MANAGEMENT SCIENCE
Vol. 59, No. 3, March 2013, pp. 573–591
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1120.1612

Biased Judgment in Censored Environments

Daniel C. Feiler
Tuck School of Business, Dartmouth College, Hanover, New Hampshire 03755, daniel.c.feiler@dartmouth.edu

Jordan D. Tong
Wisconsin School of Business, University of Wisconsin–Madison, Madison, Wisconsin 53706, jtong@bus.wisc.edu

Richard P. Larrick
Fuqua School of Business, Duke University, Durham, North Carolina 27708, larrick@duke.edu

Some environments constrain the information that managers and decision makers can observe. We examine judgment in *censored environments* where a constraint, the *censorship point*, systematically distorts the observed sample. Random instances beyond the censorship point are observed at the censorship point, whereas uncensored instances are observed at their true value. Many important managerial decisions occur in censored environments, such as inventory, risk taking, and employee evaluation decisions. In this research, we demonstrate a *censorship bias*—individuals tend to rely too heavily on the observed censored sample, biasing their belief about the underlying population. We further show that the censorship bias is exacerbated for higher degrees of censorship, higher variance in the population, and higher variability in the censorship points. In four studies, we find evidence of the censorship bias across the domains of demand estimation and sequential risk taking. The bias causes individuals to make costly decisions and behave in an overly risk-averse manner.

Key words: inference; heuristics and biases; demand estimation; inventory decisions; risky choice; learning
History: Received January 7, 2011; accepted May 9, 2012, by Teck Ho, judgment and decision making. Published online in *Articles in Advance* December 19, 2012.

> Human rational behavior . . . is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor.
> (Simon 1990, p. 7)

1. Introduction

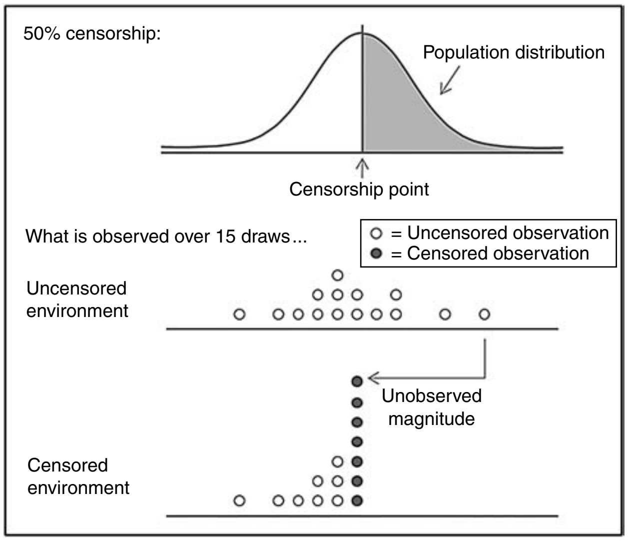
Nearly all organizations operate with incomplete and imperfect information. An informed decision often requires one to extrapolate from an imperfect sample of observations to infer the true underlying properties operating in the environment. In this paper we examine environments in which a constraint distorts the observed sample of data: *censored environments*. In a censored environment, random draws from an underlying population sometimes lie beyond a specific *censorship point* and are observed at the value of the censorship point. These instances are known to be censored, but their exact magnitude is uncertain (see Figure 1). In contrast, random draws that lie below the censorship point are uncensored and observed at their true value. The critical question in such environments is whether individuals can infer the true nature of the underlying population from their censored sample of observations.

Censored environments are interesting to study because they occur in many managerial settings. Consider the following three illustrative organizational examples. First, many firms cannot observe sales missed after an inventory stockout. If a procurement officer holds 100 units of inventory in a given period, she can observe exact values of demand less than 100, but she cannot observe the exact value of demand if demand is greater than 100. When observing sales only, which is a biased sample of demand observations, the procurement officer must infer the nature of the true underlying demand distribution.

Second, a primary task for most managers is to make inferences about the production capabilities of their employees. However, although managers can almost always observe when an employee falls short in a task, they often cannot observe how much more employees are capable of doing in periods in which they complete the work assigned to them. This asymmetry provides a biased sample of observations to the manager and makes inference about employees' capabilities difficult.

Finally, consider when a firm engages in precautionary replacement of equipment or technology after a certain amount of use (e.g., see Campbell et al. 2009). In the event that the equipment breaks before replacement, the firm observes its exact lifetime. However, in the event of replacement, the firm cannot observe the additional time the equipment could have functioned, resulting in a biased sample

Figure 1 A Depiction of Observations in an Uncensored vs. a Right-Censored Environment



of lifetime observations. In each of these examples, accurate judgment from a censored sample is crucial for effective decision making. A biased estimation of demand leads to a biased inventory policy, an inaccurate estimate of employee productivity leads to inefficient work assignments and dismissal decisions, and a biased estimation of the lifetime of a technology leads to an inefficient replacement policy.[1]

In the present research, we show that individuals in censored environments exhibit a *censorship bias*—they form beliefs about the underlying population that are biased in the direction of the observed censored sample. This occurs even with full awareness of the presence of censorship. Furthermore, we examine structural dimensions of censored environments that determine the degree to which individuals form biased beliefs. In general, this research sheds light on which censored environments are likely to yield the most biased inferences and judgments and therefore are the most important candidates for intervention and improvement.

We present results from four experiments. Study 1 examined the task of learning an unknown demand from sales, provided evidence of the censorship bias, and showed that the bias is exacerbated for higher rates of censorship and when the censorship point is variable. Study 2 again provided evidence of the censorship bias in a demand estimation task and showed that higher variance in the underlying population exacerbates the bias. Study 3 linked biased inference from censored samples to behavior by studying simultaneous judgment and decision making in a task that involved dynamic demand learning and inventory decisions. Finally, Study 4 examined a sequential risk-taking paradigm that extended our findings to a new domain and implemented a uniform rather than normal population.

[1] For a review of other examples of censorship, see Amemiya (1984).

2. Literature Review

Previous research suggests that the accuracy of an individual's inference depends on the nature and completeness of the sample of observations they experience (Gilovich 1991, Stewart et al. 2006). Hogarth (2001) suggests that inference is easiest in environments that provide immediate, clear, and unbiased observations (Fiedler 2000, Fiedler and Juslin 2006, Hogarth et al. 1991, Maddox et al. 2003). Many researchers have investigated the effects of nonrepresentative samples on individual inference. Past research has primarily examined three possible causes of nonrepresentative samples. First, an individual may cause an unrepresentative sample by using a biased collection process. For example, in hypothesis testing, individuals may only search for evidence consistent with their theories and neglect to look for disconfirming evidence (Klayman and Ha 1987, Mynatt et al. 1977, Nickerson 1998). Individuals may also face biased samples of recollections from their memories. If recency increases accessibility of memories, then more recent occurrences may be overrepresented in hindsight (MacLeod and Campbell 1992, Tversky and Kahneman 1973).

Second, an individual may face an unrepresentative sample when another individual with a specific agenda strategically presents a biased sample of observations (Brenner et al. 1996, Koehler and Thompson 2006, Larrick and Wu 2007). For example, Silverman et al. (2010) found that doctors fail to account for conflicts of interest when drawing inferences from clinical trials. Similarly, Koehler and Mercer (2009) showed that mutual fund companies only advertise their best-performing funds, but investors respond to the advertised data as if they are representative of company's overall performance. In these cases, individuals generally are not fully aware of the process by which the biased sample has been generated, but they are aware that the samples are presented with the intention of persuasion.

Finally, sometimes simple environmental constraints systematically create a biased sample. For example, if an environment permits the observation of only a chosen alternative's outcome, then an overly negative perception of a foregone alternative cannot be disconfirmed because its true value goes unobserved (Denrell 2005, 2007; Einhorn and Hogarth 1978; March 1996). Einhorn and Hogarth (1978) propose that human resource departments face this problem because they can more easily observe the performance of applicants they hire than the performance of those they reject. Empirical research has found similar results in other organizational contexts. For example, managers with overly pessimistic beliefs about employee trustworthiness and motivation institute strict controls that make it difficult for their beliefs to be disconfirmed, whereas overly optimistic managers institute lax controls and have their beliefs corrected through experience (Markle 2009, 2011).

A developing body of evidence suggests that many shortcomings in human judgment and decision making occur as a result of incorrectly treating a biased sample of observations as representative of the true population (Fiedler 2000, Fiedler and Juslin 2006, Hansson et al. 2008, Juslin et al. 2007). This perspective uses the metaphor of the decision maker as a *naïve intuitive statistician*. The metaphor proposes that individuals are optimal cognitive processors of observations, but they naïvely assume that their observed samples are representative of the population. Subsequently, when observed samples are biased, individuals may form biased inferences about the environment. This perspective has been used to provide alternative causal accounts of judgmental biases that had previously been attributed to ineffective cognitive processing (Fiedler 2000, Juslin et al. 2000). In this paper we build on a central tenet of the naïve intuitive statistician metaphor: the assumption that individuals naïvely treat observed samples as representative of underlying populations. We refer to this as Sample Naïveté Theory (SNT). Drawing on this theory, we examine the judgment and decision making of individuals faced with misrepresentative samples created by censored environments.

3. Judgment in Censored Environments

Consider the problem of estimating properties of an underlying population given a random sample. In a censored environment, at least one constraint limits the range of observable values. For simplicity, we focus primarily on judgment in "right-censored" environments in which a constraint prevents exact observations of values that fall to the right of a fixed point. We consider right-censored environments because they are common in the real world (e.g., capacity constraints), but note that the analysis is conceptually equivalent for left-censored environments.[2] Throughout the paper we will use the term censored environments to refer to right-censored environments. We also focus primarily on a normally distributed population (although we examine inferences about a uniform population in Study 4).

[2] Alternatively, an environment could be left censored when a lower bound or requirement prevents the observation of lower values. For example, Chen (2012) investigates a model in which customer type cannot be observed below a threshold. However, because the same principles should apply symmetrically, we only consider right-censored environments.

