



Mathematical modeling and Bayesian estimation for error-prone retail shelf audits

Howard Hao-Chun Chuang *

College of Commerce, National Chengchi University, Taipei 11605, Taiwan



ARTICLE INFO

Article history:

Received 5 August 2014

Received in revised form 5 October 2015

Accepted 5 October 2015

Available online 22 October 2015

Keywords:

Retail operations

Audit services

Inspection error

Risk aversion

Bayesian inference

ABSTRACT

Prevalent execution errors such as out-of-stock, inventory record inaccuracy, and product misplacement jeopardize retail performance by causing low on-shelf availability, which discourages not only retailers who have lost sales but also manufacturers who have worked hard to deliver goods into retail stores. Thus, external service companies are hired by manufacturers to conduct manual inspection regularly. Motivated by the practical need of shelf audit service providers, we use a general cost structure to develop a decision support model for periodic inspection. Some qualitative insights about the intricate relationships among inspection efficacy, cost factors, failure rate of shelf inventory integrity, and optimal decisions are derived from analytics assuming risk-neutrality. From simulation experiments we also find that managers' risk preferences have non-trivial impacts on optimal decisions. Based on a total cost standpoint high-quality inspection is predominantly preferred regardless of the level of risk aversion. Finally, we propose a Bayesian statistical model and a Markov chain Monte Carlo approach to estimate model parameters such that managers can make empirically informed decisions. Our major contribution lies in developing a mathematical model that is practically applicable and proposing a Bayesian estimation approach to rationalize unobservable model parameters, which are influential to optimal decisions but often arbitrarily assumed by decision makers.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Retail operations is composed of various tasks pertaining to assortment planning, product pricing, inventory optimization, and store execution [20]. Among those tasks, store execution is highly labor-extensive and complicated because it involves people, processes, and technology. Thus, execution errors such as shelf out-of-stock (OOS), inventory record inaccuracy (IRI), and product misplacement have become norms rather than anomalies even at financially successful retailers [47]. Store execution errors jeopardize retail performance by resulting in low on-shelf availability, which discourages not only retailers who have lost sales but also other supply chain members who have worked hard to deliver goods into the retail outlet. Being well-known for its operational excellence, Walmart recently admitted to a low on-shelf availability issue and predicted a \$3 billion opportunity in filling in empty shelves [13].

Facing prevalent issues pertaining to on-shelf availability, retailers have gradually seen the need of allocating extra labor capacity to carry out shelf audits in order to reach higher service levels [17]. However, hiring more employees who are able to execute prescribed tasks and fix shelf errors goes against the common practice in retailing to minimize labor cost [22,46]. Since low on-shelf availability is a serious problem for retailers as well as manufacturers [41], manufacturers search for

alternative approaches (rather than retailers' regular operations) to maximize the availability of their products [6,15].

A potential answer for manufacturers to fix low on-shelf availability is to ask external companies who provide shelf audit services to correct faulty items that may experience OOS, IRI, or misplacement [10]. Those external service agents are capable of working with different store formats (e.g., grocery, club, drug, convenience). Their associates excel in re-shelving or display maintenance to complement retailers' regular operations, and conduct other tasks such as placing promotional goods. Moreover, periodic shelf audits performed by those third party companies are appealing to manufacturers because they also solve the conflict of retailers' potential bias to selectively report good audit outcomes [18]. Chuang et al. [10] report a successful case in which they conduct a field experiment in a U.S. retail chain and show that external audit services is a cost-effective way for product manufacturers to improve on-shelf availability.

Even though external shelf audits seem to be a promising solution to the recurring problem of low on-shelf availability, designing a cost optimal inspection policy for those external service companies turns out to be difficult because of two issues. First, information regarding inventory transactions may not be available to the service companies who have limited/no access to point-of-sale (POS) data. Second, it is nearly impossible to achieve error-free shelf inspection because an ordinary associate usually has to audit multiple items at multiple stores within a limited time. As a result, a certain amount of inspection error is inevitable and needs to be considered by decision makers. In attempt to tackle the

* Tel.: +886 2 29393091x88147; fax: +886 2 29393754.
E-mail address: chuang@nccu.edu.tw.

forementioned issues, this paper presents a periodic inspection policy that triggers physical audits to increase on-shelf availability. We present a normative analysis of inspection decisions while taking into account inspectors' fallibility and managers' risk attitudes. Our paper addresses the question: *for external service providers with limited information about on-shelf items' status, what is the optimal frequency of shelf audits provided a level of inspection error and risk aversion?* We answer the question by deriving static analytics under risk neutrality and performing simulation studies under risk aversion.

The notion of *inspection error* and *risk aversion* is critical to our inspection policy design and makes our modeling effort relevant. On the one hand, as opposed to the commonly assumed "perfect inspection," we posit that any inspection in the real world can hardly be error-free. The reality is that inspection errors vary with human efforts and significantly increase the level of complexity surrounding the design of inspection policies [30]. Since the competencies, experiences, and motivations of inspectors are different, the probability of making mistakes will differ [5]. However, studies on the impact of error-prone inspection are scant in the context of retail shelf audits. We fill in the gap by explicitly incorporating human fallibility into our model and assessing the impact of different levels of inspection error rates.

On the other hand, numerous studies on inspection policies assume risk neutrality, which is valid only if optimal decisions are invariant with managers' risk attitudes [37]. Unfortunately, most of the earlier attempts (e.g., [28,36]) to optimize inspection decisions have not taken into account managers' risk preferences. Peecher et al. [42] point out that audit initiatives are by no means risk-free and there are different elements of risk – internal risk, control risk, and detection risk – surrounding inspection policy design. Those elements of risk lead to uncertainties in total cost of shelf audit efforts. Seeing that optimal decisions will depend on the degree of risk aversion, we take a utility-based approach to analyze how risk aversion affects the design of inspection policies. The principle of maximizing expected utility has a rich theoretical foundation [11] that enables us to explore the interaction between risk preferences and optimal decisions.

Our study makes several contributions. First, our model has a fairly general cost structure and it is built upon realistic assumptions of inspection efficacy and managerial risk preferences. Managers can adopt the proposed model to achieve cost-effective inspection and recover profit loss caused by low on-shelf availability. Our modeling effort is particularly relevant for retail service providers who need to periodically send associates into retail stores to maintain shelf inventory integrity. Second, our model considers imperfect inspection and accommodates two types of errors – the error of failing to correct faulty items and the error of miscorrecting accurate ones. Further, we assess managers' risk preferences that are found to have substantive impacts on optimal decisions. We observe that from a cost standpoint high-quality inspection (i.e., low error probability) is generally preferred regardless of the degree of risk aversion. Third, our model also captures the random degradation of on-shelf availability due to store execution errors. We find interesting dynamics among inspection efficacy, failure rate of shelf inventory, and cost factors. Our analysis shows that the ignorance of imperfect inspection and random shelf error generation would result in suboptimal audit decisions. Lastly, early papers make hypothetical assumptions about the distribution of inspection error [5,14] because no observable data can be applied to directly estimate the error distribution in a non-experimental context. We address the issue by making Bayesian inference about the level of inspection error instead of making hypothetical guesses. We adopt Bayesian hierarchical modeling and use a Metropolis-within-Gibbs sampling scheme to statistically infer unobservable human errors given observed inspection outcomes. To the best of our knowledge, there is no similar attempt reported in the literature of shelf inspection and inventory audits.

The rest of this article is organized as follows. Section 2 summarizes the relevant literature related to our work; the formulation and analysis of a periodic inspection model for shelf audit service providers under

risk-neutrality and risk-aversion are presented in section 3 and section 4 respectively. In section 5 we propose a Bayesian methodology to estimate unknown model parameters that are crucial for optimal decisions. We conclude by articulating practical implications and research limitations.

2. Related literature

A stream of literature has engaged in developing decision support models for retail shelf audits. One of the seminal studies is by Hughes [28] who formulates a Markov decision process to determine the optimal timing of audits while considering the efficacy of auditing. Morey and Dittman [36] further propose a model to calculate the optimal timing of stock audits based on pre-specified goals of inventory accuracy. More recently, Sandoh and Shimamoto [45] devise a stochastic model to find the optimal frequency of inventory counting in a supermarket. Kok and Shang [33] propose a joint inventory inspection and replenishment policy that is capable of recovering a large proportion of benefits brought by RFID adoption. DeHoratius et al. [12] develop a shelf inspection policy based on expected value of perfect information. Atali et al. [3] also work on the problem of inventory integrity within periodic review inventory systems. Our model differs from previous studies in two major aspects. First, neither sales quantity nor inventory position is known to decision makers (i.e., managers of external service firms) who typically have limited observations on on-shelf items from periodic inspection. Second, we explicitly incorporate inspection efficacy and risk preferences into inspection policy design.

Extant studies on inspection assume risk neutrality, an assumption that is not likely to be valid in our context of retail shelf audits. Peecher et al. [42] define audit risk as the product of three underlying risks: inherent risk, control risk, and detection risk. Here *inherent risk* refers to the fact that on-shelf availability could easily be compromised due to various execution errors, which are likely to persist without internal controls [43]. However, imposing internal controls (e.g., periodic inspection) has *control risk* that is related to two cost factors – a cost of inspecting/correcting faulty items and a cost of leaving faulty items unfixed. Thus, control risk involves optimizing inspection decisions to minimize the sum of those costs. Lastly, *detection risk* refers to the fact that human inspectors are not able to detect and fix all errors. More often than not, inspectors contaminate inventory data as "large errors often remain in the stock records because of inaccuracies in the counting procedure" [29].

