Chapter 5
Analytics for Operational Visibility in the Retail Store: The Cases of Censored Demand and Inventory Record Inaccuracy

Li Chen and Adam J. Mersereau

1 Introduction

A retail store is a system in which customers, associates, and merchandise interact which each other over time to produce sales and profits for the firm (Fig. 5.1). The store, however, is far from a black box from the manager's perspective. Retail managers have a number of operational levers to influence these interactions, including store design, assortment planning, pricing, inventory control, and staffing. Retail managers also have some visibility into what transpires in the store.

Historically, this visibility has been limited to inventory positions, staff schedules, and, since the emergence of barcode technologies in the 1970s, point-of-sale (POS) data. Recent years, however, have seen a heightened interest among practitioners in store visibility—how a retailer can gain clearer visibility and how it can best use this visibility for operational and marketing advantage. Citing opportunities brought by existing and new retail data sources, a recent report by the McKinsey Global Institute highlights retail's "tremendous upside potential across the industry for individual players to expand and improve their use of big data" (McKinsey Global Institute 2011).

We believe that one factor contributing to this interest in visibility is the continued rise of internet retailing (i.e., e-commerce), which continues to grow as a fraction of the overall retail industry. A commonly cited advantage enjoyed by e-commerce retailers compared with their brick-and-mortar cousins is their visibility into the sales process, given that interactions of customers with the e-commerce retail site (and with associates and inventory, when applicable) can be (and are)

L. Chen
Fuqua School of Business, Duke University, Durham, NC, USA

A.J. Mersereau (✉)
Kenan-Flagler Business School, University of North Carolina, Chapel Hill, NC, USA
e-mail: ajm@unc.edu

N. Agrawal, S.A. Smith (eds.), *Retail Supply Chain Management*,
International Series in Operations Research & Management Science 223,
DOI 10.1007/978-1-4899-7562-1_5

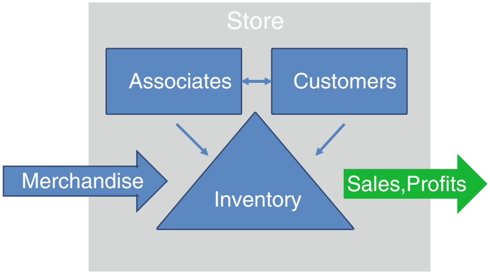

Fig. 5.1 A simplified view of in-store retail operations

recorded and mined for information. Customer clickstreams can reveal detailed insights into customer behavior. Furthermore, the customer experience is largely decoupled from firm operations in e-commerce retail, enabling tighter monitoring of inventory control and customer service. Indeed, a significant challenge of modern in-store retailing, seen in the push for "omni-channel" retailing (Brynjolfsson et al. 2013), is learning how best to compete with, complement, and learn from the e-commerce channel. Part of the answer seems to be finer visibility into and control of the in-store environment.

A second factor behind the increased interest in retail visibility is the emergence of modern technologies for in-store data collection as well as information technologies for capturing, storing, and analyzing data from these sources. These technologies bring the promise of a revolution in retail operations by offering visibility at a more granular level of detail and a finer time scale. Examples of such technologies include radio-frequency identification (RFID) and traffic counters, which have existed for a number of years but whose uses are still being explored and evaluated, and new technologies such as smart shopping carts, video monitoring, and cell phone tracking. In our investigation of these technologies and in discussions with practitioners and academics, we have encountered both optimism and skepticism about them. It is clear that new approaches are needed to translate these data sources into meaningful insights and profitable decisions and to evaluate the technologies. We will discuss some of these new visibility technologies and the associated research opportunities in Sect. 4.

The main goal of this chapter is to provide a review of two substantial literatures on in-store retail management that deal with imperfect visibility, namely demand censoring and inventory record inaccuracy. We believe that these two literatures, though largely disjoint from each other, share common features and themes that make them instructive for other problems involving in-store visibility.

Inventory Management with Censored Demand Observations Retail demand data are typically captured by POS transactions. However, POS data present an imperfect observation of true demand due to the demand censoring effect: when the actual demand exceeds the available inventory level, the excess demand is not captured by the POS data. The demand censoring problem is more prominent in brick-and-mortar stores than in online stores, because the latter can monitor and track customer purchases closely to alleviate such a problem. Academic researchers have long recognized the need to account for this censoring effect in demand forecasting and inventory management (e.g., Conrad 1976; Wecker 1978). This literature has been primarily centered on methodologies for dealing with the imperfect demand observations.

Inventory Management with Inaccurate Inventory Records Computerized inventory positions, which accumulate POS and store receipts on a daily basis, form the basis of automated replenishment policies for many retailers. There is ample evidence that such logical inventory records do not match the physical inventories on store shelves due to shrinkage, misplacement, and transaction errors (see DeHoratius and Ton 2015). In other words, retail managers have imperfect visibility into inventory in the store. A substantial literature on managing inventories given this reality has grown in the last decade, featuring diverse assumptions on error processes, decisions, and observability.

Given the common challenges in incorporating imperfect information into operational models, it is not surprising that both literatures use overlapping methodologies such as various learning and optimization paradigms. These two literatures also yield some common insights. One such insight is that lack of visibility can be costly, and if not properly accounted for can erase the gains from sophisticated, optimized policies. A second is that intelligent analytics can substitute for visibility in some cases. A third is that analytical models can help measure the return on investment of new visibility technologies by evaluating the best performance possible without visibility.

For ease of reference for readers who may be interested in only one of the topics, we have written the reviews of these two literatures to be largely self-contained. The rest of the chapter is organized as follows. In Sect. 2, we review the literature on the demand censoring problem. We discuss three types of models: Bayesian models with perishable inventory, Bayesian models with nonperishable inventory, and nonparametric models. We conclude the section by comparing the Bayesian and nonparametric models and discussing future research opportunities. In Sect. 3, we review the literature on the inventory record inaccuracy problem. Specifically, we provide a basic illustrative model for the problem, discuss the modeling issues and tradeoffs, review specific models from the literature, and conclude with a discussion of important open research questions. Section 4 discusses emerging visibility technologies and future research opportunities more generally in the general area of in-store visibility.

2 Models of Demand Censoring

In most retail environments, when inventory runs out, the unmet demand is lost and not observed. As a result, the sales data are censored by the available inventory level. When the demand distribution is known, this is a classic inventory problem involving lost sales (see Zipkin 2000 and references therein). However, if the demand distribution is not known, which is often the case for a new product introduction, one has to rely on potentially censored data to estimate the unknown demand. Intuitively, if this partial observability of demand is not factored into the estimation procedure, the demand estimate will be biased low (Wecker 1978). If the low demand estimate is subsequently used to determine inventory stocking decisions, the resulting inventory level will also be biased low and thus will lead to more lost sales and an even lower future demand estimate. To avoid this potential vicious cycle, it is important to take into account the data censoring effect in demand estimation and inventory control decisions. In other words, we need to develop "intelligent" methods to narrow the performance gap between a system with imperfect demand data and a system with full-visibility.

Consider the case in which the demands in each period, denoted by D_t, are independently and identically distributed (i.i.d.). The demand D_t here could be a residual variable after removing seasonality and promotional effects. Let us further assume that the demand probability density function, denoted by $f(\xi|\theta)$, has an unknown parameter θ, with $\theta \in \Theta$. Let $F(\xi|\theta)$ denote the cumulative distribution function (CDF) and $\overline{F}(\xi|\theta) = 1 - F(\xi|\theta)$ the complementary CDF.

Also let y_t denote the inventory level in period t. Then the sales in period t is given by $\min\{D_t, y_t\}$. If $D_t = \xi_t < y_t$, then the demand information is observed exactly, and the likelihood function is given by $f(\xi_t|\theta)$. On the other hand, if $D_t \geq y_t$, then the demand information is censored by the inventory level y_t; all we know is that the actual demand is greater than or equal to y_t, so the likelihood function is given by $\overline{F}(y_t|\theta)$.

Suppose that there are n historical sales observations. Without loss of generality, let the first j observations be the exact demand observations, i.e., $\xi_1, \ldots, \xi_j$, and let the remaining $n - j$ observations be the censored demand observations, i.e., $y_{j+1}, \ldots, y_n$. We can write the joint likelihood function as

$$\prod_{i=1}^{j} f(\xi_i|\theta) \cdot \prod_{i=j+1}^{n} \overline{F}(y_i|\theta).$$

By maximizing this expression over θ we obtain the maximum likelihood estimator (MLE) of the unknown demand parameter.

Conrad (1976) recognizes the difference between sales and demand data and proposes the above MLE method for Poisson demand. Nahmias (1994) further considers the demand censoring problem for normal demand, and provides three estimators: the MLE estimator, the best linear unbiased estimator, and a simplified estimator based on three sample statistics. He compares the performance of these three estimators by simulation. Agrawal and Smith (1996) find that the negative binomial distribution fits their empirical data significantly better than the Poisson and normal distribution, and develop estimators for the negative binomial distribution under demand censoring. Anupindi et al. (1998) apply the MLE method to estimate the Poisson demands of multiple substitutable products for a vending machine data set. In their problem, product stockouts result in only partial lost sales due to substitution. They develop an expectation-maximization (EM) method to account for *missing* stockout information in periodical inventory data. For a similar demand estimation problem, Conlon and Mortimer (2013) develop an EM method under a discrete choice model and demonstrate that failing to account for stockouts correctly can lead to biased demand estimates. Vulcano et al. (2012) further develop an efficient EM algorithm under the multinomial logit choice model, where they treat the observed demand as an incomplete observation of the primary demand (i.e., the would-be demand if all products were available for sale). Musalem et al. (2010) develop an alternative Bayesian estimation method based on data augmentation (i.e., imputing the entire sequence of sales) with Markov chain Monte Carlo methods.

To apply a Bayesian method to the estimation problem at hand, let $\pi(\theta)$ denote an initial prior belief on the unknown demand parameter θ. The posterior belief of θ given the same n historical sales observations as before can be written as

$$\pi(\theta|\xi_1,\ldots,\xi_j,y_{j+1},\ldots,y_n)=\frac{\pi(\theta)\prod_{i=1}^{j}f(\xi_i|\theta)\cdot\prod_{i=j+1}^{n}\overline{F}(y_i|\theta)}{\int_{\Theta}\pi(\theta')\prod_{i=1}^{j}f(\xi_i|\theta')\cdot\prod_{i=j+1}^{n}\overline{F}(y_i|\theta')d\theta'}.$$

As with the MLE case, the ordering of the demand observations does not affect the Bayesian posterior because of the product form of the likelihood function. For an N-point discrete demand distribution with an N-dimensional beta prior, Silver (1993) derives a recursive formula for computing the Bayesian posterior expected values of the N probability masses under demand censoring.