3.1. Judging the Mean of a Normal Population from a Censored Sample

Let $\{d_1, d_2, \ldots, d_n\}$ be a random sample of size n from a normal population with known standard deviation σ and unknown mean μ. In a right-censored environment, the *censorship point c* prevents the observation of values greater than c. Thus, define the censored sample of $\{d_1, d_2, \ldots, d_n\}$ as $\{(x_1, r_1), (x_2, r_2), \ldots, (x_n, r_n)\}$, where

$$(x_i, r_i) = \begin{cases} (d_i, 0) & \text{if } d_i < c \text{ "uncensored observation,"} \\ (c, 1) & \text{if } d_i \geq c \text{ "censored observation,"} \end{cases} \qquad i = \{1, 2, \ldots, n\}.$$

For right-censored environments, note that an observation x_i is simply the minimum of the sample d_i and the censorship point c. We say that an environment is uncensored if no observations are censored (i.e., $x_i = d_i$ for all i). Figure 1 provides a visual depiction of a right-censored and an uncensored environment. For convenience, we also define the observed sample mean $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ and the observed censorship rate $\bar{r} = (1/n)\sum_{i=1}^{n} r_i$, which is the proportion of censored observations in the sample. In this paper, we focus on the problem of estimating the population mean from a censored sample. That is, given a set of observations $\{(x_1, r_1), (x_2, r_2), \ldots, (x_n, r_n)\}$, what is μ?

It is clear that the observed sample of x's provides a biased estimation of the mean, because high draws of the population are observed at the censorship point. That is, $\bar{x}$ is a downward-biased estimate of μ. Thus, the difficulty in making judgments about the population from a censored sample is that one must correctly extrapolate from both the observed x's *and* the observed r's to make accurate inferences about the underlying population. Next, we develop our hypotheses for how individuals make such judgments by drawing on SNT.

3.2. Behavior in a Censored Sample Judgment Task

Misrepresentative samples are potential traps for decision makers. As proposed by SNT (Fiedler and Juslin 2006), individuals may naïvely rely on observed samples to make inferences as if they were representative of the truth. Even when individuals are cognizant of constraints on observations, accounting for the bias in the observed sample often remains a complex task. Individuals may understand the direction in which to adjust their beliefs, but determining the degree to which one should adjust is difficult. As a point of reference, consider the purely naïve intuitive statistician who ignores censorship altogether and treats the observed sample as fully representative. Let e^n denote such a purely naïve estimate. Then, e^n is simply the observed sample mean:

$$e^n = \bar{x}.$$

If individuals behave consistent with SNT and are naïve to the constraint that limits observations beyond the censorship point, then they will form beliefs about the population mean that are biased toward the mean of the observed sample. Let e^b denote the behavioral estimate of the population mean μ given a censored sample.

Hypothesis 1A. *In a censored environment, estimates of the population mean will be biased low, toward the naïve estimate e^n (i.e., $e^b < \mu$).*

Hypothesis 1B. *Estimates of the population mean will be lower when decision makers face a censored environment than when they face an otherwise equivalent uncensored environment.*

In censored environments, decision makers have enough information in their censored samples to form unbiased estimates of the population mean. Indeed, a prescriptive heuristic based on an approximate maximum-likelihood estimate (MLE) can accomplish this task with great accuracy (Nahmias 1994).[3] We denote the estimate of μ given a censored sample according to Nahmias (1994) as e^h:

$$e^h = \bar{y} + \frac{\sigma\phi(\Phi^{-1}(1-\bar{r}))}{1-\bar{r}},$$

where $\bar{y}$ is the sample mean of the uncensored observations, and ϕ and Φ are the probability density and cumulative distribution functions of the standard normal distribution. In short, this prescriptive heuristic uses sample statistics—it starts with the sample mean of the uncensored observations and adjusts upward according to the sample censorship rate and the population variance.[4] (When we examine the uniform distribution in Study 4, we use the MLE as a benchmark, which we describe in Appendix E.)

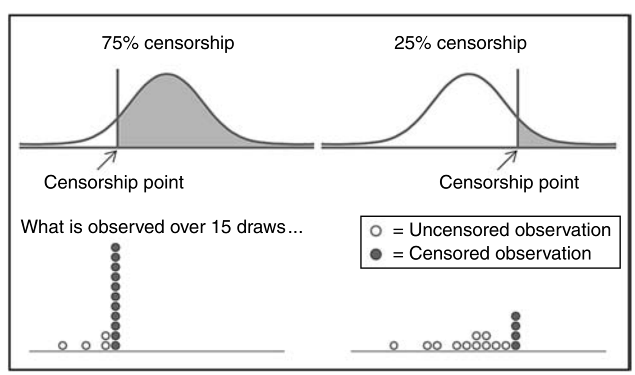
As each individual might observe a different random sample drawn from the population, it is arguably fairer to make predictions about an individual's estimate relative to the heuristic's estimate given the same sample rather than to the true population mean. Because the heuristic yields very near-optimal estimates, we predict the following hypothesis:

Hypothesis 1C. *In a censored environment, estimates of the population mean will be lower than the estimates of the prescriptive heuristic (i.e., $e^b < e^h$).*

Not all censored samples are created, or biased, equally. There are several key dimensions in censored environments that may cause individuals to form beliefs farther from the truth. The effects of these dimensions hinge on one straightforward assertion: as the distance between the observed sample mean and the true mean, $\mu - \bar{x}$, increases, the observed sample is less representative of the population, and individuals will form estimates of μ farther from the truth. We discuss two factors in censored environments that cause the observed sample mean to be farther from the true mean: the degree of censorship and population variance. First, the degree of censorship increases as the censorship point moves lower on the population distribution. This causes the observed rate of censorship to increase in expectation because a greater percentage of random instances will be restricted by the lower constraint (see Figure 2). A higher degree of censorship means that the censorship point screens observations more frequently and censored instances will be observed farther from their true values, on average. Therefore, the mean of the observed sample moves farther from the true population mean as the degree of censorship increases.

Hypothesis 2. *As the degree of censorship increases, estimates of the population mean will be biased farther from the true population mean (or equivalently, $\mu - e^b$ is decreasing in c).*

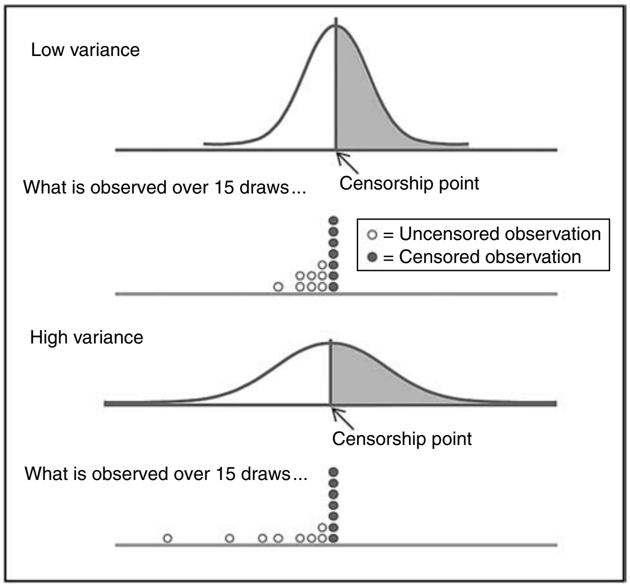
Figure 2 Censored Environments with High (75%) vs. Low (25%) Censorship



[3] In a numerical study, Nahmias (1994) evaluated the estimator's performance for an environment similar to our experimental settings: normal distribution with mean 100, standard deviation 30, censorship point 100, sample size 40. Over 1,000 replications, he found that the estimator's average estimate for the mean was 100.320 with a standard deviation of 5.827. In general, Nahmias (1994, p. 743) found that "the only case where there is a substantial difference in the performance of the two estimators [the approximate MLE estimator versus a complete MLE estimator] is for $n = 10$ and $S < \mu$" (where n is the sample size and S is the censorship point). Moreover, in Nahmias's (1994) numerical study, the standard deviation was unknown, so when the standard deviation is known (as in our studies), the estimator's performance should improve.

[4] This heuristic cannot be applied in the special case that all observations are censored.

Figure 3 A Censored Environment with Low vs. High Variance



Second, higher variance in the population causes the mean of the observed sample to lie farther from the population mean.[5] With higher variance, there exist more extreme low and high instances; however, the more extreme high instances remain censored by the environment and are still observed at the censorship point (see Figure 3). Therefore, holding the censorship point constant, the mean of the observed sample lies farther from μ as the variance of the population increases. As stated previously, we contend that as observed samples become more misrepresentative (i.e., as $\mu - \bar{x}$ increases), individuals' estimates of the population mean will be more biased.

Hypothesis 3. *As the variance of the population increases, estimates of the population mean will be biased farther from the true population mean (i.e., $\mu - e^b$ is increasing in σ).*

Censored environments are particularly interesting to study because individuals are aware of the censorship. Awareness of censorship cues decision makers into the direction of the sample bias because observations are censored in only one direction. Further, individuals can see exactly how many, and at what point, observations are censored. Therefore, we predict that individuals will use evidence of censorship to adjust their estimates of the population mean in the appropriate direction from the mean of the observed sample, $\bar{x}$. However, past research shows that adjustment from an initial anchor tends to be insufficient (Chapman and Johnson 1999; Epley and Gilovich 2001, 2004, 2006; Tversky and Kahneman 1974; Strack and Mussweiler 1997). Although we expect the naïveté prediction of Hypothesis 1 to be the dominant mechanism in censored environments, we predict that individuals will be better than "pure naïveté," at least partially accounting for the sample bias.

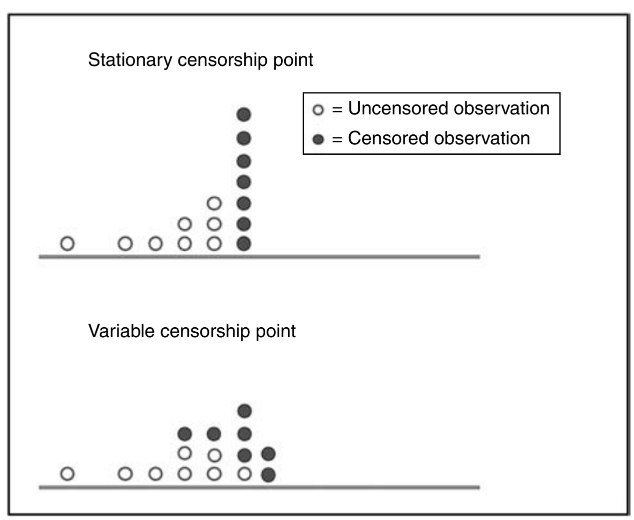
Hypothesis 4. *In a censored environment, estimates of the population mean will be greater than the mean of the observed sample (i.e., $e^b > e^n$).*

A greater degree of censorship increases the distance between the sample mean and the true mean, but it also gives individuals more evidence of the sample bias by presenting more instances of censorship. As more observations accumulate at the censorship point, it may provide a stronger perceived violation of the expected normality and randomness anticipated by the decision maker (Kahneman and Tversky 1972). This perceived violation may cause individuals to adjust for the sample bias to a greater extent. This possible secondary "cueing-to-adjust" mechanism of censored samples leads us to make the following prediction:

Hypothesis 5. *As the degree of censorship increases, individuals will adjust farther from the observed sample mean in their estimates of the population mean (i.e., $e^b - \bar{x}$ is increasing in $\bar{r}$).*

Another factor that makes the presence of censored observations more salient, and may consequently spur greater accounting for the sample bias, is the variability of the censorship point. Here, we define variability in terms of mean-preserving spreads. When the censorship point is stationary, the censored observations amass at one specific point.

