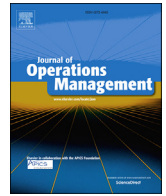




Contents lists available at ScienceDirect

Journal of Operations Management

journal homepage: www.elsevier.com/locate/jom

Erratum

Erratum to 'Inventory record inaccuracy: Causes and labor effects'

Howard Hao-Chun Chuang^a, Rogelio Oliva^{b, *}^a College of Commerce, National Chengchi University, Taipei 11605, Taiwan, ROC^b Mays Business School, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Article history:

Available online 18 March 2016

Keywords:

Retail operations

Store execution

Inventory record inaccuracy

System dynamics

Design of experiments

Bayesian inference

Econometrics

ABSTRACT

Inventory record inaccuracy (IRI) is a pervasive problem in retailing and causes non-trivial profit loss. In response to retailers' interest in identifying antecedents and consequences of IRI, we present a study that comprises multiple modeling initiatives. We first develop a dynamic simulation model to compare and contrast impacts of different operational errors in a continuous (Q, R) inventory system through a full-factorial experimental design. While backroom and shelf shrinkage are found to be predominant drivers of IRI, the other three errors related to recording and shelving have negligible impacts on IRI. Next, we empirically assess the relationships between labor availability and IRI using longitudinal data from five stores in a global retail chain. After deriving a robust measure of IRI through Bayesian computation and estimating panel data models, we find strong evidence that full-time labor reduces IRI whereas part-time labor fails to alleviate it. Further, we articulate the reinforcing relationships between labor and IRI by formally assessing the gain of the feedback loop based on our empirical findings and analyzing immediate, intermediate, and long-term impacts of IRI on labor availability. The feedback modeling effort not only integrates findings from simulation and econometric analysis but also structurally explores the impacts of current practices. We conclude by discussing implications of our findings for practitioners and researchers.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Inventory record inaccuracy (IRI) refers to the discrepancy between physical and recorded inventory levels, and is a pervasive problem in retailing. Kok and Shang (2014) conclude that IRI can be attributed to shrinkage (e.g., spoilage and theft), transaction errors, and misplacement. Because it is difficult to fully eliminate these execution errors, IRI becomes a norm rather than an anomaly in the retail sector. Kang and Gershwin (2005) report that inventory accuracy in a global retailer is on average only 51%. DeHoratius and Raman (2008) find 65% of the inventory records at a retail chain to be inaccurate, and Oliva et al. (2015) observe that more than 60% of SKUs in a European retail store have IRI. Most surprisingly, in a retail store that had not even started operating, Raman et al. (2001) found that the system had incorrect records for 29% of the items and estimated that IRI reduces a company's total profits by 10%. At the firm level, IRI can significantly distort aggregate book value of

inventory and business decisions. At the item level, IRI can delay ordering decisions because most extant inventory models do not differentiate between physical and system inventories. IRI also interrupts shelf replenishment even when there is plenty of inventory in the backroom. Consequently, retailers suffer severe out-of-stock (OOS) and significant economic loss.

To tackle IRI and associated OOS in retailing environments, radio-frequency identification (RFID) has been deemed as a promising solution (Heese, 2007; Lee and Ozer, 2007). However, issues such as cost, ownership, and privacy/security hinder the full implementation of RFID at the item-level (Kapoor et al., 2009). Even when RFID becomes cheap enough to be fully adopted like barcoding, the fact that retail operations is a complicated issue involving people, processes, and technology makes error-free operations extremely difficult to achieve. In order for retailers to enhance execution quality and data integrity, it is important for managers to understand the causes of IRI and identify the policy levers that they can use to reduce it.

While some empirical work has focused on product and store attributes that affect IRI (e.g., DeHoratius and Raman, 2008), in this work we explore the impact of store staffing levels and operational performance on IRI. Our study comprises multiple modeling initiatives. First, grounded on empirical observations and field work,

DOIs of original article: <http://dx.doi.org/10.1016/j.jom.2015.07.006>, <http://dx.doi.org/10.1016/j.jom.2016.01.001>.

* Corresponding author.

E-mail addresses: chuang@nccu.edu.tw (H.H.-C. Chuang), roliva@tamu.edu (R. Oliva).

<http://dx.doi.org/10.1016/j.jom.2016.01.002>

0272-6963/© 2015 Elsevier B.V. All rights reserved.

we formulate a dynamic model of continuous review (Q, R) inventory system and explicitly incorporate multiple execution errors into the model. The (Q, R) policy is often used for fast moving products and widely adopted by retailers, including numerous mass merchants that carry a large number of items (Kang, 2004; Kang and Gershwin, 2005) and stores that we work with. To compare and contrast the impact of different errors and their interactions, we conducted a full-factorial experimental design. We find that backroom shrinkage and shelf shrinkage errors are the dominant drivers of IRI and that, under-shelving, along with erroneous checkout and data capture, have negligible impact on IRI when compared to shrinkage. We also find that the interaction effects between error sources are non-substantial and mostly seem additive and linear. These primary findings hold under different distributional assumptions and parameter settings.

Next, we investigated the relationships between labor availability and IRI using longitudinal data from five stores in a global retail chain. After deriving a robust measure of IRI through Bayesian computation and estimating panel data models that control for store-section and time fixed effects, we find strong evidence that more full-time labor reduces IRI whereas part-time labor fails to alleviate it. Finally, we articulate the reinforcing relationships between labor and IRI by formally assessing the gain of the feedback loop based on our empirical results. We find that the work pressure introduced by IRI does further increase IRI, but the gain of feedback loop is not enough to compound its growth. We also analyze the intermediate and long-term effects of IRI on labor availability and use the developed structure to assess the impact of current staffing practices on performance.

Our paper contributes to practice and theory in four significant ways. First, the simulation model has a simple but realistic structure that addresses the issue that most retail inventory models ignore – the dynamics between the retail shelf and the backroom used for extra storage (Eroglu et al., 2013) – and allows for a joint assessment of the relative impact of operational errors in IRI. Second, despite the abundance of optimization models developed to tackle IRI, empirical investigations are scant. By econometrically estimating the effects of labor allocation on IRI, we broaden empirical knowledge of IRI and develop new research opportunities. Retail managers should be aware of labor effects on data quality, which is deemed to be an important source for competitive advantage (Redman, 1995). Third, we articulate the reinforcing relationships between labor and IRI by formally assessing the gain of the feedback loop based on our empirical findings and analyzing immediate, intermediate, and long-term impacts of IRI on labor availability. The feedback modeling effort not only integrates findings from simulation and econometric analysis but also structurally explores the implications of current practices. Last, we illustrate the utility of a joint use of system dynamics and econometrics. Such a combination widens our ability to answer questions of what-if and what-is given unobservable factors (i.e., execution errors) and limited observations of IRI over time. Using dynamic simulation, Bayesian shrinkage estimation, panel data modeling, and causal loop modeling enhances our understanding of IRI while responding to the call for adopting multiple methods (Boyer and Swink, 2008).

The rest of our article is organized as follows. Section 2 briefly reviews relevant literature to frame our contribution. Section 3 presents a continuous-time simulation analysis that enables us to identify the main drivers of IRI. We then postulate and articulate how those drivers of IRI are associated to store labor. Section 4 shows econometrical estimation results of labor availability on IRI. Section 5 presents feedback loops and behavioral dynamics associated with the impact of IRI on labor availability. We conclude by discussing managerial and theoretical implications of our findings.

2. Literature review

A significant number of studies have attempted to analyze causes and effects of IRI in recent years (e.g., Fleisch and Tellkamp, 2005; DeHoratius and Raman, 2008). Due to the randomness of errors that cause IRI and uncertainties in the distribution of IRI, simulation has been widely adopted to assess the effect of IRI on a retail supply chain (Fleisch and Tellkamp, 2005) or a retail outlet (Nachtmann et al., 2010). Among simulation studies on IRI, the continuous review (Q, R) system has been the focus of investigation. Kang and Gershwin (2005) analyzed how stock loss (shrinkage) causes IRI and severe OOS. They found that OOS increases monotonically in stock loss. Thiel et al. (2010) simulated the impact of IRI on service level and in contrast with Kang and Gershwin (2005), they observed that OOS is not a monotonic function of IRI when error rate is symmetric with a zero mean. Following Kang and Gershwin (2005), Agrawal and Sharda (2012) concentrated on IRI attributed to stock loss, and examined how the frequency of inventory audit affects OOS and average inventory. Similarly, in the first part of our paper we develop a dynamic simulation model of the (Q, R) inventory system. Our model differs from the aforementioned studies in two ways. First, while most models address a single source of error (Sahin and Dallery, 2009), Lee and Ozer (2007) point out that modeling efforts are needed to articulate the joint effect of multiple errors. Our model takes into account multiple errors (both operational and information-related) simultaneously. Although Fleisch and Tellkamp (2005) also assessed the impact of several errors using stochastic simulation, we analyze operations inside a retail store instead of flows in a three-echelon supply chain. Second, while existing simulation studies on IRI stress the consequences (e.g., inventory level, fill rate) of poor data quality (Nachtmann et al., 2010), our analysis focuses on the impacts of different antecedents of IRI.

While simulation analysis enhances our understanding about antecedents and consequences of IRI, there is still limited empirical knowledge about IRI due to the low availability of data. Few studies empirically investigate IRI through analyzing actual data on inventory discrepancies. Sheppard and Brown (1993) presented the first analysis to empirically assess how product-related factors affect IRI within a manufacturing plant. In retail stores, Raman (2000) and Oliva et al. (2015) both found that more than 60% of items had inaccurate records. Using data from a single store, Oliva et al. (2015) derived empirical estimates of an aggregate model that characterizes inventory information decay. The estimated functional form is further incorporated into inspection policy design. To our knowledge, the only cross-store econometric analysis of IRI is by DeHoratius and Raman (2008). They collected cross-sectional data on IRI from a retail chain to empirically examine IRI. The econometric analysis performed in the second part of our paper differs from DeHoratius and Raman (2008) in three important ways. First, expanding their efforts on examining how product, and store, related attributes affect IRI, we assess the association between labor decisions and IRI in each product sector. Second, we obtain longitudinal observations of IRI and labor decisions, which allow us to test labor effects while tackling unobserved factors. Third, our econometric estimation focuses on developing an operational functional form for the impact of labor on IRI (Richmond, 1993), as opposed to a correlational study to test hypotheses.

Finally, our work is also informed by system dynamics (Forrester, 1958; Sterman, 2000) efforts to assess the impact of labor and staffing levels on operational performance (e.g., Anderson, 2001; Oliva and Sterman, 2001; Lyneis and Ford, 2007). While we adopt from these articles the feedback perspective on staffing issues, our work differs in that we focus on data quality and its consequences on labor availability.