When demand is fully observable, the above Bayesian updating procedure can be greatly simplified with conjugate prior distribution families—one only needs to update the corresponding sufficient statistic of the conjugate prior (see DeGroot 1970 for a detailed discussion of this topic). However, when demand is censored due to unobserved lost sales, most common conjugate prior distribution families do not apply. In particular, Braden and Freimer (1991) conjecture that the distributions that entail a sufficient statistic under demand censoring, termed the "newsvendor distribution," are limited to the following distribution family:

$$\overline{F}(\xi|\theta)=e^{\eta(\theta)b(\xi)},$$

where $\eta(\cdot)$ and $b(\cdot)$ are real-valued functions. Examples of such distributions include the exponential distribution, the Weibull distribution, certain bounded support distributions and certain bimodal distributions (see Braden and Freimer 1991). Specifically, when $\eta(\theta)=-\theta$ and $b(\xi)=\xi^k$ with fixed $k>0$, the newsvendor distribution takes the form of the Weibull distribution. Below we use the Weibull distribution to illustrate the Bayesian updating scheme under demand censoring.

Under the Weibull distribution, the demand density function is given by

$$f(\xi|\theta) = k\theta\xi^{k-1}e^{-\theta\xi^k} \quad \text{for} \quad \xi \geq 0.$$

Let us further assume that the initial prior follows a gamma distribution with the shape parameter $a > 0$ and the scale parameter $S > 0$, i.e.,

$$\pi(\theta) = \frac{S^a\theta^{a-1}e^{-S\theta}}{\Gamma(a)} \quad \text{for} \quad \theta \geq 0.$$

Thus, given the same n historical sales observations as before, it is easy to verify that the posterior also follows a gamma distribution with the updated shape and scale parameters given by $a + j$ and $S + \sum_{i=1}^{j} \xi_i^k + \sum_{i=j+1}^{n} y_i^k$, respectively. In other words, the shape parameter increases by one only when an exact demand observation is made, and the scale parameter increases by $(\min\{\xi_t, y_t\})^k$ every period.

An advantage of the Bayesian method over the MLE method is that one can integrate demand estimation together with optimal control, and formulate the joint estimation and optimization problem as a Bayesian dynamic program. In a seminal paper, Scarf (1959) first studies such a joint estimation and optimization problem when demand information is fully observable (i.e., without demand censoring). Scarf (1960) further shows the dimensionality of the Bayesian dynamic program can be reduced for the gamma-gamma conjugate prior distribution family. Azoury (1985) extends Scarf's state-space reduction technique to various conjugate prior distribution families, such as the Pareto-uniform and the gamma-Weibull conjugate priors. Under certain suitable conditions, Lovejoy (1990) shows that the Bayesian dynamic program can be simplified to a single-period optimization problem. When demand is censored due to unobserved lost sales, the joint estimation and optimization problem becomes much more challenging. Below we provide a review of the existing literature on this subject.

2.1 Bayesian Models with Perishable Inventory

Consider a periodic-review inventory control problem for a single product. The product is stocked and sold for T periods. At the beginning of each period t ($t = 1, \ldots, T$), an inventory level y_t is chosen to minimize the total inventory holding and stockout penalty costs. The production leadtime is assumed to be negligible, so the inventory level is achieved immediately after the decision. Here we also assume the product is perishable and cannot be carried over to meet demands in subsequent periods. In this case, the on-hand inventory at the beginning of a period is always zero.

At the end of each period, a unit holding cost h or a unit penalty cost p is charged for any leftover inventory or unsatisfied demand, respectively. The purchase cost of the product is omitted in our formulation as it can be normalized to zero with the standard technique of Heyman and Sobel (1984). The terminal value at the end of the planning horizon is assumed to be zero.

Let $\pi_t(\theta)$ denote the prior belief of the unknown demand parameter θ at the beginning of period t. The predictive demand density in period t is given by $\int_\Theta f(\xi|\theta)\pi_t(\theta)d\theta$. Given the inventory level y, the single-period expected inventory holding and stockout penalty cost, denoted by $L_t(y,\pi_t)$, can be expressed as

$$\begin{aligned} L_t(y,\pi_t) &= h\mathsf{E}_{D_t|\pi_t}\left[(y-D_t)^+\right] + p\mathsf{E}_{D_t|\pi_t}\left[(D_t-y)^+\right] \\ &= h\int_\Theta\int_0^y (y-\xi)f(\xi|\theta)\pi_t(\theta)d\theta d\xi + p\int_\Theta\int_y^\infty (\xi-y)f(\xi|\theta)\pi_t(\theta)d\theta d\xi, \end{aligned}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$.

Let $V_t(\pi_t)$ denote the cost-to-go function from period t given the prior π_t. Then the Bayesian dynamic program optimality equations can be written as, for $t = 1, \ldots, T$,

$$\begin{aligned} V_t(\pi_t) &= \min_{y\geq 0}\{G_t(y,\pi_t)\} \\ &= \min_{y\geq 0}\left\{ L_t(y,\pi_t) + \int_\Theta\int_0^y V_{t+1}\left(\frac{f(\xi|\cdot)\pi_t(\cdot)}{\int_\Theta f(\xi|\theta)\pi_t(\theta)d\theta}\right) f(\xi|\theta)\pi_t(\theta)d\theta d\xi \right. \\ &\quad \left. + V_{t+1}\left(\frac{\overline{F}(y|\cdot)\pi_t(\cdot)}{\int_\Theta \overline{F}(y|\theta)\pi_t(\theta)d\theta}\right)\int_\Theta \overline{F}(y|\theta)\pi_t(\theta)d\theta \right\}, \end{aligned}$$

with $V_{T+1}(\cdot) = 0$. Let $y_t^p = \text{argmin}_{y\geq 0}\{G_t(y,\pi_t)\}$ denote the optimal inventory decision in the above problem. Also let $y_t^m = \text{argmin}_{y\geq 0}\{L_t(y,\pi_t)\}$ denote the myopic inventory decision in the problem. Note that in the case with no censoring, the myopic decision is in fact optimal in each period.

Intuitively, under demand censoring, one would stock more than the myopic inventory level to increase the chance of having an exact demand observation, i.e., $y_t^p \geq y_t^m$ for any common prior π_t. This is indeed true for arbitrary prior and demand distributions. Harpaz et al. (1982) first show this "stock more" insight under a general production output model. The same insight is shown to hold for the multiperiod newsvendor problem as described above by Ding et al. (2002), amended later by Lu et al. (2005) and Bensoussan et al. (2009). This insight is further extended to price-dependent demand models by Bisi and Dada (2007). Using the unnormalized prior technique developed in Bensoussan et al. (2005), Bensoussan et al. (2007a) show that an optimal policy exists and the "stock more" insight holds for an infinite-horizon problem.

To demonstrate this insight, let us examine the derivative of the dynamic program objective function below (Lu et al. 2008):

$$G_t^{'}(y,\pi_t) = L_t^{'}(y,\pi_t) + \left[V_{t+1}\left(\frac{f(y|\cdot)\pi_t(\cdot)}{\int_{\Theta} f(y|\theta)\pi_t(\theta)d\theta} \right) - \tilde{V}_{t+1}\left(\frac{f(y|\cdot)\pi_t(\cdot)}{\int_{\Theta} f(y|\theta)\pi_t(\theta)d\theta} \right) \right] \int_{\Theta} f(y|\theta)\pi_t(\theta)d\theta,$$

where $\tilde{V}_{t+1}(\cdot)$ is the expected cost when a suboptimal inventory policy, computed along each sample path assuming observation y was censored, is evaluated based on demand beliefs updated assuming y was uncensored. Thus, it is clear that $V_{t+1}(\cdot) \leq \tilde{V}_{t+1}(\cdot)$, and we have $G_t^{'}(y,\pi_t) \leq L_t^{'}(y,\pi_t)$. Hence, it follows that $y_t^p \geq y_t^m$ for any common prior π_t.

While this is an elegant structural result for the problem, computing the optimal inventory decision is still nontrivial. Easy-to-compute solutions are available only for certain conjugate prior distribution families. For example, Lariviere and Porteus (1999) derive a closed-form formula for the optimal inventory decision under the exponential demand distribution with a gamma prior. Bisi et al. (2011) further obtain a recursive formula for the more general Weibull demand distribution with a gamma prior. For general prior and demand distributions, Chen (2010) shows that the derivative of the dynamic program objective function can be computed by a recursive equation, but the dimensionality of the problem remains an obstacle for solving problems with relatively long time horizons.

2.2 Bayesian Models with Nonperishable Inventory

Now let us consider a more general case in which the product is nonperishable and can be carried over to meet demands in subsequent periods. In this case, the on-hand inventory at the beginning of a period is no longer zero, and we need to introduce an additional inventory state into the Bayesian dynamic program.

Let $V_t(x,\pi_t)$ denote the cost-to-go function from period t, given the on-hand inventory level x and the prior π_t. Then the Bayesian dynamic program optimality equations can be written as, for $t = 1,\ldots,T$,

$$
\begin{aligned}
V_t(x,\pi_t) &= \min_{y\geq x}\{G_t(y,\pi_t)\} \\
&= \min_{y\geq x}\left\{ L_t(y,\pi_t) + \int_{\Theta}\int_0^y V_{t+1}\left(y-\xi, \frac{f(\xi|\cdot)\pi_t(\cdot)}{\int_{\Theta} f(\xi|\theta)\pi_t(\theta)d\theta}\right) f(\xi|\theta)\pi_t(\theta)d\theta d\xi \right. \\
&\qquad \left. + V_{t+1}\left(0, \frac{\overline{F}(y|\cdot)\pi_t(\cdot)}{\int_{\Theta}\overline{F}(y|\theta)\pi_t(\theta)d\theta}\right)\int_{\Theta}\overline{F}(y|\theta)\pi_t(\theta)d\theta \right\},
\end{aligned}
$$

with $V_{T+1}(\cdot,\cdot)=0$. Let $y_t^* = \text{argmin}_{y\geq 0}\{G_t(y,\pi_t)\}$ denote the optimal inventory decision to the above problem. Bensoussan et al. (2008) show that an optimal policy also exists for the infinite-horizon problem.

Extending the derivative result of Lu et al. (2008), Chen (2010) shows that the derivative of the above objective function can be written as

$$
\begin{aligned}
G_t'(y,\pi_t) &= L_t'(y,\pi_t) + \int_{\Theta}\int_0^y V_{t+1}'\left(y-\xi, \frac{f(\xi|\cdot)\pi_t(\cdot)}{\int_{\Theta} f(\xi|\theta)\pi_t(\theta)d\theta}\right) f(\xi|\theta)\pi_t(\theta)d\theta d\xi \\
&\quad + \left[V_{t+1}\left(0, \frac{f(y|\cdot)\pi_t(\cdot)}{\int_{\Theta} f(y|\theta)\pi_t(\theta)d\theta}\right) \right. \\
&\qquad \left. -\tilde{V}_{t+1}\left(0, \frac{f(y|\cdot)\pi_t(\cdot)}{\int_{\Theta} f(y|\theta)\pi_t(\theta)d\theta}\right)\right] \int_{\Theta} f(y|\theta)\pi_t(\theta)d\theta,
\end{aligned}
$$

where $\tilde{V}_{t+1}(0,\cdot)$ is a generalization of $\tilde{V}_{t+1}(\cdot)$ in the perishable inventory case with zero starting inventory. Thus, we have $V_{t+1}(0,\cdot)\leq \tilde{V}_{t+1}(0,\cdot)$. But, on the other hand, we have $V_{t+1}'(y-\xi,\cdot)\geq 0$. Hence, $G_t'(y,\pi_t)$ can be either greater or less than $L_t'(y,\pi_t)$, implying that $y_t^* \geq y_t^m$ may not hold in this case. Thus, the "stock more" result in the perishable inventory case does not extend to the nonperishable inventory case when the optimal inventory decision is compared with the myopic decision.

