



Observation Bias: The Impact of Demand Censoring on Newsvendor Level and Adjustment Behavior

Citation

Rudi, Nils, and David Drake. "Observation Bias: The Impact of Demand Censoring on Newsvendor Level and Adjustment Behavior." *Management Science* 60, no. 5 (May 2014): 1334–1345.

Published Version

<http://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2013.1825>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:28538424>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

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

Nils Rudi

INSEAD, Boulevard de Constance, 77305 Fontainebleau, France, nils.rudi@insead.edu

David Drake

Harvard Business School, Harvard University, Boston, MA 02163 ddrake@hbs.edu

In an experimental newsvendor setting 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 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. 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, November 2011

1. Introduction and literature review

The newsvendor model is the fundamental model for managing inventory under demand uncertainty and has received significant research attention. 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.

Though the existing literature explores both level and adjustment behavior in newsvendor settings, we are, to our knowledge, the first to estimate the respective performance impacts of these behavioral patterns. We propose a method to disambiguate the overall behavioral cost (i.e., the increase of expected mismatch cost due to behavioral deviations from normative order quantity) by estimating the effect of level and adjustment behavior on expected mismatch cost, which we term *level cost* and *adjustment cost*, respectively. While level and adjustment behavior are inter-related – e.g., random errors in ordering and demand chasing can lead to level effects as shown by Su (2008) and Bostian et al. (2008), respectively – our method serves as a useful approximation of each source of behavioral cost. Understanding the relative magnitude of these costs can help inform managers as to which they should be most wary of.

Schweitzer and Cachon (2000) initiated the exploration of behavioral issues in newsvendor settings, primarily focusing on what we refer to as level behavior. They find that order quantities tend to be greater than the normative solution in low margin settings and to be less than the normative solution in high margin settings, a pattern of behavior ubiquitously reported within the experimental newsvendor literature (Bolton and Katok 2008, Lurie and Swaminathan 2009, Bostian et al. 2008, Benzion et al. 2007, etc.). Schweitzer and Cachon (2000) show that of the ten bias and preference structures considered, only anchoring and insufficient adjustment and ex-post inventory error minimization can explain such behavior. Since then, Bostian et al. (2008) show that demand chasing can lead to the this “pull-to-center” behavior, and Su (2008) proposes that random error in ordering can result in such a pattern, while Kremer et al. (2010) find empirical evidence refuting the latter as a sole explanation.

With respect to the treatment of adjustment behavior within the literature, Schweitzer and Cachon (2000) find weak support for demand chasing. Bolton and Katok (2008) segment 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. Others have also reported demand chasing and trending in level behavior, which both contribute to order adjustment (see for example Bostian et al. 2008, Benzion et al. 2008, and Lurie and Swaminathan 2009). We contribute to the understanding of adjustment behavior by providing a means to approximate its impact on performance. The fact that adjustment costs lead to significantly greater profit erosion than level costs in three out of four conditions underscores the importance of accounting for this form of deviation from the normative prescription.

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 total profits (Benzion et al. 2008, Lurie and Swaminathan 2009), with others reporting profits resulting

from level effects alone (Schweitzer and Cachon 2000) and others not reporting performance impacts at all (Feiler et al. 2011). 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.” While this is true, the efficiency metric still combines the impact of level and adjustment behavior. We extend the literature by offering a means to approximate the impact of these two aspects of ordering behavior on performance erosion. This is useful in developing the understanding that both behaviors drive costs, and in assessing the relative magnitudes of these costs within various settings.

Arguably, in most practical settings, demand feedback tends 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. We find that demand censoring induces a reduction in order quantities relative to uncensored settings, supporting censoring having significant impact on observation bias for level behavior. We also investigate the impact of observation bias on adjustment behavior, finding that the magnitude of adjustments are related to the degree of variability observed by subjects. These results can help managers prioritize order quantity improvements based on product margins and the degree of demand feedback available in the setting that they operate in.