The three types of risks found in retail shelf audits shed light on the need for incorporating risk aversion into decision support models. However, most of the models discussed above focus on mitigating *inherent* and *control* risks without explicitly examining *detection* risk. As opposed to the commonly assumed "perfect inspection" in retail operations research [33], we posit that any inspection in the real world can hardly be error-free. The reality is that inspection efficacy varies with human efforts and significantly increases the level of complexity surrounding the inspection policy design. Since the competencies, experiences, and motivations of individual auditor differ, the probability of their making inspection errors will differ [5]. The impact of error-prone inspection has been widely studied in a manufacturing environment [14,48]. That said, studies on the impact of auditor error are scant in the context of retailing. We fill in the gap by formally analyzing the costs and benefits of different levels of inspection efficacy.

Aside from the above-mentioned studies on designing cost-minimization inspection policies, our paper is related to studies that apply statistical process control (e.g., [21,25]) or acceptance sampling (e.g., [16]; [19]) approaches to improve inventory integrity. However, most of the statistical approaches require actual and/or recorded inventory levels that are not available in our setting. Moreover, with respect to unobservable inspection efficacy, early papers make hypothetical assumptions about the distribution of inspection error (e.g., [5,14]) because no observations can be used to estimate the distributions directly in a non-experimental context. We address this limitation by

developing a Bayesian estimation method model to estimate the distribution of inspection error using limited data from periodic shelf audits. Bayesian inference has been adopted by management scientists to infer demand parameters [26] and inventory level [12] in order to improve replenishment decisions. However, Bayesian hierarchical model is rarely applied due to its computational complexities, which are less of an issue nowadays due to recent advances in Monte Carlo methods [7]. Our paper stands for a compelling example of using Bayesian methods to devise a statistically grounded model. This venue is promising as the complementarity between Bayesian statistics and decision analysis is instrumental in improving decision-making with consideration to risk attitudes [11].

3. Model formulation and static optimization

We develop a cost-minimization model for external service companies who provide shelf audit services on a periodic basis (i.e., every τ days). As explained in the introduction, our decision support model is grounded on the reality faced by audit service providers (i.e., limited/unavailable inventory information and error-prone inspection). Thus, our model does not consider inventory dynamics that are entirely unknown to decision makers. Under the periodic inspection scheme, managers have to determine the optimal frequency (τ^*) of sending inspection associates into the retail store.

3.1. Model formulation

We begin with a simple discussion of the degradation in inventory integrity and the dynamics attributed to inspection frequency (τ). Let F_{j+1}^b (where the superscript b means “before”) denote the number of faulty items before the $(j + 1)$ th inspection. F_{j+1}^b is modeled as a sum of two factors:

$$F_{j+1}^b = F_j^a + D_j \quad (1)$$

The first factor in the right-hand-side (RHS) of Eq. (1) – F_j^a (where the superscript a means “after”) – is the number of faulty items remaining faulty after the j th inspection. We model the faulty items after an imperfect inventory audit as:

$$F_j^a = F_j^b - K_j(F_j^b, \alpha) + M_j(n - F_j^b, \beta) \quad (2)$$

where α denotes the probability that an inspector can identify an inaccurate item as defective and correct it, and β denotes the probability that the inspector erroneously modifies a non-faulty item. On the one hand, shelf errors fail to be fixed with probability $1 - \alpha$ due to imperfect inspection. Thus, the number of “properly corrected” items is a random variable $K_j \sim \text{binomial}(F_j^b, \alpha)$. On the other hand, accurate items are mistakenly modified with probability β due to careless inspection. So the number of miscorrected items is a random variable $M_j \sim \text{binomial}(n - F_j^b, \beta)$, where n is the total number of items to be inspected. We posit that a perfect inspection (i.e., $\alpha = 1$ and $\beta = 0$) is next to impossible when there are numerous on-shelf items to be inspected.

The second factor D_j in the RHS of Eq. (1) denotes the number of items that falls into inaccurate status between the j th and $(j + 1)$ th inspection. D_j accounts for the fact that the $(n - F_j^a)$ correct items may degrade by the time of the next inspection due to random execution errors.

$$D_j \sim \text{binomial}(n - F_j^a, P(\tau)) \quad (3)$$

where the parameter $P(\tau)$ denotes the probability that an item without error turns faulty between an inspection cycle of τ days. Following previous studies (e.g., [45]), we model the probability of an item turning faulty as time-dependent and following the exponential failure

distribution – $P(\tau) = P(T < \tau) = 1 - e^{-\lambda\tau}$ where T is a random variable denoting time to degrade and λ is a parameter denoting instantaneous failure rate. This formulation is commonly used (e.g., see [1]) as it provides a parsimonious way to characterize the likelihood of shelf error generation. A larger/smaller value of λ indicates that retailers have lower/higher store execution quality. The probability ($P(\tau)$) that an item will turn faulty increases with τ and acts as a key input to generate the binomial random variable D_j . Eqs. (1), (2), and (3) jointly govern the change in the number of faulty items before and after each inspection in the system.

We follow O'Reagan [40] who proposes a cost structure of any error detection program:

$$\begin{aligned} \text{Total cost} &= \text{Inspection costs} + \text{Correction costs} \\ &\quad + \text{Uncorrected error costs} \end{aligned}$$

The cost structure is general and applicable to characterize cost of shelf audits. Specifically, we assume the inspection costs to grow linearly with the number of faulty items (n),

$$c_i \times n \quad (4)$$

where c_i denotes inspection cost per inventory item. While inspection costs might grow with n in a non-linear fashion, for simplicity we adopt the widely used linear cost function (e.g., [33,34,40,48,49]). The linear specification is also consistent with internal accounting of an audit service provider whom we work with.

In addition to the time and efforts spent on inspecting on-shelf items, costs are incurred by correcting the physical/information status that the inspector finds erroneous. The correction costs come from both proper and improper correction. As mentioned earlier, the proper portion is incurred by the corrected number K_j and inspection efficacy (α). The improper portion is caused by the miscorrected number M_j and the rate of introducing errors (β). The correction costs are:

$$c_c \times [K_j(F_j^b, \alpha) + M_j(n - F_j^b, \beta)] \quad (5)$$

where c_c denotes correction cost per item.

The last piece of total costs is associated with the potential negative impact of unfixed shelf errors. The penalty is composed of two parts. The first part arises from F_j^a denoting the number of inaccurate items that stay faulty after the j th inspection. F_j^a is the sum of “truly” inaccurate items that inspectors are not able to fix (i.e., $F_j^b - K_j(F_j^b, \alpha)$) and “false” inaccurate items that are miscorrected (i.e., $M_j(n - F_j^b, \beta)$). Under a periodic inspection scheme, we assume that the correction of F_j^a items may happen no earlier than the next (i.e., $(j + 1)$ th) inspection such that the penalty is proportional to τ , which accounts for the elapsed time since previous inspection. The second part is attributed to D_j , which denotes accurate items turning faulty between the j th and the $(j + 1)$ th inspection. Since $D_j \sim \text{binomial}(n - F_j^a, P(\tau))$ and $P(\tau)$ is the CDF of the exponential failure distribution, we can further derive the average time of being faulty for the D_j items.

Proposition 1. Under the assumed exponential failure process, on average the newly degraded items D_j have been inaccurate for $\tau + \frac{\tau}{e^{\lambda\tau} - 1} - \frac{1}{\lambda}$ days.

Proof. Please see Appendix A.

While the exact time of being faulty for each of the D_j items is subject to random failure processes and cannot be known with certainty, the expected time derived in Proposition 1 serves as a reasonable approximation for our penalty accounting. Taken together, the uncorrected error costs that penalize poor inspections are:

$$c_u \times \left[F_j^a \tau + D_j \left(\tau + \frac{\tau}{e^{\lambda\tau} - 1} - \frac{1}{\lambda} \right) \right] \quad (6)$$

where we assume that an average cost of being inaccurate per item per day is c_u , which accounts for the economic impact of leaving errors unfixed. Since in this model we only track the binary status of an item (i.e., with or without errors), we consider the costs to be linear with respect to the number of item-days for simplicity [34].

Eqs. (4), (5), and (6) constitute the total cost within an inspection cycle of τ days. Eq. (7) illustrates the corresponding average daily cost $f(\tau)$. A key challenge for managers who aim to minimize $f(\tau)$ is to keep a balance between inspection/correction costs and the cost of leaving uncorrected shelf errors. The total daily cost in the j th cycle of a periodic shelf audit is:

$$f_j(\tau) = \frac{c_i \times n + c_c [K_j(F_j^b, \alpha) + M_j(n - F_j^b, \beta)] + c_u \times [F_j^q \tau + D_j \left(\tau + \frac{\tau}{e^{\lambda\tau} - 1} - \frac{1}{\lambda} \right)]}{\tau} \quad (7)$$

This function is calculated based on realizations of random variables and returns a real number. Note that the two cost parameters c_i and c_c can be estimated from the employee payroll and measured inspection and correction standards adopted by managers. Estimating c_u , however, is more challenging because the extra stockout, shrinkage, or spoilage costs induced by shelf errors are usually not observed. The audit service providers need to work with product manufacturers to arrive at an estimate for c_u . A rudimentary approach to calculate c_u is based on the expected daily margin. While the estimate may not be perfect, experienced managers should be able to derive a reasonable range (rather than a precise point estimate) for each cost element. A range of plausible values for cost factors should be sufficient for sensitivity analysis of optimal decisions and practical use of the model.

Given our assumption of fixed inspection interval (τ), time-invariant failure rate (λ), and stable inspection efficacy (α and β), the distribution of F_j^b soon converges to a steady state. That is, for any combination of feasible model parameters there is a number of items for which the expected number of faulty items introduced within an inspection interval is equal to the expected number of faulty items fixed. Since we model inspection efficacy and cost factors as fixed parameters, index j can be dropped when we substitute the steady state form of F^b and F^a into Eq. (1) and focus on the equilibrium condition.