Nevertheless, we can show that it is optimal to "stock more" than in a system without demand censoring. Since the myopic decision is optimal in a perishable inventory system without demand censoring, this can be seen as a generalization of the "stock more" result in the perishable inventory case. Let $V_t^o(x,\pi_t)$ denote the cost-to-go function from period t, given the on-hand inventory level x and the prior π_t for a system without demand censoring. Then the Bayesian dynamic program optimality equations can be written as, for $t=1,\ldots,T$,

$$
\begin{aligned}
V_t^o(x,\pi_t) &= \min_{y\geq x}\{G_t^o(y,\pi_t)\} \\
&= \min_{y\geq x}\left\{L_t(y,\pi_t)+\int_{\Theta}\int_0^{\infty}V_{t+1}^o\left((y-\xi)^+,\frac{f(\xi|\cdot)\pi_t(\cdot)}{\int_{\Theta}f(\xi|\theta)\pi_t(\theta)d\theta}\right)f(\xi|\theta)\pi_t(\theta)d\theta d\xi\right\},
\end{aligned}
$$

with $V_{T+1}^o(\cdot,\cdot)=0$. Let $y_t^o=\text{argmin}_{y\geq 0}\{G_t^o(y,\pi_t)\}$ denote the optimal inventory decision to the above problem.

Chen and Plambeck (2008) show that $y_t{}^* \geq y_t{}^o$ for any common prior π_t under the general discrete demand distribution. For a general continuous demand distribution, it is easy to verify that the derivative of $G_t{}^o(y,\cdot)$ is given by

$$
G_t^{o'}(y,\pi_t) = L_t^{'}(y,\pi_t) + \int_{\Theta}\int_0^{y}V_{t+1}^{o\,'}\left(y-\xi,\frac{f(\xi|\cdot)\pi_t(\cdot)}{\int_{\Theta}f(\xi|\theta)\pi_t(\theta)d\theta}\right)f(\xi|\theta)\pi_t(\theta)d\theta d\xi
$$

By backward induction, we can show that $V_{t+1}^{o\,'}(y-\xi,\cdot) \geq V_{t+1}^{'}(y-\xi,\cdot)$. Hence, it follows that $G_t^{'}(y,\pi_t) \leq G_t^{o'}(y,\pi_t)$, and we have $y_t{}^* \geq y_t{}^o$ for any common prior π_t.

Computing the optimal inventory decision for this problem is even more complex than for the perishable inventory case. Leveraging the dimensionality reduction technique developed by Scarf (1960) and Azoury (1985), Lariviere and Porteus (1999) show that this problem can be reduced to a two-dimensional dynamic program under the Weibull demand distribution with a gamma prior. Bisi et al. (2011) further show that the Weibull distribution is the only distribution that allows for such a dimensionality reduction technique for the problem. They also show that the dynamic program objective function is convex under the exponential demand distribution (a special case of the Weibull distribution when $k=1$), but is generally non-convex under other demand distributions.

From the generalized "stock more" result, a natural lower bound for the optimal inventory decision is given by $y_t{}^o$. This can be relatively easy to compute, benefiting from the fact that the corresponding Bayesian dynamic program is convex (see Scarf 1959). By the dimensionality reduction technique developed by Scarf (1960) and Azoury (1985), we can compute $y_t{}^o$ easily for an array of conjugate prior distribution families. However, for general prior and demand distributions, computing $y_t{}^o$ is still subject to the curse of dimensionality. Lu et al. (2007) derive an upper bound for the optimal inventory decision based on the first-order condition. However, their upper bound works only for certain prior and demand distributions. Chen (2010) further derives a set of upper bounds for the optimal inventory decision that works for all prior and demand distributions. For a fairly general monotone likelihood-ratio distribution family, he derives relaxed but easy-to-compute lower and upper bounds along any sample path. He also proposes two effective heuristics based on the solution bound results and the first-order condition.

2.3 Nonparametric Models

In addition to the Bayesian (parametric) models reviewed above, there is also a stream of research on the demand censoring problem based on nonparametric approaches. Under the nonparametric models, one makes no parametric assumptions on the underlying demand distribution, but employs an adaptive data-driven ordering policy that ensures the system performance converges to the optimal performance in the long run. It is worth noting here that the expected cost in each period in this literature is typically computed fixing the (unknown) true demand parameter θ. This differs from the Bayesian models, where such cost is integrated over the updated prior belief of θ, which could be influenced by the inventory decisions in the past.

Given the inventory decision y, the single-period newsvendor cost function is given by

$$L_t(y) = h \cdot \mathsf{E}_{D_t}\left[(y - D_t)^+\right] + p \cdot \mathsf{E}_{D_t}\left[(D_t - y)^+\right].$$

Let $y^* = \text{argmin}_{y \geq 0} L_t(y)$, and let L^* denote the resulting optimal cost (note that D_t is i.i.d., so the optimal decision in each period is stationary). It is easy to verify that the derivative of $L_t(y)$ is given by

$$L_t^{'}(y) = h \cdot \Pr(D_t < y) - p \cdot \Pr(D_t \geq y).$$

Thus, an unbiased sample-path estimate of the subgradient of $L_t(y)$ at y can be written as

$$H_t(y) = \begin{cases} h, & \text{if} \quad D_t < y, \\ -p, & \text{if} \quad D_t \geq y. \end{cases}$$

Using the above subgradient estimate, Burnetas and Smith (2000) propose the following simple adaptive ordering policy for the perishable inventory case:

$$y_{t+1} = y_t - \frac{y_t}{(h+p)t} \cdot H_t(y_t).$$

They show that under this ordering policy $\lim_{T \to \infty} \mathsf{E}[\Sigma_{t=1}^{T} L_t(y_t)/T] = L^*$ and y_t converges to y^* with probability one. They further extend this policy to a joint pricing and inventory ordering problem. Godfrey and Powell (2001) propose a similar sample-path subgradient estimate to successively approximate the newsvendor cost function with a sequence of piecewise-linear functions under demand censoring. A variant of their algorithm is shown to be asymptotically optimal under certain conditions (e.g., discrete demands) by Powell et al. (2004).

Huh and Rusmevichientong (2009) propose another adaptive ordering policy based on the sample-path subgradient estimate, and achieve a better rate of convergence. Specifically, assume that $\bar{y}$ is a known upper bound for the unknown optimal inventory level y^*. For some $\gamma > 0$, their adaptive ordering policy is given by

$$y_{t+1} = \max\left\{\min\left\{y_t - \frac{\gamma\bar{y}}{\max\{h,\ p\}\sqrt{t}} \cdot H_t(y_t), \bar{y}\right\}, 0\right\}.$$

They show that in the perishable inventory case, the long-run average system cost $E[\Sigma_{t=1}^{T} L_t(y_t)/T]$ under the above adaptive ordering policy converges to the optimal cost L^* at a rate of $O(1\sqrt{T})$. For the nonperishable inventory case, with an additional assumption that there is a known positive lower bound for the unknown expected demand $\mathsf{E}[D_t]$, they show that the above ordering policy achieves the same rate of convergence when γ is sufficiently small under some mild technical conditions.

For a general unknown discrete demand distribution with perishable inventory, Huh et al. (2011) propose a data-driven policy based on the Kaplan–Meier (KM) estimator (Kaplan and Meier 1958), termed the "KM-myopic" policy. To apply their policy, one needs to make the following change in the definition of demand censoring: given an inventory level y_t, one can observe the event $\{D_t = y_t\}$ distinctly from the event $\{D_t > y_t\}$. In other words, this equates to a "partial censoring" setting in which one observes an additional lost-sales indicator of whether demand *strictly* exceeds the available inventory level or not. We note that for continuous demand distributions, the notion of the lost-sales indicator is not essential because the events $\{D_t > y_t\}$ and $\{D_t \geq y_t\}$ have the same probability measure. However, for discrete demand distributions, such a notion makes a significant difference in Bayesian updating (see also Huh and Rusmevichientong 2009, Sect. 3.4).

Under this new notion of demand censoring, we provide an illustration of the KM estimator and the corresponding KM-myopic policy below. Given n sorted observations, say, $\xi_1 \leq \xi_2 \leq \xi_3^c \leq \xi_4 \leq \cdots \leq \xi_n$, where the superscript c denotes censored observations such that $D_t > \xi_t$, the KM estimator works as follows. At first, allocate probability equally among n observations. Then, starting from the left, redistribute the probability of a censored observation among higher observations iteratively. For example, in this case, the smallest censored observation is ξ_3. Thus, in the first iteration, the $1/n$ probability originally assigned to ξ_3 is shared equally among $\xi_4, \ldots, \xi_n$, each of which will hence get an updated probability of $1/n + 1/n(n-3) = (n-2)/n(n-3)$. After we pass through the observations in this way, the resulting empirical distribution is given by

$$\overline{F}_n(\xi) = \prod_{i:\xi_i \leq \xi} \left(\frac{n-i}{n-i+1}\right)^{\delta_i},$$

where $\delta_i = 0$ if ξ_i is a censored observation, and $\delta_i = 1$ otherwise. The adaptive KM myopic policy can thus be constructed as follows:

$$y_{t+1} = \min\left\{ y \geq 0 : \overline{F}_t(y) \leq \frac{p}{p+h} \right\}.$$

Huh et al. (2011) show that under the KM-myopic policy, y_t converges to the optimal inventory level y^* almost surely.

Besbes and Muharremoglu (2013) study the minimum worst-case regret for nonparametric models with perishable inventory, where they define regret as the difference between the expected cost of an adaptive policy and the full-information optimal cost L^*. They show that for a continuous demand distribution, the minimum worst-case regret under demand censoring grows logarithmically with the number of periods, as in the fully-observable demand case. On the other hand, when the demand distribution is discrete, they show that the minimum worst-case regret under demand censoring grows logarithmically with the number of periods, while regret can be bounded by a constant in the fully-observable demand case. Regret can also bounded by a constant under discrete demand in the "partially censored" setting. Thus, their finding highlights the importance of the availability of the lost-sales indicator in the existing literature of nonparametric models involving *discrete* demand distributions (e.g., Huh and Rusmevichientong 2009; Huh et al. 2011).