3. Errors in retail operations

In this section we present a dynamic simulation model to assess the impact of multiple execution errors on IRI. The model characterizes a continuous review (Q, R) inventory system in which Q denotes the order quantity and R denotes the reorder point. We articulated different types of error and identified primary drivers of IRI through design of experiments (DOE).

3.1. Continuous (Q, R) system

The core structure of the model was developed from the review of re-stocking and re-shelving policies of five different retailers, as well as extensive interviews with 18 managers responsible for the re-shelving activities for a group of SKUs (on average around 1800 SKUs per manager); twelve of these managers also had final decision rights on the re-stocking orders proposed by the firm's automated system. While the core structure of the model is highly stylized, and does not match perfectly any single retailer in our sample, the model does capture the main feedback mechanisms used by store and category managers to ensure continuous supply of products and accurate information about their holdings.

For clarity, we first present an error-free system and introduce random errors later. Similarly to previous work (Kang and Gershwin, 2005; Thiel et al., 2010), we assume single product/location, no backorder/fixed order cost, and reliable supply. We start with constant demand (δ) and fixed lead time (L). Setting the two elements deterministic allows us to clearly identify the impacts of random execution errors. The store physical inventory at time t , $I_a(t)$, can be divided into shelf inventory ($S_a(t)$) and backroom inventory ($B_a(t)$), i.e., $I_a(t) = S_a(t) + B_a(t)$. Since inventory is held in two locations (shelf and backroom) in most retail environments, explicitly modeling the two components enables us to assess operational complexity incurred by storing extra inventory in the backroom (Eroglu et al., 2013). In our model, the subscript a denotes actual physical flows, and we use it to differentiate from recorded information flows indexed by r . Table 1 summarizes the notation used in this section.

We present our model using the system dynamics convention, i.e., in continuous time and as a set of nonlinear differential equations, where the change rates represent information or material flows (measured in units/time) and the accumulation of these rates are represented as stocks (measured in units). We keep track of products in three distinct stocks: shelf, backroom, and orders-in-transit. For convenience we drop the time index t . The shelf stock

(S_a) is augmented by the shelving rate (s_a) and depleted by the merchandise withdrawal rate (w_a).

$$\frac{dS_a}{dt} = s_a - w_a \quad (1)$$

With enough items on the shelf, the withdrawal rate (w_a) will equal the constant demand rate (δ). Otherwise, the amount withdrawn will depend on remaining stock quantity and the store purchase time (τ_p), i.e., the time it takes a customer to make a purchase and walk out of the store.

$$w_a = \text{Min} \left(\delta, \frac{S_a}{\tau_p} \right) \quad (2)$$

The shelving rate (s_a) is the flow of inventories pulled from the backroom to the shelf. For simplicity, re-shelving is assumed to be continuous and proportional to the gap between desired shelf stock S^* and recorded shelf inventory S_r , and the process is adjusted by the time it takes store associates to identify the gap and replenish the shelf (τ_s). The shelving operation is also subject to a physical constraint since the retailer cannot replenish more than the amount stored in the backroom (B_a). Thus,

$$s_a = \text{Min} \left(\frac{S^* - S_r}{\tau_s}, \frac{B_a}{\tau_s} \right) \quad (3)$$

The backroom stock (B_a) is depleted by the shelving rate (s_a) and augmented by the delivery rate (d_a), which refers to the actual goods shipped to the backroom inside a retail store.

$$\frac{dB_a}{dt} = d_a - s_a \quad (4)$$

We assume a reliable supplier, and thus the delivery rate (d_a) is the order rate (o) with a fixed time delay L .

$$d_a(t) = o(t - L) \quad (5)$$

The order-in-transit stock (O) represents the orders that have been placed to the supplier but have not been delivered yet. Accordingly, it is augmented by the order rate (o) and depleted by the delivery rate (d_a):

$$\frac{dO}{dt} = o - d_a \quad (6)$$

The order rate (o) in a continuous review (Q, R) is activated whenever inventory position, i.e., the sum of recorded inventory on-hand ($I_r = B_r + S_r$) and order-in-transit (O), drops to/below the reorder point R .

$$o = \frac{Q}{dt} \times \text{Ind}(I_r + O < R) \quad (7)$$

where dt is the time step of simulation and a very small time interval to reflect the negligible time of placing an order from the automated ordering system. $\text{Ind}(\cdot)$ is an indicator function that returns one if the trigger criterion is met.

Equations (1)–(7) capture physical flows of the item from supply line to customers. Under perfect store operations, the information flows will be identical to physical flows. Therefore, the recorded shelf stock (S_r) and recorded backroom stock (B_r) change rates are just

Table 1
Notation.

Stocks			
S_a	Physical shelf inventory	S_r	Recorded shelf inventory
B_a	Physical backroom inventory	B_r	Recorded backroom inventory
O	Order in transit		
Rates			
w_a	Actual withdrawal rate	w_r	Recorded withdrawal rate
d_a	Actual delivery rate	d_r	Recorded delivery rate
s_a	Actual shelving rate	s_r	Recorded shelving rate
δ	Demand rate	o	Order rate
Parameters			
Q	Order quantity	ϵ_w	Checkout error
R	Reorder point	ϵ_c	Data capture error
L	Lead time	ϵ_s	Shelf error
S^*	Desired shelf stock	ϵ_{S-}	Shelf shrinkage
dt	Time step of simulation	ϵ_{B-}	Backroom shrinkage
τ_s	Shelving time	τ_p	Store purchase time

$$\begin{aligned}\frac{dS_r}{dt} &= s_r - w_r \\ \frac{dB_r}{dt} &= d_r - s_r\end{aligned}\quad (8)$$

$$i_a = i_r \forall i \in \{d, s, w\}$$

Equations (1)–(8) capture error-free inventory execution. The upper side of Fig. 1 exhibits stocks and flows of physical inventories. The bottom side of Fig. 1 exhibits stocks and flows of system inventories. The dashed lines in Fig. 1 reflect information processing. During normal operation the retailer continuously monitors I_r (the sum of B_r and S_r) to place orders to supply line. The retailer also monitors S_r to take reshelfing actions. The shelving follows a goal-seeking structure (Sternan, 2000) captured in (3), aiming to bring the shelf stock S back to its desired target S^* (see loop B1 in Fig. 1). The two stocks modeled in (8), together with equations (6) and (7), comprise the ordering mechanism and two feedback loops – ordering (B2) and supply (B3) – aimed at maintaining the desired inventory level in the store.

Under perfect store operations, constant demand rate, and deterministic lead time, the system in Fig. 1 reaches equilibrium and no lost sales are incurred nor does the system generate IRI (differences the values of the physical stocks and the corresponding information stocks). However, management of retail backroom and shelf involves many activities, and random errors may occur at picking, shelving, labeling, and checkout. Thus, we introduce an array of commonly reported and cited errors to represent more realistic operating conditions.

3.2. Inventory system subject to errors

We modified the equations developed earlier to capture an inventory system subject to five types of random errors, including three asymmetric errors (i.e., shelving error, shelf shrinkage, backroom shrinkage) and two symmetric errors (checkout error, data capture error). Despite being in a continuous-time setting, all errors are transaction-based and multiplicative. The multiplicative

error formulation is realistic based on field interviews and technically more robust than additive error formulation (Khader et al., 2014). The goal of our model is to formally describe the interactions between inventory policies and operational errors and allow us to simulate the effects of these interactions on IRI.

Under imperfect execution, the shelf stock (S_a), in addition of being depleted by the withdrawal rate w_a (Eq. (2)), is affected by shelf error and shelf shrinkage. Shelf error (ϵ_s) arises when less than desired quantities are put onto the shelf. Under-shelving is not uncommon since store employees could forget items in the backroom or misplace items (Eroglu et al., 2013). Since in most cases, the retail shelf space for a particular item is fixed, our model excludes the possibility of over-shelving. The exclusion of over-shelving is consistent with our conversation with retail managers and our experience of walking through aisles with retail inventory auditors. The under-shelving error in our model is captured as a fraction of the shelving rate (s_a).

Shelf shrinkage (ϵ_{s-}) refers to shelf stock loss caused by shoplifting and unrecorded damaged products (Lee and Ozer, 2007). In essence, this shrinkage should be a function of the number of items in the shelf (a fraction of the current shelf contents). However, to be consistent with industry reporting practices (Hollinger, 2009) and our interviewees, we formulate this error as a fraction of the flows out of the stocks (% of sales). Both under-shelving (ϵ_s) and shelf shrinkage (ϵ_{s-}) are one-sided (non-negative) random errors, as they cause direct reduction of available stock. The rate of change of the shelf stock (S_a) (Eq. (1)) in the presence of errors becomes

$$\frac{dS_a}{dt} = s_a(1 - \epsilon_s) - w_a - w_a(\epsilon_{s-}) \quad (9)$$

The backroom is also exposed to undetectable stock loss. Backroom shrinkage (ϵ_{b-}) can be attributed to unobserved spoilage and employee theft, which is not unusual in the retail sector (Fan et al., 2014). Like shelf-shrinkage, and for the same reasons, we model backroom-shrinkage as a fraction of the outgoing flow. Under the fallible shelving operations shown in (9), the backroom stock (B_a) change rate (Eq. (4)) becomes

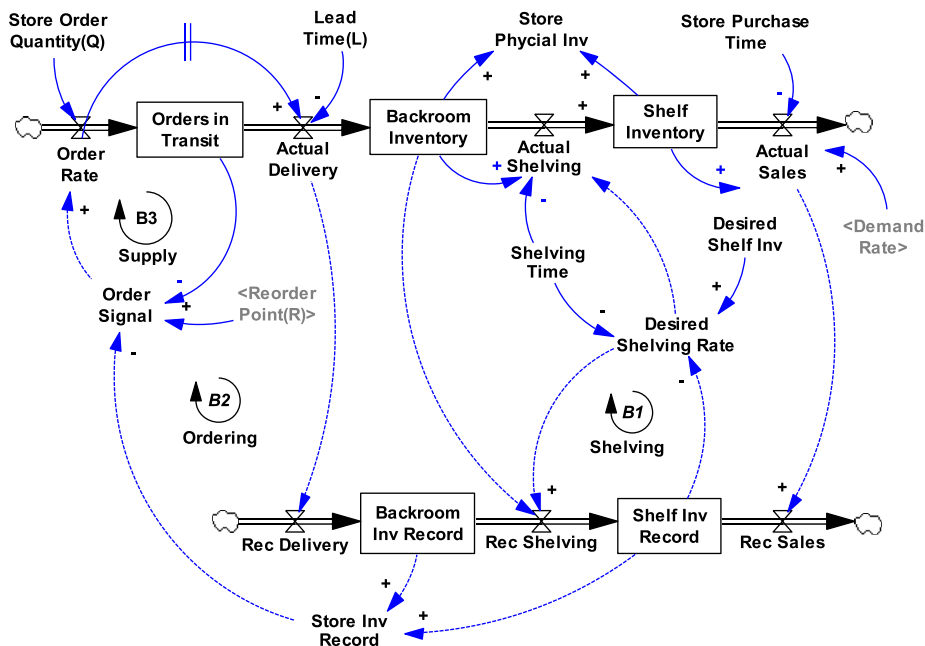


Fig. 1. Stock and flow diagram of store inventory management.

$$\frac{dB_a}{dt} = d_a - s_a(1 - \varepsilon_S) - s_a(\varepsilon_{B-}) \quad (10)$$

Similar to shelf error (ε_S) and shelf shrinkage (ε_{S-}), backroom shrinkage (ε_{B-}) is a one-sided and non-negative random error.