Proposition 2. In steady state, the expected number of faulty items before inspection is

$$E[F^b] = \frac{n(\beta + P(\tau) - \beta P(\tau))}{\alpha + \beta + P(\tau) - (\alpha + \beta)P(\tau)}$$

Proof. Please see Appendix A.

From Proposition 2, it is clear that the expected number of faulty items $E[F^b]$ before inspection is a function of inspection efficacy (α and β), number of items to be inspected (n), and the decay probability of on-shelf items ($P(\tau)$).

Proposition 3. $E[F^b]$ is non-increasing in α and non-decreasing in β , τ , and λ .

Proof. Please see Appendix A.

After verifying that $E[F^b]$ is well-behaved (according to Proposition 3), we replace $E[F^b]$ into the expectation of Eq. (7) and yield the cost function to be minimized

$$E[f(\tau)] = \frac{c_i n + c_c \frac{n\alpha[P(\tau) + 2\beta - 2P(\tau)\beta]}{\alpha + \beta - P(\tau)(\alpha + \beta - 1)} + c_u n \left[\tau + \frac{\alpha(P(\tau)k - \tau)}{\alpha + \beta - P(\tau)(\alpha + \beta - 1)} \right]}{\tau} \quad (8)$$

where $k = \tau + \frac{\tau}{e^{\lambda\tau} - 1} - \frac{1}{\lambda}$ and $P(\tau)$ is the exponential cumulative density function.

The functional form of $E[f(\tau)]$ is analytically intractable but numerically solvable. We use a one-dimensional optimization routine that searches over the positive real line to find a τ^* that minimizes the total expected cost (Eq. 8). Specifically,

$$\tau^* = \lceil \arg \min_{\tau} E[f(\tau)] \rceil \quad (9)$$

We apply the ceiling function to obtain an integer τ^* because in practice those service companies can only trigger external shelf audits on a discrete-time basis. In order to better understand the dynamics among α , β , λ , and τ^* , we perform an extensive numerical study in the next section.

3.2. Numerical study

We set the number of items to 500 and the failure rate $\lambda = 0.017$ /day. Later on in Section 5 we will show how to jointly estimate the unobservable λ , α , and β using a Bayesian hierarchical modeling approach. Following O'Regan [40], we set $c_i = 0.05$, $c_c = 0.005$, and further define $\gamma = c_u/c_i$ for ease of comparison. A full table of notation for model variables and parameters can be found in Appendix B.

Fig. 1 presents the optimized inspection frequency (τ^*) and cost ($E[f(\tau^*)]$) under various levels of inspection efficacy and $\gamma = 1$. Some points are noteworthy. First, the optimal frequency of inspection is strongly dependent on the accuracy of auditing. From the left panel of Fig. 1 we see that τ^* tends to increase with α . That is, the optimal inspections would be less frequent (i.e., higher τ^*) when the efficacy of inspection increases (i.e., higher α). Second, τ^* also tends to increase with β but the implications are different. Given the high probabilities of introducing unnecessary errors (i.e., higher β) after each inspection, aggressive inspection can be harmful rather than helpful because high inspection frequency results in higher costs than what would be obtained through correction. Lastly, the right panel of Fig. 1 shows that $E[f(\tau^*)]$ decreases with α and increases with β monotonically. Perfect inspection (i.e., $\alpha = 1$ and $\beta = 0$) results in the lowest daily cost. We also observe that the most frequent inspection ($\alpha = 0.6$ and $\beta = 0$) does not lead to the minimal cost since there is wasted effort given the error-prone inspection process.

Fig. 2 presents the optimal frequency and costs given $\gamma = 3$. The values of τ^* illustrated in the left panel are lower than those shown in Fig. 1, which makes sense as managers would prefer relatively frequent inspections when shelf errors become more costly (i.e., a higher γ). As shown in the right panel of Fig. 2, β has substantial impacts on total cost since introducing unnecessary errors undermines cost efficiency. In addition, $E[f(\tau^*)]$ under poor inspection (e.g., $\alpha = 0.6$ and $\beta = 0.4$) is several times higher than the high-accuracy scenario. So, reducing inspection errors is even more valuable provided higher uncorrected error costs.

Given a fixed λ , the variations of τ^* with α/β shown in Figs. 1 and 2 are as expected and somewhat intuitive. Nonetheless, the failure rate parameter (λ) moderates the solution behaviors of τ^* under different levels of cost and inspection efficacy. Fig. 3 shows the optimal frequency and costs given $\gamma = 1$ and λ in $[0.1, 0.5]$. The left panel of Fig. 3 indicates that when inspection efficacy is high ($\alpha = 1$ and $\beta = 0$) or moderate ($\alpha = 0.8$ and $\beta = 0.2$), τ^* tends to decrease with λ . That is, when store execution quality degrades (i.e., λ increases), inspection associates should take more aggressive initiatives (i.e., a smaller τ^*). However, τ^* stays constant when λ is even larger because more frequent inspections (a smaller τ^*) are too costly given low costs of not fixing errors ($\gamma = 1$). Interestingly, when inspection efficacy is low ($\alpha = 0.6$ and $\beta = 0.4$), τ^* first decreases with and then increases with λ . Hence, under poor store execution ($\lambda > 0.35$), managers would reduce inspection frequency (a larger τ^*) to avoid side effects of low inspection efficacy

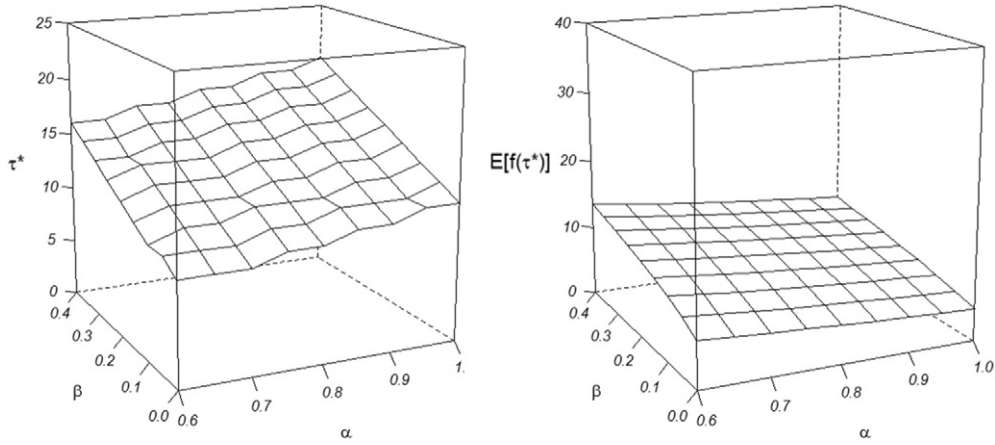


Fig. 1. τ^* and $E[f(\tau^*)]$ given $\gamma = 1$ & $\lambda = 0.017$.

(e.g., introducing more errors). The right panel of Fig. 3 illustrates $E[f(\tau^*)]$ under different levels of λ and inspection efficacy. Not surprisingly, $E[f(\tau^*)]$ increases with λ and the differences among $E[f(\tau^*)]$ caused by inspection efficacy decrease with λ , suggesting that the λ has more dominant impacts on costs when the retail store has a high intrinsic failure rate (e.g., more prevalent execution errors).

Fig. 4 presents τ^* and $E[f(\tau^*)]$ given $\gamma = 3$ and λ in $[0.1, 0.5]$. Two major observations are made from the left panel of Fig. 4. First, due to the higher cost of leaving faulty items unfixed (i.e., $\gamma = 3$), managers need to trigger shelf audits more frequently – overall τ^* becomes smaller. Second, even though low inspection efficacy ($\alpha = 0.6$ and $\beta = 0.4$) still results in relatively infrequent inspection like the $\gamma = 1$ case (Fig. 3), τ^* does not increase with λ in this $\gamma = 3$ scenario where managers cannot bear with costs incurred by not fixing shelf errors and need to be more aggressive (i.e., maintaining a small τ^*) even under poor store execution ($\lambda > 0.35$). The right panel of Fig. 4, similarly, indicates that $E[f(\tau^*)]$ increases with λ and becomes higher than under low inspection efficacy ($\alpha = 0.6$ and $\beta = 0.4$).

Figs. 3 and 4 reveal intricate dynamics among inspection efficacy (α , β), failure rate (λ), and cost ratio (γ). Apparently, ignoring human inspection error and random shelf error generation between inspections would lead to suboptimal decisions. Unlike the observable cost ratio (γ), the other three important parameters in our decision support model – α , β , and λ – are unobservable. In Section 5 we will revisit the key parameters and develop a Bayesian estimation methodology for external service companies who have limited knowledge about the n items to audit and obtain inspection reports every τ days only.

4. Monte Carlo simulation for risk analysis

4.1. Stochastic efficiency with respect to a function

Although useful in helping us better understand the dynamics between model parameters and optimal decisions, the foregoing analysis assumes risk-neutrality where the decisions just follow the expected cost. However, Jensen's inequality [11] indicates $E[U(f(\tau))] \leq U(E[f(\tau)])$ for a risk-averse decision maker who has a concave utility function U and faces stochastic costs $f(\tau)$. Since random shelf and inspection errors lead to variability (i.e., risk) in total cost $f(\tau)$, the optimal audit frequency may change with risk preferences [4]. We assess the impact of risk aversion on τ^* through stochastic efficiency with respect to a function (SERF) [35]. SERF is rooted in subjective expected utility theory and orders a set of risky alternatives in terms of certainty equivalent (CE) for a specified range of attitudes to risk [24]. Moreover, SERF does not require a prior distributional assumption on CE. Here, the risky choice is about selecting a τ that minimizes CE as we are considering cost [32].

