(Q) = E G(D, Q)$ denote expected mismatch cost, when ordering $Q$. Note that expected mismatch cost is the controllable part of expected profit, i.e., the only part which depends on the decision variable $Q$. Hence, the minimizer of $G(Q)$ will be equal to the maximizer of $\Pi(Q)$. The normative solution minimizing expected mismatch cost when demand follows a continuous distribution is the order quantity $Q^*$ which satisfies $F(Q^*) = c_u/(c_u + c_o)$, where the left-hand side $F(\cdot)$ is the Cumulative Distribution Function (CDF) of $D$ and the right-hand side is the newsvendor fractile.

3. Experimental design and notation

We conducted newsvendor experiments as one session of the core logistics course in the undergraduate program at The Norwegian School of Management. Each session started with a 90-minute
lecture on the newsvendor model that included explanation of the key features of the problem, motivation of its relevance, explanation of the tradeoff between underage and overage, including which side of the mean to order on, and insight into the effects of parameters. The lecture did not, however, include the newsvendor fractile nor a method for computing the normative quantity. The lecture also specifically covered the issue of censored demand realizations, and subjects were informed that some of them, but not all, would face such situations – and that all of them would be informed about the actual underlying demand distribution which was i.i.d. across periods. Similarly, they were informed that some would face conditions where margin exceeded cost, while others would face conditions where cost exceeded margin. After the lecture, students participated in a 2x2, between-subject experiment with treatments to manipulate the margin of “wodgets” sold (low vs. high) and the degree of demand feedback (uncensored vs. censored).

In the experiment, we used normally distributed demands with mean $\mu = 1000$ and standard deviation $\sigma = 400$, so that the coefficient of variation was large enough to make an impact and small enough for a normal distribution to be reasonable. All subjects experienced realizations from a common demand path. Since all subjects also upfront received complete information regarding the demand distribution, we can focus exclusively on the impact of observation bias, rather than confounding its effect with that of demand learning. Furthermore, a known demand distribution makes a myopic policy optimal, thus avoiding complications that would result from dynamics.

Following Schweitzer and Cachon (2000), we fixed unit revenue at $r = 12$ while we manipulated unit cost as one treatment, with $w = 3$, which we denote high margin good (HMG), and $w = 9$, which we denote low margin good (LMG). To study observation bias, we manipulated demand feedback, with uncensored demand, where subjects observed actual demand realizations $D$, and censored demand, where subjects were limited to observing quantities sold, $\min(D, Q)$. Subjects in the uncensored condition were also informed of underage costs, $c_u (D - Q)^+$, and overage costs, $c_o (Q - D)^+$, for each period. Within the censored condition, subjects were still informed of demand outcomes and overage costs for periods following overage outcomes, but they were only informed that a stock-out occurred during periods in which they experienced underage.
A total of 310 subjects participated in the experiment, with each randomly assigned to one of the four conditions. We include the results from 269 of these subjects in the results presented within Section 4. Of the 41 remaining subjects, 30 were dropped from the study because they did not complete the 50 trials within the allotted time, two were unable to play the game due to technical problems, four were excluded due to strong evidence suggesting that they had obtained demand information, and the remaining five were excluded for severe over-ordering. (Of the latter five, four had orders in excess of 5000 and one had an average order quantity 8.5 standard deviations greater than the condition’s average.) Finally, we corrected three specific order quantities of the remaining subjects where a mistyping was obvious.

Subjects completing all 50 trials entered a lottery where each “euro” of profit they accumulated throughout the experiment earned them a chance to win an iPod. One iPod was raffled in each of the six course sections. The Appendix provides details on the information and interface presented to the subjects.

We use expected mismatch cost (cost of demand uncertainty) rather than profit to estimate subject performance. We do this for two reasons: (i) It represents the *controllable part* of expected profit, i.e., the part which depends on $Q$, and (ii) it enables us to compare the alternatives with low and high margin goods on equal terms. The latter stems from the fact that the expected mismatch cost functions of the two margin scenarios are mirror images of each other around mean demand. This results from unit underage (overage) cost in LMG (HMG) equalling unit overage (underage) cost in HMG (LMG), and from the demand distribution in all conditions being identical and symmetric around the mean. In the general model, and when making statements which are valid independent of margin, we use capital and bold-face notation for quantity decisions $Q$ and mismatch cost $G$. When making margin-specific statements, we use regular upper-case notation for HMG and lower-case notation for LMG (a notation mnemonic: big notation corresponds to big margin and small notation corresponds to small margin). The normative order quantities are then $q^* = 730$ and $Q^* = 1270$ for LMG and HMG, respectively. Furthermore, we use superscript $u$ for the uncensored case and superscript $c$ for the censored case. We index a specific subject (newsvendor)
by $i$, and consider the situation in which there are $T = 50$ repeated periods, indexed by $t$, and $D_t$’s are normal i.i.d.. When the subscript $t$ is dropped, it indicates an average value over $t$. So subject $i$’s average quantity when facing HMG is $Q_i = \sum_{t=1}^{T} Q_{i,t}/T$, subject $i$’s standard deviation of order quantities when facing LMG is $\sigma_{q_i}$, and subject $i$’s average expected mismatch cost when facing LMG is $g_i = \sum_{t=1}^{T} g(q_{i,t})/T$.

Based on the 2x2 design employed in the experiment, subjects are members of a set determined by condition, $LU$, $LC$, $HU$, and $HC$, which are four disjoint sets where $L$ denotes a low margin good, $H$ denotes a high margin good, $U$ denotes uncensored demand, and $C$ denotes censored demand. (Each subject plays one of these conditions for the entire duration of the experiment.) We denote condition average measures by dropping the subscript $i$, e.g., $Q^c = \sum_{i \in HC} Q_i / \langle HC \rangle$ and $\sigma_{q^u} = \sum_{i \in LU} \sigma_{q_i} / \langle LU \rangle$, where $\langle A \rangle$ denotes the cardinality of the set $A$ (i.e., the number of subjects in the condition). Table 1 summarizes this general notation as well as notation specific to experimental conditions. (Note that level and adjustment costs indicated in Table 1 are introduced in Section 4.3.)