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 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). Feiler et al. (2011) study biased judgment in censored settings. However, their first three studies focus on the estimation of the demand distributions (potentially unknown) mean, while the fourth study consider optimal stopping time. In that sense, Feiler et al (2011) provides a complement to the present paper: they focus on the demand estimation task under censoring, while we focus on the order quantity decision. In the one out of four studies where Feiler et al. (2011) do include the order decision, they do so with a critical fractile of $1/2$. The central issue of the newsvendor model is the balancing of overage and underage costs while facing uncertain demand. A fractile of $1/2$ with a symmetric distribution is a special case of the newsvendor problem where the quantity decision results in trivially ordering expected demand. Our study follows the tradition of newsvendor experiments in looking at non-equal overage and underage costs, where the quantity decision becomes non-trivial.

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 4.4 presents additional findings related to learning in terms of both level and adjustment behavior and the relationship between both behaviors. Finally, Section 5 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

$$\Pi(D, \mathbf{Q}) = (r - w) D - \mathbf{G}(D, \mathbf{Q}), \quad (1)$$

where $\mathbf{G}(D, \mathbf{Q}) = c_u (D - \mathbf{Q})^+ + c_o (\mathbf{Q} - D)^+$ is the mismatch cost when ordering $\mathbf{Q}$, 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 $\Pi(\mathbf{Q}) = \mathbb{E}\Pi(D, \mathbf{Q})$ denote expected profit and $\mathbf{G}(\mathbf{Q}) = \mathbb{E}\mathbf{G}(D, \mathbf{Q})$ denote expected mismatch cost, when ordering $\mathbf{Q}$. Note that expected mismatch cost is the *controllable* part of expected profit, i.e., the only part which depends on the decision variable $\mathbf{Q}$. Hence, the minimizer of $\mathbf{G}(\mathbf{Q})$ will be equal to the maximizer of $\Pi(\mathbf{Q})$. The normative solution minimizing expected mismatch cost when demand follows a continuous distribution is the order quantity $\mathbf{Q}^*$ which satisfies $F(\mathbf{Q}) = c_u / (c_u + c_o)$, 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.¹

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). **These parameters result in a HMG critical fractile of 0.75 and a LMG critical fractile of 0.25, which aligns with several studies (e.g., Schweitzer and Cachon 2000; Bolten and Katok 2008; Bostian et al. 2008; and Lurie and Swaminathan 2009), but notably differs from Feiler et al. (2011) who use a fractile of 0.5. We study HMG and LMG settings – i.e., fractiles above and below 0.5 – for two reasons: 1) the central issue of the newsvendor model is balancing overage and underage costs while facing uncertain demand, which becomes trivial when overage and underage costs are equal (because the normative quantity reduces to expected demand); and 2) we expect level and adjustment behavior and costs, and the impact of censoring on those costs, to vary between HMG and LMG settings.**

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

¹ For robustness, we also ran an experiment with a known and an unknown demand distribution described in Section 4.4.

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.

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 $\mathbf{Q}$, and (ii) it enables us to compare the alternatives with low and high margin goods on equal terms (since the expected mismatch cost functions of the two margin scenarios are mirror images of each other around mean demand). In the general model, and when making statements which are valid independent of margin, we use capital and bold-face notation for quantity decisions $\mathbf{Q}$ and mismatch cost $\mathbf{G}$. 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 σ_{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).

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.

Conditions		LU	LC	HU	HC
		I = 64	I = 68	I = 66	I = 71
Subjects' average order quantities	Normative: Q^*	730	730	1270	1270
	Mean	836.72	690.31	1080.76	964.13
	Median	844.05	687.70	1053.63	1043.54
	Std. deviation	118.04	146.66	140.69	236.70
	95% conf. int.	[807.24, 866.21]	[654.81, 725.81]	[1046.17, 1115.34]	[908.11, 1020.16]
	Min	526.26	387.00	808.84	515.90
	Max	1156.98	1089.48	1466.20	1445.00
Within-subject standard deviation	Mean	245.67	176.13	270.54	211.65
	Median	246.78	157.97	271.74	195.92
	Std. deviation	82.33	81.15	85.10	90.73
	95% conf. int.	[225.10, 266.23]	[156.49, 195.77]	[249.62, 291.46]	[190.17, 233.13]
	Min	40.28	56.67	71.17	4.24
	Max	445.83	479.11	462.39	519.42
Expected mismatch cost	Normative: $G(Q^*)$	1525.33	1525.33	1525.33	1525.33
	Mean	1993.34	1800.87	2230.64	2602.65
	Median	1875.58	1728.86	2260.59	2336.81
	Std. deviation	341.65	262.26	404.70	889.46
	95% conf. int.	[1907.99, 2078.68]	[1737.39, 1864.35]	[2131.15, 2330.12]	[2392.12, 2813.18]
	Min	1542.19	1547.75	1595.08	1525.41
	Max	3489.14	3101.20	3097.31	4665.81

Table 1 Descriptive statistics.