Since the stochastic cost $f(\tau)$ in Eq. (7) is analytically intractable, we use Monte Carlo simulation to investigate the sample paths of $f(\tau)$ and feed the simulated $f(\tau)$ into a utility function that is monotonically decreasing in $f(\tau)$ and exhibits concavity within the risk aversion bounds. We adopt an exponential utility function $U(C) = -\exp(-C/r_a)$, where C is the monetary cost and r_a is the coefficient of absolute risk aversion ($r_a = 0$ if risk neutral) [37]. The exponential utility function belongs to the class of utility functions with constant absolute risk aversion (CARA), and is appealing in our case because the cardinal coefficient r_a

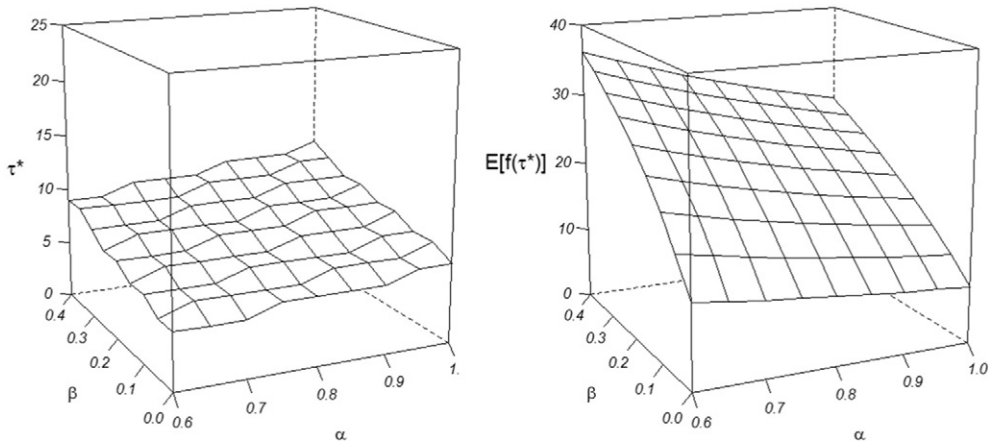
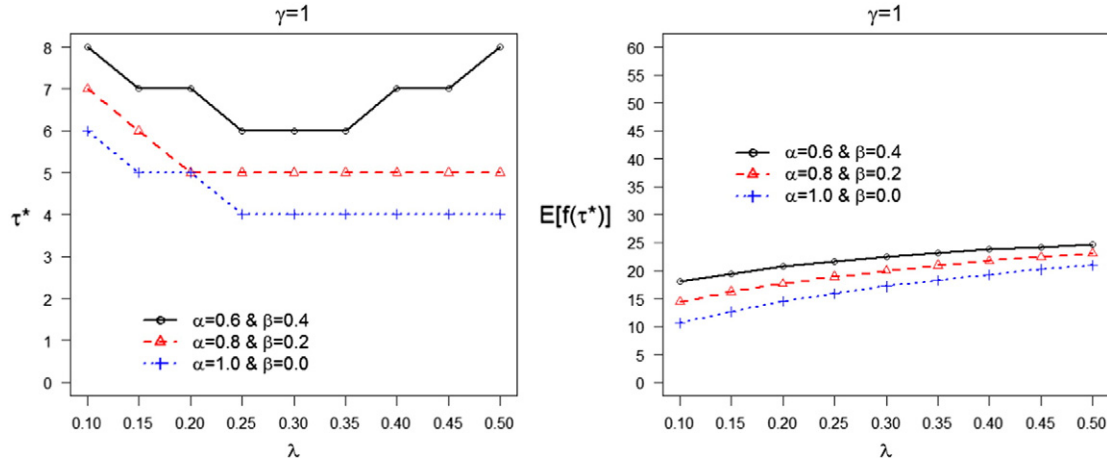


Fig. 2. τ^* and $E[f(\tau^*)]$ given $\gamma = 3$ & $\lambda = 0.017$.

Fig. 3. Impacts of λ on τ^* and $E[f(\tau^*)]$ given $\gamma = 1$.

gives an effective measure of risk aversion. The expected utility $E[U]$ is calculated as:

$$E[U(C, r_a)] \approx \sum_{i=1}^m U(C_i, r) P(C_i) \quad (10)$$

To simplify the computation, we use a discrete approximation to $E[U]$ with m replications, where each run i has the same probability $P(C_i)$ in the Monte Carlo simulation [35]. After 1000 runs we elicit $E[U]$ and convert it into CE to find a τ that minimizes $CE(\tau)$ through numerical search. For the sake of variance reduction, we replicate the computation 100 times and take the average of the 100 optimized τ_s to obtain the final τ^* in a particular scenario.

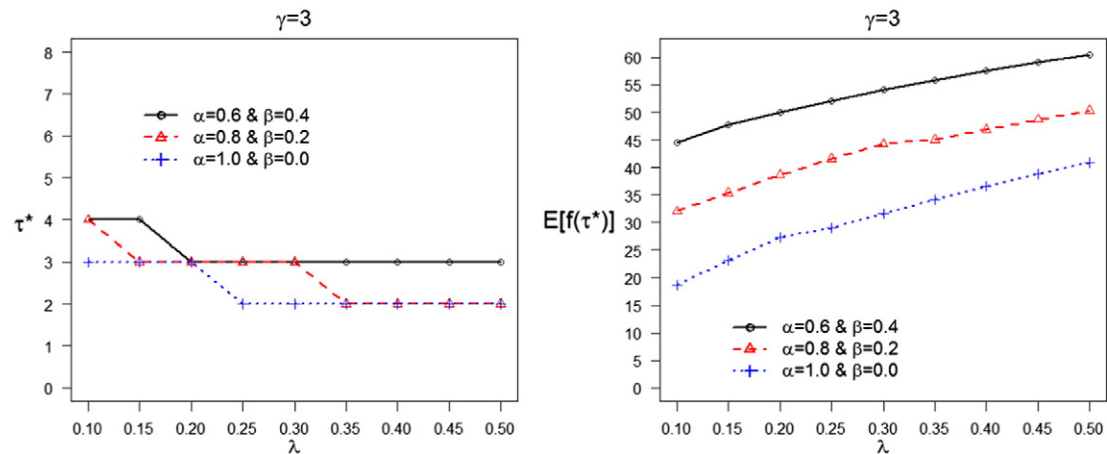
We derive the functional form of CE, $\log(-E[U])/r_a$, using the property: $CE(C, r) = U^{-1}(C, r)$ [35]. Although CE minimization is equivalent to $E[U]$ maximization, the CE is expressed in monetary terms and thus much easier to interpret than the utility. If CE is known for different risky alternatives (i.e., inspection frequencies), it is easy to make a choice and estimate the risk premium, which is the difference between the expected cost under risk-neutrality and the CE under risk-aversion. Here the most preferred alternative is the one resulting in the lowest CE.

4.2. Simulation experiment

We program the model and perform Monte Carlo simulation using R. Fig. 5 exhibits optimal decisions (τ^*) and CE given $\gamma = 3$ and $\beta = 0.2$. The left panel of Fig. 5 suggests that no single τ^* optimizes the

inspection policy across the whole range of r_a . Interestingly, even though we expect that a risk-averse manager would prefer intensive audits (i.e., lower τ^*), we find the opposite. For instance, when $\alpha = 1$, τ^* increases from 9 in the risk neutral case ($r_a = 0$) to 11 under high risk-aversion ($r_a = 5$). The finding implies that a highly risk-averse manager would reduce audit frequency under error-prone inspection ($\alpha < 1$ and $\beta = 0.2$). In addition, we see that τ^* tends to increase with α . This is consistent with the risk-neutral analysis showing that audits do not need to be so frequent provided better inspection efficacy. Modifying the cost structure so that $\gamma = 1$, we find the impact of r_a to be weaker, resulting in a τ^* range of [14,16] (results not shown in Figure). An explanation is that the decision-maker becomes less sensitive to risks because the inspection process does not incur as many costs as the large γ scenario ($\gamma = 3$).

Assuming $\gamma = 3$ and $\alpha = 0.8$, Fig. 6 shows that optimal decisions still vary with the degree of risk aversion. The left panel shows that when $\beta = 0.4$, τ^* increases from 10 in the risk neutral case ($r_a = 0$) to 12 under high risk-aversion ($r_a = 5$). A simple explanation for this result is at a certain point the manager becomes concerned about the negative consequences of frequent inspection (due to high β) so that a larger τ^* is economically more favorable. Interestingly, τ^* stays at 6 under $\beta = 0$ and risk aversion has no impact on optimal decisions. Hence, when there is no risk of miscorrecting non-faulty items, managers need not be as concerned about negative outcomes associated with β and τ^* becomes stable. The right panel of Fig. 6 shows that high-quality inspection results in the minimum CE, similar to the risk-neutral case where high α and low β are preferred in terms of total costs.

Fig. 4. Impacts of λ on τ^* and $E[f(\tau^*)]$ given $\gamma = 3$.