2.4 Open Research Areas

We have reviewed both Bayesian and nonparametric models for the demand censoring problem. Each type has its own strengths and limitations. For example, the Bayesian models entail an elegant Bayesian dynamic program formulation of the joint estimation and optimization problem. One can rely on these models to derive interesting structural results that shed light on the value of information and Bayesian learning. However, computing the optimal policy for the Bayesian models is nontrivial for relatively long time-horizon instances due to the curse of dimensionality. To overcome the dimensionality challenge, one typically has to resort to a fairly restrictive newsvendor distribution family that preserves the conjugate prior structure under demand censoring. This limits the applicability of the Bayesian models. The nonparametric models, on the other hand, work well for long time-horizon problems, and there is no need for any prior knowledge of the underlying demand distribution. As illustrated in our review, the adaptive ordering policies are often quite intuitive and easy to implement. The main challenge here, however, is to ensure the adaptive ordering policies converge quickly to the true optimal policy. Otherwise, the system performance in relatively short time horizons could be poor.

Despite the plethora of studies on demand censoring as reviewed above, there remain many open problems for future research. Below we discuss several of them.

1. *Demand Substitution:* Many retailers implicitly rely on demand substitution to mitigate the out-of-stock effect of a particular item at a particular store. There is an extensive literature on demand substitution, which is discussed in the chapters

concerning retail assortment planning in this volume. In our censored demand context, incorporating demand substitution among multiple products into the learning model could be of great practical value. Chen and Plambeck (2008) present a Bayesian model to jointly estimate the demand rate and the substitution probability. However, to keep their problem tractable, they make a simplifying assumption that the excess demand and the resulting substitution quantity are observable. It would be interesting to relax this assumption to investigate how demand censoring would affect the optimal inventory decisions under substitution. This is an open research problem that can be addressed by both the Bayesian and nonparametric approaches.

2. *Non-Stationary Demand:* Another practical consideration is non-stationary demand, which is common in many retail environments. Most of the censored demand models reviewed above assume the demand distribution is stationary. If the systematic variations in demand are deterministic (e.g., known seasonality), then one can simply normalize the demand observation by removing the deterministic variation components, so as to convert the problem to an equivalent stationary-demand one. However, if the systematic variations follow a random process, the problem becomes more complicated. Chen (2013a) shows that some of the results obtained under stationary demand can be extended to the Markov-modulated demand processes when the state transition probabilities are known. The case involving unknown transition probabilities is an open problem, as it is not clear how demand censoring would affect the learning of the unknown probabilities.
3. *Sales Transaction Timing Information:* One could further improve learning under censored demand by incorporating the timing of sales transactions. Jain et al. (2015) study such a Bayesian inventory control problem. They find that, when stockout timing information is available, the system performance improves significantly compared with the case without such information. Given that modern POS data include transaction timestamps, it would be interesting to further understand how timing information impacts some of the results reviewed here.
4. *Pricing Decisions:* One could also incorporate pricing decisions into the demand learning models. Burnetas and Smith (2000) propose an adaptive pricing and ordering policy for a price-dependent demand model with demand censoring. Bisi and Dada (2007) consider the joint pricing and ordering problem for price-dependent models in the Bayesian framework. Chen (2013a) studies a Bayesian dynamic pricing problem with an unknown customer willingness-to-pay distribution. In this case, if a customer buys a product, her willingness to pay must be greater than or equal to the posted price; if she does not buy the product, her willingness to pay must be below the posted price. Thus, the posted price serves as either a left- or right-censoring point of the customer's willingness to pay. Chen (2013a) proposes several approximation techniques to tackle this two-sided censoring problem. Applying the nonparametric approach to this two-sided censoring problem could be another interesting future research direction.

5. *Positive Replenishment Lead Times:* Both the Bayesian and nonparametric models in the literature assume zero lead time. Extending the existing models to the case of positive lead times would be an interesting and important contribution to the literature. However, we envision that such an extension could be technically challenging, because the lost-sales problem with a positive lead time is a known hard problem even when the demand distribution is known (see Zipkin 2000).

3 Models of Inventory Record Inaccuracy

There is ample evidence that the inventory available to customers on retail shelves is not correctly reflected in the retailers' computerized inventory records. In other words, retail managers have imperfect visibility into inventory in the store. DeHoratius and Raman (2008) examine the physical audit of a large, anonymous retail chain and observe that only 35 % of the retailer's inventory records match the physical inventory in the store. The extent of the problem is corroborated by other authors. Kang and Gershwin (2005) observe only 51 % record accuracy at a second anonymous retailer, and Gruen and Corsten (2008) find 32 % record accuracy at a third. We do not review in detail the literature on empirical measurement of inventory record inaccuracy; we refer the interested reader instead to the survey of DeHoratius and Ton (2015) in this volume.

Our focus instead is on potential analytical responses to the record inaccuracy phenomenon. Nearly all classical research on inventory management research assumes that the customer-available inventory level is known at every point in time, and landmark results in inventory theory rely on known inventory positions as a core (if not always explicit) assumption. A few analytical models of record inaccuracy date to the 1970s (e.g., Iglehart and Morey 1972), motivated by warehouse applications. There has been a surge of interest in inventory record inaccuracy in the past decade, particularly specialized to retail contexts, coinciding with new empirical studies and the rise of inventory tracking technologies—most prominently, item-level RFID tags which potentially offer real-time information on inventory locations and movements.

DeHoratius et al. (2008) outline three possible, non-exclusive responses of a retailer to inventory record inaccuracy: prevention, correction, and integration. Prevention refers to the elimination of root causes of inventory record inaccuracy, correction refers to inspection efforts, and integration refers to decision tools that account for the possible presence of inventory record inaccuracies. Our focus here is on "integrative" analytical approaches to inspection and replenishment, which we view as complementary to efforts towards prevention.

Analytical models are valuable for a few reasons. First, record inaccuracy is a significant feature of real inventory systems, and accounting for it has the potential to improve the matching of supply with demand and reduce inventory-related costs. Automated replenishment systems that assume accurate inventory records may not live up to their billing when this assumption is violated. Second, modeling record inaccuracy helps measure the return on investment of inventory tracking technologies such as RFID. By comparing the inventory management cost of a "full-visibility" retailer equipped with an inventory tracking technology (an idealized model of which affords perfect inventory visibility) with the best-possible performance of an "intelligent" or "informed" retailer with distributional information about errors, one obtains a measure of the value of inventory visibility (Rekik et al. 2008; Kök and Shang 2007; Lee and Özer 2007). In addition, many papers also consider as a benchmark the performance of a "naive" or "ignorant" retailer who is oblivious to errors. Models of "intelligent" retailers are the focus of this review. A common theme in the literature on inventory record inaccuracy is that "intelligent" inventory models that account for record inaccuracy can recapture a significant fraction of the benefits of visibility without the substantial physical investment in tracking technologies.

The purpose of this section is to review the analytical literature on inventory record inaccuracy with an eye towards how analytical models can make best use of available information in the absence of inventory visibility afforded by tracking technologies or process improvement initiatives. We begin by presenting an example model of inventory record inaccuracy to illustrate some basic insights and challenges. We then discuss key modeling considerations before discussing relevant papers in more detail. We conclude the section with a discussion of open research directions.

3.1 A Basic Model

Consider a basic, single period inventory model in which a decision maker (DM) chooses an inventory quantity to stock in the face of uncertain demand. As a benchmark, assume a newsvendor setup in which the DM has full knowledge of an initial stock x. The DM places an order for y items at unit cost c and the items arrive immediately with no lead time. Random demand D then arrives according to probability distribution F, yielding sales $S = \min\{D, x+y\}$. A penalty cost of p per unit is charged for unsatisfied demand $D - S$, and leftover inventory $x + y - S$ is salvaged for $c_s - h$ per unit. If inventory records are perfect and the initial inventory x is known, we can write the problem as

$$\min_{y \geq 0} \; L(x, y) - c_s \mathsf{E}_D\left[(x + y - D)^+\right], \tag{5.1}$$

where

$$L(x, y) = cy + p\mathsf{E}_D\left[(D - x - y)^+\right] + h\mathsf{E}_D\left[(x + y - D)^+\right].$$

The solution is well-known to be of the critical fractile type:

$$y^{*} = \min\left\{ y \geq 0 : F(x+y) \geq \frac{p-c}{p+h-c_s} \right\}. \tag{5.2}$$

Now suppose that inventory records are inaccurate, which we model by replacing the initial inventory position x by a random variable X with distribution P. We can write the new problem as

$$\min_{y \geq 0} \overline{L}(P,y) - c_s \mathsf{E}_{X,D}\left[(X+y-D)^{+}\right], \tag{5.3}$$

where

$$\overline{L}(P,y) = cy + p\mathsf{E}_{X,D}\left[(D-X-y)^{+}\right] + h\mathsf{E}_{X,D}\left[(X+y-D)^{+}\right].$$

The solution retains the critical fractile form (see Mersereau 2013),

$$\bar{y}^{*} = \min\left\{ y \geq 0 : W(y) \geq \frac{p-c}{p+h-c_s} \right\}, \tag{5.4}$$

but the demand distribution F is replaced by a new distribution $W(y) = \Pr(D - X \leq y)$. The distribution W reflects demand less available inventory and can be computed as a convolution of F and P.

It is intuitive that in many realistic cases the solution to (5.4) should exceed that of (5.2) in order to make up for inventory lost in the error process (assuming that $\mathrm{E}[X] \leq x$) and to buffer the additional newsvendor uncertainty introduced by the distribution P (assuming the fractile $\frac{p-c}{p+h-c_s}$ is sufficiently large). Indeed, a number of authors (e.g., Kök and Shang 2007; DeHoratius et al. 2008; Atali et al. 2011) observe either analytically or numerically that record inaccuracy does indeed tend to increase stocking quantities under reasonable assumptions on demand and/or error distributions.[1] We revisit this "uncertainty effect" on replenishment in Sect. 3.3.1.

We note that this single-period model can also be viewed as a random yield model with additive yield uncertainty. See Yano and Lee (1995) for a detailed review of the literature on inventory management with random yield. In the random yield literature, errors are typically connected to incoming replenishments and are typically immediately observed by the DM. Therefore, inventory uncertainty does not persist or accumulate over time. With record inaccuracy, however, errors generally persist until the retailer performs an inspection. This is a significant

[1] The result is difficult to prove generally. Song (1994) includes a detailed analysis of the conditions required to rank newsvendor stocking quantities for different probability distributions, and these conditions are difficult to verify here.

challenge in moving from a single-period model to a multiperiod model of inventory record inaccuracy.