Figure 4 A Censored Environment with Stationary vs. Variable Censorship Points



[5] The actual property we are considering here is second-order stochastic dominance (see Muller and Stoyan 2002), which for the normal distribution is equivalent to the variance.

A stationary censorship point, therefore, yields a single large spike of censored observations at that point. A variable censorship point, on the other hand, will produce censored instances at a number of different points (see Figure 4). Even if the variability is small, it will prevent observations from compiling into a single spike. We posit that when censored observations are allowed to accrue at a single point (i.e., with a stationary censorship point), they become more salient to the decision maker. This causes the decision maker to account for the sample bias to a greater extent.

Hypothesis 6. *When facing a stationary, as opposed to variable, censorship point, individuals will adjust farther from the observed sample mean in their estimates of the population mean (i.e., $e^b - \bar{x}$ is larger if c is stationary).*

Next, we describe and present results from four studies used to empirically examine human judgment in censored environments. Studies 1 and 2 examined inference from censored samples and the factors that mitigate or exacerbate the censorship bias. Studies 3 and 4 linked inference to choice and examined how biased beliefs about the population can lead to biased decisions.

4. Study 1: Degree of Censorship and Variability of Censorship Point

Study 1 was designed to test for the existence of a censorship bias (Hypotheses 1A–1C). It was also designed to test whether the censorship bias is exacerbated as the degree of censorship increases (Hypothesis 2) and when censorship points are variable (Hypothesis 4). In this task, individuals faced a normally distributed demand with an unknown mean and a known standard deviation. They observed randomly generated sales of a newspaper company and a binary indication of whether the company sold out in each period. They then made an estimate of the underlying mean demand and, subsequently, observed the sales of the next period. The sales feedback for all past periods always remained visible, and the task included 30 periods. In this task, the demand distribution represents the underlying population about which an individual needs to draw inferences. The inventory of the paper company acts as a censorship point for observing demand, and the actualized sales are the observed sample.

To manipulate the degree of censorship, the inventory levels (i.e., censorship points) were centered at the 25th, 50th, or 75th fractile on the demand distribution. These fractiles were chosen to be consistent with the large body of literature on stocking decisions with known demand that has primarily studied the 25th and 75th critical fractiles (e.g., Schweitzer and Cachon 2000, Bolton and Katok 2008, Bostian et al. 2008, Lurie and Swaminathan 2009, Ho et al. 2010, Chen et al. 2013). The presence of the censorship point at different fractiles changes the degree of censorship and the expected rate of censorship.

The variability of the censorship point was also manipulated. In the stationary condition, the inventory available for sale was constant across all 30 periods. In the variable condition, the inventory was selected randomly each period from a uniform distribution, ± 25 units from the relevant fractile determined by the degree of censorship.

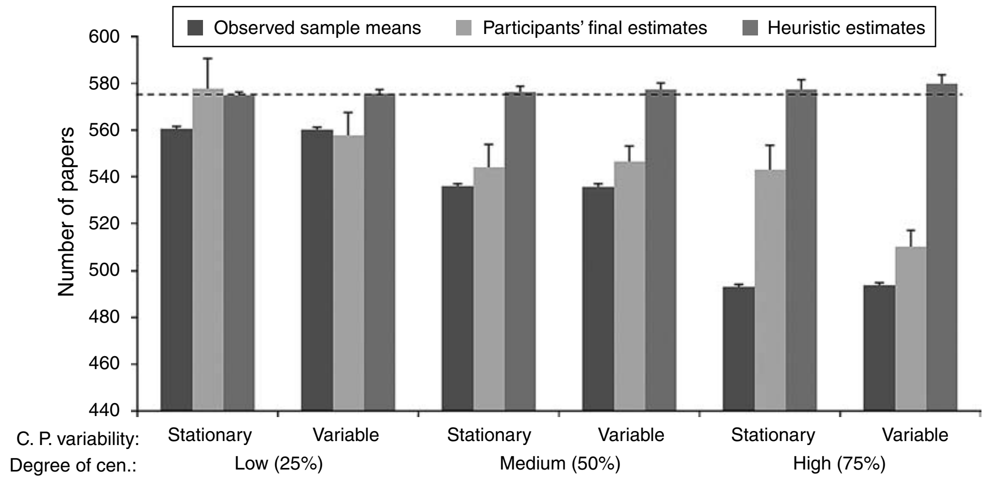
This study was a three (degree of censorship: high, medium, low) by two (censorship point variability: stationary, variable) between-subjects design with within-subject repeated measures for the 30 periods. The underlying unknown demand function for all participants was normal with $\mu = 575$ and $\sigma = 100$. Therefore, in the high-, medium-, and low-censorship conditions, the mean inventory levels were 507, 575, and 643, respectively. To control for some of the noise across conditions, each participant was yoked with a participant from each of the other conditions (there were six total cells in the study design). These yoked participants faced the same sequence of 30 demand instances. In this manner, 18 demand sequences were randomly generated before the study, and participants were randomly assigned to a condition and a demand sequence. Because each condition contained the same 18 demand sequences, the possibility of differences across conditions emerging as a result of noise in the random demand instances is reduced.

4.1. Methods

4.1.1. Participants. One hundred and eight undergraduates at a major American university signed up for the study through an online scheduling system and participated in a computer lab. A stated prerequisite for participation was having completed at least one college-level statistics course. This prerequisite ensured that participants had previously been exposed to normal distributions. They received \$5 for participation and could earn up to an additional \$5 based on the mean absolute deviation (MAD) of their demand estimates from the true mean demand. For every 15 units of MAD, participants lost \$1 from the bonus until it was exhausted. This relationship was linear, and individuals could be awarded fractions of dollars.

4.1.2. Instructions and Procedures. The user interface was programmed in Microsoft Excel. Participants read an information sheet explaining the details of the game. They were informed with text and a figure that the demand distribution was normal with a stationary mean, *m*, and a standard deviation of 100. They were told that the mean *m* was

Figure 5 **Study 1 Mean Final Demand Estimates, Observed Sample Means, and Prescriptive Heuristic Estimates Displayed by Condition, with Standard Error Bars**



Note. The dashed line indicates the true mean demand, 575.

equally likely to be anywhere between 400 and 800. Before the commencement of the study, the instructions were reviewed with each participant to ensure comprehension.

Appendix A contains a snapshot of the user interface. In each period, participants observed sales and a binary indication of whether the paper company sold out. They then made an estimate of the underlying mean demand, m. This process continued for a total of 30 periods.

4.2. Results

4.2.1. Comparisons to Truth. The initial analyses were done in a repeated measures model using residual maximum-likelihood estimation and an autoregressive covariance structure.[6] Degree of censorship and censorship point variability were specified as fixed effects in the model. An interaction between the experimental manipulations and period was also included in the model. Period was not significant as a main effect or as a moderator ($p > 0.25$ for each). The average estimates for low, medium, and high censorship were significantly lower than the true population mean ($t(102) > 2.3$, $p < 0.05$ for each). This provides substantial evidence of a censorship bias as predicted in Hypothesis 1. Also, consistent with Hypothesis 2, there was a significant main effect of degree of censorship ($F(2, 102) = 3.83$, $p < 0.03$), such that mean perceived demand declined as censorship increased ($M_{\text{low}} = 567.7$, $M_{\text{medium}} = 547.6$, and $M_{\text{high}} = 530.0$; $\mu = 575$).

4.2.2. Comparisons to Heuristic Estimate. An alternative analytical approach is to compare their estimates to what a prescriptive approximate maximum-likelihood heuristic would have estimated given the same observed sample (Nahmias 1994). To use this approach, we compared their final estimates (i.e., their estimates with the full sample of 30 observations) to the heuristic estimates given the same sample (see Figure 5). Heuristic estimates were not significantly different from the true mean across degrees of censorship or censorship point variability conditions ($t(102) < 1.7$, $p > 0.1$ for each). However, participants' final estimates were significantly lower than heuristic estimates given the same sample with a high (75%) or medium (50%) censorship ($t(102) = 7.07$ and 4.29, respectively, $p\text{s} < 0.001$). Final estimates were not significantly lower than heuristic estimates with a low (25%) censorship ($t(102) = 1.03$, $p > 0.3$).

4.2.3. Adjustments from Sample Mean. Although we predicted that individuals would be naïvely biased toward the observed sample mean, we also predicted that they would use cues of censorship to adjust their beliefs. To test this, we compared their final estimate to their observed sample mean across conditions. There was no interaction between degree of censorship and censorship point variability. In the high-, medium-, and low-censorship conditions, the final estimates were higher than their observed sample means by 33.2, 9.5, and 7.2. They adjusted significantly more with high (75%) censorship than with medium (50%) or low (25%) censorship ($t\text{s}(102) > 2.39$, $p\text{s} < 0.02$); however, the latter two were not significantly different ($t(102) = 0.22$, $p > 0.8$). Finally, when individuals faced a stationary, rather than variable, censorship point, they adjusted 16.7 units farther from the observed sample mean, a significant difference ($t(102) = 2.07$, $p < 0.05$). These findings are consistent with our predictions that individuals are partially sensitive to censorship and do

[6] All analyses and simple effects tests were performed with SAS proc mixed.

use information in the sample to account for the sample bias (Hypothesis 4). Specifically, more censored observations and a stationary censorship point cued greater adjustment from the observed sample mean (Hypotheses 5 and 6).

Study 1 provides support for several predictions. First, estimates of the population mean (mean demand) tended to be downward biased relative to the true mean and the prescriptive heuristic estimate given the same sample. Second, the censorship bias worsened as the degree of censorship increased. Third, individuals were sensitive to both the degree of censorship and censorship point variability cues causing them to account for the sample bias to a greater extent and adjust farther from the observed sample mean.