<table>
<thead>
<tr>
<th>Condition notation</th>
<th>LU</th>
<th>LC</th>
<th>HU</th>
<th>HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margin condition</td>
<td>LMG</td>
<td>LMG</td>
<td>HMG</td>
<td>HMG</td>
</tr>
<tr>
<td>Demand feedback</td>
<td>Uncensored</td>
<td>Censored</td>
<td>Uncensored</td>
<td>Censored</td>
</tr>
<tr>
<td>Parameter</td>
<td>$r$</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Values</td>
<td>$w$</td>
<td>9</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>$c_u$</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>$c_c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General quantity</td>
<td>$Q$</td>
<td>$Q_{i,t}$</td>
<td>$Q_i$</td>
<td>$Q^u$</td>
</tr>
<tr>
<td>Individual order quantity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual avg. quantity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition avg. quantity</td>
<td>$q^u$</td>
<td>$q^c$</td>
<td>$Q_u$</td>
<td>$Q_c$</td>
</tr>
<tr>
<td>General avg. within std. deviation</td>
<td>$\sigma_{q^u}$</td>
<td>$\sigma_{q^c}$</td>
<td>$\sigma_{Q^u}$</td>
<td>$\sigma_{Q^c}$</td>
</tr>
<tr>
<td>Condition avg. within std. deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General individual level cost</td>
<td>$\delta^u$</td>
<td>$\delta^c$</td>
<td>$\Delta^u$</td>
<td>$\Delta^c$</td>
</tr>
<tr>
<td>Condition avg. level cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General individual adjustment cost</td>
<td>$\psi^u$</td>
<td>$\psi^c$</td>
<td>$\Psi^u$</td>
<td>$\Psi^c$</td>
</tr>
<tr>
<td>Condition avg. adjustment cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Experimental condition notation and parameters, and general notation.

4. Development and analysis of hypotheses

In this section, we develop the theory and hypotheses related to level and adjustment behavior, particularly as they pertain to observation bias under demand censoring. We present the corresponding results within each subsection to facilitate the exposition’s cohesiveness.
It is important to note the unit of analysis on which we generally measure and test ordering behavior in this paper. When performing tests on level behavior (average order quantities), we do so on the population of subjects’ average quantities \( Q_i \). The rationale behind this is that it is individual subjects, not groups of subjects, that exhibit a particular behavioral pattern. Similarly, when performing tests on the effect of adjustment behavior on decisions (variability of order quantities), we do so on the population of within-subject standard deviation of quantities \( \sigma_{Q_i} \) since within-subject variability represents the extent to which a subject adjusts her order quantities, while between-subject variability reflects the extent to which subjects differ in level behavior. The descriptive statistics resulting from the experiment are given in Table 2.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>LU ( I = 64 )</th>
<th>LC ( I = 68 )</th>
<th>HU ( I = 66 )</th>
<th>HC ( I = 71 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects’ average order quantities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normative: ( Q^* )</td>
<td>7.30</td>
<td>7.90</td>
<td>12.70</td>
<td>12.70</td>
</tr>
<tr>
<td>Mean</td>
<td>836.72</td>
<td>690.31</td>
<td>1080.76</td>
<td>964.13</td>
</tr>
<tr>
<td>Median</td>
<td>844.05</td>
<td>687.70</td>
<td>1053.63</td>
<td>1043.54</td>
</tr>
<tr>
<td>Std deviation</td>
<td>118.04</td>
<td>146.66</td>
<td>140.69</td>
<td>236.70</td>
</tr>
<tr>
<td>95% conf. int.</td>
<td>[807.24, 866.21]</td>
<td>[654.81, 725.81]</td>
<td>[1046.17, 1115.34]</td>
<td>[908.11, 1020.16]</td>
</tr>
<tr>
<td>Min</td>
<td>526.26</td>
<td>387.00</td>
<td>808.84</td>
<td>515.90</td>
</tr>
<tr>
<td>Max</td>
<td>1156.98</td>
<td>1089.48</td>
<td>1466.20</td>
<td>1445.00</td>
</tr>
<tr>
<td>Within-subject standard deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normative: ( G(\bar{Q'}) )</td>
<td>1525.33</td>
<td>1525.33</td>
<td>1525.33</td>
<td>1525.33</td>
</tr>
<tr>
<td>Mean</td>
<td>1993.34</td>
<td>1800.87</td>
<td>2230.64</td>
<td>2602.65</td>
</tr>
<tr>
<td>Median</td>
<td>1875.58</td>
<td>1728.86</td>
<td>2260.59</td>
<td>2336.81</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>341.65</td>
<td>262.26</td>
<td>404.70</td>
<td>889.46</td>
</tr>
<tr>
<td>95% conf. int.</td>
<td>[1907.99, 2078.68]</td>
<td>[1737.39, 1864.35]</td>
<td>[2131.15, 2330.12]</td>
<td>[2392.12, 2813.18]</td>
</tr>
<tr>
<td>Min</td>
<td>1542.19</td>
<td>1547.75</td>
<td>1595.08</td>
<td>1525.41</td>
</tr>
<tr>
<td>Max</td>
<td>3489.14</td>
<td>3101.20</td>
<td>3907.31</td>
<td>4665.81</td>
</tr>
</tbody>
</table>

Table 2 Descriptive statistics.

4.1. Level behavior

Theory and hypotheses: level behavior in uncensored environments. Here we study subject decisions (i.e., order quantities) with respect to level behavior when demands are uncensored.

Schweitzer and Cachon (2000) find support for the pull-to-center effect, i.e., a tendency to order between the normative solution and expected demand, which is consistent with anchoring and insufficient adjustment as well as ex-post inventory error minimization. We expect that our experiment will yield corresponding results when demand is uncensored.
Hypothesis 1. The order quantities with uncensored demands will fall between the normative quantity and mean demands, (a) \( q^* < q^u < \mu \) and (b) \( \mu < Q^u < Q^* \).

Results. The average quantity values are shown in Figure 1(a).