It is natural to formulate a multiperiod inventory problem as a Markov decision process (MDP). With perfect inventory records, we can formulate a T-period lost-sales version of problem (5.2) as a MDP with a one-dimensional state representing the current inventory position. Let x_{t-1} indicate the inventory position at the beginning of period t, let D_t denote random demand in period t (drawn from a potentially time-varying demand distribution F_t), and indicate by $V_t(\cdot)$ the cost-to-go from period t through the end of the horizon. The Bellman equation is as follows for $t = 1, \ldots, T$:

$$V_t(x_{t-1}) = \min_{y \geq 0} \{L(x_{t-1}, y)\} + \mathsf{E}_{D_t}[V_{t+1}(U(x_{t-1} + y - D_t))], \tag{5.5}$$

where $V_{T+1}(x_T) = -c_s x_T$ Here, $U(x) = (x)^+$ is an update function that specifies the inventory carried to the next period. With record inaccuracy, the inventory record is no longer a sufficient summary of the system state and the inventory optimization becomes a "partially observed" MDP (POMDP). Define $P_t(x) = \Pr(X_t \leq x | \mathcal{H}_t)$ as the probability distribution of the inventory random variable X_t conditional on the observed process history $\mathcal{H}_t$. We may consider P_t to be the system state of a modified dynamic programming formulation for $t = 1, \ldots, T$,

$$\begin{aligned}\overline{V}_t(P_{t-1}) = \min_{y \geq 0} \left\{\overline{L}(P_{t-1}, y)\right\} \\ + \mathsf{E}_{X_{t-1}, D_t}\left[\overline{V}_{t+1}\left(\overline{\mathcal{U}}_t(P_{t-1}, y, \min\{X_{t-1} + y, D_t\})\right)\right],\end{aligned} \tag{5.6}$$

where $\overline{V}_{T+1}(P_T) = -c_s \mathsf{E}[X_T]$. Here, the update operator $\overline{\mathcal{U}}_t$ transforms P_{t-1} to P_t given replenishment y and observed sales $S_t = \min\{X_{t-1} + y, D_t\}$. We do not express the $\overline{\mathcal{U}}_t$ operator here explicitly, but we note that it can be complicated, in general depending on probability distributions of both paying demand and unobserved errors. It must shift the inventory distribution up and down to reflect observed inventory inflows (replenishments) and observed outflows (sales). It must accumulate potential errors occurring in period t. Finally, as we discuss later, the update may also account for inferences the DM can make about customer-available inventory based on sales or other side observations. DeHoratius et al. (2008) and others derive $\overline{\mathcal{U}}_t$ using Bayes law. In other models (e.g., Kök and Shang 2007) the classical inventory record and the number of periods of error accumulation serve as sufficient statistics for P_{t-1}, in which case the update operator is simpler to express.

POMDPs are provably difficult to solve in general (Papadimitriou and Tsitsiklis 1987), suggesting that a problem like (5.6) is unlikely to be solvable without restrictions or approximations.

3.2 Modeling Considerations

While we have attempted in Sect. 3.1 to frame a fairly general model of replenishment under inventory record inaccuracy, this model already makes a number of strong assumptions, in particular about the decisions available to the DM, the modeling of errors, and the DM's observations of the system. These three dimensions represent key distinctions among papers in the literature, and we briefly discuss each one in turn.

1. **Decisions:** Section 3.1 formulates the problem of *replenishment* under inventory record inaccuracy, but inventory *inspection* (also referred to as counting or auditing) is another control available to a decision-maker operating with inventory record errors. Traditionally, retailers do periodic (often annual) inventory counts for accounting purposes. More frequent inspections, referred to as "cycle counts," may follow a fixed schedule based on an ABC-type categorization of stock-keeping units (SKUs), in which SKUs judged to be particularly at risk of inaccurate records, or of strategic or financial importance, are scheduled for cycle counts more frequently. An alternative to such static counting schedules are dynamic versions of cycle counts in which the retailer chooses which SKUs to inspect each day based on real-time information.

   Conceptually, it is straightforward to extend (5.6) to dynamically trigger inspections. We add a binary decision variable z_t each period which is an input to the update operator $\overline{\mathcal{U}}_t$. An inspection in period t resolves the uncertainty around X_t, which we model with an update that sets P_t to a distribution with all its weight at a single value (or to an appropriate probability distribution that represents an imperfect inspection).
2. **Error Process:** A key distinction among models of inventory record inaccuracy is the modeling of the error process. Most authors work in a periodic review setting and assume an error random variable (sometimes referred to as "invisible" or "non-paying" demand) that contributes to the discrepancy between available and recorded inventory each period. These discrepancies are not directly observed, and they accumulate over time between inventory inspections. A modeler of inventory record inaccuracy must make a number of decisions about the error process. Errors can be modeled as additive (e.g., DeHoratius et al. 2008) or multiplicative (e.g., Rekik et al. 2008) relative to the inventory level, and dependent on or independent of demand, replenishment, and/or inventory levels. Errors can be modeled as occurring before or after demand within a period, or interleaved with demand (e.g., Atali et al. 2011). Errors themselves may be directly costly in that they imply a physical loss or gain of saleable units (e.g., Kang and Gershwin 2005) or costless (e.g., Camdereli and Swaminathan 2010). Errors may be modeled as deterministic or associated with a probability distribution. Typical assumed probability distributions are one-sided (e.g., Huh et al. 2010), implying that customer-available inventory

is always less than or equal to recorded inventory, or symmetric around zero (e.g., Kök and Shang 2007).

In order to appreciate these modeling decisions, we can categorize the sources of inaccurate inventory levels, following Atali et al. (2011), into shrinkage (i.e., physical loss of inventory, typically through theft or damage), transaction errors (i.e., scanning, receipt, or counting errors that impact inventory records but not physical inventory), and misplacements (i.e., in which inventory is temporarily unavailable to the customer but still physically present in the store).[2] These different error sources suggest different assumptions about inventory dynamics and cost accrual. For example, it is common to model shrinkage using a one-sided error process inducing direct stock losses, and transaction errors using an additive, symmetric error process that incurs no direct cost.

Modeling all sources of errors in detail (as in Atali et al. 2011) is arguably truest to retail realities, given that all three types of errors are presumably present in retail settings. (DeHoratius and Raman 2008 report discrepancies of both signs in the audit data they analyze, with 58 % of errors such that physical inventory is less than recorded inventory.) However, such a model may be difficult to estimate from data, and it may require additional state variables for tracking the different types of error accumulations to allow for proper accounting of costs. Instead, most authors model a single error process that either reflects a single error source (shrinkage, transaction errors, or misplacement) or an aggregation of error sources.

Assuming that demand and errors occur interleaved within a period is also desirable but complicates modeling because of the different accounting of lost sales and "lost errors." Such a model must account for all possible sequences of demand and errors within a period. Instead, many authors model errors as occurring together, either before or after demand within a period.

Another common simplification is to assume errors arise from a stochastic process independent of demand and inventory levels. In many retail contexts, we would expect this not to be the case; for example, the same underlying factors leading to high or low demand would seem to also impact the volume of shrinkage, misplacement, and transaction errors. Because demand and inventory levels are not directly observed, modeling this dependency can bring complications that destroy problem structure. In some models, these complications take the form of an additional layer of conditioning in the update operator (e.g., DeHoratius et al. 2008). In others, the dynamic program state may need

[2] Here we depart slightly from DeHoratius and Ton (2009) in terminology. DeHoratius and Ton (2009) define "inventory record inaccuracy" as the difference between a store's recorded inventory position and the physical inventory in the store. Misplaced inventory, which is physically present in the store, does not contribute to inventory record inaccuracy in this definition. In our discussion, we will liberally use the term "inventory record inaccuracy" to refer to the difference between customer-available inventory and recorded inventory. That is, we consider misplaced inventory to be part of inventory record inaccuracy.

to include the history of sales observations, leading to a curse of dimensionality (e.g., Kök and Shang 2007).

3. **Observability:** A critical modeling choice is what the DM observes about sales and stockouts. In the lost sales model of Sect. 3.1, sales are the minimum of demand and customer-available inventory and are therefore statistically dependent on customer-available inventory. In such a model, the DM can in theory use sales observations to make inferences on available inventory; for example, if a DM observes a sequence of periods with zero sales, this may signal that there is no customer-available stock. Such inferences can be modeled using Bayes law. While potentially powerful, these inferences yield a complicated $\overline{\mathcal{U}}_t$ operator in problem (5.6) (DeHoratius et al. 2008) that depends on sales observations and demand distributions.

   An alternative, which seems reasonable especially when stockouts are rare, is to ignore the signalling potential of sales observations. Such an assumption greatly simplifies the $\overline{\mathcal{U}}_t$ operator, as errors accumulate independently of sales. In such cases, the inventory record and the number of periods since the last inspection typically serve as sufficient statistics for the multiperiod dynamic optimization (Kök and Shang 2007).

   A third possibility is to assume that customer-available inventory levels become observed whenever they reach zero (e.g., Bensoussan et al. 2007b). This can be practically motivated by assuming that customers who find an empty shelf request a "rain check" that is recorded, or by the practice of "zero-balance walks" in place at some retailers, in which employees periodically look for empty shelves in the store.

3.3 Review of Existing Literature

With this backdrop, we now review the operations management literature on store-level analytical models of inventory record inaccuracy. Given the challenges inherent in problems like (5.6), we believe that a fruitful way to categorize the existing literature on inventory record inaccuracy is by the modeling assumptions and analytical approximations employed to enable tractable analysis and computation. We put the literature into four categories: single-period models, classical multiperiod models, multiperiod models featuring low-dimensional sufficient statistics for P_t, and "partially observed" multiperiod models employing Bayesian updating.

3.3.1 Single Period Models

Single-period models of optimal stocking under inventory record inaccuracy yield some basic insights while avoiding some of the complexities inherent in multiperiod POMDP formulations like (5.6). For this reason, single-period models are often employed as starting points upon which more complex models are built (e.g., Kök and Shang 2007; Huh et al. 2010; Mersereau 2013), or as stylized building blocks within more complex systems. For example, Heese (2007), Gaukler et al. (2007), Sahin and Dallery (2009) and Camdereli and Swaminathan (2010) employ single-period models to study the impact of inventory record inaccuracy on supply chain coordination. As our focus is on in-store operations, we do not review the supply chain aspects of these papers.

Rekik et al. (2008) analyze a single-period model that modifies the classical newsvendor problem by allowing for multiplicative misplacement errors to occur before paying demand arrives. A parameter θ is defined as the ratio between customer-available inventory and the inventory record, and is considered to be both deterministic as well as uniformly distributed on [0,1]. The authors explicitly look at the profit of naive, intelligent, and full-visibility retailers and conclude that the intelligent retailer achieves significant benefits over the naive retailer. The authors also examine stocking quantities: for the deterministic case stocking quantities first increase with θ (to make up for reduced yield) and then decrease with θ (to reduce misplaced inventory and associated overage charges).