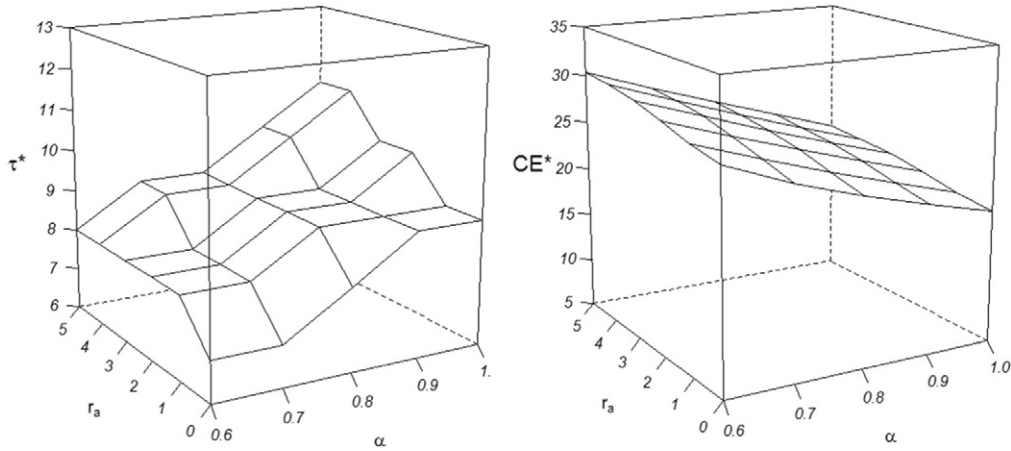


Fig. 5. τ^* and $CE(\tau^*)$ given $\gamma = 3$, $\lambda = 0.017$, & $\beta = 0.2$.

As a final test, we explore the interaction between risk aversion and shelf inventory failure rate under low ($\alpha = 0.6$ and $\beta = 0.4$) and high ($\alpha = 1$ and $\beta = 0$) inspection efficacy. The left panel of Fig. 7 suggests that τ^* tends to decrease in λ for slightly risk-averse managers ($r_a = 1$). However, due to the low-quality inspection, managers with moderate ($r_a = 3$) and high risk aversion ($r_a = 5$) become more concerned about inspection efficacy and would not trigger as frequent audits (i.e., a larger τ^*) even when shelf error generates quickly ($\lambda \geq 0.2$). The right panel of Fig. 7 indicates that risk preferences have a comparatively small impact on τ^* under perfect inspection ($\alpha = 1$ and $\beta = 0$). Similar to the risk neutral case in Fig. 4, τ^* first decreases in λ and then stays constant when store execution quality is too low ($\lambda \geq 0.35$). Also, Fig. 7 suggests that regardless of the level of risk aversion, managers are more willing to take shelf audit initiatives (i.e., a smaller τ^*) provided error-free inspection processes.

Summarizing, our analysis reveals that τ^* is fairly sensitive to inspection efficacy, manager's risk preferences, and store execution quality. So, optimal decisions for shelf audit service providers must be a function of the foregoing factors. From a total cost standpoint high-quality inspection is predominantly preferred regardless of risk attitudes. Note that in the analysis above we take (α, β, λ) as a given although those model parameters that significantly affect τ^* are not observable in the course of normal operations. In the following section, we propose a Bayesian approach to more precisely

estimate the three parameters based on the directly observable shelf audit reports.

5. Bayesian estimation of unknown parameters

In the foregoing analysis we set α , β , and λ as fixed parameters. While critical to our modeling framework, those parameters, however, are unobservable. Without experimental data, we cannot estimate the three uncertain quantities using the frequentist methodology directly. Nonetheless, our early analysis of both risk-neutral and risk-averse cases shows that optimal decisions vary significantly with unknown inspection efficacy and failure rate of shelf inventory. Having a good knowledge about (α, β, λ) will be highly helpful for managers to make cost-effective decisions. Therefore, instead of imposing peculiar assumptions on those unknown parameters ([5,19,38]), we propose a method to derive statistical inferences about (α, β, λ) using the data observed from error-prone periodic shelf-audits. The key idea is to devise a hierarchical Bayes model that enables us to infer the posterior distributions of α , β , and λ so that we can make our best guess about the parameters.

Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ be a data vector that contains observed outcomes in shelf audit reports generated from a periodic inspection cycle of $\bar{\tau}$ days. $Y_i = 1$ if the i th item is reported to be inaccurate and $Y_i = 0$ if no error is reported. For each observed Y_i there is an

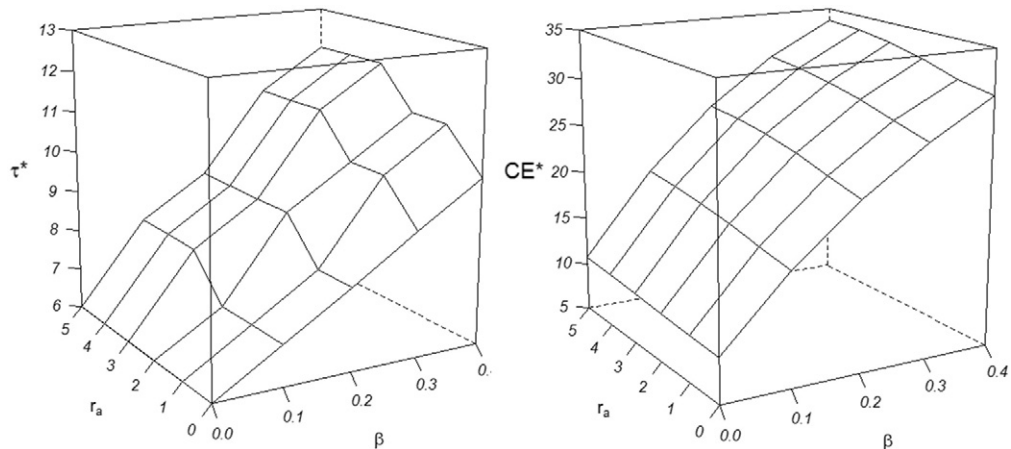


Fig. 6. τ^* and $CE(\tau^*)$ given $\gamma = 3$, $\lambda = 0.017$, & $\alpha = 0.8$.

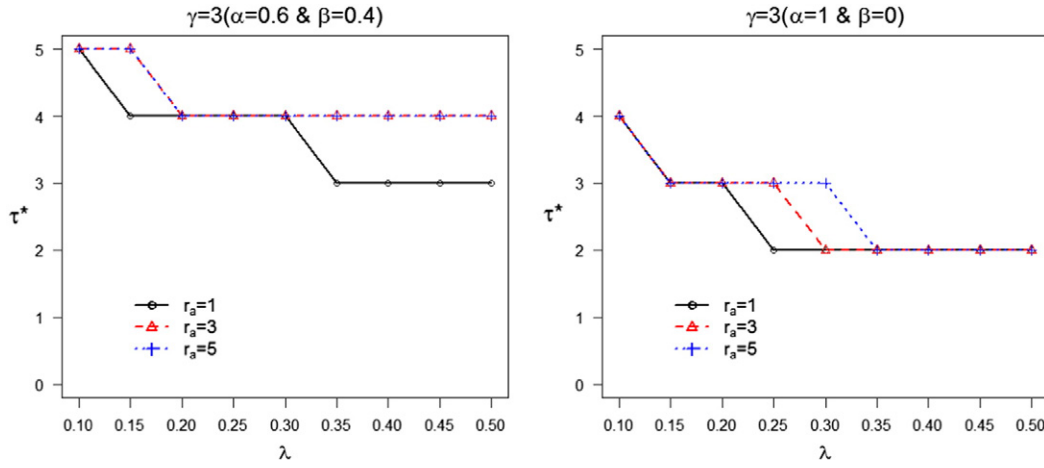


Fig. 7. τ^* under different levels of risk aversion and inspection quality.

unobservable variable X_i that reflects the “true status” of the item. The variable X_i is equal to 1 (i.e., the item “really” is faulty) or equal to 0 (i.e., the item “really” is accurate) with probability $1 - e^{-\lambda\bar{\tau}}$ and $e^{-\lambda\bar{\tau}}$ according to the exponential failure distribution. Assuming imperfect inspection, if $X_i = 1$, the corresponding $Y_i \sim \text{Bernoulli}(\alpha)$. If $X_i = 0$, the corresponding $Y_i \sim \text{Bernoulli}(\beta)$. In the context of Bayesian inference, X_i is the upper-level latent variable. Conditional on X_i , the observable Y_i is independent of the failure rate λ . The Bayesian hierarchical modeling framework is:

$$\begin{aligned} \alpha &\sim f(\cdot) \\ \beta &\sim g(\cdot) \\ \lambda &\sim h(\cdot) \\ X_i|\lambda &= \begin{cases} 1 \text{ with probability } 1 - e^{-\lambda\bar{\tau}} \\ 0 \text{ with probability } e^{-\lambda\bar{\tau}} \end{cases} \\ Y_i|X_i, \alpha, \beta &\sim \begin{cases} \text{Bernoulli}(\alpha) & \text{if } X_i = 1 \\ \text{Bernoulli}(\beta) & \text{if } X_i = 0 \end{cases} \end{aligned} \quad (11)$$

The prior distributions of α , β , and λ (i.e., $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$) can be any parametric distributions that reflect a manager's belief ex ante. We adopt beta priors for α and β because they naturally fit error probabilities ranging between $[0, 1]$. Moreover, the conjugacy between beta distributions and Bernoulli sampling models makes posterior distributions analytically tractable. We adopt gamma prior for λ to account for strict positivity of λ . The gamma prior is popular and useful due to its great flexibility.