To test Hypotheses 1(a) and (b), we construct 99% confidence intervals around \( q^u \) and \( Q^u \). The null hypotheses (i.e., that pull-to-center will not be observed) will be rejected at the \( p < 0.01 \) level if the confidence limits are contained within their respective pull-to-center regions, which has bounds established by expected demand and the normative order quantity (i.e., between 730 and 1000 in LMG, and between 1000 and 1270 in HMG). From Table 3, we see that there is statistically significant evidence for subjects’ average order quantities falling within the pull-to-center boundaries (\( p < 0.01 \) for both LMG and HMG). These results are consistent with Schweitzer and Cachon (2000) as well as the newsvendor experiments we refer to in our literature review.\(^1\)

![Figure 1](image-url)  
*Figure 1* For the four conditions, (a) average order quantities and (b) average standard deviation of order quantities.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>s.e.</th>
<th>99% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1(a)</td>
<td>LU</td>
<td>836.72</td>
<td>14.76</td>
</tr>
<tr>
<td>H1(b)</td>
<td>HU</td>
<td>1080.76</td>
<td>17.32</td>
</tr>
</tbody>
</table>

Table 3 99% confidence interval of uncensored quantities testing pull-to-center effect.

\(^1\)Schweitzer and Cachon (2000) also model other preference functions with a clear prediction of effect on order quantities. Specifically, risk seeking and stock-out averse preferences would lead to over-ordering, while risk aversion, underestimation of mismatch cost, waste aversion and loss aversion preferences would lead to under-ordering. The latter four preferences as secondary effects would lead us to conjecture that the pull-to-center effect for LMG is weaker than for HMG which in notation would translate to \( (Q^* - Q^u) - (q^u - q^*) > 0 \). This is supported here (\( p < 0.01 \)), and is in contrast to Schweitzer and Cachon (2000), although in their case not with significance.
Theory and hypotheses: the effect of observation bias on level behavior. Most practical news-vendor settings involve demand censoring. Firms typically have visibility to sales, but rarely have the luxury of observing true demand realizations. Although the literature explores demand censoring from a normative perspective, to our knowledge, no prior work has explored its behavioral impacts. We posit that demand censoring will impact decision-making by inducing bias, which we label observation bias, that reduces order quantities.

We propose that demand censoring can induce observation bias through salience differences that arise from the asymmetric awareness of overage and underage costs. Lewis (1969) defines salience as an entity or property that “stands out from the rest by its uniqueness in some conspicuous respect.” Under demand censoring, decision-makers receive more precise demand feedback following overage events than following underage events: i.e., they receive information that is asymmetric in its salience. We posit that this asymmetry can impact decision-making through the salience effect, which Tversky and Kahneman (1974) offer as one factor that can influence subjective probabilities under the availability heuristic. Tversky and Kahneman (1973) define the availability heuristic as “a judgmental heuristic in which a person evaluates the frequency of classes or the probability of events by availability, i.e., by the ease with which relevant instances come to mind.”

Through the availability heuristic, decision makers place greater subjective probabilities on events or outcomes that are more easily recalled. It is important to note that Tversky and Kahneman (1973) points out that the ease with which relevant information could be recalled from memory affects subjective probabilities; the cognitive operations of retrieval (recall from memory) and construction (solving the task at hand) are not necessary for the heuristic to impact decisions. Salience effect bias arises when the availability heuristic leads to subjective probability distortion relative to objective probabilities due to asymmetries in the richness or cognitive impact of informational cues. Taylor et al. (1979) argued that salience effects impact decision-making on the subconscious level, i.e., that they are “an automatic perceptual bias (not unlike optical illusions), which occur without intention.” They have since softened that stance, stating that salience effects are not fully automatic because they can be overcome through training, coaching, and other “forms
of involvement” (Fiske and Taylor 1991). This implies that subjects do not have to explicitly consider the subjective probabilities that they place on various demand realizations for the salience effect to impact their decision making. Simply being exposed to the asymmetric information cues is sufficient.

Given that subjects have full knowledge of the demand distribution and other system parameters required to solve the newsvendor problem, they should (normatively) ignore demand realization signals. However, in accordance with Massey and Wu’s (2005) system neglect hypothesis, they can be expected to place inordinate weight on the signal (i.e., the overage and underage outcomes) that they observe relative to the system parameters that should drive normative decision-making (i.e., unit overage cost, unit underage cost, and the demand distribution). The system neglect hypothesis reflects a more general finding – that salient, but logically irrelevant (or less valuable), information often impacts judgment and decision-making to a greater extent than it normatively should, a result supported throughout various literatures (see Taylor and Fiske 1978 for examples in social psychology; and Chapman and Johnson 2002 for examples in cognitive psychology). Within newsvendor settings, the commonly reported result of demand chasing is a further example of such an effect.

In light of the above discussion on observation bias, we expect demand censoring to impact level behavior according to the following hypotheses:

**Hypothesis 2.** *Demand censoring will lead to a reduction in order quantity levels, (a) $q^c < q^u$ and (b) $Q^c < Q^u$. The difference due to demand censoring will be more prominent for LMG than for HMG, (c) $Q^u − Q^c < q^u − q^c$. 

Hypothesis 2(c) arises from the conjecture that order quantities will be greater for HMG than for LMG. This level difference implies that demand feedback will more frequently be censored for LMG than for HMG. Hence, we expect a stronger effect due to censoring for LMG than for HMG.

Note that, should these hypotheses hold, demand censoring will cause subjects to order further from the normative quantity in HMG settings. In LMG settings, however, demand censoring would
cause subjects to adjust quantities toward (and possibly beyond) the normative quantity $q^*$, i.e., it would partially counter-balance the pull-to-center effect on $q^u$. We will discuss this further in subsection 4.3, where we discuss performance impacts of level behavior, adjustment behavior, and observation bias.

**Results.** The average quantity values of the four conditions are given in Table 2 and shown in Figure 1(a).

From Table 4, we see that for both LMG and HMG, order quantities tend to be significantly lower with demand censoring than without it (by 141.41 and 116.62 units, respectively), providing support for H2(a) and (b) ($p < 0.01$ for both LMG and HMG). In fact, we find the effect of demand censoring so strong that subjects’ ordering violates the lower bound of the pull-to-center regions – a robust result within the literature – for both LC (two tailed t-test, $p < 0.05$) and HMG, although not significantly so in the latter case. We do not find significance for H2(c), but it holds directionally. (Note that unlike the other tests here, the test of H2(c), a difference-in-difference test, is performed using ANOVA with contrasts.)