Heese (2007) uses a multiplicative error model with uniformly distributed yields and makes similar observations about stocking quantities to Rekik et al. (2008). (His "centralized" model can be viewed as a single-location model.) Furthermore, even when setting the mean error ratio to one, Heese (2007) finds that the DM orders more than without inventory uncertainty for sufficiently high target service levels. We alluded to this "uncertainty effect" on stocking quantities in Sect. 3.1. Mersereau (2013) suggests an uncertainty effect in a model similar to (5.3). Mersereau (2013) also finds that optimal stocking levels can decrease if the DM anticipates physical errors to occur after stocking levels are chosen. This "direct loss" effect can be understood as reducing the stock available for theft or damage.

Single-period models therefore yield three insights into the effects of inventory record inaccuracy on optimal stocking levels: (1) optimal stocking levels may increase to make up for reduced yield, (2) they may also increase to buffer additional uncertainty brought by record inaccuracy; and (3) they may decrease in order to reduce the inventory available for misplacement or shrinkage.

3.3.2 Classical Multiperiod Models

A prevalent approach to modeling inventory record inaccuracy is to assume that inventory errors follow a pre-determined probability distribution that is independent of sales observations. As mentioned in Sect. 3.2, this greatly simplifies the update operator $\overline{\mathcal{U}}_t$ in (5.6), because the number of periods of error accumulation often serves as a sufficient statistic for the shape of P_t. Despite this simplification, optimal policies appear to be difficult to characterize in these systems except in specific cases.

An early stream of literature on inventory record inaccuracy, dating to Iglehart and Morey (1972), views error accumulation in the inventory system as a renewal process and seeks an auditing trigger that achieves a pre-specified probability of a "warehouse denial;" i.e., an event in which there is a physical stockout even though the inventory record appears sufficient to cover demand. This is an appropriate service metric in a warehouse context in which denials are observed by the firm and trigger a reconciliation of the inventory record with the physical inventory state.

Iglehart and Morey (1972) assume that errors are additive, stationary, mean zero random variables and that the DM maintains a buffer stock to account for them. Given a fixed buffer stock, the authors derive an asymptotic normal distribution approximation for the probability of cumulative errors exceeding the buffer stock. Their model decouples the classical safety and cycle stocks from the buffering of inventory inaccuracies, and the payoff is a joint inspection and replenishment policy expressed in closed form. The model of Rekik et al. (2009) is related in that the DM minimizes holding cost subject to a constraint on the probability of stockout during a finite horizon.

Morey (1985) uses a similar framework to Iglehart and Morey (1972) to establish "back-of-the-envelope" expressions for service levels as functions of error parameters, buffer stocks, and audit frequencies. Morey and Dittman (1986) generalizes Iglehart and Morey (1972) to determine audit frequencies in more general internal control settings, not necessarily inventory-related.

Kang and Gershwin (2005) present a detailed motivation for the problem of inventory record inaccuracy, including empirical evidence from an anonymous retailer. The paper's analysis is largely based on a numerical simulation of a (Q, R)-based stochastic inventory model with additive one-sided errors (called "stock loss" in the paper). One insight is that "freezing" of replenishment is possible; this occurs when the inventory record is above the reorder point yet there is no customer-available inventory on the shelf, in which case no sales occur and an automated replenishment system places no orders. The authors conclude that inventory inaccuracy may be especially costly in naive lean systems which carry little stock to buffer the additional uncertainty. This can be viewed as a corollary of the "uncertainty effect" discussed in Sect. 3.3.1. The paper goes on to numerically evaluate several remediation heuristics.

3.3.3 Multiperiod Models Featuring Sufficient Statistics

A number of papers analyzing multiperiod inventory optimization problems feature conditions or assumptions under which the multidimensional state P_t of a POMDP like (5.6) can be represented by a low-dimensional set of sufficient statistics. While such representations can incur a cost in terms of model generality, they have significant analytical and computational benefits.

Kök and Shang (2007) focus on joint replenishment and dynamic inspection triggering in a model in which errors are additive and have mean zero. They assume that both demand and errors are backlogged and that errors accumulate irrespective of backlogs. As a result, the error process decouples from inventory levels, as in Iglehart and Morey (1972), and the accumulated discrepancy between physical and recorded inventory is the sum over periods of individual errors. That is, if an error ε_t occurs each period, the accumulated error after j periods of no inspections is $\overline{\varepsilon}_j \equiv \sum_{t=1}^{j} \varepsilon_t$. Its distribution is a j-fold convolution of the one-period error distribution. As a result, the authors are able to formulate the joint replenishment-inspection problem using a two-dimensional state (z_t, j_t), where z_t is the inventory record at time t (maintained by adding replenishments and subtracting observed sales each period) and j_t is the number of periods since the last audit.

Unfortunately, even with these simplifications the authors show that the multiperiod problem is non-convex. The authors suggest an "inspection adjusted base-stock" (IABS) policy that replenishes according to a j_t-dependent base-stock policy and inspects when the inventory record falls below a j_t-dependent cutoff. An IABS policy is optimal for the single-period problem, and an IABS policy seems to perform well as a heuristic for the multiperiod problem.

Atali et al. (2011) provide a detailed model of inventory errors, explicitly distinguishing among shrinkage, transaction errors, and misplacements in their model. Furthermore, they model demand and errors using a "random disaggregation" approach that splits an overall demand random variable into components for paying demand and various error sources. As a result, their model allows for demand and errors to be interleaved within a period. In solving their intelligent ("informed") retailer model, the authors approximate the distribution of total errors by a distribution that depends only on the inventory record and the number of periods since the last audit, as in Kök and Shang (2007). A state-dependent base-stock replenishment policy results from this approximation. A numerical study shows that the intelligent retailer achieves cost close to a full-visibility one and that detailed modeling of errors can achieve significant gains over aggregate error models for some parameter choices. A related model appears in Avrahami et al. (2012), who find through a numerical study that a "static" informed policy that knows only mean error information does nearly as well as an intelligent policy based on distributional error information.

Huh et al. (2010) show that a similar two-dimensional state to the one in Kök and Shang (2007) is sufficient for a particular model in which inventory inaccuracy is driven by additive shrinkage only, replenishments are only possible immediately after an inspection is made, and stockouts induce automatic inspections (akin to a "zero-balance walk"). In a given period, the DM knows that the true inventory level has only decreased since the last inspection (since errors only reduce physical inventory and since replenishments require inspections). If a stockout has not occurred, then the most recent post-inspection inventory level less recorded demand must exceed the accumulated errors (whose distribution is determined by the number of periods since the last inspection). The inventory distribution *conditional* on there being no stockout can therefore be computed given the inventory record, the number of periods since the last audit, and the error distribution. The authors present a rigorous dynamic programming formulation based on this result and show that a threshold-based inspection policy, coupled with an order-up-to replenishment policy, is optimal for an infinite-horizon problem satisfying a number of technical assumptions.

3.3.4 Multiperiod Models Using Bayesian Updating

A set of authors studying inventory record inaccuracy has chosen to consider the partial observability of inventory levels more directly. These models require minimal assumptions on the inventory error distribution. In particular, the DM's belief P_t around inventory positions can be updated based on POS data. These models are more complex, however, in that the state space of the MDP is the space of possible distributions on X_t. Because of this complexity, optimal policies have only been computed for some simplified cases; otherwise, results are limited to heuristics and approximations.

DeHoratius et al. (2008) consider a multiperiod lost sales inventory system with discrete additive errors drawn from an arbitrary discrete distribution. The authors propose maintaining an explicit inventory belief P_t they call a "Bayesian inventory record" or "BIR." P_t is updated according to Bayes rule, using sales observations as signals of the underlying inventory levels. In particular, the Bayes update reflects that no sales may indicate a stocked out situation, and positive sales indicate that the inventory could not have been fewer than what was sold. The authors prove that such a solution avoids the problem of inventory "freezing" identified by Kang and Gershwin (2005).

DeHoratius et al. (2008) suggest a myopic replenishment policy and a BIR-based heuristic for dynamic triggering of inspections. The authors discuss the estimation of necessary parameters and report on a simulation study calibrated with retailer data that compares the performance of naive, intelligent ("Bayes"), and full-visibility ("Full") retailers. They demonstrate that the intelligent solution achieves a service-inventory tradeoff that captures a substantial portion of the benefits of the full-visibility solution.

DeHoratius et al. (2008) demonstrate that the updates can be performed efficiently in closed form when inventories and demands are discrete, but partial observability of inventory levels clearly adds analytical and computational complexity as discussed in Sect. 3.1. Mersereau (2013) analyzes in detail the problem of replenishment optimization for the model of DeHoratius et al. (2008), identifying both uncertainty and loss effects in a single-period version of the model. In a two-period version of the model, the author also identifies an "information effect:" stocking less can actually reduce the variance of the BIR and enhance information content for future periods. Mersereau (2013) proceeds to approximate the POMDP using an approach borrowed from the machine learning literature. A key finding is that an intelligent myopic policy is near-optimal in numerical trials.

Bensoussan et al. (2011b) formulate a related model to DeHoratius et al. (2008) in that excess demand is unobservable. Errors are one-sided, and demand and inventory are permitted to be continuous. Continuous inventory and demand complicates the updating process; the resulting inventory belief is a mixture of continuous and discrete distributions. The authors prove the existence and uniqueness of an optimal policy, present a lower-bounding approach, and propose an iterative approximation algorithm.

A separate series of papers considers similar "partially observed" inventory systems with continuous inventory levels where the DM only observes whether or not the physical inventory level is strictly positive. In particular, sales are not observed. Errors are not explicitly modeled but can be assumed to be a component of the (unobserved) demand process. Bensoussan et al. (2007b) considers such a model with lost sales. As in DeHoratius et al. (2008), the state of the system is represented by a distribution around the customer-available inventory level that is updated in a Bayesian fashion. The resulting replenishment problem is therefore defined on a functional state space, and the authors focus on finding conditions for an optimal solution to exist and to be unique. Bensoussan et al. (2008) perform related analyses for a variation of the Bensoussan et al. (2007b) model in which backorders (i.e., "rain checks") are permitted and the DM only observes the inventory level when it is negative. Bensoussan et al. (2011a) use a value function approximation to approximate the problem of Bensoussan et al. (2008). In a numerical study, they observe both an uncertainty and an information effect with interpretations related to those in Mersereau (2013).

Finally, Chen (2013b) considers the problem of dynamic cycle count triggering using a simplified POMDP in which the system can switch from a "normal" state in which the inventory level is known to a "faulty" state in which the system is stocked out. This results in a partial decomposition of the replenishment and inspection decisions. The inspection policy is an easily computed threshold policy based on the number of consecutive zero-sales periods, and the optimal replenishment is a base-stock policy with base-stock levels depending on the time since the last positive sale. The author finds a loss effect; the error process drives the retailer to stock less to limit the inventory made unavailable by errors. Chuang and Oliva (2013) also use a two-state model of record accuracy to determine the inspection frequency in a fixed inspection policy.

3.4 Open Research Areas

Despite numerous and varied analytical approaches to modeling retail inventory inaccuracy in recent years, there remain a number of open opportunities for future research.