We first derive the full conditional distribution of X_i from the Bayes' theorem:

$$\begin{aligned} P(X_i|\lambda, \alpha, \beta, Y_i) &\propto P(Y_i|X_i, \alpha, \beta)P(X_i|\lambda) \\ &\Rightarrow \frac{P(X_i = 1|\lambda, \alpha, \beta, Y_i)}{P(X_i = 0|\lambda, \alpha, \beta, Y_i)} = \frac{\text{dbern}(Y_i, \alpha)(1 - e^{-\lambda\bar{\tau}})}{\text{dbern}(Y_i, \beta)e^{-\lambda\bar{\tau}}} \quad (\text{from Baye's rule}) \end{aligned} \quad (12)$$

where $\text{dbern}(\cdot)$ refers to the Bernoulli probability mass. We then derive the full conditional distribution of α given a beta prior (i.e., $\alpha \sim \text{beta}(a_1, b_1)$).

$$\begin{aligned} P(\alpha|\beta, \lambda, \mathbf{X}, \mathbf{Y}) &\propto P(Y_1, \dots, Y_n, X_1, \dots, X_n, \lambda, \alpha, \beta) \\ &\propto \alpha^{\sum I(Y_i = 1, X_i = 1)} (1 - \alpha)^{\sum I(Y_i = 0, X_i = 1)} \alpha^{a_1 - 1} (1 - \alpha)^{b_1 - 1} \\ &\sim \text{beta}\left(\sum I(Y_i = 1, X_i = 1) + a_1, \sum I(Y_i = 0, X_i = 1) + b_1\right) \end{aligned} \quad (13)$$

The first line above states that the full conditional distribution of α is proportional to the joint distribution of data (\mathbf{Y}), latent variable (\mathbf{X}), λ , α ,

and β . After dropping out the distribution not involving α , it can be shown that $p(\alpha|\beta, \mathbf{X}, \mathbf{Y})$ conforms to a beta distribution.

Similarly, the full conditional of β given a beta prior ($\beta \sim \text{beta}(a_2, b_2)$) is:

$$P(\beta|\alpha, \lambda, \mathbf{X}, \mathbf{Y}) \sim \text{beta}\left(\sum I(Y_i = 1, X_i = 0) + a_2, \sum I(Y_i = 0, X_i = 0) + b_2\right) \quad (14)$$

Lastly, the full conditional of λ given a gamma prior ($\lambda \sim \text{gamma}(a_3, b_3)$) is only known to a certain proportionality:

$$P(\lambda|\mathbf{X}, \alpha, \beta, \mathbf{Y}) \propto P(\mathbf{X}|\lambda)P(\lambda) \quad (15)$$

To construct the posterior distributions numerically, one can use a Metropolis-within-Gibbs sampler [27], which is a popular Markov chain Monte Carlo (MCMC) algorithm because of its ability to accommodate a sampling scheme with high dimensionality. The Metropolis steps for sampling λ are:

1. Define a proposal distribution $J(\theta_\lambda|\theta_\lambda^{(s)})$
2. Sample a proposal value θ_λ^* from $J(\theta_\lambda|\theta_\lambda^{(s)})$
3. Compute the acceptance ratio $r = \frac{P(\theta_\lambda^*|\theta_\lambda^{(s)})}{P(\theta_\lambda^{(s)}|\theta_\lambda^{(s)})} = \frac{P(\theta_\lambda^*|\theta_\lambda^*)p(\theta_\lambda^*)}{P(\theta_\lambda^{(s)}|\theta_\lambda^*)p(\theta_\lambda^{(s)})}$ (16)
4. Let $\theta_\lambda^{(s+1)} = \begin{cases} \theta_\lambda^* & \text{with probability } \min(r, 1) \\ \theta_\lambda^{(s)} & \text{with probability } 1 - \min(r, 1) \end{cases}$

For the proposal distribution of MCMC, one can employ a random walk proposal or others [7] to initialize the Bayesian simulation. The other three full conditional distributions of X_i , α , and β jointly constitute the Gibbs sampler that starts with the vector $\theta^{(s)} = (\theta_\lambda^{(s)}, \theta_\alpha^{(s)}, \theta_\beta^{(s)}, \theta_\beta^{(s)})$ and transits to $\theta^{(s+1)}$ in the following way:

1. Sample $\theta_\lambda^{(s+1)}$ from $p(\theta_\lambda|\theta_\lambda^{(s)}, \theta_\alpha^{(s)}, \theta_\beta^{(s)}, \mathbf{Y})$
2. Sample $\theta_\alpha^{(s+1)}$ from $p(\theta_\alpha|\theta_\lambda^{(s+1)}, \theta_\alpha^{(s)}, \theta_\beta^{(s)}, \mathbf{Y})$
3. Sample $\theta_\beta^{(s+1)}$ from $p(\theta_\beta|\theta_\lambda^{(s+1)}, \theta_\alpha^{(s+1)}, \theta_\beta^{(s)}, \mathbf{Y})$
4. Sample $\theta_\beta^{(s+1)}$ from $p(\theta_\beta|\theta_\lambda^{(s+1)}, \theta_\alpha^{(s+1)}, \theta_\beta^{(s+1)}, \mathbf{Y})$

After S iterations, the sequences $\theta = \{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(S)}\}$ are expected to form a stationary Markov Chain that has the desired Markovian behaviors: irreducible, aperiodic, and recurrent [27]. The simulation convergence can be evaluated by standard metrics for MCMC (e.g., stationarity and no stickiness). For practical implementation of Eq. (17), we recommend the use of *thinning* to improve the convergence of the Markov chain [27]. Specifically, one should assign a large number to S to MCMC scans in which only every s scan is saved. For instance, a

thinning of $s = 80$ would reduce the size of a 48,000-scan Markov chain down to a quality sample of 600 observations. Then one can drop the first 100 out of the 600 observations to account for the burn-in period. A large S and a long burn-in period help achieve convergence.

Given the audit frequency of $\bar{\tau}$ days, we can use the Bayesian estimation method to obtain the posterior distribution of λ from $\theta_\lambda = (\theta_\lambda^{(1)}, \theta_\lambda^{(2)}, \dots, \theta_\lambda^{(S)})$, which informs us the failure rate of on-shelf items. Also, the sampled sequences $\theta_\alpha = \{\theta_\alpha^{(1)}, \theta_\alpha^{(2)}, \dots, \theta_\alpha^{(S)}\}$ and $\theta_\beta = \{\theta_\beta^{(1)}, \theta_\beta^{(2)}, \dots, \theta_\beta^{(S)}\}$ constitute the posterior distributions $P(\alpha|Y)$ and $P(\beta|Y)$ that we are looking for. To sum up, the Bayesian estimation methodology generates robust estimates of (α, β, λ) to which optimal solutions are very sensitive. Obtaining robust estimates of unknown parameters makes differences to total costs.

6. Concluding remarks

6.1. Validation

After obtaining a thorough understanding of model behaviors (in Sections 3 and 4) and model parameters (in Section 5), here we detail three major tasks involved in model validation, which helps decision makers build confidence in the proposed model. In addition to testing key model assumptions upon which the cost function is built, the first and second tasks are aimed for validating important model structures and parameters. The last task is focused on examining optimal decisions constructed from the model and cost implications pertaining to optimal decisions. Even though the three tasks may not be exhaustive, they have already covered core assumptions and purposes of our model. Finishing those tasks will be a crucial step before the proposed policy can be applied to actual audit service operations.

The first task is to validate the assumed exponential failure distribution that is critical to our penalty accounting for unfixed errors (Eq. (6)) as well as total cost function (Eq. (8)). The exponential distribution implies a *constant* hazard rate. That is, conditional on that an item has stayed accurate up to time t , the instantaneous failure probability between time t and time $t + \Delta$ is time-invariant. One quick and easy way to test this modeling assumption is to apply time-to-event analysis techniques to assess the hazard rate [31]. The decision maker can use available data (e.g., historical audit reports) and calculate time-since-last-correction for each observation. The censored time-to-failure observations then can be used to estimate the hazard rate (through maximum likelihood estimation) and assess the statistical significance of parameter estimates. If one finds no evidence to reject assumed constant hazard rate, our formulation predicated on exponential distribution will suffice. A more stringent test is to test other distributions with time-variant hazard rate (e.g., Weibull) and assess differences (in terms of information criterion) between the exponential and its alternatives. If the differences are not substantial, one should stay with the current formulation based on the exponential distribution as other distributions with extra parameters will lead to a complexed cost function that is less tractable and applicable.

The second task is to validate estimation results of unknown yet important model parameters. After running the Bayesian hierarchical model and MCMC (see Section 5), one is supposed to examining the acceptance ratio r (in Eq. (16)) of the proposal distribution. A rule of thumb is that acceptance rates should fall between 25% and 50% [44]. If acceptance rates are too low or too high, parameters of adopted proposal distributions need to be fine-tuned to ensure an effective transition of Markov chains. Moreover, the decision maker must examine trace plots and autocorrelation of simulated distributions of (α, β, λ) and ensure that no stickiness/high-order autocorrelation exhibits given the nature of Markov chains [27]. Finally, the decision maker ought to perform a visual comparison between prior and posterior

distributions. When the sample size for estimation increases, one should expect to see tighter posterior distributions.

After model functions and parameters are validated, the last task is to validate optimal decisions (τ^*). The τ^* (computed from Eq. (9)) has to be compared to τ^{actual} (i.e., *actual* audit cycle time of the audit service provider), τ^{min} (i.e., *minimum* audit cycle time that could be executed by the audit service provider), and τ^{max} (i.e., *maximum* audit cycle time that could be tolerated by the audit service provider). Specifically, given estimated model parameters, decision makers are supposed to calculate $(E[f(\tau^*)] - E[f(\tau^*)]) / E[f(\tau^*)]$ (where $E[f(\tau)]$ is Eq. (8) and $\tau \in \{\tau^{actual}, \tau^{min}, \tau^{max}\}$). For instance, when τ^* is less than τ^{min} , the ratio $(E[f(\tau^{min})] - E[f(\tau^*)]) / E[f(\tau^*)]$ allows decision makers to evaluate whether it is feasible/sensible to acquire extra labor capacity for shelf audits (as such τ^* can be executed). The cost differential will give managers a clear idea of economic benefits (in terms of expected daily cost over an audit cycle) and operational feasibility of τ^* constructed from the decision support model.