5. Study 2: Population Variance

Study 2 was designed to provide a second test of censorship bias and to test whether higher population variance exacerbates the censorship bias (Hypothesis 3). Several methodological changes were made to test the robustness of the effect in a new experimental paradigm. In this study, participants observed the 30 periods of randomly generated sales information simultaneously in a graph (for an example, see Appendix B) and were asked to estimate the mean of the underlying demand distribution. This new paradigm arguably achieves greater external validity because many businesses assess demand by observing summaries of sales in reports and in review meetings. Individuals were informed that the demand distribution was approximately normal with an unknown mean and known variance. The true underlying mean demand was 745 for all participants. Participants were told the variance of the underlying demand distribution, which depended on the condition; σ was either 100 or 200. In the graph of sales, stockouts were indicated by white bars and nonstockouts were indicated by black bars.

Each demand instance was independent and randomly drawn from the underlying demand distribution. Therefore, the set of instances observed by each participant was unique. We chose to use a variable censorship point because this is more representative of what is faced by inventory managers in the real world as a result of changing orders and availability. To do so, the exogenous inventory levels were determined by a uniform distribution ranging from 720 to 770 (i.e., ± 25 units from the true mean demand). All participants faced the same 30 levels of inventory, but their sales results were different because each period's draw of demand was random and independent.

5.1. Method

5.1.1. Participants. Participants were 65 executive M.B.A. students at an American business school. They were recruited to participate via a post on an electronic bulletin board. All participants had previously taken both a statistics and an operations course; in the latter, they were taught the newsvendor solution with a known demand distribution. By participating, the students raised money to be donated to a nonprofit organization providing medical aid in Haiti, following the 2010 earthquake. For each person that participated, \$10 was donated to the charity. If an estimate of the underlying mean demand was within 20 units of the true mean, the charity earned a bonus donation of \$3.

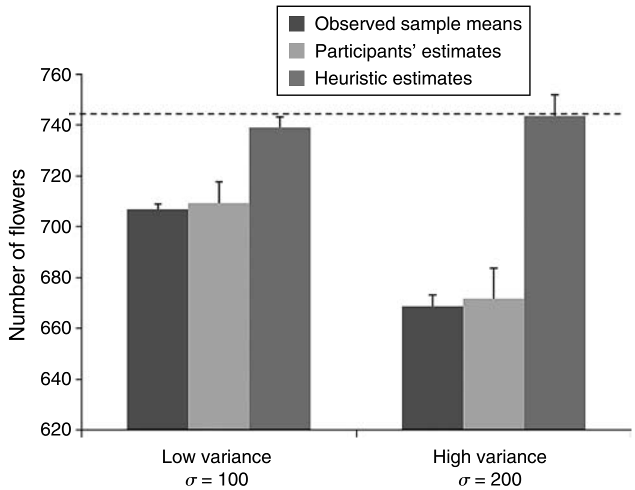
5.1.2. Instructions and Procedure. The study was run through an online survey system, Qualtrics. Participants were told to imagine that they would be helping Franny's Flower Shop determine their mean demand. First, they read instructions for the estimation task, saw a depiction of the shape of the demand distribution, and observed a bar graph of sales and stockouts from the previous 30 periods. The depiction of demand showed a normal distribution with the appropriate standard deviation, centered at an unknown mean m. Participants then estimated the mean of the underlying demand distribution. Last, participants answered the following comprehension check question: "In the graph of sales, a white bar indicated that Franny sold out that day. If a white bar showed sales of 700, then you know demand that day was (a) less than or equal to 700, (b) exactly 700, or (c) greater than or equal to 700."

5.2. Results and Discussion

There were two very extreme estimates of mean demand, 350 and 1,100, both in the high-variance condition. The externally studentized residuals for these two observations were 3.8 and 5.0, respectively. Additionally, the Cov ratios of these two observations were 0.64 and 0.39, suggesting that including them in the analyses significantly decreased the stability of the model. Therefore, as recommended by Cohen et al. (2003), we did not include these two extreme observations in the following analyses (Belsley et al. 1980), although the same trends hold with them included. Nine participants did not answer the comprehension check correctly. Although the same findings hold with them included, they are excluded from the results presented here.

5.2.1. Comparisons to Truth. Using analysis of variance (ANOVA), we tested whether variance of demand moderated the downward bias of demand estimates. As predicted in Hypothesis 3, the mean demand estimates in the high-variance condition were significantly lower than those in the low-variance condition ($F(1, 54) = 6.51$, $p < 0.02$; low variance: $M = 709.3$, $SD = 43.4$, $N = 27$ versus high

Figure 6 Study 2 Mean Demand Estimates, Observed Sample Means, and Prescriptive Heuristic Estimates Displayed by Variance Condition, with Standard Errors Bars



Note. The dashed line indicates the true underlying mean demand, 745.

variance: $M = 671.6$, $SD = 64.4$, $N = 29$; see Figure 6). The estimates were biased significantly below the true underlying mean demand in both variance conditions ($ts(54) > 3$, $ps < 0.01$).

5.2.2. Comparisons to Heuristic Estimate. We also compared the estimates of individuals to the estimate of a prescriptive approximate MLE heuristic (Nahmias 1994) given the same sample, a test of Hypothesis 3. Participant estimates in the low- and high-variance conditions were 29.8 and 71.9 units lower than the heuristic estimates given the same observed sample ($ts(54) > 2.5$, $ps < 0.05$), a significantly greater difference with high variance ($F(1, 54) = 7.06$, $p < 0.05$). The heuristic estimates were not significantly different than the underlying population mean of 745 in either variance condition (low variance: $e_h = 739.1$; high variance: $e_h = 743.5$).

5.2.3. Adjustment from Sample Mean. In neither variance condition were individuals' estimates significantly higher than their observed sample means ($ts(54) < 0.3$, $ps > 0.75$). The mean estimates in the low- and high-variance conditions were 2.6 and 2.9 units above their observed sample means, respectively. The findings of this study were more consistent with pure naïveté and did not support Hypothesis 4.

In sum, these findings provide additional evidence that inferences in censored environments are biased toward the observed sample mean. Furthermore, with higher variance, the observed sample was more misrepresentative, which, as expected, exacerbated the censorship bias. Unlike Study 1, however, individuals did not appear to significantly adjust from the observed sample mean in their estimates of the population mean. There are several possible reasons for this difference. In this study, participants made only one estimate, whereas participants in Study 1 made 30 estimates, allowing them to observe and experience the censored environment to a greater extent. Second, this study presented the sample in a graph, without participants being able to see the exact values of each outcome. Therefore, individuals may have had difficulty inferring exactly what the average of the observed sample was, in addition to having difficulty accounting for censorship.

6. Study 3: Censored Demand Feedback with Known vs. Unknown Demand

In Studies 3 and 4, participants acted as both judges and decision makers. The purpose of these studies was twofold. First, they linked biased inference from censored samples to actual decision making. Second, they explicitly compared judgment in censored versus uncensored environments. The focus of Study 3 was to examine biased inventory orders resulting from biased beliefs about demand.

Much behavioral research in the inventory context has focused on the newsvendor ordering task with a *known* demand distribution (e.g., Schweitzer and Cachon 2000). In these studies, participants were informed of a demand distribution and the cost parameters for their good. They were then asked to make a sequence of stocking decisions with incentives to maximize profit. In contrast, this study examines how individuals perform when demand is *unknown*. In Study 3, individuals needed to update their beliefs about demand to inform their inventory decisions in an effort to maximize their expected profit. Based on Studies 1 and 2, we expected that stocking decisions would be significantly lower when facing censored, as opposed to uncensored, demand observations. However, we expected this stocking bias to be driven by downward-biased beliefs about demand and therefore predicted to find this censorship bias only when participants needed to learn an unknown demand, but not when demand was known.

Study 3 was a two (environment: censored, uncensored) by two (demand knowledge: known, unknown) between-subjects design with within-subject repeated measures for the 30 periods. Censorship was manipulated by allowing some participants to observe only sales each period (censored environment) and others to additionally observe actual demand each period (uncensored environment). Also, we manipulated whether individuals knew the true underlying mean demand. Some participants were told the true underlying mean of demand (known demand), while others were told that their true underlying mean demand was equally likely to be anywhere between 400 and 800 (unknown demand).

In the task, participants purchased newspapers for \$1, sold them for \$2, and discarded excess inventory at the end of each period at no cost. Participants were told the overage and underage costs were both equal to \$1. Given the symmetry of the demand distribution, the optimal policy[7] with known demand was simply to stock the mean of demand.

As in Study 1, participants were yoked across conditions with common demand sequences. Each cell of the design had the same 19 demand sequences, reducing the probability that differences across conditions could emerge due to randomness. Each of the 19 demand sequences were generated from a different underlying demand, with means randomly selected between 500 and 700. Although participants were given a prior of $U(400, 800)$ for the selection of the underlying mean, we actually used $U(500, 700)$ to avoid ceiling or floor effects that might limit the potential bias in their beliefs.

6.1. Methods

6.1.1. Participants. Participants were 76 M.B.A. or business Ph.D. students at an American business school and were randomly assigned to one of the four experimental conditions. All participants had previously taken a basic statistics course and also either an operations management course or an advanced statistics course, making the subject pool quite statistically sophisticated. For each person that participated, \$8 was donated to a club or charity of their choice. Furthermore, additional money could be earned based on profit earned in the game that could be either kept by the participant or also donated. For every \$2,000 earned in the game, they earned \$1 in bonus money and fractions of dollars could be earned. In playing the game, participants generally earned \$4–\$9 in bonus money.

6.1.2. Instructions. Participants were given an instruction sheet explaining the details of the game. They were informed that demand was normally distributed with a mean of m and standard deviation of 100, and they were shown a picture of this distribution. They were told that before beginning the game, some participants would get to learn their exact m, and others would learn a range of where m might be. They were informed that in either case the mean of demand was stationary and did not change over the course of the study. The cost parameters were then explained, and it was explicitly stated that it was just as costly to order one unit too many as it was to order one unit too few.

[7] Optimal here assumes expected profit maximization with full information about the demand distribution. The optimal ordering policy without full information is, in fact, even higher due to the value of "stalking information" (e.g., see Lariviere and Porteus 1999).

6.1.3. Procedures. The study was run in a computer lab and was created in a Microsoft Excel interface. At the top of the user interface, participants were informed either of the m of their demand distribution (known demand condition) or that the m of their demand distribution was randomly chosen with equal likelihood from the range 400 to 800 (unknown demand condition). In each period in the inventory game, participants estimated the underlying mean demand *and* made a stocking decision. Their own stocking decision, therefore, acted as the censorship point.

See Appendix C for a depiction of the study interface. First, in two columns participants estimated or reported their underlying mean demand and then made a stocking decision for the period. Sales feedback was then automatically generated, and participants in the uncensored condition also observed the actual demand instance that period. The cumulative average of sales (and demand in the uncensored condition) was updated at the bottom of the screen as participants updated demand beliefs, made stocking decisions, and observed sales feedback for 30 periods.