The formulations of (9) and (10) do not capture the case that the item is taken out of the backroom but never makes it onto the shelf due to ignorance or low engagement. Nonetheless, this error is emulated by backroom shrinkage (ε_{B-}) in (10) where the outflow of the backroom stock (B_a) is different from the inflow to the shelf stock (S_a).

We assume that the processing of inventory information is also subject to errors. The recorded shelf stock (S_r) can deviate from the shelf stock (S_a) because the recorded shelving rate (s_r) is distorted by the data capture error (ε_c) and the withdrawal rate (w_r) is distorted by checkout error (ε_w). A typical cause of checkout error is misclassification – scanning two different items as two of the same item. For instance, a customer purchases a chicken-flavored soup and a beef-flavored one. When the employee uses POS scanner to record sales, he may scan one flavor twice instead of scanning the items separately. Such time-saving behavior causes IRI (Raman, 2000). Following how Nachtmann et al. (2010) operationalize errors regarding point-of-sale data records in their simulation, we model information errors as a symmetric fraction of the withdrawal rate with a uniform distribution around a mean of zero. Under imperfect checkout, the recorded shelf stock change rate (S_r) (Eq. (8)) becomes

$$\frac{dS_r}{dt} = s_r - w_r = s_a(1 + \varepsilon_c) - w_a(1 + \varepsilon_w) \quad (11)$$

Note that the recorded shelving rate considers the actual shelving rate (without the shelving errors included) as the base for adjustment for the data capture error. Not only is this assumption consistent with the information that would be available to data entry personnel – assuming that the shelving errors were not intentional – but it also allows for an exploration of the interaction effects of operational and information-processing errors.

The rates affecting the recorded backroom stock (B_r) shown in equation (8) are also subject to information processing errors. Specifically, the outflow – recorded shelving rate (s_r) – can deviate from actual shelving rate (s_a) because of data capture error (ε_c), which occurs when store employees do not correctly record reshelving quantities, so the store fails to keep track of the true flow of products. While the recorded delivery rate could also deviate from the actual delivery rate, we left it unaffected for this analysis. In the presence of data capture error, the rate of change of B_r in (8) is modified into

$$\frac{dB_r}{dt} = d_r - s_r = d_a - s_a(1 + \varepsilon_c) \quad (12)$$

Similar to checkout error (ε_w), we model data capture error (ε_c) as a symmetric fraction of the corresponding rate with uniform distribution with zero mean. While our model could capture random errors with any known distribution, the uniform distribution setup is consistent with prior simulation studies that model random errors in system inventory levels (Angulo et al., 2004; Waller et al., 2006; Nachtmann et al., 2010).

Note that consistent with the goals for the model, i.e., to assess the impact of different sources of error, the model does not contain the structure to capture the traditional mechanisms to correct IRI: inventory auditing policies and discovery of shelf stockouts by walking around. While these efforts certainly have an effect on the resulting magnitude of IRI, we justify this model boundary choice by observing that any reasonable implementation of these efforts

would make them proportional to the IRI level, regardless of its source. Fixing IRI through audits only sets IRI back to zero at a point of time (i.e., interrupts the growth of IRI) and does not affect the underlying execution errors that persist to make IRI re-accumulate after record reconciliation. Thus, the inventory correction efforts would not have an impact on the relative impact of different operational error sources.

The performance metric – IRI – is the *average absolute difference* between physical (I_a) and system inventory levels (I_r) over the simulation horizon. Our focus on absolute deviation is consistent with prior studies (e.g., DeHoratius and Raman, 2008), and the average is more reliable than a single snapshot of IRI at a point of time. The IRI metric is calculated as

$$\frac{\int_0^T |I_a(t) - I_r(t)| dt}{T} \quad (13)$$

where T is the length of simulation. We also keep track of shelf-OOS (out-of-stock) ratio – the total duration of S_a being empty relative to T . Shelf-OOS ratio is a relevant metric that helps us better explain the impact of errors on IRI and defined as

$$\frac{\int_0^T \text{Ind}(S_a(t) < 1) dt}{T} \quad (14)$$

3.3. Experimental design

We built the model using VENSIM DSS (Ventana Systems, 2010) and simulated the model using Euler integration method for 360 days with an integration interval (dt) of 0.03125. Based on interviews with retail executives who provided data for our study, we set the model parameters as follows: $\delta = 10$ units/day, $Q = 100$ units, $S_a(0) = S^* = 50$ units, $\tau_p = 0.125$ day, and $\tau_s = 0.25$ day. The reorder point R is $S^* + \delta \times (\text{Safety Coverage} + L) = (50 + 10 \times (3 + 3)) = 110$. We assume initial inventory records to be accurate (i.e., $B_r(0) = B_a(0)$ and $S_r(0) = S_a(0)$) and $I_a(0) = S^* + R$. Note that those parameter values were not set for optimal restocking decisions (there are no cost considerations in our model), but selected to eliminate shelf-OOS under normal operations. Specifically, the desired shelf level (S^*) is set to be five times the daily demand, and reorder point (R) includes safety coverage despite the fact that our model has deterministic demand and reliable suppliers. The findings and insights are qualitatively the same given different parameter values. We tested the model under extreme conditions and a variety of scenarios (Forrester and Senge, 1980; Sterman, 2000). The model behavior is not affected by changes in simulation interval, and the main results are robust to changes in the assumptions for demand – i.e., introducing a random element on the demand, or changing its relative magnitude.

As an illustration of model behavior, we report in Fig. 2 the simulation results of a store that experiences shelf and backroom shrinkage. The figure shows the behavior of actual (I_a) and recorded inventory levels (I_r), as well as the shelf inventory position (S_a). As a result of the shrinkage, the actual inventory position drops below the recorded inventory. Since the recorded position is used to make shelving (and purchasing) decisions, the amount of inventory in the shelf drops. Eventually the shelf empties (day ~320) and the system falls into the “freezing” scenario that has been reported in previous works (Kang, 2004; Kang and Gershwin, 2005), i.e., the ordering mechanism is not triggered (frozen) as I_r is still greater than the reorder point R and there are no additional sales to diminish it. The rise in I_r over time is another known attribute of “freezing” in which each time shrinkage errors occur, the gap between I_r and I_a grows.

Since the reorder quantity is a fixed Q , the accumulation of error makes I_r rise to a point where I_r stays above R and remains unchanged (Kang and Gershwin, 2005).

In line with prior simulation studies related to IRI (e.g., Fleisch and Tellkamp, 2005; Waller et al., 2006; Nachtmann et al., 2010), we employ a full factorial design with three levels of experimental factors to evaluate principal and interaction effects of multiple errors on IRI. Documentation of the full model and simulations runs, according to system dynamics standards (Martinez-Moyano, 2012; Rahmandad and Sterman, 2012), are available in the electronic supplement of this paper.

Table 2 reports the levels of the errors that create variations in the system. We set the experimental values of the five random errors based on interviews with retail managers. For the shelf error, we could not find reference values from the literature and set the values following managers' estimates. Although Kang and Gershwin (2005) and Rekik et al. (2009) tested the impact of shrinkage in a range from 1% to 7% of inventory flows, we found these estimates to be too high for our research sites. In line with our interviewees' responses and the National Retail Security Survey (Hollinger, 2009), we restricted the range of each shrinkage error within 0%–2% of sales, for a combined expected shrinkage as high as 4% of sales. As for the two information-related errors, checkout and data capture, we specify symmetric and uniform distributions following prior studies that use error terms from $\pm 5\%$ to $\pm 15\%$ (Waller et al., 2006; Nachtmann et al., 2010), although we limited the worst case scenario to $\pm 10\%$ error, which was deemed more reasonable by our interviewees. This reduced error range is consistent with the checkout inaccuracy estimated by Zabriskie and Welch (1978) and the standard deviation of the resulting distribution is similar to the one used by Sahin and Dallery (2009). The full factorial design has a total of $3^5 = 243$ cases and we replicated each scenario 50 times to ensure stable statistical results.

The introduction of random errors on continuous time models like ours, however, requires special treatment since the numbers generated by random functions are independently and identically distributed (white noise). However, calling these functions on every simulation time step results on an oversampling of the independent process and yields stochastic disturbances with constant power spectral density, that, while having the appropriate distribution properties (correct mean and standard deviation), might be changing too quickly relative to the underlying process generating the disturbances. In reality, the processes generating the disturbances have some inertia, as real quantities cannot change infinitely fast and the future values of the disturbance depend on its history. Realistic stochastic processes with persistence are called

Table 2

Factors and experimental levels for simulation.

Factors	Corr. time	Level
Shelf error (ϵ_S)	5 days	$B = 0$ $L \sim \text{Uniform}[0, 0.01]$ $H \sim \text{Uniform}[0, 0.02]$
Shelf shrinkage (ϵ_{S-})	1 day	$B = 0$ $L \sim \text{Uniform}[0, 0.01]$ $H \sim \text{Uniform}[0, 0.02]$
Backroom shrinkage (ϵ_{B-})	3 days	$B = 0$ $L \sim \text{Uniform}[0, 0.01]$ $H \sim \text{Uniform}[0, 0.02]$
Checkout error (ϵ_W)	1/3 day	$B = 0$ $L \sim \text{Uniform}[-0.05, 0.05]$ $H \sim \text{Uniform}[-0.10, 0.10]$
Data capture error (ϵ_C)	5 days	$B = 0$ $L \sim \text{Uniform}[-0.05, 0.05]$ $H \sim \text{Uniform}[-0.10, 0.10]$

B: Base; L: Low; H: High.