<table>
<thead>
<tr>
<th></th>
<th>Uncensored</th>
<th>Censored</th>
<th>Difference</th>
<th>$t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>LMG $q$</td>
<td>836.72</td>
<td>690.31</td>
<td>146.41</td>
<td>$t = 6.34$</td>
</tr>
<tr>
<td>H2 (b)</td>
<td>HMG $Q$</td>
<td>1080.76</td>
<td>964.13</td>
<td>116.62</td>
<td>$t = 3.53$</td>
</tr>
<tr>
<td>(c)</td>
<td>Difference in differences</td>
<td>29.79</td>
<td></td>
<td>0.72</td>
<td>p-value=0.2348</td>
</tr>
</tbody>
</table>

Table 4  t-tests of the effect of censoring on level behavior

4.2. Adjustment behavior

As discussed previously, we focus on within-subject variability to assess the magnitude of adjustment behavior. We do so since subjects – not groups – are the unit of analysis.

**Theory and hypotheses: the effect of observation bias on adjustment behavior.** It is reasonable to expect that subjects’ quantity adjustments will be influenced by what they observe from period to period, which, in turn, depends on demand realizations. When demand feedback is uncensored, subjects are able to observe full demand realizations, $D$. However, subjects’ period-to-period observations when demand feedback is censored are restricted to sales realizations, min($D, Q$). The variability of demand in our setting is $\sigma = 400$, while the variability of sales is bounded above by
this figure. Choi et al. (2008) give the following expression for the standard deviation of sales when demands are normally distributed and $Q$ is fixed:

$$\sigma_{\min(D,Q)} = \sigma \sqrt{z\phi(z) + \Phi(z) + z^2\Phi(z) - (\phi(z) + z\Phi(z))^2},$$

where $z = (Q - \mu)/\sigma$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the unit normal PDF and CFD, respectively. Since the standard deviation of sales will also be affected by the degree of order adjustments, we consider the case where $Q$ follows a normal distribution with mean $\mu_Q$ and standard deviation $\sigma_Q$ (where we resort to numerical estimation of the resulting standard deviation of sales). Table 5 gives the corresponding standard deviations of sales when fixing $\mu_Q$ at expected demand and the two pull-to-center boundaries.

<table>
<thead>
<tr>
<th>$\sigma_{\min(D,Q)}$</th>
<th>$\mu_Q = 730$</th>
<th>$\mu_Q = 1000$</th>
<th>$\mu_Q = 1270$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_Q = 0$</td>
<td>142.6</td>
<td>233.5</td>
<td>316.8</td>
</tr>
<tr>
<td>$\sigma_Q = 100$</td>
<td>165.8</td>
<td>241.0</td>
<td>317.2</td>
</tr>
<tr>
<td>$\sigma_Q = 200$</td>
<td>216.7</td>
<td>261.4</td>
<td>318.9</td>
</tr>
<tr>
<td>$\sigma_Q = 300$</td>
<td>276.7</td>
<td>292.1</td>
<td>321.8</td>
</tr>
</tbody>
</table>

Table 5 Standard deviation of sold quantity $\min(D,Q)$ for $\mu_Q$ fixed at $\mu$ and pull-to-center boundaries.

As was the case with level behavior, subjects should normatively ignore sales and demand signals given full knowledge of the system parameters. However, as Massey and Wu (2005) illustrated through their system neglect hypothesis, subjects are likely to place inordinate weight on signals. Therefore, we conjecture that the more variability there is in what subjects observe, the more they will adjust their order quantities. As a result, we expect that demand censoring will lead to a lower degree of adjustment of order quantities. Further, we expect that this effect will be stronger for LC than for HC, since we expect the former to have lower order quantities and therefore less variable sales. Note that this relationship should be interpreted with care due to its natural endogeneity, i.e., we conjecture that the variability of order quantities depends on the variability of sales, but as demonstrated, the variability of sales also depends on the variability of order quantities.

**Hypothesis 3.** Orders will be less variable with censored demand feedback than with uncensored demand feedback, (a) $\sigma_{Q^e} < \sigma_{Q^u}$ and (b) $\sigma_{Q^e} < \sigma_{Q^u}$. The effect of demand censoring on order variability will be more prominent for LMG than for HMG, (c) $\sigma_{Q^u} - \sigma_{Q^e} > \sigma_{Q^u} - \sigma_{Q^e}$. 

Results. The average within-subject quantity standard deviations of the four conditions are shown in Table 2 and displayed in Figure 1(b) at the beginning of this section.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Uncensored</th>
<th>Censored</th>
<th>Difference</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) LMG</td>
<td>245.67</td>
<td>176.12</td>
<td>69.54</td>
<td>4.88</td>
<td>0.0000</td>
</tr>
<tr>
<td>(b) HMG</td>
<td>270.54</td>
<td>211.65</td>
<td>58.88</td>
<td>3.92</td>
<td>0.0001</td>
</tr>
<tr>
<td>(c) Difference in differences</td>
<td>10.65</td>
<td>t = 0.51</td>
<td>p-value = 0.3041</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6 t-tests of the effect of censoring on adjustment behavior.

From Table 6, we see that demand censoring leads to significantly less order variability for both LMG and HMG (by 69.54 and 58.88 units, respectively), providing support for H3(a) and (b) ($p < 0.01$ for both cases). To test H3(c) we used ANOVA with contrasts. While the results directionally support the hypothesis that demand censoring has a stronger attenuating effect on order variability in LC than in HC, that support is not statistically significant.

While the data support H3(a) and (b), it is unclear whether demand censoring tends to attenuate order adjustment as a result of sales variability being less than demand variability (i.e., as posited), or due to an alternative explanation. As an alternative explanation, order adjustment could tend to increase in order level. By H2(a) and (b), there was strong support for the conjecture that demand censoring tends to reduce order levels. Therefore, if order adjustment does tend to increase in order level, demand censoring could have an indirect attenuating effect on order adjustment – i.e., demand censoring tends to reduce order levels, and reduced order levels could tend to reduce order adjustment.

To test for such a level effect as an alternative, we focus on the uncensored demand feedback conditions. Subjects in both HU and LU observe demand realizations, so they both are exposed to full demand variability (as opposed to sales variability), and hence sales variability would not explain differences observed in order adjustment between HU and LU. However, HU and LU do differ in average order level. Therefore, we isolate a potential level effect on order adjustment by comparing these two conditions. We find that order adjustment in LU is significantly less than HU (from Table 2, the within subject standard deviation difference is 24.87, which is significant
This indicates support for a level effect on adjustment behavior: a subject’s order adjustment tends to increase in their average order level. This, in turn, provides support for an indirect effect of demand censoring on order adjustment – demand censoring tends to reduce average level, and this reduction in level, in turn, tends to attenuate order adjustment.