1. *Multi-SKU and Multi-Location Models:* As with much of classical inventory theory, single-SKU models dominate the analytical literature on inventory record inaccuracy. Kök and Shang (2014) consider coordinated inspection policies in a serial supply chain. We are aware of little research, however, on models that use data across stores or SKUs. Consider the following inspection

trigger policy: inspect a SKU at a store when its recent sales fall significantly below sales for the same SKU at neighboring stores. It is intuitive that similar SKUs and stores, used in this way, could serve as useful benchmarks for detecting deviations from normal operations. Substitution is also potentially relevant to include in models of inventory record inaccuracy. For example, a retailer might suspect that a SKU has too little customer-available inventory after detecting increased sales of substitute SKUs. Extending models like (5.6) to multiple SKUs adds considerable complexity to the update operator and dimensionality to the state space, however.

2. *Estimation of Model Parameters:* Despite a fairly rich body of empirical research into the presence of record inaccuracy, there remain a number of open questions surrounding the estimation of the daily or weekly error processes assumed by most analytical models. DeHoratius et al. (2008) present a basic estimation approach, and Chuang and Oliva (2013) provide a structural approach for estimating error incidence at the SKU level. Nevertheless, we believe that detailed estimation of error processes remains an unresolved issue. As a result, the existing papers make use of a wide range of assumptions on error distributions. Furthermore, estimation of other model parameters may be confounded by record inaccuracy. Mersereau (2015) shows that the presence of inventory record inaccuracy can introduce biases into the estimation of paying demand.
3. *Analytical and Computational Tractability:* Efficient solutions, much less complete characterizations, of problems like (5.6) have proved elusive without approximations or restrictive assumptions. There is apparent in the existing literature a tradeoff between model realism and tractability, with no clear dominant approach. This leaves room for continued analytical and algorithmic work on both optimal solutions and useful approximations and heuristics.
4. *Comparison of Models and Prescriptions:* Despite the large number of competing models of inventory inaccuracy and solutions for replenishment and inspection, we are not aware of any efforts to compare them. One advantage of Bayesian models like DeHoratius et al. (2008) and Chen (2013b) is that they make use of sales information as signals about inventory levels. It is intuitive that this information should be most useful when stockouts are relatively common. It would be interesting to examine under what conditions a POMDP-based model like DeHoratius et al. (2008) outperforms a sufficient statistic model like Kök and Shang (2007), and vice versa.
5. *Pilot Testing of Policies:* Given the eminent practicality of inventory models integrating inventory inaccuracy, implementations of responses to inventory record inaccuracy would be especially interesting. Such reports have started to emerge. Chuang et al. (2012) report on a field experiment in which a data-driven heuristic was used to trigger inspections. Hardgrave et al. (2013) report on two controlled field experiments measuring the reduction in record inaccuracy enabled by real RFID implementations. Both papers suggest that the potential improvements to retail operations can be substantial.

4 Visibility Technologies and Research Opportunities

Both the literatures on demand censoring and inventory record inaccuracy formulate and solve problems of decision-making under uncertainty, and it is therefore not surprising that these literatures pull from a common set of methodologies including statistical decision theory, stochastic (and partially observed) dynamic programming, and Bayesian and nonparametric inference. We have proposed several specific research directions related to demand censoring and inventory record inaccuracy in Sects. 2 and 3, respectively. We add that demand censoring and inventory record inaccuracy tend to occur simultaneously in many retail stores, and their interaction leads to additional challenges. For example, when records are inaccurate, the retailer no longer receives a reliable indicator of when stockouts occur. Mersereau (2015) is the one paper we are aware of that considers both features together. One unique insight is that if demand censoring is accounted for but inventory record inaccuracy is not, then the retailer will tend to underestimate demand over time. We believe that there is room for further examination of this interaction as well as other interactions involving multiple sources of uncertainty, even though considering multiple uncertainties together brings obvious modeling complications.

We conclude the chapter by looking to other interesting directions for future research on in-store visibility that extend beyond demand censoring and inventory record inaccuracy. We believe that exciting research opportunities abound if we consider other types of information made available by new in-store visibility technologies. Below we discuss some of the modern and emerging technologies developed for the retail industry, categorized by the three main components of the store as illustrated in Fig. 5.1.

1. *Inventory Information.* We introduced RFID in Sect. 3. As the price of RFID tags decreases, attaching RFID tags to individual items (as opposed to cases or pallets) becomes increasingly feasible. The application of RFID technology has received strong interest among individual retailers, technology providers (e.g., Tyco Retail Solutions), trade journals (e.g., RFID Journal), and academics (e.g., the University of Arkansas Walton College's RFID Research Center). Waller et al. (2011) list a full 60 uses of RFID in apparel retail supply chains. Fisher and Raman (2010), who use RFID as a case study to illustrate the opportunities and risks inherent in new retail technology, call RFID "revolutionary." Beyond RFID, new crowdsourcing platforms such as Quri and Gigwalk enlist shoppers to report the status of inventory levels and displays via smartphone, offering retailers a true customer view of their store operations. Interestingly, these technologies also appear to be used by brand managers to monitor retailers' execution and adherence to the brand's promotion plans.
2. *Customer Flow Information.* Traffic counters—sensors that measure traffic in retail stores (e.g., ShopperTrak)—have become common in retail. Knowing how many potential customers are in the store at a time enables retailers to estimate conversion from traffic to sales and to match staffing with customer traffic.

Technologies are increasingly able to track customer movements within the store; for example, by detecting "pings" of customer cell phones (e.g., Euclid Analytics), by attaching RFID tags and mobile devices to shopping carts (e.g., MediaCart!), and by "seeing" customer bodies using infrared technology (e.g., Irisys). Video footage is increasingly analyzed by software to detect and record customer locations and customer engagement (e.g., SCOPIX Solutions, Envysion, RetailNext). By identifying highly trafficked areas of the store these technologies can assist with store layout decisions, and by measuring queue lengths and wait times they can inform queue management. Mobile devices also offer the opportunity for retailers to address individual customers as they shop with store maps, inventory information, and promotions (e.g., Apple's iBeacon).

3. *Store Associates Task Information.* Store associates increasingly carry mobile devices (i.e., smartphones and tablets) to communicate with each other, to give them real-time access to product, sales, and inventory information and to enable them to perform checkout, inspection, and replenishment functions (e.g., Motorola Retail 2008). Such devices offer the possibility of enhanced visibility to associates on the store floor in addition to management.

One possibility is that some of the estimation and inference problems reviewed in Sects. 2 and 3 may become less important as these visibility technologies become more reliable and inexpensive and retailers learn to make use of the information they provide. Nevertheless, we believe that new data sources will also inspire new research problems, and that visibility technologies and analytical methodologies may complement each other in many cases. For example, perhaps a retailer's response to demand censoring can be enhanced by using customer traffic data to make inferences about lost sales in the event of a stockout. Perhaps models of inventory record inaccuracy can be improved using information from an RFID reader that detects whether items are in the front- or backroom of a store. Ultimately, analytical methodologies form the link between new visibility technologies and better decisions. Below we suggest two broad categories of new analytical research opportunities in store operations that could complement the new visibility technologies.

New Insights from Combining Data Sources While it is common to simplify analytical operations management models by assuming a single location, SKU, or customer segment, we believe that there may be significant gains from leveraging data across stores and SKUs to impute missing in-store data. For example, as discussed in Sect. 2.4, sales data from multiple SKUs can be used to estimate substitution probabilities and to determine the optimal stocking policy for multiple SKUs. Another example was suggested in Sect. 3.4: data from other stores and SKUs may be used as benchmarks against which deviations can be detected for the purpose of process control. Given the large number of emerging visibility technologies listed above, there may also be significant value to considering multiple visibility technologies together; for example, recall from Sect. 3 that POS data can be used to make inferences on uncertain inventory levels. By modeling the interactions between different processes in a store, we believe that both better empirics and improved analytical decisions may be possible. To give two recent examples from the literature, Perdikaki et al. (2012) and Mani et al. (2015) use traffic counting and conversion data to measure the impact of labor staffing on sales performance, with clear implications on labor planning. Lu et al. (2013) use video data to measure queue lengths and thereby quantify the impact of queue lengths on customer purchase behavior, with clear implications on queue design and staff scheduling.

New Parameters and New Decisions As in any operational context, the parameters of an analytical model must be estimated before a model can be used for decision-making. Retail environments are especially complex and non-stationary, heightening the need for estimation. Though we have not attempted to review it in this chapter, there is a growing empirical literature gaining ever finer insights into retail operations from richer datasets using more sophisticated methodologies. The rise of new visibility technologies expands the set of operational parameters that can conceivably be estimated. As an example, customer tracking technologies, by identifying more and less trafficked locations in the store, potentially allow for more detailed, location-specific assortment planning. Furthermore, new technologies offer retail managers new levers in the store. To give just one example, new digital price tags (e.g., Altierre Corp.) and customized mobile phone offers (e.g., Retailigence's adPop) allow for dynamic pricing that can potentially depend on real-time traffic and inventory states.

In conclusion, we believe that the study of visibility in retail stores exemplifies the trend towards business analytics more generally. Inventory management with censored demand observations and record inaccuracy represent just two examples of what is possible. The interplay between information, technology, inventory optimization, customer behavior, and human resources suggest a range of fresh analytical questions that have the potential to make a real impact on practice. Our hope is that our surveys and discussion here encourage further research on these topics.

Acknowledgements The authors thank the editors and an anonymous reviewer for constructive comments that greatly improved the chapter. They thank Jan Davis, Nicole DeHoratius, Saravanan Kesavan, Marcelo Olivares, Ariel Schilkrut, and participants in the 2013 Consortium for Operational Excellence in Retailing annual conference and in the 2013 UNC Retail Conference for valuable discussions and input.

References

Agrawal, N., & Smith, S. (1996). Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics, 43*, 839–861.

Anupindi, R., Dada, M., & Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science, 17*(4), 406–423.

Atali, A., Lee, H., & Özer, Ö. (2011). *If the inventory manager knew: Value of visibility and RFID under imperfect inventory information*. Stanford University, Working Paper.

Avrahami, A., Tzimerman, A., Herer, Y. T., & Shtub, A. (2012). *The value of inventory accuracy in supply chain management*. Technion: Israel Institute of Technology, Working Paper.

Azoury, K. S. (1985). Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science, 31*(9), 1150–1160.

Bensoussan, A., Cakanildirim, M., Sethi, S. P., & Shi, R. (2011a). Calculation of approximate optimal policies in a partially observed inventory model with rain checks. *Automatica, 47*, 1589–1604.

Bensoussan, A., Cakanyildirim, M., Li, M., & Sethi, S. P. (2011b). *Inventory control with a cash register: Sales recorded but not demand or shrinkage*. University of Texas at Dallas, Working Paper.