“pink noise” and are characterized by a correlation time that determines the historic window that anchors the future values of the stochastic distribution (i.e., the inertia). We adopt the pink noise structure to model random execution errors as suggested by Sterman (2000).¹ We determine the correlation time τ_e (i.e., inertia) in the pink noise structure for the errors in our model based on each error's process characteristics. We set the correlation time of 5 days for the shelving (ϵ_S) and data capture (ϵ_C) errors since these operations take place simultaneously and with this expected frequency. Given that checkout errors are related to the operators, we assume a correlation time of one shift for the checkout errors (ϵ_W). As for the two shrinkage errors, we posit that the correlation time for shelf shrinkage (ϵ_{S-}) is smaller than the correlation time for backlog shrinkage (ϵ_{B-}) since the former is likely to be attributed to mishaps on the shop floor with intense customer contact, while the latter is caused by staff theft or operational errors. Accordingly, we set these correlation times to one and three days respectively. We discuss the effect of these distribution and correlation time assumptions when analyzing our experimental results below.

3.4. Results and analysis

We analyze simulation results using analysis of variance (ANOVA) by least-square fit (Ye et al., 2000) and assess the significance of each experimental factor using an F -test, which is widely used in simulation studies (Nachtmann et al., 2010). Table 3 shows the F -Ratios and p -values for the main effect of the five types of errors and the second order interactions on Average $|IRI|$.² All of the main effects are statistically significant and the adjusted R^2 for the full model is 0.90. Fig. 3 shows the response profile, with 95% confidence intervals, for the main effects of each error under the three treatment levels. To illustrate the interaction effects, we plot

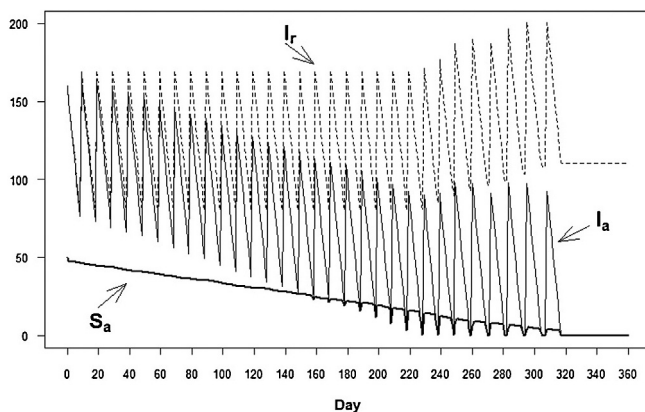


Fig. 2. Effect of shrinkage on inventory and shelf stocks.¹ Illustration assumptions $\epsilon_S \sim \text{Uniform}(0, 0.03)$ and $\epsilon_{B-} \sim \text{Uniform}(0, 0.04)$.

¹ To generate pink noise error that are uniformly distributed, we first generate standard normal random errors (i.e., white noise) every dt and filter the white noise to generate standard normal pink noise. By using standard normal white noise instead of uniform white noise as inputs, this structure improves Sterman (2000) and no longer relies on central limit theorem to generate normally distributed pink noise (Fiddaman, 2010). We transform the standard normal pink noise into uniform $[0, 1]$ pink noise numerically, using the probability integral transform (Rice, 2006) and then rescale into uniform $[0, a]$ or uniform $[-a, a]$.

² We also assessed performance using final IRI (i.e., $|I_a - I_r|$ at $t = 360$) and the results were statistically similar.

Table 3
Observed effects for each factor.

Source	F ratio	Prob > F
Shelving Error (SE)	226.45	<0.0001
Shelf Shrinkage (SS)	24458.78	<0.0001
Backroom Shrinkage (BS)	28520.67	<0.0001
Checkout Error (CO)	53.86	<0.0001
Data Capture Error (DC)	184.81	<0.0001
SE × SS	61.51	<0.0001
SE × BS	13.16	<0.0001
SE × CO	0.29	0.8861
SE × DC	3.69	0.0052
SS × BS	72.01	<0.0001
SS × CO	55.50	<0.0001
SS × DC	27.58	<0.0001
BS × CO	45.92	<0.0001
BS × DC	10.97	<0.0001
CO × DC	0.70	0.5938

the response profile under the three experimental levels for the two dominant errors, i.e., shelf and backroom shrinkage.

Inspection of results reveals that the shrinkage errors (ϵ_S and ϵ_B) dominate the response profile as they account for 98% of the observed variance. Both error terms are highly significant ($F = 24458.78$, p -value<0.0001 and $F = 28520.67$, p -value<0.0001 for shelf and backroom shrinkage) and have the expected direction of increasing the IRI as the shrinkage rates increases. The interaction term between these two errors is also the largest interaction term ($F = 72.01$, p -value<0.0001) and accounts for an additional 0.27% of the explained variance. Note that this dominance of the shrinkage errors is despite the fact that our worst-case scenario for shrinkage rates (0.02) is much lower than those tested in previous studies (e.g., Kang and Gershwin, 2005; Rekik et al., 2009). These two errors represent outflows from the physical supply line (one from the shelf stock and one from the backroom stock) that are not reflected on the information system thus directly affecting product availability but leaving inventory information unchanged. These two errors also account for 38% of the variance observed in shelf-OOS. Finally, it is interesting to note that the magnitude of the impact of these two errors is almost identical, indicating that there is no operational difference on where the shrinkage takes place.

The third operational error – shelving error (ϵ_S) – is statistically significant ($F = 226.45$, p -value<0.0001) but much less impactful than the two asymmetric shrinkage errors (it explains only 0.42% of the observed variance in IRI). The impact of the two information-processing errors is also statistically significant ($F = 53.86$, p -value<0.0001 and $F = 184.81$, p -value<0.0001 for checkout error and data capture error) but not substantial (0.44% of the observed variance). To our surprise, these three often-mentioned errors (Rekik et al., 2008; Sahin and Dallery, 2009) are dominated by the two shrinkage errors and reveal interesting interaction effects. As shown in the top panel of Fig. 3, when shrinkage errors are absent (B), the other three errors have positive (although weak) effects on IRI. However, their impacts are in the opposite direction (i.e., average IRI is decreasing in the error rate) when shrinkage errors are present (see the middle and bottom of Fig. 3). The negative interactions are explained by the fact that the simulated SKU ‘freezes’ more frequently (see Fig. 2 above) under higher shrinkage rates. The shelf-OOS likelihood jumps from 0.15 when there are no shrinkage errors (B) to 0.36 when shrinkage rates are set to 0.01 (L), and to 0.63 when both shrinkage rates are set to 0.02 (H) and these differences are highly significant when assessed using the *relative risk* (Morris and Gardner, 1988) (p -value<0.001 for L/B and p -value<0.001 for H/L). When shrinkage errors result in freezing and there are no inventory transactions, the associated shelving, checkout, and data capture errors are inactive, thus truncating the

growth of IRI. Even if the system does not reach persistent shelf-OOS, both sales and shelving flows will still be reduced by both shrinkage errors, thus weakening the overall impact of other three errors on IRI.

Note that the results reported above are robust to changes in the distribution assumptions for the random errors, and to changes in the correlation time for the asymmetric errors. While the average IRI does decrease with the increase in correlation time of the symmetric errors, the relative magnitude of the effects does not change. Please refer to this paper's electronic supplement for detailed results and discussion of these tests.

The observed results are consistent with the claim that shrinkage is more difficult to tackle than other sources of IRI (Lee and Ozer, 2007) and suggest that retail managers aiming to improve inventory data quality should pay attention to shelf as well as backroom shrinkage. In retrospect, this result is not surprising, as operational errors represent a deviation from the desired product flow and the disinformation related to these flows, whereas information-processing errors represent only information deviations. IRI is directly associated with information processing and as such, managers may tend to focus on fixing the symptoms by demanding information personnel to be more careful about data capture. However, our findings suggest that errors arising from physical operations can actually be the underlying causes of IRI and should be focal improvement targets. The electronic supplement also reports results to changes in the error distribution symmetry assumptions that support conjectures about error interactions put forward by Sahin and Dallery (2009) and Nachtmann et al. (2010).

Finally, the interaction profile graphs reveal that there are no strong interaction effects among operational errors in the system. This is confirmed by the fact that the eight (out of ten) statistically significant interaction terms explain only 1% of the observed variance in IRI. By specifying a model that explicitly accounts for conservation of matter and information distortions, we uncover that the effects of all the errors' interactions on IRI are almost linear. That said, a potential explanation for this lack of interactions is the fact that the model does not contain the reinforcing relationships between labor availability and existing IRI. We explore labor effects on data integrity in the next two sections.

4. Impact of labor availability on IRI

The prevalence of execution errors reflects the retailer's inability to manage not only its floor but also labor (DeHoratius and Raman, 2007). Evidently, the behaviors and skills of labor affect the occurrence of those five errors that are significantly associated with IRI. For instance, Zabriskie and Welch (1978) show that checkout accuracy is significantly associated with labor experience and attitude. Also, while transaction errors may be fixed through better scanning technologies, and misplacement/under-shelving may be tackled by shelf audits, the two most impactful shrinkage errors identified from simulation are closely related to employees' behaviors and monitoring efforts. Shrinkage (especially in the backroom) must be fixed from inside by employees who are capable of meeting process conformance standards. In line with Ton (2009), we posit that an easy way to tackle the primary drivers of IRI is to enhance process conformance by increasing labor capacity in retail stores. Allocating enough labor and retaining a stable mix of full-timers and part-timers relieves employees' workload and reduces poor execution of prescribed tasks that cause aforementioned errors. However, deploying sufficient labor with adequate skills is something fundamental but unfortunately ignored by retail managers who pursue payroll minimization (Fisher et al., 2009). In this section, we attempt to empirically assess the impact of labor availability on inventory data integrity. We explore this relationship

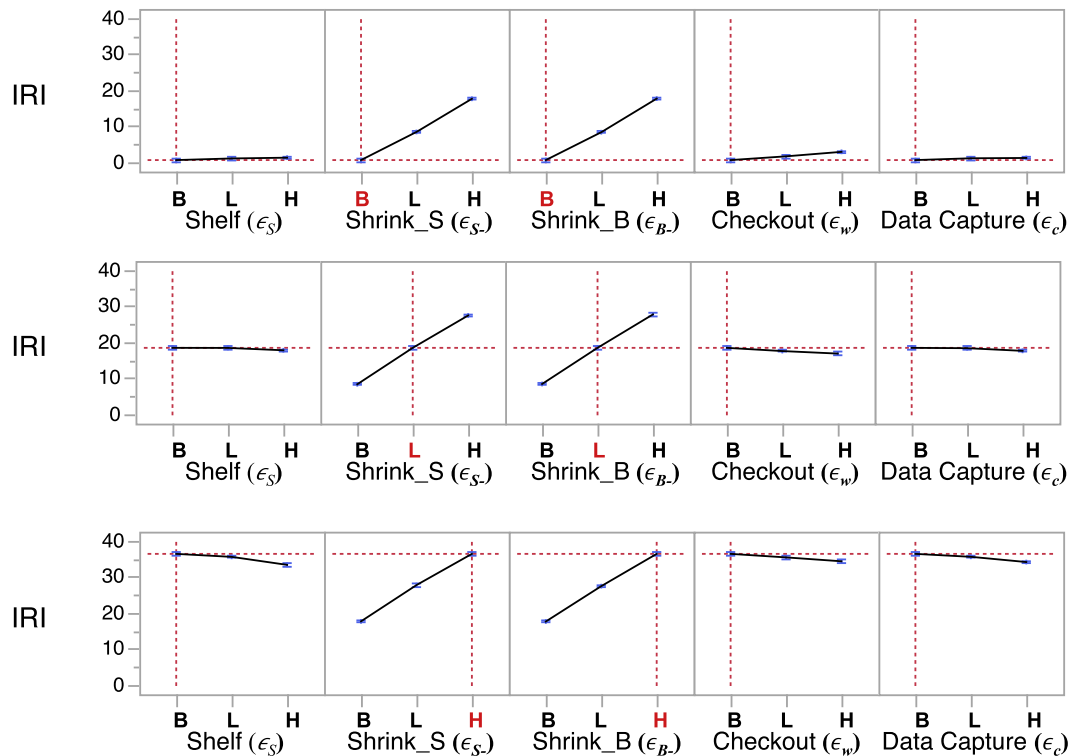


Fig. 3. Effects of errors on IRI. (Top: SS=B & BS=B; Middle: SS=L & BS=L; Bottom: SS=H & BS=H).

by measuring labor availability in terms of staffing levels and the employees' experience and training.