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 σ_{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 1.

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).

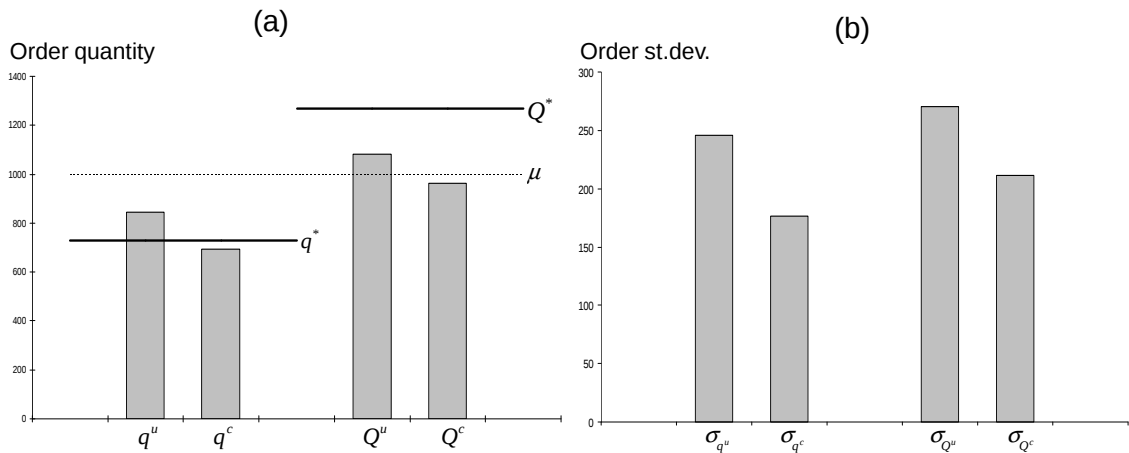


Figure 1 For the four conditions, (a) average order quantities and (b) average standard deviation of order quantities.

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 2, 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.²

		Mean	s.e.	99% C.I.
H1(a)	LU	836.72	14.76	[797.53, 875.91]
H1(b)	HU	1080.76	17.32	[1034.80, 1126.71]

Table 2 99% confidence interval of uncensored quantities testing pull-to-center effect.

Theory and hypotheses: the effect of observation bias on level behavior. Most practical newsvendor 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 censoring will impact observation bias resulting in reduced order quantities relative to when demand is uncensored.

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,

² 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.

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.

In light of the discussion on observation bias above and in the introduction, 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 order 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 1 and shown in Figure 1(a).

From Table 3, 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.)

		Uncensored	Censored	Difference		
H2	(a) LMG q	836.72	690.31	146.41	$t = 6.34$	p-value= 0.0000
	(b) HMG Q	1080.76	964.13	116.62	$t = 3.53$	p-value= 0.0003
	(c) Difference in differences			29.79	$t = 0.72$	p-value= 0.2348

Table 3 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, $\min(D, \mathbf{Q})$. The standard deviation of demand in our setting is $\sigma = 400$, while the standard deviation of sales is increasing in $\mathbf{Q}$ and is bounded above by 400.

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) $\sigma_{q^c} < \sigma_{q^u}$ and (b) $\sigma_{Q^c} < \sigma_{Q^u}$. The effect of demand censoring on order variability will be more prominent for LMG than for HMG, (c) $\sigma_{q^u} - \sigma_{q^c} > \sigma_{Q^u} - \sigma_{Q^c}$.*

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

		Uncensored	Censored	Difference	
H3	(a) LMG σ_q	245.67	176.12	69.54	$t = 4.88$ p-value= 0.0000
	(b) HMG σ_Q	270.54	211.65	58.88	$t = 3.92$ p-value= 0.0001
	(c) Difference in differences			10.65	$t = 0.51$ p-value= 0.3041

Table 4 t-tests of the effect of censoring on adjustment behavior.