The presence of this indirect effect of demand censoring on order variability, however, does not preclude the existence of the posited direct effect as well. To test for the possibility that demand censoring has both a direct and indirect effect on order variability, we estimated the following model through OLS:

$$\sigma_{Q_i} = \beta_0 + \beta_1 Cen_i + \beta_2 HMG_i + \beta_3 Cen_i HMG_i + \beta_4 Q_i + \varepsilon_i,$$

where $Cen_i$ is an indicator taking the value of 1 if subject $i$ participated in LC or HC and $HMG_i$ is a control indicator taking the value of 1 if subject $i$ participated in HU or HC. We found strong significance for $Cen_i$ ($\beta_1 = -51.94, p < 0.01$) and for $Q_i$ ($\beta_4 = 0.120, p < 0.01$), while $HMG_i$ and $Cen_i HMG_i$ were insignificant. This suggests that demand censoring has both a direct attenuation effect on order variability as well as an indirect effect through order level.

It is also interesting to note that an informal investigation of the data from the experiment suggests that while subjects facing uncensored demand feedback tend to adjust their quantity up immediately following the first occurrence of underage, subjects facing censored demand feedback tend to hesitate after the first occurrence of underage and rather adjust their quantity up only after the second subsequent underage. However, since this was not among our original conjectures, we have not pursued formal analysis of this phenomenon.

4.3. Impact of behavior on performance

Theory and hypotheses: performance impacts. To investigate the effect of level and adjustment behavior with and without demand censoring, we consider their impacts on expected mismatch

---

2 We do not include a corresponding analysis comparing HC to LC because the average observed variability (sales variability) is less in LC than HC (a result of average order quantity being less in LC than HC). Therefore, we can not determine if the decreased order adjustment in LC relative to HC (a difference of 35.52, significant at the $p < 0.01$ level) is due to reduced signal variability or reduced average level. Since demand variability is independent of order quantity, the comparison between HU and LU is not confounded in this manner.
cost. In this respect, the expected additional cost due to subject \(i\)’s level behavior is denoted\(^3\) \(\Delta_i\) and is captured by the difference between the expected mismatch cost when ordering the subject’s average quantity and the expected mismatch cost when ordering the normative quantity, \(\Delta_i = G(Q_i) - G(Q^*)\). Further, the expected cost of subject \(i\)’s adjustment behavior is denoted \(\Psi_i\) and is captured by the difference between the subject’s average expected mismatch cost and the expected mismatch cost when ordering the average order quantity, \(\Psi_i = G_i - G(Q_i)\). It follows trivially that \(\Delta_i \geq 0\), with equality iff \(Q_i = Q^*\). By Jensen’s inequality, it follows that \(\Psi_i \geq 0\), with equality iff \(Q_{i,t} = Q_i\), \(\forall t\). To summarize, subject \(i\)’s average expected mismatch cost is disambiguated into the normative cost, the level cost and the adjustment cost, \(G_i = G(Q^*) + \Delta_i + \Psi_i\). We refer to \(\Delta_i + \Psi_i\) as the total behavioral cost.

For HMG, based on H1(b) and H2(b), we expect that the level cost will be higher in censored than uncensored demand feedback conditions. This effect is further magnified in terms of the effect on expected mismatch cost since, by convexity of \(G\), the further \(Q\) is from the normative order quantity \(Q^*\), the steeper \(G(Q)\) becomes. For LMG, based on H1(a) and H2(a), we conjecture that demand censoring will partially counter-balance the pull-to-center effect, but may reduce orders to levels below the normative solution. Consider then the case where demand censoring leads to order quantities below the normative quantity, i.e., \(q^c_i < q^*\). By the fact that for LMG,

\[
\frac{d}{dx} [g(q^* + x) - g(q^* - x)] = (c_u + c_o) [F(q^* - x) + F(q^* + x) - 2F(q^*)] > 0, \ \forall x > 0,
\]

it follows that \(g(q^* - x) < g(q^* + x), \forall x > 0\), i.e., the mismatch cost is “flatter” to the left of \(q^*\) than to its right. Therefore, if demand censoring reduced order quantities beyond the normative solution for LMG (i.e., if \(q^c < q^*\)), then \(\delta^c\) would be less than \(\delta^u\) as long as the demand censoring effect were not so severe as to reduce quantities to a level more distant from \(q^*\) than they were without demand censoring. On the other hand, if demand censoring reduces order level as expected, but not to levels below \(q^*\), then \(\delta^c < \delta^u\) would hold trivially. Hence, it is reasonable to assume that censoring will reduce level costs.

\(^3\) Note that we continue to follow the convention of using upper-case and bold-face notation when not being margin specific, and non-bold upper-case for HMG and lower-case for LMG.
The effect on adjustment cost will predominantly depend on the degree of order adjustment (i.e., within-subject variability of order quantities) and the degree of convexity of the expected mismatch cost in the region of the order quantities (more convexity will lead to a stronger effect on adjustment cost). The degree of convexity of the expected mismatch cost is quantified by its second derivative, which takes its maximum value at a quantity equal to mean demand $\mu$, and is decreasing symmetrically at both sides of $\mu$. In both margin treatments considered here, $G''(Q)$ reduces by about 20% from its maximum at the pull-to-center boundary $Q = \mu$ to the other pull-to-center boundary $Q = Q^*$.

When comparing adjustment cost between the censored and uncensored demand feedback conditions, we conjecture that the effect of the difference in order variability will dominate the relatively small effect of the degree of convexity, as discussed above. For LMG, from H2(a), we expect that the mismatch cost will be less convex in the region of $q^c$ than in the region of $q^u$. Furthermore, from H3(a), we expect that subjects’ order quantities when demand feedback is censored, $q^c$, will be less variable than their quantities when demand feedback is uncensored, $q^u$. Both of these effects suggest that the adjustment cost will be lower under censored demand feedback than uncensored demand feedback. For HMG, on the other hand, from H1(b) and H2(b) we expect that the effect of the degree of convexity of $G$ will contribute to demand censoring having a larger adjustment cost than when demands are uncensored. However, H3(b) suggests that the effect of quantity variability will be in the opposite direction. From the discussion above, we expect that the latter of these effects will dominate the first for HC, and conjecture that the adjustment cost will be lower in HC than in HU.