Sufficient staffing levels ensure seamless handling of customer services and in-store logistics, both of which are determinants of sales performance. Ingene (1982) finds a positive and linear association between staffing levels and sales volume per store. Three more recent studies (Fisher et al., 2006; Perdikiaki et al., 2012; Chuang et al., 2016) suggest that sales volume is a non-decreasing, concave function of staffing levels. While staffing levels are found to positively affect sales performance, understaffing is a norm rather than an anomaly in the retail sector due to the common practice of wage minimization (Fisher et al., 2009). Oliva and Sterman (2001) show that understaffing in a service context leads to fatigue, corner-cutting, and service quality erosion. Staffing levels also affect conformance quality that reflects how well store employees execute the prescribed processes. Ton (2009) finds that increasing staffing levels is positively associated with process conformance. In addition to its negative impact on service quality and conformance quality, understaffing potentially can undermine data quality. Since understaffing can cause personnel fatigue, hurried/mindless actions, and poor conformance, all the foregoing consequences of understaffing could lead to execution errors discussed in the simulation analysis. Those errors in information-processing and material handling increase the occurrence and magnitude of IRI. Finally, a fundamental approach to tackle IRI is to frequently inspect the shelves — through predefined cycle-counting or intensive random aisle walking. Frequent data audits are only possible when the organization has sufficient workforce so that employees are not overwhelmed with other tasks.

Simple head-count, however, is not enough to account for labor availability, as individuals' capabilities have an impact on their ability to perform a task. This is particularly salient in the retail industry where, due to volatile customer demands, retailers often bring in less-expensive and more flexible part-time-employees

(PTEs) to cover the short-term mismatches between required labor and the capacity available through full-time-employees (FTEs) (Kesavan et al., 2014). However, unlike FTEs, most PTEs receive limited training and do not accumulate as much experience on the job since they are the last to be hired and the first to be dismissed. As a result, PTEs are consistently assigned to supporting tasks such as re-stocking shelves. To date, there is scant evidence on how the mix of FTEs and PTEs affects store performance. A notable exception is Netessine et al. (2010) who find that allocating more FTEs as opposed to PTEs has positive effects on basket values.

The idea that more PTEs may be counterproductive or detrimental to store performance is not only because of lack of training/experience but also due to the fact that PTEs often obtain minimum benefits, have unstable schedules, and consequently, are less committed to their job (Netessine et al., 2010). Such a lack of commitment causes high turnover of PTEs. The high turnover, unfortunately, is exacerbated by the retailers' efforts to cut wages and eventually causes low process conformance (Ton and Raman, 2010). Ton (2014) calls the practice of cutting labor budgets by myopically hiring more PTEs a "bad job" strategy that creates a vicious cycle of low quality/quantity of people, poor operational execution, and lost sales/profits.

In the following subsections we describe the research setting, data collection, and operationalization of variables we used to explore the structural functional relationship between labor availability, considering labor mix, and IRI.

4.1. Data and measures

4.1.1. Inventory record inaccuracy

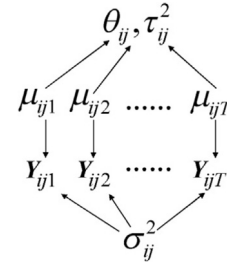
We obtained observations over a period of 10 months on IRI, labor, and time-variant (e.g., number of transactions) and time-invariant (e.g., store area) store characteristics from 5 stores of a global retailer of DIY supplies. A typical store has 100,000 square-

feet of retail area and carries 33,000 SKUs. Each store has 13 different product sections in which section managers coordinate the stores' internal operations. A section is a group of related SKUs that are normally placed in close proximity within the store. Section managers are responsible for maintaining the shelves, displaying the product (i.e., order, access, price labels, etc.), re-stocking the shelves from the backroom, and making adjustments to system-generated order decisions. Each section manager also has the autonomy to make labor allocations and variations on those decisions allow us to examine the effect labor availability has on data quality. Each section manager is also responsible for setting a monthly target of inventory counts to ensure data integrity, as all restocking decisions in the store are driven by an automated replenishment system.

Each of the 13 product sections had its own inspections in which a portion of SKUs with faulty inventory records (i.e., $|I_a - I_r| > 0$) was detected and corrected on a nearly daily basis. The inspection intensity varied from day-to-day and was subject to managers' orders as well as slack in labor capacity. Understandably, these inventory counts have low priority relative to the task of restocking shelves or helping customers and it is not infrequent that the actual counts are only a fraction of the stated goals. When performing shelf inspections, store associates only recorded error corrections along with I_a and I_r that were useful for managers to diagnose the severity of IRI.

One of the challenges of working with the data was to derive a robust measure for the dependent variable IRI for each store-section-month as category managers gave inspection different priority. Even within the same section, there was high variance in the inspection rate from month to month. For the 5 stores \times 13 sections \times 10 months = 650 units (from now on a unit refers to a store-section-month), we had a total of 66,525 corrections that can be used to estimate time-variant metrics of IRI. The number of corrections at month t in section j of store i (k_{ijt}) ranges from 0 to 1242 with a mean of 102 and a standard deviation of 152. One potential metric of IRI for each unit is mean absolute deviation (MAD), a simple and useful measure to capture the mean and spread of IRI (DeHoratius and Raman, 2008). However, the small number of correction records in many units (i.e., 200 out of 650 units have less than 30 correction records) makes MAD (i.e., raw IRI average) unreliable. Furthermore, all observed IRI are zero-truncated as only corrections, i.e., $|I_a - I_r| > 0$, were recorded. That is, we do not have record of all the inspections that were performed but resulted in no correction.

To tackle small-sample biases in many units and the censoring issue, we adopted Bayesian shrinkage³ estimation to make robust mean inferences given a small sample. This method has been used in marketing studies (e.g., Johnson et al., 2004; Shin et al., 2012) and aims to take advantage of information across subsamples. We defined $\mathbf{Y}_{ijt} = (Y_{ijt,1}, \dots, Y_{ijt,m(t)})$ as a vector of IRI corrections at month t in section j of store i . We assumed the sampling model $p(\mathbf{Y}_{ijt} | \mu_{ijt}, \sigma_{ij}^2) \sim$ zero-truncated Poisson-inverse Gaussian(μ_{ijt}, σ_{ij}^2) following Oliva et al. (2015) who showed that the Poisson-inverse Gaussian distribution is flexible enough to characterize the distribution of IRI at store, section, or subsection levels. Based on the fact that SKUs within a section are more similar and closely related to each other across months than to SKUs randomly sampled from the entire store, we set $\mu_{ijt} \sim \text{Normal}(\theta_{ij}, \tau_{ij}^2)$ (Hoff, 2009) and developed a Bayesian hierarchical model for each store-section (5 stores \times 13 sections). The proposed model does not allow between-



store interactions since it is unreasonable to assume that IRI of a store affects IRI of another store. Below is a visualization of the hierarchical model where θ_{ij} and τ_{ij}^2 are hyper-parameters that define the distribution of the means of the Poisson-inverse Gaussians.

The objective of each store-section model above is to estimate a posterior mean μ_{ijt} for a unit. The posterior mean is robust in that it allows μ_{ijt} with a small sample in one month to “borrow” some information from a larger sample of corrections in the same store-section but from other months. The information sharing also accommodates unknown inspection frequency since it is reasonable to assume proportionality between the number of observed corrections and inventory counts. The model also assumes a common variability parameter σ_{ij}^2 for \mathbf{Y}_{ijt} . Since all corrections are within the same year and from stores in the sample company with standard procedures, this assumption is reasonable and avoids more complicated parameterization (Hoff, 2009). Note that it is possible to devise a more complicated model that shares information content across 13 sections within each store. However, the model becomes cumbersome as the number of parameters to be estimated in a single model increases from $(T + 3)$ to $(T + 3) \times 13 + m$ additional hyper-parameters. Our model strikes a balance between quality of estimation and complexity of computation.

We derived the joint conditional distribution based on the set up – for convenience we drop the store (i) and section (j) indices:

$$P(\mu_1, \dots, \mu_T, \theta, \tau^2, \sigma^2 | \mathbf{Y}_1, \dots, \mathbf{Y}_T) \propto P(\theta)P(\tau^2)P(\sigma^2) \prod_{t=1}^T p(\mu_t | \theta, \tau^2) \prod_{t=1}^T \prod_k p(Y_{tk} | \mu_t, \sigma^2) \quad (15)$$

Based on equation (15), we developed a Markov Chain Monte-Carlo (MCMC) (Jackman, 2009) sampling scheme that allowed us to approximate the Bayes IRI average (i.e., posterior means of IRI $E[\mu_{ij} | \mathbf{Y}_{ij}]$) using simulation and derive the full conditional distribution for all model parameters – θ , τ , σ , and the key parameter μ (see the detailed derivation in the Appendix). The posterior mean vectors $E[\mu_{ij} | \mathbf{Y}_{ij}]$ are robust measures of data quality for all units and serve as dependent variables of regression later. Fig. 4 shows that the posterior mean differs more from the raw mean when the sample size of a unit is small (e.g., < 30). This correction of small sample bias comes from incorporating information from other months and reflects the effect of Bayesian shrinkage. The Bayesian and raw means of IRI become nearly identical when the sample size increases and more information is available.