6.2. Discussion

Retailers and product manufacturers have come to the realization that store execution errors and the consequently low on-shelf availability hampers operational as well as financial performance. Practitioners claim an urgent need for improved shelf audits/asset tracking within retail stores [2]. When a shelf stock-out occurs, manufacturers typically lose a nontrivial fraction of their customers to their competitors. Therefore, external service companies are hired by manufacturers to conduct manual inspection regularly. Motivated by the practical need of audit service providers, we adopt a fairly general cost structure to develop a decision support model for periodic inspection. Unlike prior studies that focus on deriving inspection policies for retailers, our policy for external audits is designed to run alongside retailers' inspection effort and to address problems that may not be fully eliminated by retailers.

The proposed decision support model not only captures random failure of shelf inventory integrity but also considers human errors in audit initiatives. The inspector fallibility deserves more investigation since inspection can hardly be perfect. By explicitly modeling human errors in inspection processes, we allow decision makers to adjust inspection frequencies under different levels of inspection efficacy. In addition to revealing a non-negligible impact of imperfect inspection on optimal decisions, we uncover sophisticated relationships among error-prone inspection, on-shelf items' failure rate, and cost factors.

The notion of error-prone inspection is highly relevant to labor-intensive service operations and has important implications for managers. Human errors (i.e., α and β) are difficult to avoid due to behavioral factors (e.g., training, experiences, and fatigue). In reality, inspection associates may fail to do the job right simply because they have to examine too many items during a limited time. As a result of fatigue and pressure, employees may decide to cut corners and eventually cause operational quality erosion [39]. Researchers should incorporate human fallibility into decision support models and investigate incentives that elicit human efforts to improve service, conformance, and data quality.

Our analysis also investigates into the influence of risk preferences on optimal decisions. Grounded on subjective expected utility theory and Jensen's inequality, our simulation analysis makes the proposed decision support model more comprehensive and favorable to risk-averse managers. By using the CARA exponential utility function, our model considers the level of risk aversion that does affect the allocation of optimal audit efforts. Based on our findings managers should realize that inspection decisions that merely consider expected cost would be myopic and suboptimal. On top of model analysis under risk neutrality and risk aversion, we employ

Bayesian methods to derive statistical estimates of unobservable model parameters (α, β, λ).

When the proposed model is implemented, under an audit cycle of τ days, the decision maker (e.g., external audit service provider) will execute shelf audits and collect audit outcomes (i.e., the observed \mathbf{Y} vector discussed in Section 5). After that he/she will need to update decision criteria. Specifically, the decision maker has to re-estimate $P(\alpha|\mathbf{Y})$, $P(\beta|\mathbf{Y})$ and $P(\lambda|\mathbf{Y})$ using collected audit outcomes, and he/she can use the mode/mean of the posterior distributions as new model parameters. After that the decision maker needs to re-assess cost parameters (c_i, c_e, c_u) using the latest information regarding employee payroll and loss of leaving shelf errors unfixed (as discussed in Section 3). Given empirically grounded parameter estimates, new audit decisions will be computed and then executed.

While analytical modeling approaches ensure cost optimality, parameters of analytical models may not be entirely observable or estimable, and hence heuristic approaches are seemingly more useful for supporting shelf audit decisions [41]. Despite taking an analytical approach, our model is still practically applicable and resulting decisions from our model are empirically informed. The aforementioned tasks – execute shelf audits, collect audit outcomes, and update decision criteria – constitute a continuous improvement process that can be incorporated into a decision support system for error-prone shelf audits. Different from RFID-enabled cases with perfect information (e.g., [9]), our case is under limited information and nonetheless, all parameters of our decision model can be empirically rationalized and continuously updated when it comes to practical implementation.

Several limitations of our study pinpoint opportunities for future research. First, we adopt a linear inspection cost specification in line with previous studies and the accounting scheme of practitioners we work with. Subsequent studies can explore the impact of a nonlinear inspection cost function on optimal decisions. Second, due to information availability, our model does not consider the number of inventory transactions between inspections and applies the exponential distribution to capture the degradation of shelf inventory integrity over time. If manufacturers or retailers are willing to share POS/inventory data to external service providers, inventory dynamics could be further incorporated into our model. Third, for the simulation analysis under risk aversion we adopt a CARA exponential utility function. Future studies could employ utility functions that exhibit constant relative risk aversion to assess whether the functional form for utility makes results of analysis qualitatively different. Last, we propose a Bayesian model and a MCMC scheme to infer unobserved human error that contaminates data generated from error-prone inspection processes. A potentially interesting extension would be to compare and contrast our Bayesian statistical model to the set of measurement error models – most of them are rooted in the frequentist paradigm – developed by statisticians [23].

Despite these limitations, our work delivers a pragmatic decision model to managers who have a strong interest in fixing shelf errors through inspection or launching periodic shelf audit services. Even though RFID-enabled automatic counting seems to be an attractive alternative to error-prone manual counting [8], a full deployment of item-level RFID is still hard to achieve for numerous retailers due to various concerns related to cost, privacy, etc. For firms that may not be willing or able to adopt RFID, periodic inspection is still the most common and effective approach to fix execution errors and increase product availability. With appropriate modification, our decision support model is potentially applicable to manufacturing, healthcare, and military (where inventory integrity is paramount). In short, our modeling effort hopes to remind researchers and practitioners of the importance of high-quality shelf inspection that could potentially recover a significant amount of unnecessary loss. The goal of improving on-shelf availability cannot be overemphasized given the critical role inventory plays in retail operations.

Acknowledgments

The author is grateful for the suggestions provided by James Marsden and three referees. Their comments helped improve the content and presentation of the paper. The author also wishes to thank Rogelio Oliva as this work could not have been done without his guidance in the early stage of this research. Special thanks go to James W. Richardson who introduced the SERF technique to the author.

Appendix A. Proofs

Proof of Proposition 1. By construction, the D_j inventory items are accurate after the j th inspection and turn faulty before the $(j + 1)$ th inspection. We first define a random variable T that denotes the time to fall into inaccurate status between the time interval $(0, \tau)$. So, the expected time of being inaccurate during an inspection cycle of τ days is $\tau - E[T]$, which can be derived as follows. Given the exponential failure process, the cumulative density of T is

$$F(t) = P(T \leq t | 0 < T < \tau) = \frac{P(T \leq t \cap 0 < T < \tau)}{P(0 < T < \tau)} = \frac{1 - e^{-\lambda t}}{1 - e^{-\lambda \tau}}, \quad 0 < t < \tau$$

Accordingly, the probability density of T is

$$f(t) = \frac{dF(t)}{dt} = \frac{e^{\lambda(\tau-t)}\lambda}{e^{\lambda\tau}-1}, \quad 0 < t < \tau$$

$$E[T] = \int_0^\tau 1 - F(t) dt = \int_0^\tau t * f(t) dt = \frac{1}{\lambda} - \frac{\tau}{e^{\lambda\tau}-1}$$

\therefore The expected time of being inaccurate is $\tau - E[T] = \tau + \frac{\tau}{e^{\lambda\tau}-1} - \frac{1}{\lambda}$ \square

Proof of Proposition 2.

$$\begin{aligned} E[F_{j+1}^b] &= E[F_j^a] + E[D_j] = E[F_j^a] + (n - E[F_j^a])P(\tau) = nP(\tau) + (1 - P(\tau))E[F_j^a] \\ E[F_j^a] &= E[F_j^b] - E[K_j(F_j^b, \alpha)] + E[M_j(n - F_j^b, \beta)] \\ &= E[F_j^b] - E[F_j^b]\alpha + (n - E[F_j^b])\beta = (1 - \alpha - \beta)E[F_j^b] + n\beta \\ \therefore E[F_j^b] &= nP(\tau) + (1 - P(\tau))\{(1 - \alpha - \beta)E[F_{j-1}^b] + n\beta\} \\ \text{In steady state } E[F_j^b] &= E[F_{j+1}^b] = E[F^b] \text{ and solve for } E[F^b] \\ \Rightarrow E[F^b] &= \frac{n(\beta + P(\tau) - \beta P(\tau))}{\alpha + \beta + P(\tau) - (\alpha + \beta)P(\tau)} \quad \square \end{aligned}$$

Proof of Proposition 3.

$$\frac{\partial E[F^b]}{\partial \alpha} = -\frac{n(-1 + e^{\tau\lambda} + \beta)}{(-1 + e^{\tau\lambda} + \alpha + \beta)^2} \text{ where the denominator is non-negative.}$$

Since $e^{\tau\lambda} > 1$, the numerator is also non-negative $\Rightarrow \frac{\partial E[F^b]}{\partial \alpha} \leq 0$.

$$\frac{\partial E[F^b]}{\partial \beta} = \frac{n\alpha}{(-1 + e^{\tau\lambda} + \alpha + \beta)^2} \text{ where the denominator is non-negative.}$$

The numerator is also non-negative $\Rightarrow \frac{\partial E[F^b]}{\partial \beta} \geq 0$.

$$\frac{\partial E[F^b]}{\partial \tau} = \frac{n\alpha\lambda e^{\lambda\tau}}{(-1 + e^{\tau\lambda} + \alpha + \beta)^2} \text{ where the denominator is non-negative.}$$

The numerator is also non-negative $\Rightarrow \frac{\partial E[F^b]}{\partial \tau} \geq 0$.

$$\frac{\partial E[F^b]}{\partial \lambda} = \frac{n\alpha\tau e^{\lambda\tau}}{(-1 + e^{\tau\lambda} + \alpha + \beta)^2} \text{ where the denominator is non-negative.}$$