Since the effect of demand censoring is expected to be in the same direction for both LMG behavioral costs, we clearly expect the overall behavioral cost to be smaller in LC than in LU. When comparing HC and HU, however, the effect of demand censoring on the two behavioral costs are expected to be in different directions and it is unclear which one will dominate. Hence, we cannot make a hypothesis regarding the direction of overall behavioral cost.
Hypothesis 4. Demand censoring will lead to a smaller level cost for LMG and a larger level cost for HMG, (a) \(\delta^u > \delta^c\) and (b) \(\Delta^u < \Delta^c\). Demand censoring will lead to lower adjustment costs, (c) \(\psi^u > \psi^c\) and (d) \(\Psi^u > \Psi^c\). Demand censoring will lead to a smaller total behavioral cost for LMG, (e) \(\delta^u + \psi^u > \delta^c + \psi^c\).

Results. Figure 2 shows the expected normative mismatch cost, average level cost and average adjustment cost for all four conditions.

As indicated in Table 7, we find directional support for H4(a); demand censoring leads to lower level cost for LMG (by 27.21), but not with significance. For HMG, however, demand censoring leads to significantly larger level cost (by 517.95), supporting H4(b) \((p < 0.01)\). Supporting H4(c) and (d), adjustment costs are significantly smaller under demand censoring for both LMG and HMG by 165.26 and 145.94, respectively \((p < 0.01 \text{ for both LMG and HMG})\). Interestingly, in three of the four conditions, namely LU, LC and HU, we find that adjustment cost exceeds the corresponding level cost significantly \((p < 0.01 \text{ for LU}, p < 0.05 \text{ for LC and HU})\). For HC, we find the opposite to be true, level cost exceeds adjustment cost with significance \((p < 0.01)\). This stresses the importance of explicitly considering adjustment behavior in newsvendor settings in addition to the more commonly emphasized level behavior.
Finally, supporting H4(e), demand censoring leads to a lower behavioral cost by 191.50 for LMG ($p < 0.01$). Though we could not formulate a hypothesis regarding HMG total behavioral cost due to competing level and adjustment effects, using a two-sided test we find that demand censoring leads to a higher behavioral cost by 372.01 for HMG ($p < 0.01$).

<table>
<thead>
<tr>
<th></th>
<th>Uncensored</th>
<th>Censored</th>
<th>Difference</th>
<th>$t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) LMG $\delta$</td>
<td>131.34</td>
<td>104.12</td>
<td>27.21</td>
<td>0.95</td>
<td>0.1704</td>
</tr>
<tr>
<td>(b) HMG $\Delta$</td>
<td>295.93</td>
<td>813.88</td>
<td>-517.95</td>
<td>-4.22</td>
<td>0.0001</td>
</tr>
<tr>
<td>(c) LMG $\psi$</td>
<td>336.67</td>
<td>171.42</td>
<td>165.26</td>
<td>4.76</td>
<td>0.0000</td>
</tr>
<tr>
<td>(d) HMG $\Psi$</td>
<td>409.38</td>
<td>263.44</td>
<td>145.94</td>
<td>3.72</td>
<td>0.0001</td>
</tr>
<tr>
<td>(e) LMG $\delta + \psi$</td>
<td>468.01</td>
<td>275.54</td>
<td>191.50</td>
<td>3.64</td>
<td>0.0002</td>
</tr>
<tr>
<td>HMG $\Delta + \Psi$</td>
<td>705.31</td>
<td>1077.32</td>
<td>-372.01</td>
<td>-3.11</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Table 7: t-tests for behavioral costs. (*Two-sided test.)

5. Learning and individual relationships of level and adjustment behavior

Within the previous section we studied level and adjustment behavior, comparing both across demand feedback conditions to test for the posited presence and effects of observation bias. Here we study subject learning, again comparing across demand feedback conditions. We then explore the behavioral relationship between level and adjustment behavior on an individual level.

5.1. Learning

Here we explore learning, which we define as a systematic change in behavior with respect to the normative solution. We consider learning in terms of level behavior (i.e., systematic change in subjects’ order quantity level) and learning in terms of adjustment behavior (i.e., systematic change in subjects’ order quantity adjustments). We consider both learning effects under each of the four conditions. Positive learning in terms of level behavior would mean that the level of a subject’s order quantities trends in the direction of the normative order quantity $Q^*$, while positive learning in terms of adjustment behavior would mean that the degree of a subject’s order variability decreases over time.

**Learning and level behavior** A natural measure for learning with respect to level behavior is the degree of subjects’ tendency to get closer to the normative quantity over time, i.e., a dependent variable expressed by $|Q_{it} - Q^*|$ tend to decrease in time. Since changes in level behavior occur on
an individual level, we treat the data as a panel rather than collapsing subject orders into period averages. Consider the following example: If in HMG, one subject changes quantity from 1300 to 1500, and another changes quantity from 900 to 800, we would say that they have negative learning in terms of avoiding level behavior, as both subjects move away from the normative quantity. Considering only the period averages for the two subjects, however, the average quantity would increase from 1100 to 1150, incorrectly indicating positive learning (i.e., indicating a trend toward the normative quantity when, in fact, both subjects diverged from it).

When investigating learning, it is natural to also reflect particular phenomena of the setting. The first of these is that subjects are different in their behavior, reflected by using a fixed effects panel regression that treats each subject’s orders like a cluster. Second, as indicated in previous work, subjects’ ordering behavior tends to be influenced by their mismatch between demand and supply in the previous period (see for example Schweitzer and Cachon 2000 and Bolton and Katok 2008). We therefore include terms for positive imbalance $\text{Overage}_{i,t} = (Q_{i,t} - D_t)^+$ and negative imbalance $\text{Underage}_{i,t} = (D_t - Q_{i,t})^+$ as control variables to capture the roller-coaster effect due to demand chasing. This leads to the following models for LU and HU:

$$|Q_{i,t+1} - Q^*| = \beta_0 + \beta_1 t + \beta_2 \text{Overage}_{i,t} + \beta_3 \text{Underage}_{i,t} + \nu_i + \epsilon_{i,t}, t = 1, \ldots, 49, \quad (2)$$