Descriptive statistics in Table 4 show significant variability among units. Compared to Raw IRI average, the Bayes IRI average has a higher mean but a significantly smaller variance (p -value < 0.001), as would be expected from the shrinkage procedure (Hoff, 2009), which reduces variability by sharing observed information across periods within each store-section. The higher mean is also a direct consequence of shrinkage as a large number of corrections in a unit have more influence on the resulting value

³ This is a statistical shrinkage that refers to a Bayesian estimator and is not to be confused with inventory shrinkage described in §3.

than those units with low number of corrections. Note that the sample size of Raw IRI average and Bayes IRI average is only 617 because for 33 units we have no observed corrections ($k_{ijt} = 0$).

4.1.2. Labor availability and controls

We obtained monthly data on *total labor hours* (LabHrs) allocated to each store-section as well as the number of full-time employees (FTEs) and the part-time employees (PTEs) in a store-month. We used the staffing numbers to calculate the fraction of FTEs hours in store labor (FullFr) as $FTEs / (FTEs + PTEs \times \theta)$. Such a fractional operationalization based the number of persons is common in panel data econometrics and labor economics (McKee and West, 1984; Baltagi, 2006). The parameter θ ($0 \leq \theta \leq 1$) reflects the fact that a PTE may not work for as many hours as a FTE does. Based on our interviews with the director of operations and store managers, we set $\theta = 0.5$ and estimated the labor availability as the addition of full-time labor (FullTime) as $LabHrs \times FullFr$ and part-time labor (PartTime) as $LabHrs \times (1 - FullFr)$.

Finally, we identified relevant control variables from DeHoratius and Raman (2008) and secured their corresponding data. The first two controls are *product variety* and *inventory density*, both of which cause higher environmental complexity and are considered as drivers of IRI. We approximated product variety as the number of SKUs and inventory density as *inventory value* (€)/store area. Finally, since increased transactions may lead to more checkout errors, we also controlled for the number of transactions. All these controls were available at the store-month unit of analysis. Table 2 presents descriptive statistics and pairwise correlations of variables used for empirical estimation.

4.2. Results and analysis

Considering that IRI is the consequence of deviations in the stores' operating procedures, we assume that the improvement resulting from any efforts applied to improving the execution of those procedures would follow the well-established-empirical-finding of reducing defects at a constant fractional rate (Schneiderman, 1988; Keating et al., 1999). That is, IRI should show an exponential decline as incremental resources are applied to store's operations. In our case, we assumed that the ability to perform the store operations correctly is a function of the staffing level, and we specified a non-linear function that characterizes the impact of full-time and part-time labor on IRI

$$IRI = \alpha \exp(\beta_1 FullTime + \beta_2 PartTime + \pi X) \quad (16)$$

where the coefficients β_1 and β_2 indicate the impact of full-time and

part-time labor allocation on IRI. A negative coefficient indicates that IRI is reduced by the availability of that type of labor. X is an array of other factors that may affect IRI.

To econometrically estimate equation (16), we took natural logarithms of (16) and specified a two-way fixed effects (FE) model that considers store-section FE (α_{ij}) as well as time effects (D_t) (Cameron and Trivedi, 2010). FE modeling allows for the store-section-specific FE (α_{ij}) to be correlated with other regressors. Specifying α_{ij} helps absorb store- and section-specific time-invariant factors that may affect IRI. The model is also robust to the unbalanced panel data we have (since we did not have observations of IRI for all units).

$$\begin{aligned} \ln(BayesIRIAvg_{ijt}) = & \alpha_{ij} + \beta_1 FullTime_{ijt} + \beta_2 PartTime_{ijt} \\ & + \pi_1 ProductVar_{it} + \pi_2 InvDensity_{it} \\ & + \pi_3 Transactions_{it} + D_t + \omega_{ijt} \end{aligned} \quad (17)$$

Table 5 shows the results of de-mean estimation of (17) (Cameron and Trivedi, 2010). The standard errors shown in parentheses are robust to heteroskedasticity and allow for intra-panel correlations. We performed the Wooldridge test (Wooldridge, 2001) and find no evidence of autocorrelation in residuals (ω_{ijt}). To test the proposed functional form, we compared our log-linear model specification to the linear and inverse forms. The log-linear model in eq. (17) achieved a better fit than the other models, and performing the Box-Cox model specification test (Box and Cox, 1964) we failed to reject the log-linear model specification (p -value = 0.425) while rejecting the other two.

In model I we included full-time labor hours (FullTime) and part-time labor hours (PartTime) as well as time dummies. The regression is significant (p -value < 0.001), explains 60% of the observed variance and coefficients of both variables are significant at the 0.05 level. The coefficient β_1 for full-time labor is, as expected, negative, indicating that additional FTEs reduce the IRI. β_2 , however, is positive indicating that more PTEs increase the IRI. In models II–IV we controlled for product variety (ProductVar), inventory density (InvDensity), and number of transactions (Transactions). For all models only the control for Transactions is significant. For completeness we also performed random effects (RE) estimation (model V). RE modeling is more efficient but assumes α_{ij} to be a random number uncorrelated with other regressors. We conducted the Hausman test (Wooldridge, 2001) and find no systematic differences between FE and RE estimates. Further, we performed RE estimation (model VI) with Mundlak correction (Mundlak, 1978) for heterogeneity bias by adding two terms for the unit mean of FullTime and PartTime (Bell and Jones, 2015). The results of model VI are consistent with those from ordinary RE modeling (model V) and Transactions is still the only significant control.

The estimates for the two labor components are stable across all models, but part-time labor becomes insignificant when InvDensity and Transactions are included in the model. According to our field interviews, most PTEs were brought in to assist shelf restocking operations in busy season and the number of PTEs varied month by month, which is supported by a high ratio of *within* variance to *total* variance (0.60). However, each section has a stable base level of FTEs and the variance between sections is much larger than the variance within sections, yielding a within-to-total variance ratio of 0.02. The drop in significance of PartTime is explained by the inclusion of controls with a relatively high within-to-total variance ratio, i.e., InvDensity (0.63) and Transactions (0.15). Nevertheless, the remaining positive estimates of PartTime are consistent with store managers' perceptions that PTEs are not as well trained and more likely to make mistakes that cause IRI (e.g., grab the wrong

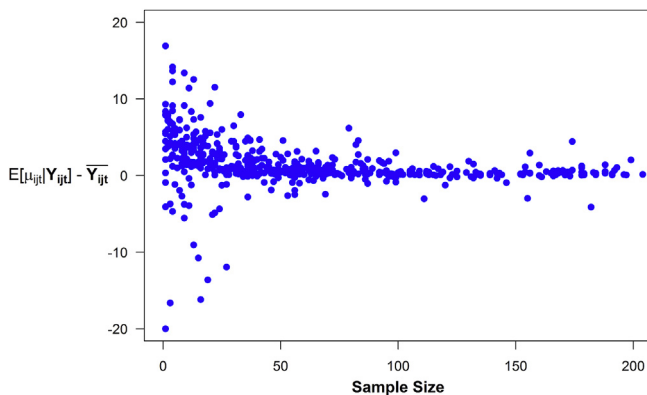


Fig. 4. Effect of Bayesian shrinkage.

Table 4
Descriptive statistics and pairwise correlations.

	Store-section-Month (<i>ijt</i>) variables									
	N	Mean	Stdev.	Min	Max	Labhrs	RawIRIAvg			
LabHrs	650	1071.09	546.10	217.88	3014.39					
RawIRIAvg	617	5.92	5.96	1.00	61.24	−0.11**				
BayesIRIAvg	617	7.09	4.66	1.11	38.99	−0.22**	0.77**			
	Store-Month (<i>it</i>) variables									
	N	Mean	Stdev.	Min	Max	FTEs	PTes	FullFr	ProductVar	InvDensity
FTEs	50	147.12	43.76	92.00	207.00					
PTes	50	46.92	16.71	18.00	85.00	0.26**				
FullFr	50	0.86	0.05	0.77	0.96	0.52**	−0.68**			
ProductVar	50	32346.95	3536.72	26131.00	40591.00	0.33**	0.62**	−0.28**		
InvDensity	50	1049.06	247.89	614.63	1668.73	0.14**	0.21**	−0.09*	0.27**	
Transactions	50	523298.20	159715.50	289732.00	876240.00	0.79**	0.56**	0.10*	0.61**	0.17**

*sig. 0.05 level; **sig. 0.01 level.

Table 5
Parameter estimates of panel data models.

Variable	I (FE)	II (FE)	III (FE)	IV (FE)	V (RE)	VI (RE_M)
Intercept	2.1717*** (0.1432)	1.8040*** (0.4456)	1.3467*** (0.5066)	0.4021 (0.6622)	1.1026** (0.4827)	1.3256** (0.5225)
FullTime	−0.0004** (0.0002)	−0.0004** (0.0002)	−0.0004** (0.0002)	−0.0004** (0.0001)	−0.0005*** (0.0001)	−0.0004*** (0.0002)
PartTime	0.0007** (0.0003)	0.0007** (0.0003)	0.0005 (0.0004)	0.0003 (0.0004)	0.0003 (0.0004)	0.0005 (0.0003)
ProductVar		0.0132 (0.0149)	0.0141 (0.0148)	0.0250 (0.0159)	0.0220 (0.0142)	0.0212 (0.0143)
InvDensity			0.0005 (0.0003)	0.0005 (0.0003)	0.0002 (0.0002)	0.0001 (0.0002)
Transactions				0.0016** (0.0006)	0.0009** (0.0003)	0.0007** (0.0003)
Time dummies	Included	Included	Included	Included	Included	Included
R-square	0.607	0.608	0.609	0.612	0.610	0.610
Adj. R-square	0.600	0.600	0.601	0.603	0.601	0.600
F-statistic	2.770***	2.540***	2.550***	2.650***		
Wald χ^2					44.350***	56.040***
No. of groups	65	65	65	65	65	65
No. of Obs	617	617	617	617	617	617

sig. 0.05 level; *sig. 0.01 level.

SKU from the backroom or place a particular SKU in the wrong place).

As a final check, we tested alternative values of θ for $\text{FullTime} = \text{FTEs}/(\text{FTEs} + \text{PTes} \times \theta)$ as different θ s would cause slight variations in the FullTime and PartTime estimated coefficients. We re-estimate the FE and RE models for each θ in the relevant range of [0.3, 1.0] with an increment of 0.005. The results remain unchanged and still provide significant evidence of the negative effect of full-time labor on IRI and the non-significant effect (when controls are included) of part-time labor on IRI.