6.2. Results

Because each of the 19 demand distributions faced by individuals had a different randomly determined μ, all estimates were analyzed relative to their respective μ That is, μ was set equal to 0, and estimates greater than their respective μ were positive and estimates less than μ were negative. As in Study 1, analyses were first done with a repeated measures model using residual maximum-likelihood estimation and an autoregressive covariance structure. Demand knowledge and censorship environment were specified as fixed effects, and an interaction between the experimental manipulations and period was also included in the model. The repeated variable period was not significant as a main effect or as a moderator for demand beliefs or stocking decisions. See Table 1 for a summary of the results.

Table 1 **Average Estimates of Mean Demand and Stocking Decisions by Condition**

Demand knowledge	Censorship environment	Avg. estimates of mean demand		Avg. stocking decisions	
		M	*SD*	*M*	*SD*
Known demand	Uncensored	0.0	0.3	1.7	34.5
	Censored	−1.3	6.0	4.1	16.1
Unknown demand	Uncensored	−4.4	25.7	−7.0	36.1
	Censored	−27.9*	34.6	−31.1*	40.6

Note. The estimates and stocking decisions are relative to true mean demand.
*Indicates a significant bias below the true mean ($p < 0.01$).

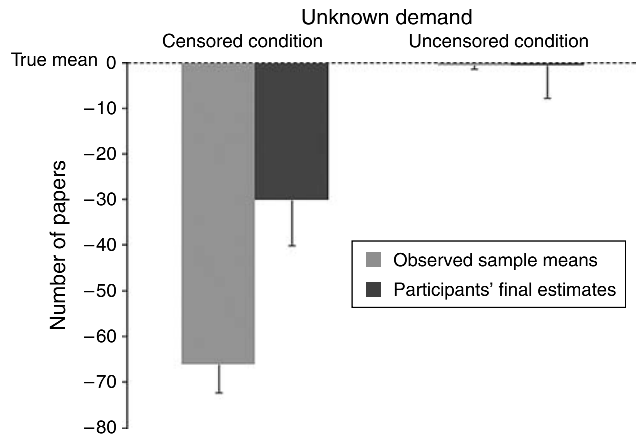
6.2.1. Demand Beliefs. There was a significant interaction between demand knowledge and censorship environment ($F(1, 72) = 11.65$, $p < 0.002$). When demand was unknown, censoring demand feedback significantly biased demand beliefs downward ($M_{\text{unknown-censored}} = -27.9$ versus $M_{\text{unknown-uncensored}} = -4.4$; $F(1, 72) = 26.17$, $p < 0.001$). As one would expect, with known demand, censoring demand feedback had no effect on demand beliefs ($M_{\text{known-censored}} = -1.3$ versus $M_{\text{known-uncensored}} = 0.0$; $F(1, 72) = 0.08$, $p > 0.75$). Furthermore, estimates of the underlying mean demand were not significantly different from the true mean when demand was either known or uncensored ($p > 0.15$ for each). However, when demand was unknown in a censored environment, average estimates were significantly lower than true mean demand ($M = -27.9$, $t(72) = 8.59$, $p < 0.001$).

6.2.2. Stocking Decisions. The results for stocking decisions were almost identical to those for demand beliefs. There was a significant interaction between demand knowledge and censorship environment ($F(1, 72) = 9.02$, $p < 0.005$). When demand was unknown, stocking decisions were significantly lower in a censored environment than an uncensored environment ($M_{\text{unknown-censored}} = -31.1$ versus $M_{\text{unknown-uncensored}} = -7.0$; $F(1, 72) = 14.96$, $p < 0.001$). When demand was known, stocking decisions were not significantly different across censored and uncensored environments ($F(1, 72) = 0.14$, $p > 0.7$). Furthermore, stocking decisions were not significantly different from the optimal level when demand was either known or uncensored ($p > 0.1$ for each). As expected, when demand was unknown in a censored environment, stocking decisions were significantly lower than optimal ($M = -31.1$; $t(72) = 7.07$, $p < 0.001$).

6.2.3. Adjustments from Sample Mean. Next, we tested whether individuals made estimates of mean demand higher than the mean of their final observed sample (Hypothesis 4). The results are illustrated in Figure 7. With censorship, both participants' final estimates of mean demand and their final stocking decisions were significantly higher than their 30-period observed sample mean in both demand knowledge conditions (unknown demand: $\bar{x} = -66.1$, final estimate $= -30.0$, final stock $= -28.9$; known demand: $\bar{x} = -41.6$, final estimate $= -0.8$, final stock $= 6.7$; ts$(72) > 4.5$, ps < 0.001). This evidence suggests that individuals were not purely naïve and did significantly account for censorship by adjusting above the sample mean in their estimates of the population mean.

In summary, censorship significantly biased estimates of mean demand below the true mean when individuals needed to infer an unknown demand.

Figure 7 **Participants' Mean Final Estimates of Mean Demand and Mean Observed Sample Means with Standard Error Bars Displayed by Censorship Condition**



Notes. Data are from the unknown demand conditions only. Estimates are shown relative to true mean demand.

Further, stocking decisions were unaffected by censorship when demand was known, but when demand needed to be learned, censorship caused stocking decisions to be significantly lower. In other words, when individuals formed an accurate understanding of demand—with known demand or with unknown demand in an uncensored environment—their stocking decisions were approximately optimal on average. However, when they formed a biased belief about demand, their stocking decisions were consequently also biased.[8]

7. Study 4: Risk Perceptions in Sequential Risk Taking

The empirical work that we have presented thus far has focused on inferences about the mean of a normal distribution in the domain of learning demand from sales. Study 4 extends our findings in two ways: (1) it tests the censorship bias in a different context, sequential risk taking (Pleskac 2008, Wallsten et al. 2005); and (2) it examines inferences about a uniformly distributed population, as opposed to a normally distributed population. Sequential risk taking consists of a repeated risky choice in which risk changes dynamically and systematically. The risky choice initially has a high probability of a positive payoff (or *gain*) that continues play and a small probability of a negative payoff (or *bust*) that ends play. After each risky choice that results in a gain, the individual may choose to take another risk in hopes of yet another gain or to quit the round to avoid a potential bust. This decision depends on two critical factors: the risk preference of

[8] See the work by Rudi and Drake (2011) for some evidence that in some cases censorship may bias stocking decisions even with known demand.

the decision maker, and the perceived probability of a bust. In this study, we examine how censorship affects the latter of these two factors.

Censorship occurs in sequential risk taking when individuals cannot observe the precise turn on which they would have busted if they had continued taking more risks. With censorship, individuals cannot observe how many total gains they *could have* achieved before busting. On the other hand, in rounds that do end in a bust, they can observe exactly how many gains they were able to achieve before busting. This asymmetry generates censored samples that may cause individuals to underestimate the mean number of risks they can take before a bust.

Clinical psychologists have used sequential risk-taking tasks to simulate risk taking with drugs, alcohol, and other public health risks. Behavior on these laboratory tasks has been shown to correspond well with risk-taking behavior in the real world (Lejuez et al. 2002, 2003; Wallsten et al. 2005). Sequential risk taking also occurs in organizations, such as when setting replacement policies (see §1). Social dynamics involving potential conflict may also involve sequential risks. For example, an ambitious negotiator may continually fight for as much value as possible at the bargaining table while risking offending the opposing party and ruining a potential deal altogether. Similarly, managers must promote disagreement and the sharing of diverse opinions among their employees while risking the emergence of interpersonal conflict. This is one of the biggest challenges facing managers today (Behfar et al. 2008, Jehn et al. 1999, Simons and Peterson 2000). The present study examines how individuals infer risk from the outcomes of their sequential risky choices.

7.1. The Task

The sequential risk-taking task was adapted from Pleskac's (2008) Angling Risk Task.[9] Participants played a fishing game in which they made 25 trips to a pond (see Appendix D for a picture of the task). Participants were informed that all trips began with the same number of fish in the pond (N), but were not told the precise value of N. One of the fish in the pond was blue and the rest, $N-1$, were red. Catching the red fish gave them money (i.e., a gain), but catching the blue fish ruined that trip (i.e., a bust). Participants were given a uniform prior for N between 5 and 45. In reality, for all participants, the number of fish in the pond was set to 24.

Each trip to the pond entailed the following process. Participants could use a "cast and catch a fish" button to randomly catch one fish from the pond. Each time they caught a red fish, 5 cents was placed into a temporary *trip bank*. If they ever caught the blue fish, they lost the money from their trip bank and that trip ended. However, at any point after catching a red fish, they could use a "quit trip and collect" button to end the trip and move their trip bank money to their *permanent bank*. This is how they earned money in the game.

Each trip started with the same number of fish in the pond (24 fish). The chance of catching the blue fish on the next cast increased with each red fish caught, because there was then one fewer red fish in the pond (i.e., sampling without replacement). Also, the amount of money that they would forfeit with a bust increased as they accumulated red fish on a given trip. At the end of each trip, the number of fish in the pond was always reset to 24 and always with only one blue fish. After 25 trips, the game ended and participants could keep the money in their permanent banks. At this point, we asked participants to estimate N, the number of fish in the pond at the start of each trip. In this task, busts followed a discrete uniform distribution between 1 and N (or in this case 1 and 24). Therefore, participants essentially estimated the maximum of the uniform distribution of blue fish.

This environment is censored: In trips where participants chose "quit trip and collect" to cash in their trip earnings, they could not observe how many more red fish they could have caught before busting. This served as the censored condition. In the uncensored condition, after they quit a trip to collect their trip money, we simulated more casts and told them on which cast they would have caught the blue fish if they had continued casting. Therefore, when uncensored, at the end of each round the participant could observe on which cast the blue fish was or would have been caught. Participants were aware of both possible conditions.

7.2. Methods

7.2.1. Participants. The participants were 39 undergraduate students at an American university. They were randomly assigned to a censorship condition. Participants earned a base pay of \$4 plus bonus money based on performance in the task. The bonus money was comprised of the money earned from the fishing game and an additional dollar if their estimate of N was within 1 of the correct answer. Most participants earned \$8–\$15.

7.2.2. Procedures. Participants were given an instruction sheet that explained the rules of the game as described above. They were not informed of the estimation task (guessing the number of fish in the pond) until after the fishing game was completed. The task was programmed in Authorware, and the study was

[9] The censored condition in this study was very similar to Pleskac's (2008) "catch and keep" fishing tournament with cloudy conditions.