where $\nu_i$ is the individual effect and $\epsilon_{i,t}$ is the error term. A similar model is used for LC and HC, except $\text{Underage}_{i,t}$ is replaced with the indicator variable $\text{Stockout}_{i,t} = 1_{D_t > Q_{i,t}}$. Table 8 gives the results for the regression for the four conditions. Based on the time coefficient, while there is no support for significant learning in LMG, there is significant support for learning in HMG. This result relates to Bolton and Katok (2008), who found learning more prevalent in their high margin conditions. The difference in level learning between uncensored and censored demand feedback conditions is slight and insignificant for both LMG and HMG (both the difference of 0.0633 between LU and LC and the difference of -0.3025 between HU and HC are insignificant).
Learning and adjustment behavior Learning with respect to adjustment behavior can be measured by a decreasing tendency to change the order quantities from period to period. Here we use the dependent variable $|Q_{i,t+1} - Q_{i,t}|$ to represent this adjustment, and learning would be represented by a decrease over time. The change in dependent variable is the only change from the model in equation 2 used to study learning in terms of level behavior. Therefore, we estimate the following models to explore adjustment learning in LU and HU:

$$|Q_{i,t+1} - Q_{i,t}| = \beta_0 + \beta_1 t + \beta_2 Overage_{i,t} + \beta_3 Underage_{i,t} + \nu_i + \epsilon_{i,t}, \ t = 1, \ldots, 49.$$ 

Again, in LC and HC we replace $Underage_{i,t}$ with $Stockout_{i,t}$. Fixed effect panel results are given in Table 9. The data support positive learning with respect to adjustment behavior in all four conditions: the sign on the time coefficient is negative with strong significance in all conditions. We observe that the magnitude of this learning is smaller in censored than in uncensored demand feedback conditions (the difference of -0.3676 between LU and LC is insignificant, while the difference of -0.7956 between HU and HC is weakly significant at $p < 0.10$). This directionally corresponds to the latter having more potential for learning with respect to adjustment behavior as posited by H3(a) and H3(b) and supported by the results in Table 6.
5.2. Subjects’ level and adjustment behavior relationships

We here investigate if there is a tendency for subjects’ level and adjustment behaviors to be related. Figure 3 tells an interesting story. At a first look (without taking the bubble sizes into account, i.e., thinking of them as scatter plots) it illustrates the positive relationship between standard deviation of sales $\sigma_{\min(D,q_i)}$ and standard deviation of order quantities $\sigma_{\bar{q}_i}$ discussed in subsection 4.2.

![Figure 3](image)

**Figure 3** Bubble chart of subjects’ standard deviation of sales and of quantity in censored demand feedback conditions. Bubble size representing the difference between subject’s and condition’s average order quantity $Q_i - \overline{Q}$ (with white bubbles representing negative difference and black bubbles representing positive difference).

(a) LC (b) HC.

Then, take the bubble color and size into account, where size represents subjects’ deviation from their condition average quantity $Q_i - \overline{Q}$, with white indicating a negative difference and black indicating a positive difference. Strikingly, the cluster of points in the HC condition (Figure 3(b)) consists of subjects who tend to order well above the condition average $Q^c = 964.13$ (a casual inspection indicates that the subjects in the cluster on average order in the low 200’s units above their condition average, which brings them, on average, quite close to the normative $\bar{Q}^* = 1270$). This suggests that the cluster consists of subjects who “get it” in terms of both level behavior (by ordering closer to the normative quantity) and adjustment behavior (by having low $\sigma_{\bar{Q}_i}$), and hence outperform the others in the condition on both of these dimensions. Taking a closer look at Figure 3(a) suggests that there is actually a similar cluster in LC: the small black bubbles of
relatively low standard deviation of sales (between 100 and 200 on the x-axis). A casual inspection indicates that the subjects in this cluster on average order a bit less than 50 units above their condition average $q^c = 690.31$, which brings them on average very close to the normative $q^* = 730$. (Note that this cluster does not stand out as clearly as in HC since (i) it falls nicely into the straight line and (ii) its bubble sizes are relatively small as these subjects do not order very much above the condition average.) This suggests that in both HC and LC there are subjects who “get it” in terms of avoiding both level behavior and adjustment behavior. We explore that behavior relationship further here.

In general, we expect that subjects who perform well in terms of level behavior will also perform well in terms of adjustment behavior as a main effect. There are, however, two additional effects that might be prevalent. First, when considering the decisions (i.e., $Q_i$ and $\sigma Q_i$), the effect of both level and adjustment costs are rather small around the normative order quantity as the expected mismatch cost function is rather flat in this region. This might lead to a less significant relationship between level and adjustment behaviors for the conditions where subjects tend to be close to the normative quantity. Second, for HC, subjects who order further below the normative quantity are more exposed to demand censoring, which, based on the discussion related to H3, we expect will lead to less variability in the subject’s order quantities. For HC, this second effect will draw in the opposite direction to the first: those subjects who perform worse in terms of level behavior will benefit in that they will be exposed to reduced variability in sales, and therefore are expected to vary their order quantities less. While we conjecture that each subject’s adjustment behavior and level behavior will be positively related for the other three conditions, for the reasons stated above, we do not have a clear prediction of the sign of the relationship in HC.

<table>
<thead>
<tr>
<th></th>
<th>LU</th>
<th>LC</th>
<th>HU</th>
<th>HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{Q_i - Q^*</td>
<td>\sigma Q_i}$</td>
<td>0.4006***</td>
<td>0.1060</td>
<td>0.3220***</td>
</tr>
<tr>
<td>$\rho_{\Delta_i, \Psi_i}$</td>
<td>0.4547***</td>
<td>0.3009**</td>
<td>0.2518**</td>
<td>-0.4222***</td>
</tr>
</tbody>
</table>

Table 10 Correlation, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The data support the above conjectures. As Table 10 indicates, there is a positive correlation between level and adjustment behaviors – both in decisions and costs – for LU, LC and HU
Rudi and Drake: Observation bias effects on level and adjustment behavior

(although not significant for LC in terms of decisions). This suggests that, in these three conditions, the subjects who do well in terms of level behavior also do well in terms of adjustment behavior. For HC, however, the opposite effect occurs, with a negative correlation between level and adjustment behaviors for decisions and costs. This suggests that for the subjects with more level behavior, the corresponding effect of more frequent censoring reduces the degree of order variability more than not “getting it” (in terms of understanding that adjustments are bad) increases it.