5. Impact of IRI on labor availability

The results from the previous section provide clear evidence that labor availability has an effect of IRI. Furthermore, it is reasonably easy to identify scenarios where IRI, or its consequence, shelf-OOS, could have an impact on the workload of store employees. For example, as discussed in §3, once IRI is present, the restocking system becomes unreliable, increasing the probability of shelf-OOS. Empty shelves translate into confused or angry customers that interrupt normal operations and require special attention through help with searches (either physical or in the

computer system), trips to the back room, or placing orders for the missing SKUs. Eventually, high IRI will trigger a response from management that will require employees to perform inventory counts or walk-by shelf inspection. All these activities represent distractions from 'regular' operations, thus reducing the labor available to perform them. As per §4, lower labor availability translates into even higher IRI, thus creating a reinforcing loop. Fig. 5 operationalizes this reinforcing loop (R1) through the construct of work pressure – the ratio of labor required to effective

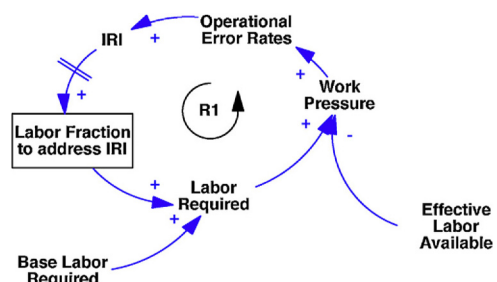


Fig. 5. IRI-Induced work pressure.

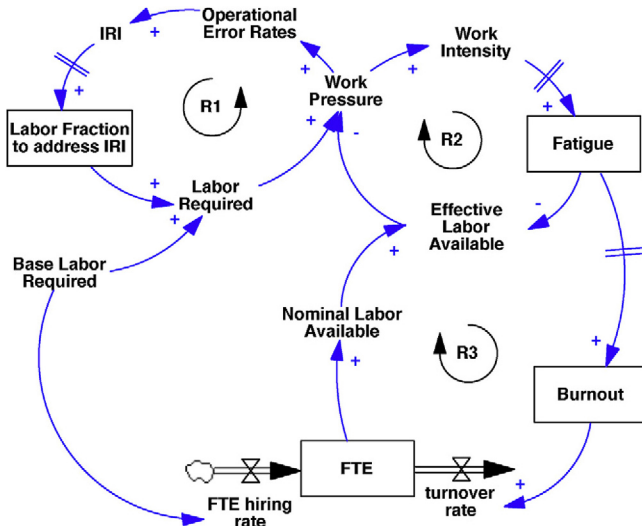


Fig. 6. IRI-Induced fatigue and burnout.

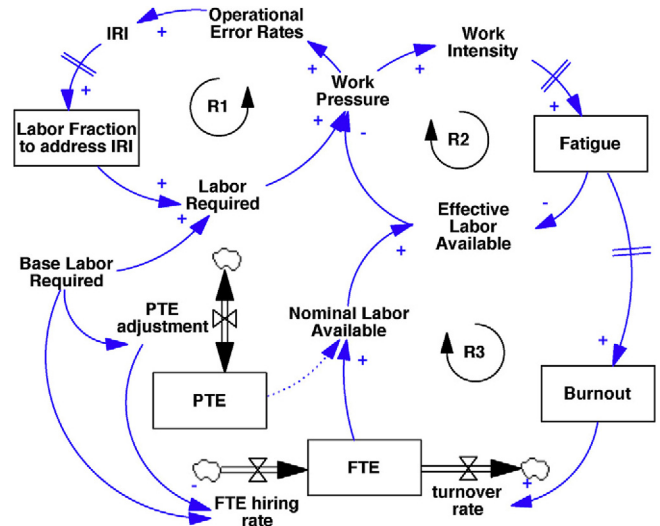


Fig. 7. Myopic PTE adjustment to labor availability.

labor available (Oliva, 2001).

Nevertheless, it is hard to imagine a situation where this reinforcing loop would run out of control and all employees will be responding to IRI-created work. Customers, for one, would quickly realize that the establishment is not the best option for their purchases and would search for alternative retailers. This would reduce the amount of base labor required to handle regular traffic and would bring work pressure back to normal operating range. However, a quick inspection of the shape and strengths of the links in loop R1 reveals that the gain of this loop saturates very quickly as only a limited fraction of total labor goes into IRI activities.⁴ Specifically, to assess the gain of the loop around the backlog-shrinkage error, we first used the model described in §3 to estimate the impact of the error rate on IRI. Simple OLS regressions explained 99% of the observed variance and yielded tight estimates for the error rate coefficients on average IRI ($\beta_{\epsilon_{B-}} = 907.0, p\text{-value} < 0.0001$). We further assumed that work pressure (WP) has a multiplicative effect on the operational error rates defined in §3, that is, $\epsilon_{B-}^{WP} = \epsilon_{B-} \times WP$, and that IRI can take up to 10% of the base labor required per SKU once it reaches the magnitude of the desired shelf level (S^*). Finally, we assumed that the impact of IRI is not felt instantaneously by the workforce, but takes time to accumulate. We approximated this adjustment through a first order exponential process with a time constant of one week. With those assumptions the gain of the loop around ϵ_{B-} is only 0.004 at the average observed IRI and assuming a normal initial work pressure.⁵ Even if all the other error rates were similarly activated, the addition of the five loop gains will still be less than one, and it would take a work pressure greater than 50, i.e., a severe understaffing, for loop R1 to have a gain greater than one. Note finally, that as formulated, the gain of the loop decreases with IRI as only a limited fraction of labor can realistically be allocated to handling IRI issues.

However, the feedback loop depicted in Fig. 5 is not working in

isolation and poor data integrity could induce negative consequences with different time delays. Fig. 6 shows two simple, yet relevant, feedback loops that characterize the intermediate and long-term impacts of work pressure on labor availability. The loop R2 suggests that IRI-induced work pressure could lead to higher work intensity. Extended periods of work intensity results in employees' fatigue, thus reducing their effectiveness and reinforcing work pressure. Extended periods of fatigue, in turn, results in employee burnout, which eventually increases FTE turnover rate, not only reducing headcount (R3), but also affecting the nominal labor productivity as new FTEs will require time to assimilate (not shown in the figure) (see Oliva and Sterman, 2001, 2010 for evidence and calibration of these feedback loops in service settings). Of course, IRI is not the only source of increasing work pressure: normal variations in customer arrivals, delays in the recruiting processes, and introduction of new product lines are just few examples of normal variations that could trigger the cascading effects on operational error rates and compound the growth of IRI.

The situation described above can easily be prevented if the effective labor available is maintained above the base labor required and work pressure is maintained below one. However, common hiring practices in the retail industry do not seem to be in line with this basic tenant. First, as shown in §4, there is no evidence that PTEs have an effect on improving IRI performance – if anything; there is weak evidence that they make things worse. Second, addressing IRI issues (helping customers, placing orders or realizing audits) is most likely done by FTEs, a reasonable assumption given the low training of PTEs. Consequently, the current practice of using PTEs to cover temporary staffing gaps has no substantial effect on the labor available for the purposes of the structure depicted in Fig. 6, as PTEs are not capable of containing the IRI from regular operations and are not capable of supporting the customer to deal with the consequences of IRI. We integrate this structural element in Fig. 7 and show in a dashed link the weak effect that PTEs have on nominal labor availability. While the effect of PTEs on the described dynamics is inconsequential, the fact that management perceives PTEs as being helpful and reduce the FTEs' hiring rate does have an effect on the dynamics as it sustains the work pressure that triggers all the reinforcement loops described above.

6. Discussion

Our study takes an empirically grounded multiple-modeling

⁴ The gain of a link is defined as the ratio of the output to the input, and the gain of a loop is the product of all the link gains in the loop. A reinforcing loop has a positive gain, but to create exponential growth the gain has to be greater than one.

⁵ The gain of the loop is given by the expression $\beta_i \epsilon_i e^{-\gamma IRI / S^*} \gamma \delta / \tau S^*$, where IRI is the average $|I_t - I_d|$ (~20 units), S^* is the desired shelf inventory level (50 units), ϵ_i is the base error rate for error type i , β_i is the estimated impact of the error rate on IRI (from regression of model output), τ is the time constant for the impact of IRI to affect labor requirements (7 days), δ is the maximum effect of IRI on labor requirements (10%) and γ is a parameter that controls the shape of the impact of IRI on labor requirements in the range $[0, \delta]$, in our case $\gamma = 5$.

approach to derive insights that help managers prevent occurrence of IRI. We began with modeling a continuous (Q, R) system with explicit considerations of recorded and actual inventories in the backroom and on the shelf. We used this model, through a full-factorial simulation experiment, to assess the impact five different operational errors that are thought to drive IRI. We set our error rates to values observed in our research context, but the impact of those rates in IRI was inferred through the model. We found strong support for the premise that inventory accuracy is vulnerable to operational errors. More importantly, while the literature (Kok and Shang, 2007; Chen and Mersereau, 2015) identifies three operational sources of IRI – shrinkage, transaction errors, and misplacement – no formal effort had been made to assess their relative impact. Our experimental design identified that all operational errors were significant, but we also found that transaction errors (e.g., checkout and data capture) and misplacement (e.g., under-shelving) have negligible impact on IRI and that shrinkage, either shelf or backroom, is the largest contributor to IRI.

We then empirically estimated the relationship between labor availability and IRI. We operationalized labor availability as a function of the level as well as mix of store workers. Our estimation model focused on a structural explanation of IRI, as opposed to a correlational study to test hypotheses, and we were able to identify significant labor effects on operational outcomes. While the negative association between FTEs and IRI conforms to the advocate that increasing FTEs improves operational execution (Ton, 2012, 2014), the partially significant positive impact of PTEs on IRI reveals the potential drawback of increasing temporary workers. Finally, we articulated the reinforcing relationships between labor and IRI by formally assessing the gain of the feedback loop based on our empirical findings and presented multiple feedback loops with the intermediate and long-term impact of IRI on labor availability. The feedback modeling exercise not only illustrates unanticipated consequences of IRI with different time delays but also reinforces the need for more effort and dedication to tackle IRI. Allocating adequate labor to prevent operational errors identified in §3 would be instrumental in enhancing process conformance and data accuracy. Our analysis of labor and IRI, however, does not rule out alternative drivers of IRI (e.g., those identified by DeHoratius and Raman, 2008). We simply aim to inform retail managers that employees matter when it comes to IRI and add to the newly emerging stream of studies on labor effects, store execution, and retail performance.

A major limitation of our paper is the generalizability of labor effects on retail data quality. Similar to DeHoratius and Raman (2008), we analyze secondary data from stores of a single retailer. Although the significance and appropriateness of random effects modeling in §4 help us generalize beyond this organization, statistically we may not be able to claim that our finding will hold beyond this retail chain. Another limitation of our analysis is that we only focus on quantities of labor. Labor quality could also moderate the effects of full-time and part-time labor on IRI. Although we could not obtain data on education, age, experience, engagement, and attitude of employees to assess such moderations, our conversation with the senior manager gives us no reason to suspect that significant differences in labor quality/training exist among the five stores. While we focused more on labor allocation, it would also be interesting to investigate the impact of section manager skills and background on the relationships identified in our paper.