From Table 4, 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 1, the within subject standard deviation difference is 24.87, which is significant at $p < 0.05$).³ 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

³ 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.

censoring has both a direct and indirect effect on order variability, we estimated the following model through OLS:

$$\sigma_{\mathbf{Q}_i} = \beta_0 + \beta_1 \text{Cen}_i + \beta_2 \text{HMG}_i + \beta_3 \text{Cen}_i \text{HMG}_i + \beta_4 \mathbf{Q}_i + \varepsilon_i,$$

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 $\mathbf{Q}_i$ ($\beta_4 = 0.120$, $p < 0.01$), while HMG_i and $\text{Cen}_i \text{HMG}_i$ 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.

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 estimated impacts on expected mismatch cost. In this respect, the expected additional cost due to subject i 's level behavior is denoted⁴ Δ_i and is estimated 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, $\Delta_i = \mathbf{G}(\mathbf{Q}_i) - \mathbf{G}(\mathbf{Q}^*)$. Further, the expected cost of subject i 's adjustment behavior is denoted Ψ_i and is estimated by the difference between the subject's average expected mismatch cost and the expected mismatch cost when ordering the average order quantity, $\Psi_i = \mathbf{G}_i - \mathbf{G}(\mathbf{Q}_i)$. It follows trivially that $\Delta_i \geq 0$, with equality iff $\mathbf{Q}_i = \mathbf{Q}^*$. By Jensen's inequality, it follows that $\Psi_i \geq 0$, with equality iff $\mathbf{Q}_{i,t} = \mathbf{Q}_i, \forall t$. To summarize, subject i 's average expected mismatch cost is disambiguated into the normative cost, and the estimates of the level cost and the adjustment cost, $\mathbf{G}_i = \mathbf{G}(\mathbf{Q}^*) + \Delta_i + \Psi_i$. 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 $\mathbf{G}$, the further $\mathbf{Q}$ is from the normative order

⁴ 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.

quantity $\mathbf{Q}^*$, the steeper $\mathbf{G}(\mathbf{Q})$ 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., $q_i^c < q^*$. 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 δ^c would be less than δ^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.

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 μ , and is decreasing symmetrically at both sides of μ . In both margin treatments considered here, $G''(Q)$ reduces by about 20% from its maximum at the pull-to-center boundary $\mathbf{Q} = \mu$ to the other pull-to-center boundary $\mathbf{Q} = \mathbf{Q}^*$.

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.

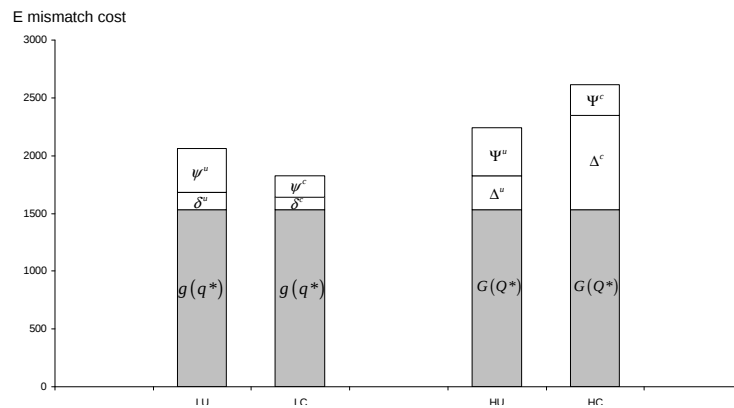


Figure 2 Separation of behavioral costs for the four conditions.

As indicated in Table 5, 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 ($p < 0.01$ 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 (by 205.30, 67.30 and 113.46, respectively, with two-tailed t-test, $p < 0.01$ for LU, $p < 0.05$ for LC and HU). For HC, we find the opposite to be true, level cost exceeds adjustment cost with significance (by 550.44, with two-tailed t-test, $p < 0.01$). 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 ($p < 0.01$).

		Uncensored	Censored	Difference		
H4	(a) LMG δ	131.34	104.12	27.21	$t = 0.95$	p-value= 0.1704
	(b) HMG Δ	295.93	813.88	-517.95	$t = -4.22$	p-value= 0.0001
	(c) LMG ψ	336.67	171.42	165.26	$t = 4.76$	p-value= 0.0000
	(d) HMG Ψ	409.38	263.44	145.94	$t = 3.72$	p-value= 0.0001
	(e) LMG $\delta + \psi$	468.01	275.54	191.50	$t = 3.64$	p-value= 0.0002
	HMG $\Delta + \Psi$	705.31	1077.32	-372.01	$t = -3.11$	p-value= 0.0023*

Table 5 t-tests for behavioral costs. (*Two-sided test.)