6. Discussion and managerial implications

The newsvendor model has been a primary building block in the advancement of Supply Chain Management models for the past 20 years. Two major forms of managerial implications in operations management are through insights and impact.

Managerial insights are generated by increased awareness, or greater cognizance, of the phenomena of interest. This paper provides new results into how decision-makers make newsvendor decisions by (i) separating the behavioral decision patterns into level and adjustment behaviors and costs, and (ii) studying how censored demands, a situation arguably faced in most newsvendor settings, impacts these behavioral patterns and costs – with selected results summarized in Table 11. By mapping the behavioral patterns to their effects on expected mismatch cost, we are able to disambiguate their respective effects on performance. The fact that adjustment cost exceeds the corresponding level cost in three of the four conditions highlights its importance as a factor in understanding the behavioral aspects of the newsvendor model. In light of this, we would like to emphasize that one of the contributions of this paper is the demonstration of how different aspects of the newsvendor model, a rather complex managerial decision setting, result in an intricate combination of behavioral deviations from the normative solution. This suggests that in addition to understanding cognitive causes of behavioral patterns in such settings, it is important to understand the aspects of their co-existence.

We also explore the important effects of demand censoring on level and adjustment behaviors vis-à-vis normative decisions and mismatch costs. Despite subjects having full knowledge of the
Table 11  Summary of selected results using symbols: '*' significance, '<'> significantly smaller/larger, with one, two and three symbols indicating significance at levels $p = 0.10$, $p = 0.05$ and $p = 0.01$, respectively, and a symbol in parentheses indicating directional support (but not statistically significant).

demand distribution, demand censoring suppressed order quantities significantly in both LMG and HMG settings, so much so that ordering in these conditions violated the pull-to-center effect found throughout the behavioral newsvendor literature (and tending to order outside the pull-to-center boundaries with significance in the LMG setting). This demand censoring effect led to a significant increase of level costs in the HMG setting equal to 175.0% of the level cost in HU (34.0% of the normative mismatch cost) and a non-significant reduction in level costs equal to 20.7% of level cost in LU (1.8% of the normative mismatch cost) in the LMG setting. In terms of adjustment behavior, we found that censoring suppressed subjects’ order variability, and conjectured that the reduction in observed variability due to censoring is a significant driver of this reduced adjustment behavioral. Our results support the conjecture that the magnitude of adjustments is driven, at least in part, by the degree of variability that subjects observe. As a result, demand censoring led to a significant reduction in adjustment cost equal to 35.7% of adjustment cost with uncensored demand feedback (9.6% of normative mismatch cost) in the HMG setting, and a significant reduction in adjustment cost equal to 49.1% of adjustment cost with uncensored demand feedback (10.8% of normative mismatch costs) in the LMG setting. The overall effect of demand censoring was a significant increase in behavioral costs in HMG of 52.7% (24.4% of mismatch costs), and a significant decrease of behavioral costs in LMG of 41.1% (12.6% of mismatch costs). These results indicate that more information through full demand feedback can actually lead to worse performance due to behavioral
tendencies.

Managerial impact can either come directly through Decisions Support Systems (DSS) or indirectly by changing how current or future managers and executives actually makes decisions. While this work does not provide directly implementable content for DSS, it does advocate the use of DSS when practical by illustrating that these decisions are, in general, complex and costly, even in simplified experimental settings. In terms of changing how decisions are made, this preferably comes by building cognizance through experiential learning as an outreach of research. Fiske and Taylor (1991) found that salience effects, the posited mechanism driving observation bias, can be overcome through training and coaching. In conjunction with the web based game, this research has also resulted in a two-part teaching case. The first part, Ludo (A) by Rudi and Drake (2009), describes a setting that involves LMG, HMG and censored demand, and introduces the newsvendor concept. After completing a series of periods in the web-based game, students do an assignment to calculate the normative solutions based on formulas given in Ludo (B) by Rudi (2009), which, together with the results from the game, forms the basis for a debrief session. This has been adopted for MBA core operations classes, and Executive MBA programs, as well as for Executive Education Programs. By this outreach, this research has already had widespread impact by training about 3,000 current and future managers and executives in the key concepts of the newsvendor model and its central behavioral issues.

Acknowledgement

This paper significantly benefited from discussions with Neil Bearden, Serguei Netessine, Anne Grete Preus, Taylor Randall and Arnstein Aassve. Eirill Bø facilitated access to subjects for the experiment, the web-based game was programmed by Stig Palmquist, and Agnese Vitali provided assistance for statistical tests. This research has been partially funded by the INSEAD Alumni Fund (IAF).

References


University Press.


Corbett, C.J. and J.C. Fransoo. 2007. Entrepreneurs and newsvendors: Do small businesses follow the newsvendor logic when making inventory decisions? University of California, Los Angeles, working paper.


Tong, J., D. Feiler and R. Larrick. 2009. Out of sight, out of mind: Asymmetric learning when
demand is censored by inventory. Duke University, working paper.


Appendix: Experiment process and interface

Each subject was provided with a log-in code for the web-site hosting the experiment and subjects accessed the site from their own laptop computers. After logging in with the code provided, subjects were provided with instructions describing the newsvendor task they faced. They were reminded that each “euro” of profit they earned would improve their chances of winning an iPod raffled at the end of the session. Each student then completed four practice periods with the option to complete an extra four rounds of practice if desired. After completing the practice rounds, another set of instructions reminded students of their task and of their chances to win the iPod. Subjects then continued to the 50 experimental trials.

During each trial, subjects received price, cost, margin and demand information. The demand distribution was displayed graphically and described as “normal” with the mean and standard deviation provided. As the rounds progressed, subjects in the uncensored condition received information on their order quantity, the demand realization, period overage costs, period underage costs, period profits and cumulative profits for all previously completed periods. Subjects in the uncensored condition were also provided a graph that updated after the completion of each period and
Rudi and Drake: Observation bias effects on level and adjustment behavior

displayed their order quantity and realized demand. Subjects in the censored condition received sales rather than demand information within the table and within the graph provided. For periods with underage or no error, subjects in the censored condition were notified that a stock-out occurred, but received no information regarding underage costs. Subjects were also presented each period with the unit revenue, cost and margin information pertinent to their value condition.

Figure 4 displays a screenshot of the uncensored condition interface, and Figure 5 displays the interface presented in the censored conditions.

![Figure 5 Interface for censored condition.](image-url)