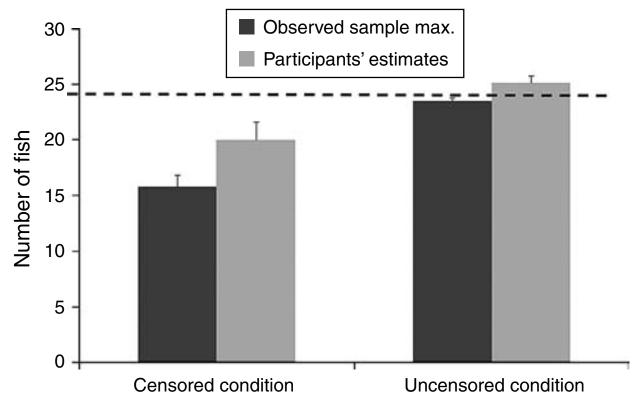
run on computers in a lab. The permanent bank, temporary bank, trip number, result of the most recent cast, and a count of red fish caught on the current trip were displayed on the screen (see Appendix D). When a participant caught the blue fish, a notice popped up showing them the blue fish, the number of casts they made that trip, and the amount of money that they forfeited from their temporary bank. Between trips, participants were reminded that the number of fish in the pond had returned to N with one blue fish. When participants clicked "quit trip and collect," they were shown how many casts they made on that trip and how much money they earned. In the uncensored condition, we simulated more casts and showed them on which cast they would have caught the blue fish had they continued casting. In the censored condition, they did not observe this simulation of additional casts.

After the 25th trip, participants were asked to estimate N, the number of fish in the pond at the beginning of each trip. Subsequently, we asked them to imagine that they were going to play the game again, but this time with a *known* number of fish in the pond. However, instead of making individual choices, they would need to set a "decision policy" of how many times they would cast before clicking "quit trip and collect." We asked how many casts they would want to do before quitting the trip if they knew that there were 24 fish in the pond (including one blue fish). Participants were then paid based on their permanent bank status and the accuracy of their estimate of N (see specific incentives above).

7.3. Results

The true total number of fish in the pond was 24. Individuals in the censored environment on average estimated that there were 20.0 total fish in the pond ($SD = 6.4$),[10] whereas the average estimate of those in the uncensored environment was 25.1 ($SD = 2.8$), a significant difference between conditions ($F(1, 36) = 10.0$, $p < 0.01$; refer to Figure 8). The former was significantly lower than the true number of fish in the pond ($t(36) = 3.54$, $p < 0.002$), but the latter was not ($t(36) = 0.93$, $p > 0.35$). As an indication of their decision policy for when to stop taking risks, we examined the average number of casts made in trips where participants quit the trip before a bust (Pleskac 2008). Individuals averaged 20% fewer casts

Figure 8 **Study 4 Mean Participant Estimates of the Number of Fish in the Pond (N) and the Mean Maximum Number of Fish Observed by Participants Displayed by Censorship Condition, with Standard Error Bars**



Note. The dashed line indicates the true number of fish, 24.

in these trips when censored than when uncensored (9.8 versus 12.2 casts; $F(1, 36) = 7.67$, $p < 0.01$). Overall, including busts, individuals averaged fewer casts when censored than when uncensored (8.2 versus 9.6; $F(1, 36) = 5.85$, $p < 0.03$). Furthermore, when facing censorship, individuals caught fewer blue fish (11.7 versus 13.6 blue fish; $F(1, 36) = 2.8$, $p = 0.10$), but earned less money across the 25 trips ($5.90 versus $6.75; $F(1, 36) = 2.47$, $p = 0.12$) because they quit earlier, although these differences were not significant at the $\alpha = 0.05$ level, two-tailed.

Had they known that there were 24 fish in the pond, individuals across conditions demonstrated very similar preferences. Individuals in the censored and uncensored conditions, on average, indicated that they would have set a policy of casting 13.3 and 12.9 times before quitting, respectively (not significantly different, $p > 0.7$). For individuals in the censored environment, their self-reported preferred policy with 24 fish was to take significantly more risks than the level of risk revealed in their actual behavior in the game (9.4 casts; $t(37) = 4.16$, $p < 0.001$). However, for individuals in the uncensored environment, their self-reported preferred policy with 24 fish was not significantly different from their behavior in the game (12.2 casts; $t(37) = 0.77$, $p > 0.4$).

7.3.1. Comparisons to Maximum-Likelihood Estimate. Next, we compared participants' estimates of the number of fish in the pond to the MLE given the same censored samples (see Appendix E for details on the MLE calculation). In the censored environment, participants' estimates were significantly lower than the respective MLEs given the same samples ($M_{\text{estimate-censored}} = 20.0$ versus $M_{\text{MLE-censored}} = 23.0$; $t(18) = 2.27$, $p < 0.05$). The MLEs in the censored environment were not significantly different from the true number of fish in the pond, 24 ($t(18) = 1.61$, $p > 0.12$).

[10] There was one very unusual observation in the censored condition where a participant caught the blue fish on one of the first three casts in each of their first four trips to the pond. The chance of this happening is approximately 1 in 4,000. This participant proceeded to never cast more than three times on a single trip, and then estimated that there were four fish in the pond. To be conservative, we omitted this outlier from the analysis, although including it would only increase the magnitude of the bias in the censored condition.

7.3.2. Comparisons to Observed Sample Maximum. We also compared the censored estimates to a purely naïve estimate given the same sample. The naïve estimate, e^n, in this case was the maximum number of fish observed in any round. If the trip with the highest number of caught fish did not result in a blue fish, then we set the naïve estimate to be one greater than the number of fish caught on that trip because the naïve judge knows that there is at least one more fish. For example, if the round with the most caught fish involved quitting a trip after 15 casts, then $e^n = 16$. In the censored condition, individuals significantly adjusted from their maximum number of fish observed in their estimates of the number of fish in the pond ($M_{\text{naïve-censored}} = 15.8$ ($SD = 4.7$) versus $M_{\text{estimate-censored}} = 20.0$; $t(18) = 4.92$, $p < 0.001$; refer to Figure 8).

In sum, censorship caused individuals to take fewer risks and to underestimate the total number of fish in the pond. This decrease in risk taking caused them to catch fewer blue fish (i.e., achieve fewer busts), but also to earn less money than individuals playing without censorship. Had they known that there were 24 fish in the pond at the start of each trip, they reported that they would have preferred to take more risks. In other words, the censorship bias caused individuals to take fewer risks than they actually would have liked to take. Lastly, although censorship caused individuals to underestimate the number of fish in the pond, individuals were less biased than a purely naïve estimate based on the observed sample maximum, suggesting that they did partially account for censorship.

When making decisions from description, the probabilities tied to possible outcomes are known, and choices are driven by risk preferences. However, when making decisions from experience, the probabilities tied to outcomes must be learned over time, and choices are driven by both risk preferences *and* risk perceptions (Hau et al. 2008, Hertwig et al. 2004). This study demonstrates how censorship can bias risk perceptions when making decisions from experience leading to overly conservative choices.

8. General Discussion

In this research we demonstrated a *censorship bias*—individuals in censored environments tend to rely too heavily on their observed sample, biasing their beliefs about the underlying population. We found evidence of such sample naïveté in each of four empirical studies. Study 1 examined the task of learning the mean of a normally distributed unknown demand from sales. This study provided evidence of the censorship bias: Individuals with censorship underestimated mean demand. Furthermore, the censorship bias was exacerbated for higher degrees of censorship and when the censorship point was variable. Study 2 demonstrated the bias in a one-shot estimation task and found that higher variance in the underlying population exacerbated the bias. Study 3 linked biased inferences from censored samples to behavior by studying simultaneous judgment and decision making in a task that involved dynamic demand learning and inventory choices. Censorship caused individuals to underestimate demand and stock less inventory than optimal. Finally, Study 4 examined censorship in a sequential risk-taking task where individuals attempted to avoid a negative outcome that was uniformly distributed with an unknown upper bound. Individuals with censorship took fewer risks and underestimated the upper bound of the distribution. Had they known the upper bound, they reported that they would have preferred to take more risks than they did. Censorship caused them to form biased beliefs about the environment and led them to behave in an overly risk-averse way.

Consistent with the naïve intuitive statistician metaphor (Fiedler and Juslin 2006, Juslin et al. 2007), the censorship bias was greater when the censored sample was less representative of the true population. In some instances, individuals performed almost as if they were completely naïve to the sample bias created by censorship. In other circumstances, individuals did use evidence of censorship to adjust from the observed sample to form their beliefs about the underlying population. Nevertheless, their adjustments largely fell short of theoretically attainable heuristic strategies. In Studies 1 and 2, individuals' estimates could be compared to the estimates of a simple MLE-based heuristic estimate (Nahmias 1994) given the same censored samples. The heuristic estimate given the same censored samples greatly outperformed the estimates of individuals, suggesting that individuals could benefit from simple decision aids in censored environments.

These empirical results suggest that judgment in censored environments may be driven by the use of the censored sample as an initial anchor for estimating characteristics of the underlying population (Chapman and Johnson 1999; Epley and Gilovich 2001, 2004, 2006; Tversky and Kahneman 1974; Strack and Mussweiler 1997). Because individuals tend to naïvely believe biased samples are more representative of populations than they actually are (Fiedler 2000, Juslin et al. 2007), they insufficiently adjust from the observed sample and form biased beliefs about the population. Fiedler and Juslin (2006) speculated that the severity of sample naïveté may vary across several sampling conditions: passive (being exposed to a sample) versus active sampling (taking part in creating a sample), and simultaneous versus sequential sampling. In our studies, we used four different experimental paradigms to test the robustness of the effect, and several trends emerged across these paradigms in the degree of the censorship bias. Participants demonstrated less adjustment for sample bias when passively sampling (Studies 1 and 2) than when actively sampling (Studies 3 and 4). Behavioral estimates of the mean were closer to the observed sample mean when individuals were simply exposed to a sample rather than actively taking part in shaping the sample (for comparison, see Figures 5 and 6 versus Figures 7 and 8). Furthermore, estimates of individuals tended to be more biased when observing a simultaneously drawn sample with a one-shot estimation (Study 2) than when observing sequential observations (Studies 1, 3, and 4), although we did not find significant improvement over time with sequential sampling. These qualitative observations suggest that sample naïveté may be worse when passively observing samples and when making one-shot judgments.

We demonstrated that a greater degree of censorship and higher variance in the population can make beliefs about the population more biased. We believe other factors also affect judgment in censored environments by making the observed sample less representative of the underlying population. A negative correlation between censorship points and population draws makes the observed sample more misrepresentative of the underlying population because high draws become censored at even lower points. For example, a negative correlation may occur if a firm has more limited access to inventory in periods where customer demand is high. On the other hand, a positive correlation between censorship points and sample observations makes the sample more representative of the population because high draws coincide with high censorship points. For example, if a firm can partially anticipate random demand realizations and appropriately adjust inventory levels, then we expect beliefs about the mean of the demand distribution to be less biased.