4.4. Additional results and robustness

Several additional results and robustness tests were performed as a part of this research, which we describe here.

In terms of adjustment behavior, in line with existing literature, we find that subjects tend to chase demand. We also find that in the censored demand environments (where subjects experience different observed variability in terms of sales), higher observed variability tends to lead to higher adjustment behavior. Furthermore, there tends to be a positive association between level behavior

and adjustment behavior, indicating that “those who get it, really get it”. We also study learning in terms of level behavior (tending to get closer to normative quantity) and in terms of adjustment behavior (tending to reduce order variability). While this only holds directionally for low margin products in terms of level behavior, for all others it holds with significance.

On a referee’s request, we also studied data from a second experiment which included an unknown and a known demand distribution as an additional treatment, resulting in a 2x2x2 study design. As discussed in the introduction, normative theory gives a clear prescription that one should over-order relative to the myopic solution when facing an unknown demand distribution and censored demand. The results presented throughout this section were highly robust when testing the same set of hypotheses in the unknown demand settings. Furthermore, when comparing subjects’ level and adjustment behavior, as well as the components of behavioral costs, no results were statistically different between the known and unknown demand distribution in the new experiment. This suggests that our findings are robust to unknown demand distribution settings, and also that subjects tend to behave contrary to what normative theory suggests when faced with an unknown demand distribution and censored demands.

5. Discussion and managerial implications

Discussion. This paper provides insight into newsvendor decision-making by (i) proposing a method to approximate mismatch costs resulting from level and adjustment behavior, and (ii) studying how censored demand, a situation arguably faced in most newsvendor settings, impacts level and adjustment behavior and costs – with selected results summarized in Table 6 (together with results mentioned in Section 4.4).

By providing a method to approximate portions of expected mismatch costs related to behavioral patterns, we are able to disambiguate their respective effects on performance. The fact that estimated adjustment cost significantly exceeds the corresponding level cost in three of the four conditions highlights its importance as a factor in understanding the impact of behavioral patterns on the newsvendor model. In light of this, we would like to emphasize that one of the contributions

Test	LU	LC	HU	HC
Within pull-to-center region	***		***	
Below pull-to-center region		**		(*)
Comparing average quantities	>>>		>>>	
Comparing quantity st.dev.	>>>		>>>	
Pos. relationship sales st.dev. and quantity st.dev.	NA	***	NA	***
Pos. relationship neg. mismatch and update	***	***	***	***
Pos. relationship pos. mismatch and update	***	***	***	***
Comparing level cost	(>)		<<<	
Comparing adjustment cost	>>>		>>>	
Comparing behavioral cost	>>>		<<<	
Learning to avoid level bias	(*)	(*)	***	***
Learning to avoid adjustment bias	***	***	***	***

Table 6 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).

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 effects of demand censoring on level and adjustment behaviors vis-a-vis the corresponding measures for uncensored demand. Despite subjects having full knowledge of the demand distribution in all settings, censored demand leads to significantly lower order quantities than uncensored demand in both LMG and HMG settings, with the results being robust to settings where the demand distribution is unknown. This demand censoring effect led to a significant increase relative to the uncensored setting of level costs in the HMG setting equal to 175.0% (34.0% of the normative mismatch cost) and a non-significant reduction relative to the uncensored setting in level costs in the LMG setting equal to 20.7% (1.8% of the normative mismatch cost). In terms of adjustment behavior, we found that censoring suppressed subjects' order variability relative to uncensored settings, 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 significantly lower adjustment cost (by 35.7%) than