The numerator is also non-negative $\Rightarrow \frac{\partial E[F^b]}{\partial \lambda} \geq 0$ \square

Appendix B. Table of notations

Variables and parameters			
τ	Frequency of inspection	K_j	Number of corrected items
F^b	Faulty items before inspection	M_j	Number of mis-corrected items
F^a	Faulty items after inspection	D_j	Number of newly degraded items
τ^*	Optimal frequency of inspection	γ	Ratio of c_u -to- c_i
n	Number of items to be inspected	c_i	Inspection cost per item
λ	Exponential failure rate	c_c	Correction cost per item
α	Probability of correcting an item	c_u	Cost of not fixing errors per item/day
β	Probability of mis-correcting an item	r_a	Coefficient of absolute risk aversion

References

- [1] A.T. de Almeida, Multicriteria decision making on maintenance: spares and contracts planning, *European Journal of Operational Research* 129 (2) (2001) 235–241.
- [2] S. Anand, C. Cunnane, Minimizing Retail Stock-Out and Over-Stock: Optimizing Inventory for Customer Satisfaction, Research Preview, Aberdeen Group, Boston, 2009.
- [3] A. Atali, H. Lee, O. Ozer, If the inventory manager knew: value of visibility and RFID under imperfect inventory information, Working Paper, Stanford University, 2011.
- [4] R. Baker, Risk aversion in maintenance: a utility-based approach, *IMA Journal of Management Mathematics* 21 (4) (2010) 319–332.
- [5] D.P. Ballou, H.L. Pazer, The impact of inspector fallibility on the inspection policy in serial production systems, *Management Science* 28 (4) (1982) 387–399.
- [6] M. Boyle, Wal-Mart Brings in Consultants to Help Keep Its Shelves Stocked, *Bloomberg News* 2011 (<http://www.bloomberg.com/news/2011-10-06/wal-mart-s-empty-shelves-erode-walton-s-legacy.html>).
- [7] S.P. Brooks, Markov Chain Monte Carlo method and its application, *Journal of the Royal Statistical Society: Series D (The Statistician)* 47 (1) (1998) 69–100.
- [8] O.E. Cakici, H. Groenevelt, A. Seidmann, Using RFID for the management of pharmaceutical inventory — system optimization and shrinkage control, *Decision Support Systems* 51 (4) (2011) 842–852.
- [9] C. Condea, F. Thiesse, E. Fleisch, RFID-enabled shelf replenishment with backroom monitoring in retail stores, *Decision Support Systems* 52 (4) (2012) 839–849.
- [10] H.H. Chuang, R. Oliva, S. Liu, On-shelf availability, retail performance, and external audits: a field experiment, *Production and Operations Management* (2015) (in press).
- [11] M.H. DeGroot, *Optimal Statistical Decisions*, Wiley-Interscience, New York, 2004.
- [12] N. DeHoratius, A.J. Mersereau, L. Schrage, Retail inventory management when records are inaccurate, *Manufacturing & Service Operations Management* 10 (2) (2008) 257–277.
- [13] R. Dudley, Wal-Mart Sees \$3 Billion Opportunity Refilling Empty Shelves, *Bloomberg News* 2014 (<http://www.bloomberg.com/news/2014-03-28/wal-mart-says-refilling-empty-shelves-is-3-billion-opportunity.html>).
- [14] S.O. Duffuaa, Impact of inspection errors on performance measures of a complete repeat inspection plan, *International Journal of Production Research* 34 (7) (1996) 2035–2049.
- [15] R. Ernst, S.G. Powell, Manufacturer incentives to improve retail service levels, *European Journal of Operational Research* 104 (3) (1998) 437–450.
- [16] R. Ernst, J. Guerrero, A. Roshwalb, A quality control approach for monitoring inventory stock levels, *Journal of the Operational Research Society* 44 (11) (1993) 1115–1127.
- [17] Y. Ettouazi, N. Yates, C. Mena, Examining retail on shelf availability: promotional impact and a call for research, *International Journal of Physical Distribution & Logistics Management* 42 (3) (2012) 213–243.
- [18] J. Fernie, D.B. Grant, On-shelf availability: the case of a UK grocery retailer, *The International Journal of Logistics Management* 19 (3) (2008) 293–308.
- [19] W.G. Ferrell Jr., A. Chhoker, Design of economically optimal acceptance sampling plans with inspection error, *Computers & Operations Research* 29 (10) (2002) 1283–1380.
- [20] M.L. Fisher, Rocket science retailing: the 2006 Philip McCord Morse lecture, *Operations Research* 57 (3) (2009) 527–540.
- [21] M.L. Fisher, A. Raman, *The New Science of Retailing: How Analytics are Transforming the Supply Chain and Improving Performance*, Harvard Business Press, Boston, 2010.
- [22] M.L. Fisher, J. Krishnan, S. Netessine, Are your staffing levels correct? *International Commerce Review* 8 (2–4) (2009) 110–115.
- [23] W.A. Fuller, *Measurement Error Models*, John Wiley & Sons, Inc., 1987.
- [24] J.B. Hardaker, R.B.M. Huirne, J.R. Anderson, G. Lien, *Coping with Risk in Agriculture*, CABI, 2004.
- [25] G. Hausrucker, Approaches to Measuring On-Shelf Availability at The Point of Sale, ECR Europe White Paper, 2006 (http://ecr-all.org/content/ecropedia_element.php?ID=12181).
- [26] Hill, Applying Bayesian methodology with a uniform prior to the single period inventory model, *European Journal of Operational Research* 98 (3) (1997) 555–562.
- [27] P.D. Hoff, *A First Course in Bayesian Statistical Methods*, Springer, New York, 2009.
- [28] J.S. Hughes, Optimal internal audit timing, *The Accounting Review* 52 (1) (1972) 56–68.
- [29] D.L. Iglehart, R.C. Morey, Inventory systems with imperfect asset information, *Management Science* 18 (8) (1972) B388–B394.
- [30] J. Juran, Inspector's errors in quality control, *Mechanical Engineering* 57 (1) (1935) 643–644.
- [31] N.M. Kiefer, Economic duration data and hazard functions, *Journal of Economic Literature* 26 (2) (1988) 646–679.
- [32] C. Kirkwood, *Strategic Decision Making: Multiple Objective Decision Analysis with Spreadsheets*, Duxbury Press, Pacific Grove, CA, 1997.
- [33] A.G. Kok, K.H. Shang, Inspection and replenishment policies for systems with inventory record inaccuracy, *Manufacturing & Service Operations Management* 9 (2) (2007) 185–205.
- [34] S. Kumar, Development of internal audit and cycle-counting procedures for reducing inventory miscounts, *International Journal of Operations and Production Management* 12 (1) (1992) 61–70.
- [35] G. Lien, S. Stodal, J.B. Hardaker, L.J. Asheim, Risk aversion and optimal forest replanting: a stochastic efficiency study, *European Journal of Operational Research* 181 (3) (2007) 1584–1592.
- [36] R.C. Morey, D.A. Dittman, Optimal timing of account audits in internal control, *Management Science* 32 (3) (1986) 272–282.
- [37] H. Moskowitz, R. Plante, Effect of risk aversion on single sample attribute inspection plans, *Management Science* 30 (10) (1984) 1226–1237.
- [38] V.M. Ng, On an estimation problem in multiple inspections, *Journal of the Royal Statistical Society: Series D (The Statistician)* 38 (4) (1989) 217–219.
- [39] R. Oliva, J.D. Sterman, Cutting corners and working overtime: quality erosion in the service industry, *Management Science* 47 (7) (2001) 894–914.
- [40] R.T. O'Reagan, Relative costs of computerized error inspection plans, *Journal of the American Statistical Association* 64 (December) (1969) 1245–1255.
- [41] D. Papakiriakopoulos, K. Pramatar, G. Doukidis, A decision support system for detecting products missing from the shelf based heuristic rules, *Decision Support Systems* 46 (3) (2009) 685–694.
- [42] M.E. Peecher, R. Schwartz, I. Solomon, It's all about audit quality: perspectives on strategic-systems auditing, *Accounting, Organizations and Society* 32 (4–5) (2007) 463–485.
- [43] A. Raman, Retail-data quality: evidence, causes, costs, and fixes, *Technology in Society* 22 (1) (2000) 97–109.
- [44] C.P. Robert, G. Casella, *Introducing Monte Carlo Methods with R*, Springer, New York, 2009.
- [45] H. Sandoh, H. Shimamoto, A theoretical study on optimal inventory-taking frequency for retailing, *Journal of Retailing and Consumer Services* 8 (1) (2001) 47–52.
- [46] Z. Ton, Why good jobs are good for retailers? *Harvard Business Review* (2012) 124–131 (January–February).
- [47] Z. Ton, A. Raman, The effect of product variety and inventory levels on retail sales: a longitudinal study, *Production and Operations Management* 19 (5) (2010) 546–560.
- [48] S.A. Vander Wiel, S.B. Vardeman, A discussion of all-or-none inspection policies, *Technometrics* 36 (1) (1994) 102–109.
- [49] H. Wan, X. Xu, Technical note: reexamination of all-or-none inspection policies in a supply chain with endogenous product quality, *Naval Research Logistics* 55 (3) (2008) 277–282.

Howard Hao-Chun Chuang is an assistant professor at the College of Commerce, National Chengchi University. He holds a Ph.D. in information and operations management from the Mays Business School at Texas A&M University. Dr. Chuang's research interests include retail service operations, Internet commerce, and supply chain management. His research papers have been published in *Journal of Operations Management*, *Production and Operations Management*, *Decision Sciences*, etc. Prior to his academic career, he served as a supply corporal for a year in Penghu defense command of Taiwan army.