Our results suggest that skewness in the population may also exacerbate the censorship bias. A right-censored environment would censor the long tail of a right-skewed population. Therefore, the true value of the high observations in the long tail would go unobserved causing the observed sample to be less representative of the population. Future research may explore when these and other factors determine the extent to which individuals form biased beliefs in censored environments.

Future research could also explore how censorship affects inferences about *nonstationary* stochastic processes (Bloomfield and Hales 2002, Massey and Wu 2005, Kremer et al. 2010). If individuals are naïve to the misrepresentativeness of an observed censored sample, then they will be slower to adjust their beliefs when a shift in the underlying stochastic process yields more censorship than when it yields less censorship. In the case of demand learning, our results suggest that individuals may be slower to adjust their beliefs for upward shifts in demand compared to downward shifts.

Van Nieuwerburgh and Veldkamp (2006) suggested that some large-scale economic patterns may actually be driven by asymmetric learning. Specifically, they argued that the asymmetry in business cycles (Sichel 1993, Veldkamp 2005)—rapid recession and slow growth—emerges as a consequence of what can be learned from productivity. During a boom, firms engage in high investment and productivity. When the boom ends, firms have precise evidence of the downturn and decisively decrease investment, yielding low productivity. However, when growth resumes, the low productivity allows only noisy signals of improvement, which slows learning and makes recovery more gradual. In our terms, the productivity level censors what firms can observe about the market. Downturns yield more uncensored observations allowing swift recognition of a shift, whereas upturns produce more censored observations that only provide imprecise evidence of improvement. In this way, it is conceivable that our microlevel findings relate to macrolevel learning processes (Van Nieuwerburgh and Veldkamp 2006).

8.1. Managerial Implications

Our empirical findings on learning an *unknown* demand complement the large amount of inventory-ordering research that assumes a *known* demand distribution (e.g., Schweitzer and Cachon 2000; Bolton and Katok 2008; Croson and Donohue 2003, 2006; Su 2008; Ho et al. 2010, Özer et al. 2011[11]). Demand beliefs directly inform inventory decisions. Therefore, even if an inventory policy is already determined and optimized for certain cost parameters, firms using past sales data may underorder because the censorship bias causes them to underestimate demand. Furthermore, our results suggest that the censorship bias will be more problematic in several predictable circumstances. For instance, the censorship bias in demand estimation is likely to be a greater problem with low-profit margin products, for which it is in the manager's interest to maintain lower inventory and incur more frequent stockouts; the resulting higher

[11] These authors consider a setting in which the demand distribution is known to the retailer but only partially known to the manufacturer.

degree of censorship is likely to cause greater underestimation of demand. Conversely, for high profit-margin products, greater inventory is maintained and fewer stockouts occur, so a censorship bias is likely to be smaller. Similarly, a cash- or space-constrained firm may be forced to hold lower levels of inventory, which is likely to cause greater underestimation of demand. The effect of higher variance in the population suggests that high demand variance may be costly not only because of expected underage and overage costs, but also because of a greater demand estimation bias. Finally, the effect of a variable censorship point suggests that demand inference is likely to be more accurate in cases where inventory levels remain constant for multiple periods of time before adjustment. Consistency in inventory levels may facilitate better demand learning by increasing the salience of stockouts as they amass at a single point (for related research, see Bolton and Katok 2008 and Lurie and Swaminathan 2009).

In domains of risky choice, individuals often cannot observe what would have happened if they had taken more risks (March 1996). Because individuals tend to rely on their observed sample of experiences to make judgments, they fail to appreciate potential benefits that were missed as a result of conservative decision making. As in Study 4, censorship may cause people to overestimate risk and consequently avoid risky options that they actually would have liked to take. Therefore, it may be particularly important for managers to reconsider their perceptions of risk when they use conservative risky-choice policies, which censor outcomes to a greater extent. Similarly, having insurance when taking risks not only reduces the variance of outcomes, but also realigns incentives to make exploratory risk taking more attractive. Therefore, moral hazard may improve learning.

Our predictions can also be extended to other organizational domains not studied in this paper. For instance, the production capacity of a person or a system may be probabilistic, and the amount actually produced may be censored by the amount of work available to be done. For example, the amount of work a manager assigns may censor the amount of work an employee can complete (Feiler 2012). Similarly, in sequential production processes, such as assembly lines, the work available to one station may be limited by an earlier lower-capacity station. In these cases, when production capacity falls short of the amount of work available to be done, the manager can observe precisely what was completed and what was unfinished. However, when production capacity exceeds the amount of work available, the manager receives a censored observation: she cannot observe how much more could have been completed. Our findings suggest that in contexts where production capacity is censored by work available, individuals tend to underestimate production capacity, which could lead to suboptimal work allocation, production line design, or supplier selection decisions.

Our findings also suggest that censored environments may be good candidates for intervention with cognitive repairs (Heath et al. 1998), decision aids, or, if practically and financially possible, optimization tools. For example, in follow-up research, we have found that asking individuals to explicitly estimate the true value of each censored observation can significantly improve the accuracy of their estimates of the population mean. Furthermore, the highly robust performance of Nahmias's (1994) prescriptive heuristic in our studies demonstrates the value of implementing even very simple decision tools for improving judgment in censored environments, although more complicated solutions should be implemented when censorship points change dramatically across periods.

Finally, there is a large body of organizational research arguing that managers and firms experiment too little, leading them to persist with a narrow set of beliefs and strategies (March 1991). When there are constraints that systematically limit what one can see, managers may consider encouraging greater experimentation to increase learning. In the case of censored environments, optimal experimentation requires sacrificing some short-term profitability by acting in a way that reduces censorship and reveals more true values of otherwise censored instances (e.g., Harpaz et al. 1982, Lariviere and Porteus 1999). However, our empirical findings suggest that such experimentation may be more effective for human decision makers when censorship points are systematically set at a stationary value for multiple periods than when they are adjusted each period. An important direction for future research is to examine whether decision makers recognize the need to explore in environments with constraints on information. The evidence in this paper suggests that individuals do not do so optimally. Furthermore, research could examine whether individuals are capable of exploring effectively even when experimentation is of little or no cost.

8.2. Conclusion

Although previous research has developed statistical tools for coping with censored data, little attention has been given to how managerial intuition may be biased by censorship. This paper provides insights into how individuals make judgments in censored environments, which can be applied to various managerial settings. Individuals in censored environments tend to rely too heavily on their observed sample, causing them to form biased beliefs about the underlying population. Systematic aspects of the environment—such as the degree of censorship, population variance, and censorship point variability—increase the degree of bias. The censorship bias can cause suboptimal decision making that may be costly for organizations. An important challenge faced by managers is the need to build flexible organizations that can recognize and harvest gains in high potential-performance periods, while avoiding excessive vulnerability in low potential-performance periods.

Acknowledgments
The authors thank Dan Ariely, Li Chen, Nicholas Epley, Gerd Gigerenzer, Reid Hastie, Teck Ho, Gurhan Kök, Albert Mannes, John Payne, Devin Pope, Lael Schooler, Sim Sitkin, Jack Soll, Jeannette Song, Richard Thaler, George Wu, and three anonymous reviewers for their helpful feedback and conversations. The authors are grateful for the financial support provided by the Fuqua School of Business at Duke University.

Appendix A. Depiction of the Task Interface for Study 1

Period	Sales	Sold out?	Estimate *m* (mean demand)
1	580	Yes	600
2	502	No	558
3	573	Yes	578
4	521	No	554
5	545	No	?

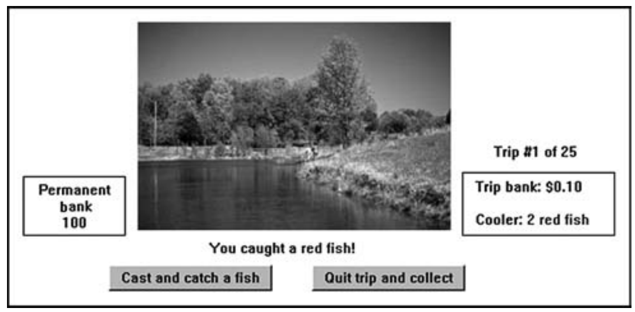
Note. This participant next needs to input a best estimate for the mean of underlying demand (m) for period 5.

Appendix C. Depiction of the Game Interface in Study 3

Period	What is *m*?	Order	Sales	Demand	Profit
1	600	600	600	673	600
2	650	645	525	525	405
3	575	580	580	630	580
4	615	615	604	604	593
5	?	?			

Notes. The demand column (in the dashed outline) only appeared for participants in the uncensored condition. This participant next needs to input a best guess for underlying mean demand, m, and an ordering decision for period 5.

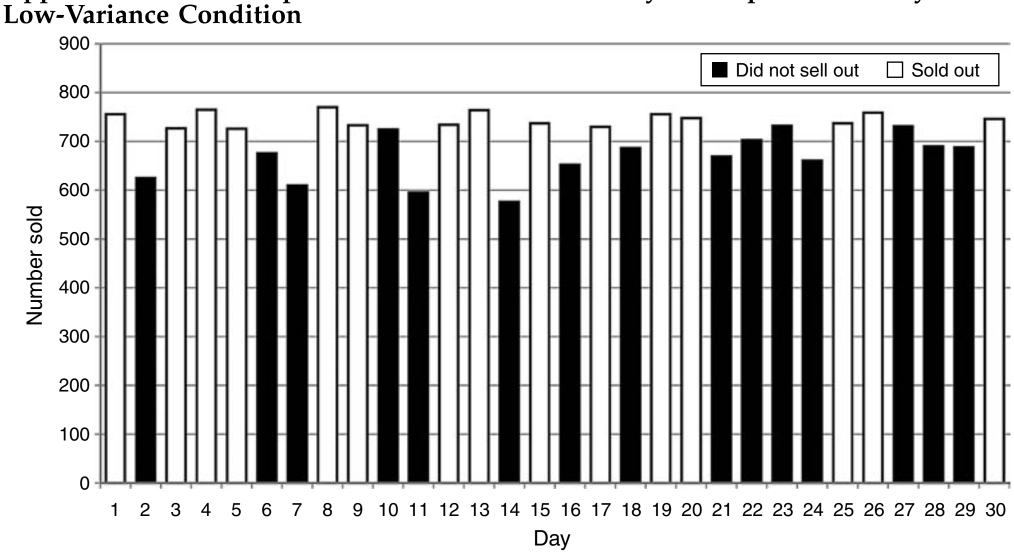
Appendix D. A Screenshot of the Sequential Risk-Taking Task in Study 4



Appendix E. Numerical Procedure for MLE Calculation Given Censored Sample in Study 4
Let $f(x)$ and $F(x)$ be the PDF and CDF of the chosen underlying distribution, which depend both on x and on the unknown distribution parameter(s). The likelihood function L can be shown to be

$$L = C\left\{\prod_{j \mid r_j=0} f(x_j)\right\}\left\{\prod_{j \mid r_j=1} [1-F(x_j)]\right\}.$$

Appendix B. An Example of Sales Data Observed by Participants in Study 2 in the Low-Variance Condition



Note. Mean demand was 745; $SD = 100$.