with uncensored demand feedback (9.6% of normative mismatch cost) in the HMG setting, and a significantly lower adjustment cost (by 49.1%) than 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 relative to the uncensored setting in HMG of 52.7% (24.4% of mismatch costs), and a significant decrease of behavioral costs relative to the uncensored setting 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 implications. This research provides managers with insight into how both order adjustment and order level impact behavioral mismatch costs. Further, our results provide guidance as to which of these sources of behavioral cost managers are likely to be most exposed to given the per unit profitability of the product that they manage – LMG or HMG – and given the setting that they operate under – uncensored or censored demand feedback. As adjustment behavior is addressed by improving the consistency with which an order policy is applied, and order level is addressed by improving the accuracy of that policy with respect to the normative quantity, these results can help managers prioritize improvement efforts. Our results suggest that managers in LMG settings would tend to have greater exposure to adjustment costs regardless of the demand censoring setting. Managers in HMG uncensored settings would also tend to have greater exposure to adjustment costs while managers in HMG censored settings would tend to be more exposed to level costs. Improvements to the consistency with which an order policy is applied, and its accuracy relative to the normative quantity could be prioritized accordingly.

In general, changing how order quantity decisions are made preferably comes by building awareness of level and adjustment costs and their sources. Fiske and Taylor (1991) found that salience effects, the posited mechanism driving the difference in observation bias between uncensored and censored settings, can be overcome through training and coaching. Further, Bolton and Katok (2008) found that personal experience led to the greatest improvement in performance within their study. In this vein, the research presented here has also resulted in a pedagogical case, Ludo (A)

and (B). The case describes a setting that involves LMG, HMG and censored demand, and introduces the newsvendor concept. After completing a series of periods through the web-based game, students are led in a discussion on level and adjustment costs using the decisions that they made. At the time of this writing, this case has been adopted for MBA core operations classes, Executive MBA programs, and for Executive Education Programs, training more than 3,000 managers and executives in the key concepts of the newsvendor model and its central behavioral issues.

References

- Benzion, U., Y. Cohen, R. Peled and T. Shavit. 2008. Decision-making and the newsvendor problem: an experimental study. *Journal of the Operational Research Society*. **59**, 1281-1287.
- Bolton, G.E. and E. 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, G.E., A. Ockenfels and U. Thonemann. 2008. Managers and students as newsvendors – how out-of-task experience matters. University of Cologne, working paper.
- Bostian, A.A., C.A. Holt and A.M. Smith. 2008. Newsvendor “pull-to-center” effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*. **10**(4), 590-608.
- Cachon, G. and C. Terwiesch. 2009. *Matching Supply with Demand: An Introduction To Operations Management*. McGraw-Hill/Irwin, New York.
- Ding, X., M.L. Puterman and A. Bisi. 2002. The censored newsvendor and the optimal acquisition of information. *Operations Research*. **50**(3), 517-527.
- Fiske, S.T. and S.E. Taylor. 1991. *Social cognition*. Singapore: McGraw-Hill Book Company.
- Harpaz, G., W.Y. Lee and R.L. Winkler. 1982. Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Science*. **28**(6), 589-603.
- Kremer, M., S. Minner and L.N. Van Wassenhove. 2010. Do random errors explain newsvendor behavior? *Manufacturing & Service Operations Management*, forthcoming.

- Lariviere, M.A. and E.L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*. **45**(3), 346-363.
- Lewis, D.K. 1969. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA.
- Lurie, N.H. and J.M. Swaminathan. 2009. Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human Decision Processes*. **108**(2), 315-329.
- Massey, C. and G. Wu. 2005. Detecting regime shifts: The causes of under- and overreaction. *Management Science*. **51**(6), 932-947.
- Nahmias, S. 1994. Demand estimation in lost sales inventory systems. *Naval Research Logistics Quarterly*. **41**, 739-757.
- Schweitzer, M.E. and G.P. Cachon. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*. **46**(3), 404-420.
- Su, X. 2008. Bounded Rationality in Newsvendor Models. *Manufacturing & Service Operations Management*. **10**(4), 566-589.
- Taylor, S.E. and S.T. Fiske,.1978. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.) *Advances in Experimental Social Psychology*. Vol 11. New York: Academic Press, 1978.
- Taylor, S.E., J. Crocker, S.T. Fiske, M. Sprinzen and J. Winkler. 1979. The generalizability of salience effects. *Journal of Personality and Social Psychology*. **37**(3), 357-368.
- D. Feiler, J. Tong, and R. Larrick. 2011. Biased judgment in censored environments. Duke University, working paper.
- Tversky, A. and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*. **5**, 207-232.
- Tversky, A. and D. Kahneman. 1974. Judgment under uncertainty: heuristics and biases. *Science*. **185**(4157), 1124-1131.