Bensoussan, A., Cakanyildirim, M., Minjárez-Sosa, J. A., Sethi, S. P., & Shi, R. (2008). Partially observed inventory systems: The case of rain checks. *SIAM Journal on Control and Optimization, 47*(5), 2490–2519.

Bensoussan, A., Cakanyildirim, M., & Sethi, S. (2005). On the optimal control of partially observed inventory systems. *Comptes Rendus Mathematique, 341*, 419–426.

Bensoussan, A., Cakanyildirim, M., & Sethi, S. (2007a). A multi-period newsvendor problem with partially observed demand. *Mathematics of Operations Research, 32*(2), 322–344.

Bensoussan, A., Cakanyildirim, M., & Sethi, S. (2007b). Partially observed inventory systems: The case of zero balance walk. *SIAM Journal on Control and Optimization, 46*(1), 176–209.

Bensoussan, A., Cakanyildirim, M., & Sethi, S. (2009). A note on 'the censored newsvendor and the optimal acquisition of information'. *Operations Research, 57*(3), 791–794.

Besbes, O., & Muharremoglu, A. (2013). On implications of demand censoring in the newsvendor problem. *Management Science, 59*(6), 1407–1424.

Bisi, A., & Dada, M. (2007). Dynamic learning, pricing, and ordering by a censored newsvendor. *Naval Research Logistics, 54*, 448–461.

Bisi, A., Dada, M., & Tokdar, S. (2011). A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing Service Operations Management, 37*(11), 1390–1405.

Braden, D. J., & Freimer, M. (1991). Informational dynamics of censored observations. *Management Science, 37*(11), 1390–1405.

Brynjolfsson, E., Hu, Y. J., & Rahman, M. S. (2013). Competing in the age of omnichannel retailing. *MIT Sloan Management Review, 54*(4), 23–29.

Burnetas, A., & Smith, C. (2000). Adaptive ordering and pricing for perishable products. *Operations Research, 48*(3), 436–443.

Camdereli, A. Z., & Swaminathan, J. M. (2010). Misplaced inventory and radio-frequency identification (RFID) technology: Information and coordination. *Production Operations Management, 19*(1), 1–18.

Chen, L. (2010). Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Operations Research, 58*(2), 396–413.

Chen, L. (2013a). *Bayesian dynamic pricing with two-sided censored customer willingness-to-pay data*. Duke University, Working Paper.

Chen, L. (2013b). *Fixing phantom stockouts: Optimal data-driven shelf inspection policies*. Duke University, Working Paper.

Chen, L., & Plambeck, E. L. (2008). Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing Service Operations Management, 10*(2), 236–256.

Chuang, H., & Oliva, R. (2013). *Empirical modeling of inventory record audit policies*. Texas A&M University, Working Paper.

Chuang, H., Oliva, R., & Liu, S. (2012). *On-shelf availability, retail performance, and external audits: A field experiment*. Texas A&M University, Working Paper.

Conlon, C. T., & Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics, 5*(4), 1–30.

Conrad, S. A. (1976). Sales data and the estimation of demand. *Operational Research Quarterly, 27*(1), 123–127.

DeGroot, M. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

DeHoratius, N., Mersereau, A., & Schrage, L. (2008). Retail inventory management when records are inaccurate. *Manufacturing Service Operations Management, 10*(2), 257–277.

DeHoratius, N., & Raman, A. (2008). Inventory record inaccuracy: An empirical analysis. *Management Science, 54*(4), 627–641.

DeHoratius, N., & Ton, Z. (2015). The role of execution in managing product availability. In N. DeHoratius & Z. Ton (Eds.), *Retail supply chain management. Quantitative models and empirical studies*. New York: Springer.

Ding, X., Puterman, M., & Bisi, A. (2002). The censored newsvendor and the optimal acquisition of information. *Operations Research, 50*(3), 517–527.

Fisher, M., & Raman, A. (2010). *The new science of retailing*. Boston: Harvard Business Press.

Gaukler, G. M., Seifert, R. W., & Hausman, W. H. (2007). Item-level RFID in the retail supply chain. *Production and Operations Management, 16*(1), 65–76.

Godfrey, G., & Powell, W. (2001). An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science, 47*(8), 1101–1112.

Gruen, T. W., & Corsten, D. (2008). *A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry*. Arlington (VA): Food Marketing Institute.

Hardgrave, B. C., Aloysius, J. A., & Goyal, S. (2013). RFID-enabled visibility and retail inventory record inaccuracy: Experiments in the field. *Production and Operations Management, 22*(4), 843–856.

Harpaz, G., Lee, W., & Winkler, R. (1982). Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Science, 28*, 589–603.

Heese, H. S. (2007). Inventory record inaccuracy, double marginalization, and RFID adoption. *Production Operations Management, 16*(5), 542–553.

Heyman, D., & Sobel, M. (1984). *Stochastic models in operations research, volume II: Stochastic optimization*. New York: McGraw-Hill.

Huh, W. T., Levi, R., Rusmevichientong, P., & Orlin, J. B. (2011). Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Operations Research, 59*(4), 929–941.

Huh, W. T., Olvera-Cravioto, M., & Özer, Ö. (2010). *Joint audit and replenishment decisions for an inventory system with unrecorded demands*. Columbia University, Working Paper.

Huh, W. T., & Rusmevichientong, P. (2009). A non-parametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research, 34*(1), 103–123.

Iglehart, D. L., & Morey, R. C. (1972). Inventory systems with imperfect asset information. *Management Science, 18*(8), B338–B394.

Jain, A., Rudi, N., & Wang, T. (2014). Demand estimation and ordering under censoring: Stockout timing is (almost) all you need. *Operations Research* (forthcoming).

Kang, Y., & Gershwin, S. (2005). Information inaccuracy in inventory systems: Stock loss and stockout. *IIE Transactions, 37*, 843–859.

Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481.

Kök, A. G., & Shang, K. (2007). Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing Service Operations Management, 9*(2), 185–205.

Kök, A. G., & Shang, K. (2014). Evaluation of cycle-count policies for supply chains with inventory inaccuracy and implications on RFID investments. *European Journal of Operational Research, 237*(1), 91–105 (Note: This reference was formerly titled "Evaluation of supply chains with inventory record inaccuracy and implications on RFID investments").

Lariviere, M., & Porteus, E. (1999). Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science, 45*(3), 346–363.

Lee, H., & Özer, Ö. (2007). Unlocking the value of RFID. *Production and Operations Management, 16*(1), 40–64.

Lovejoy, W. (1990). Myopic policies for some inventory models with uncertain demand distributions. *Management Science, 36*(6), 724–738.

Lu, X., Song, J.-S., & Zhu, K. (2005). On 'the censored newsvendor and the optimal acquisition of information'. *Operations Research, 53*(6), 1024–1027.

Lu, X., Song, J.-S., & Zhu, K. (2007). *Inventory control with unobservable lost sales and Bayesian updates*. Duke University, Working Paper.

Lu, X., Song, J.-S., & Zhu, K. (2008). Analysis of perishable-inventory systems with censored demand data. *Operations Research, 56*(4), 1034–1038.

Lu, Y., Musalem, A., Olivares, M., & Schilkrut, A. (2013). Measuring the effect of queues on customer purchases. *Management Science, 59*(8), 1743–1763.

Mani, V., Kesavan, S., & Swaminathan, J. M. (2014). Estimating the impact of understaffing on sales and profitability in retail stores. *Production and Operations Management* (forthcoming).

McKinsey Global Institute. (2011). Big data: The next frontier for innovation, competition and productivity. Accessed April 9, 2013, from www.mckinsey.com/mgi

Mersereau, A. J. (2013). Information-sensitive replenishment when inventory records are inaccurate. *Production and Operations Management, 22*(4), 792–810.

Mersereau, A. J. (2014). Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing and Service Operations Management* (forthcoming).

Morey, R. C. (1985). Estimating service level impacts from changes in cycle count, buffer stock, or corrective action. *Journal of Operations Management, 5*(4), 411–418.

Morey, R. C., & Dittman, D. A. (1986). Optimal timing of account audits in internal control. *Management Science, 32*(3), 272–282.

Motorola Retail. (2008). *The next revolution in retail technology*. Accessed July 23, 2013, from www.motorolasolutions.com

Musalem, A., Olivares, M., Bradlow, E. T., Terwiesch, C., & Corsten, D. (2010). Structural estimation of the effect of out-of-stocks. *Management Science, 56*(7), 1180–1197.

Nahmias, S. (1994). Demand estimation in lost sales inventory systems. *Naval Research Logistics, 41*, 739–757.

Papadimitriou, C., & Tsitsiklis, J. N. (1987). The complexity of Markov decision processes. *Mathematics of Operations Research, 12*, 441–450.

Perdikaki, O., Kesavan, S., & Swaminathan, J. M. (2012). Effect of traffic on sales and conversion rates of retail stores. *Manufacturing Service Operations Management, 14*(1), 145–162.

Powell, W., Ruszczynski, A., & Topaloglu, H. (2004). Learning algorithms for separable approximations of discrete stochastic optimization problems. *Mathematics of Operations Research, 29*(4), 814–836.

Rekik, Y., Sahin, E., & Dallery, Y. (2008). Analysis of the impact of RFID technology on reducing product misplacement errors at retail stores. *International Journal of Production Economics, 112*(1), 264–278.

Rekik, Y., Sahin, E., & Dallery, Y. (2009). Inventory inaccuracy in retail stores due to theft: An analysis of the benefits of RFID. *International Journal of Production Economics, 118*, 189–198.

Sahin, E., & Dallery, Y. (2009). Assessing the impact of inventory inaccuracies within a newsvendor framework. *European Journal of Operational Research, 197*, 1108–1118.

Scarf, H. (1959). Bayes solution of the statistical inventory problem. *Annals of Mathematical Statistics, 30*, 490–508.

Scarf, H. (1960). Some remarks on bayes solutions to the inventory problem. *Naval Research Logistics, 7*, 591–596.

Silver, E. (1993). Bayesian updating of an arbitrary discrete distribution under a special case of partial information. *Communications in Statistics Stochastic Models, 9*(4), 615–638.

Song, J.-S. (1994). Leadtime uncertainty in a simple stochastic inventory model. *Management Science, 40*(5), 603–613.

Vulcano, G., van Ryzin, G., & Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Operations Research, 60*(2), 313–334.

Waller, M. A., Cromhout, D. B., Patton, J. B., Williams, B. D., & Hardgrave, B. C. (2011). *An empirical study of potential uses of RFID in the apparel retail supply chain*. University of Arkansas Information Technology Research Institute, Working Paper ITRI-WP156-0111.

Wecker, W. E. (1978). Predicting demand from sales data in the presence of stockouts. *Management Science, 24*(10), 1043–1054.

Yano, C. A., & Lee, H. L. (1995). Lot sizing with random yields: A review. *Operations Research, 43*(2), 311–334.

Zipkin, P. (2000). *Foundations of inventory management*. Boston: McGraw-Hill.