Here, C is a constant, the second product relates to those observations that are uncensored, and the third product relates to those observations that are censored. Using a numerical search (it is useful to use the log-likelihood), it is not difficult to find the parameter value that maximizes the likelihood function.

References

Amemiya T (1984) Tobit models: A survey. *J. Econometrics* 24:3–61.

Behfar K, Peterson R, Mannix EA, Trochim W (2008) The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes. *J. Appl. Psych.* 93:170–180.

Belsley DA, Kuh E, Welch RE (1980) *Regression Diagnostics: Identifying Influential Data and Source of Collinearity* (John Wiley & Sons, New York).

Bloomfield RJ, Hales J (2002) Predicting the next step of a random walk: Experimental evidence of regime-shifting beliefs. *J. Financial Econom.* 65(3):397–414.

Bolton GE, Katok E (2008) Learning-by-doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing Service Oper. Management* 10(3):519–538.

Bostian AJ, Holt CA, Smith AM (2008) Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing Service Oper. Management* 10(4):590–608.

Brenner LA, Koehler DJ, Tversky A (1996) On the evaluation of one-sided evidence. *J. Behavioral Decision Making* 9(1):59–70.

Campbell A, Whitehead J, Finkelstein S (2009) Why good leaders make bad decisions. *Harvard Bus. Rev.* 87(2):60–66.

Chapman GB, Johnson EJ (1999) Anchoring, activation, and the construction of values. *Organ. Behav. Human Decision Processes* 79(2):115–153.

Chen L (2012) Dynamic pricing with active learning under two-sided censoring. Working paper, Fuqua School of Business, Duke University, Durham, NC.

Chen L, Kök AG, Tong JD (2013) The effect of payment schemes on inventory decisions: The role of mental accounting. *Management Sci.* 59(2):436–451.

Cohen J, Cohen P, West SG, Aiken LS (2003) *Applied Multiple Regression/Correlation Analysis for Behavioral Sciences* (Lawrence Erlbaum Associates, Mahwah, NJ).

Croson R, Donohue K (2003) Impact of POS data sharing on supply chain management: An experimental study. *Production Oper. Management* 12(1):1–11.

Croson R, Donohue K (2006) Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Sci.* 52(3):323–336.

Denrell JC (2005) Why most people disapprove of me: Experience sampling in impression formation. *Psych. Rev.* 112(4):951–978.

Denrell JC (2007) Adaptive learning and risk taking. *Psych. Rev.* 114(1):177–187.

Einhorn HJ, Hogarth RM (1978) Confidence in judgment: Persistence of the illusion of validity. *Psych. Rev.* 85(5):395–416.

Epley N, Gilovich T (2001) Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psych. Sci.* 12(5):391–396.

Epley N, Gilovich T (2004) Are adjustments insufficient? *Personality Soc. Psych. Bull.* 30(4):447–460.

Epley N, Gilovich T (2006) The anchoring and adjustment heuristic. *Psych. Sci.* 17(4):311–318.

Feiler DC (2012) Managerial underestimation of employee capability. Working paper, Dartmouth College, Hanover, NH.

Fiedler K (2000) Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psych. Rev.* 107(4):659–676.

Fiedler K, Juslin P (2006) Taking the interface between mind and environment seriously. Fiedler K, Juslin P, eds. *Information Sampling and Adaptive Cognition* (Cambridge University Press, New York), 3–32.

Gilovich TR (1991) *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life* (Free Press, New York).

Hansson P, Juslin P, Winman A (2008) The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *J. Experiment. Psych.: Learn., Memory, Cognition* 34(5):1027–1042.

Harpaz G, Lee WY, Winkler RL (1982) Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Sci.* 28(6):589–603.

Hau R, Pleskac TJ, Kiefer J, Hertwig R (2008) The description-experience gap in risky choice: The role of sample size and experienced probabilities. *J. Behav. Decision Making* 21(5): 493–518.

Heath C, Larrick RP, Klayman J (1998) Cognitive repairs: How organizational practices can compensate for individual shortcomings. *Res. Organ. Behav.* 20:1–37.

Hertwig R, Barron G, Weber EU, Erev I (2004) Decisions from experience and the effect of rare events in risky choice. *Psych. Sci.* 15(8):524–539.

Ho T-H, Lim N, Cui TT (2010) Reference-dependence in multi-location newsvendor models: A structural analysis. *Management Sci.* 56(11):1891–1910.

Hogarth RM (2001) *Educating Intuition* (University of Chicago Press, Chicago).

Hogarth RM, McKenzie CRM, Gibbs BJ, Marquis MA (1991) Learning from feedback: Exactingness and incentives. *J. Experiment. Psych.: Learn., Memory, Cognition* 17(4):734–752.

Jehn K, Northcraft G, Neale M (1999) Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Admin. Sci. Quart.* 44(4):741–763.

Juslin P, Winman A, Hansson P (2007) The naive intuitive statistician: A naive sampling model of intuitive confidence intervals. *Psych. Rev.* 114(3):678–703.

Juslin P, Winman A, Olsson H (2000) Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psych. Rev.* 107(2):384–396.

Kahneman D, Tversky A (1972) Subjective probability: A judgment of representativeness. *Cognitive Psych.* 3(3):430–454.

Klayman J, Ha YW (1987) Confirmation, disconfirmation, and information in hypothesis testing. *Psych. Rev.* 94(2):211–228.

Koehler JJ, Mercer M (2009) Selection neglect in mutual fund advertisements. *Management Sci.* 55(7):1107–1121.

Koehler JJ, Thompson WC (2006) Mock jurors' reactions to selective presentation of evidence from multiple-opportunity searches. *Law Human Behav.* 30(4):455–468.

Kremer M, Moritz B, Siemsen E (2011) Demand forecasting behavior: System neglect and change detection. *Management Sci.* 57(10):1827–1843.

Lariviere MA, Porteus EL (1999) Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.* 45(3):346–363.

Larrick RP, Wu G (2007) Claiming a large slice of a small pie: Asymmetric disconfirmation in negotiation. *J. Personality Soc. Psych.* 93(2):212–233.

Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, et al. (2002) Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J. Experiment. Psych.: Appl.* 8(2):75–84.

Lejuez CW, Aklin WM, Jones HA, Richards JB, Strong DR, Kahler CW, et al. (2003) The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Experiment. Clinical Psychopharm.* 11(1):26–33.

Lurie NH, Swaminathan JM (2009) Is timely information always better? The effect of feedback frequency on decision making. *Organ. Behav. Human Decision Processes* 108(2):315–329.

MacLeod C, Campbell L (1992) Memory accessibility and probability judgments: An experimental evaluation of the availability heuristic. *J. Personality Soc. Psych.* 63(6):890–902.

Maddox WT, Ashby FG, Bohil CJ (2003) Delayed feedback effects on rule-based and information-integration category learning. *J. Experiment. Psych.: Learn., Memory, Cognition* 29(4):650–662.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

March JG (1996) Learning to be risk averse. *Psych. Rev.* 103(2): 309–319.

Markle A (2009) Asymmetric disconfirmation in managerial beliefs about employee motivation. Unpublished doctoral dissertation, University of Chicago, Chicago.

Markle A (2011) Dysfunctional learning in decision processes: The case of employee reciprocity. *Strategic Management J.* 32(13):1411–1425.

Massey C, Wu G (2005) Detecting regime shifts: The causes of under- and overreaction. *Management Sci.* 51(6):932–947.

Muller A, Stoyan D (2002) *Comparison Methods for Stochastic Models and Risks* (John Wiley & Sons, New York).

Mynatt CR, Doherty ME, Tweney RD (1977) Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quart. J. Experiment. Psych.* 29(February): 85–95.

Nahmias S (1994) Demand estimation in lost sales inventory systems. *Naval Res. Logist.* 41(6):739–757.

Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. General Psych.* 2(2):175–220.

Özer Ö, Zheng Y, Chen KY (2011) Trust in forecast information sharing. *Management Sci.* 57(6):1111–1137.

Pleskac TJ (2008) Decision making and learning while taking sequential risks. *J. Experiment. Psych.: Learn., Memory, Cognition* 34(1):167–185.

Rudi N, Drake D (2011) Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. Working Paper 12-042, Harvard Business School, Boston.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Sichel DE (1993) Business cycle asymmetry: A deeper look. *Econom. Inquiry* 31(2):224–236.

Silverman GK, Loewenstein GF, Anderson BL, Ubel PA, Zinberg S, Schulkin J (2010) Failure to discount for conflict of interest when evaluating medical literature: A randomized trial of physicians. *J. Medical Ethics* 36(5):265–270.

Simon HA (1990) Invariants of human behavior. *Annual Rev. Psych.* 41(1):1–19.

Simons TL, Peterson RS (2000) Task conflict and relationship conflict in top management teams: The pivotal role of intragroup trust. *J. Appl. Psych.* 85(1):102–111.

Stewart N, Chater N, Brown GDA (2006) Decision by sampling. *Cognitive Psych.* 53(1):1–26.

Strack F, Mussweiler T (1997) Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *J. Personality Soc. Psych.* 73(3):437–446.

Su X (2008) Bounded rationality in newsvendor models. *Manufacturing Service Oper. Management* 10(4):566–589.

Tversky A, Kahneman D (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psych.* 5(2):207–232.

Tversky D, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1130.

Van Nieuwerburgh S, Veldkamp LL (2006) Learning asymmetries in real business cycles. *J. Monetary Econom.* 53(4):753–772.

Veldkamp LL (2005) Slow boom, sudden crash. *J. Econom. Inquiry* 124(2):230–257.

Wallsten TS, Pleskac TJ, Lejuez CW (2005) Modeling behavior in a clinically-diagnostic sequential risk-taking task. *Psych. Rev.* 112(4):862–880.