Despite the limitations, our modeling efforts carry pragmatic implications for retail managers. Our findings suggest that management should re-direct their efforts to address shelf and backroom shrinkage. Extensive surveys from both US (Hollinger and

Davis, 2003) and European retail companies (Bamfield, 2004) suggest that employee, customer, and supplier theft cause approximately 80% of shrinkage – the other 20% being caused by internal errors (e.g., poor compliance, spoilage, accidentally damaging goods). While a technical fix for shrinkage would be item-level RFID (Lee and Ozer, 2007), this solution is too expensive to be adopted by most retailers. Instead, our findings suggest that improving labor availability could be an effective solution, at least in reducing employee theft and internal errors, which account for ~63% of shrinkage in US and ~47% of shrinkage in Europe (Bamfield, 2004). Ensuring sufficient staffing levels and adequate labor mix may be the easiest thing for retailers to do in order to tackle execution failures and IRI induced by those errors. That said, retail managers should not just expand labor capacity as doing so may lead to over-staffing. Managers ought to adopt prescriptive labor planning models that avoid myopic wage minimization and consider costs of errors/IRI associated with staffing levels and labor mix. For example, Chuang et al. (2016) propose a data-driven staffing heuristic based on observed store traffic. Following the principle of profit maximization, traffic-based labor planning significantly outperforms actual staffing decisions that are primarily budget- or cost-driven. Moreover, labor decisions should explicitly strike the balance between FTEs and PTEs rather than naively increasing PTEs to address staffing shortfalls. As mentioned earlier, lack of commitment and training may cause PTEs to commit errors that hamper data integrity more easily. For retailers who are not able/willing to increase staffing levels or change labor mix, labor quality can be improved by aligning incentives (Gino and Pisano, 2008), building awareness (Fisher, 2004), or developing the appropriate processes (Oliva and Watson, 2011).

Our study also carries theoretical implications for researchers. First, the proposed simulation model can be used to assess the impact of random errors on not only IRI but also lost sales. By expanding our model, researchers can reassess the backroom effects (Eroglu et al., 2013) in a generic setting subject to different execution errors while exploring the impact of decisions on shelf space, reorder point, order quantity/case pack size on inventory performance. Researchers can also design and test the efficacy of different shelf inspection policies (e.g., zero-balance walk, cycle-counting, daily counts) after adequately modifying the model. Moreover, since the errors investigated through simulation are common enough, the assumption that the decision maker knows the actual inventory level is not credible (Cachon, 2012). Our findings support the necessity to infer erroneous inventory records and incorporate statistical estimates into inventory control (DeHoratius et al., 2008; Mersereau, 2013).

Second, while previous studies show that sufficient staffing levels help improve service quality (Oliva and Stermann, 2001) and conformance quality (Ton, 2009), our empirical estimation suggests that full-time store workforce enhances data quality as well. The dependence between labor availability and data integrity warrants future investigation. More efforts are needed to articulate the impacts of labor mix as well. Even though PTEs could be an effective means of responding to demand spikes (Kesavan et al., 2014), inexperienced PTEs often need assistance from FTEs and this mentoring requires increasing the workload of FTEs. As a result, it becomes more difficult for FTEs to prioritize assignments and causes productivity loss (Oliva and Stermann, 2010). These dynamics arisen from labor allocation deserve further examination since it is important for managers in service settings to understand the pros and cons of building operational flexibility through mixed workforce.

Lastly, our combination of multiple methods to analyze IRI creates advances in empirical research. We enrich research angles

and insights applying simulation of differential equations, Bayesian statistics, panel data econometrics, and causal loop diagrams. On the empirical front, we are fairly cautious about imposing assumptions on the data generating process. Instead of following the predominant Gaussian thinking criticized by Singhal and Singhal (2012), we take an empirically-grounded approach to set up a sampling model in Bayesian shrinkage, which results in robust econometric estimation. From the system dynamics approach (Forrester, 1958; Sterman, 2000), we deploy two ideas that hopefully we will find broader adoption in operations management research. First, our empirical exploration of the relationship between labor and IRI is operational. That is, we focused on identifying the functional form of the relationship and providing operational explanations and consequences of our findings (Richmond, 1993). We believe that this type of analysis is more meaningful than hypotheses-testing correlational studies and is more conducive to generating operational recommendations. Second, our study expanded the boundary of the IRI problem to include the feedback mechanisms and managerial decisions that cause and sustain the problem. We believe that this broader perspective and the formal assessment of the feedback mechanism are becoming increasingly relevant as operations management is tackling ever broader issues and problems e.g., supply chain management, strategic operations, etc.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jom.2016.01.002>.

Appendix B. Bayesian shrinkage estimation

For convenience we drop the store (i) and section (j) indices. We first specify the following prior distributions for the four parameters in the Bayesian hierarchical model.

$$p(\theta) \sim \text{normal}(\theta_0, \gamma_0^2)$$

$$p\left(\frac{1}{\tau^2}\right) \sim \text{gamma}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right)$$

$$p\left(\frac{1}{\sigma^2}\right) \sim \text{gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

$$p(\mu_t | \theta, \tau^2) \sim \text{normal}(\theta, \tau^2)$$

Using equation (15) and the property of conjugacy, the full conditional distributions of θ and τ^2 belong to known parametric distributions (Hoff, 2009).

$$p(\theta | \cdot) \propto p(\theta) \prod_{t=1}^T p(\mu_t | \theta, \tau^2) \Rightarrow$$

$$p(\theta | \cdot) \sim \text{normal}\left(\frac{T\bar{\mu}/\tau^2 + \theta_0/\gamma_0^2}{T/\tau^2 + 1/\gamma_0^2}, \frac{1}{T/\tau^2 + 1/\gamma_0^2}\right)$$

$$p(\tau^2 | \cdot) \propto p(\tau^2) \prod_{t=1}^T p(\mu_t | \theta, \tau^2) \Rightarrow$$

$$p\left(\frac{1}{\tau^2} | \cdot\right) \sim \text{gamma}\left(\frac{\eta_0 + T}{2}, \frac{\eta_0 \tau_0^2 + \sum_{t=1}^T (\mu_t - \theta)^2}{2}\right)$$

The full conditional distributions of σ^2 and μ_t are only known proportionally.

$$p(\sigma^2 | \cdot) \propto p(\sigma^2) \prod_{t=1}^T \prod_k p(Y_{tk} | \mu_t, \sigma^2)$$

$$p(\mu_t | \cdot) \propto p(\mu_t | \theta, \tau^2) \prod_k p(Y_{tk} | \mu_t, \sigma^2)$$

Given the two closed-form full conditionals and the two proportionality conditions, we use a Metropolis-Hasting within-Gibbs sampler (Hoff, 2009) to derive posterior estimates. Take the mean parameter μ_t for example, the algorithm works in the following way.

1. Sample a proposal value μ_t^* from $J(\mu_t | \mu_t^{(s)})$
2. Compute the acceptance ratio $r = \frac{J(\mu_t^{(s)} | \mu_t^*) P(\mu_t^* | Y)}{J(\mu_t^* | \mu_t^{(s)}) P(\mu_t^{(s)} | Y)}$
3. Let $\mu_t^{(s+1)} = \begin{cases} \mu_t^* & \text{with probability } \min(r, 1) \\ \mu_t^{(s)} & \text{with probability } 1 - \min(r, 1) \end{cases}$

The selection of proposal distributions is critical to the implementation of the algorithm. For $p(\sigma^2 | \cdot)$ and $p(\mu_t | \cdot)$ that cannot be directly sampled from known distributions, we use *random walk* proposals (Hamada et al., 2008).

$$\log((\mu_t)^*) = \log((\mu_t)^{s-1}) + \delta_1 \text{ where } \delta_1 \sim \text{Normal}(0, s_1^2)$$

$$\log((\sigma^2)^*) = \log((\sigma^2)^{s-1}) + \delta_2 \text{ where } \delta_2 \sim \text{Normal}(0, s_2^2)$$

The parameters s_1 and s_2 need to be fine-tuned to ensure the effective transition of Markov chains. The acceptance rates of the two proposal distributions are important performance measures of the Metropolis–Hastings step. A rule of thumb is that acceptance rates should fall between 25% and 50%. We set $s_1 = 1$ and $s_2 = 2$, which result in successful convergence of MCMC according to different metrics (e.g., acceptance rates, trace plots, autocorrelation).

References

- Agrawal, P.M., Sharda, R., 2012. Impact of frequency of alignment of physical and information system inventories on out of stocks: a simulation study. *Int. J. Prod. Econ.* 136 (1), 45–55.
- Anderson, E.G., 2001. The nonstationary staff-planning problem with business cycle and learning effects. *Manag. Sci.* 47 (6), 817–832.
- Angulo, A., Nachtmann, H., Waller, M.A., 2004. Supply chain information sharing in a vendor-managed inventory partnership. *J. Bus. Logist.* 25 (1), 101–120.
- Baltagi, B.H., 2006. *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*. Emerald Group Publishing, Bingley, UK.
- Bamfield, J., 2004. Shrinkage, shoplifting and the cost of retail crime in Europe: a cross-sectional analysis of major retailers in 16 European countries? *Int. J. Retail Distrib. Manag.* 32 (5), 235–241.
- Bell, A., Jones, K., 2015. Explaining fixed effects: random effects modeling of time-series cross-sectional and panel data. *Political Sci. Res. Methods* 3 (1), 133–153.
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *J. R. Stat. Soc. Ser. B* 26 (2), 211–252.
- Boyer, K.K., Swink, M.L., 2008. Empirical elephants-why multiple methods are essential to quality research in operations and supply chain management. *J. Op. Manag.* 26 (3), 337–348.
- Cachon, G.P., 2012. What is interesting in operations management? *Manuf. Serv. Op. Manag.* 14 (2), 166–169.
- Cameron, A.C., Trivedi, P.K., 2010. *Microeconometrics Using STATA*, second ed. Stata Press, College Station, TX.
- Chen, L., Mersereau, A.J., 2015. Analytics for operational visibility in the retail store: the cases of censored demand and inventory record inaccuracy. In: Agrawal, N., Smith, S.A. (Eds.), *Retail Supply Chain Management*. International Series in Operations Research and Management Science, vol. 223. Springer, New York, pp. 79–112.

- Chuang, H.H.C., Oliva, R., Perdikaki, O., 2016. Traffic-based labor planning in retail stores. *Prod. Op. Manag.* 25 (1), 96–113.
- DeHoratius, N., Raman, A., 2007. Store manager incentive design and retail performance: an exploratory investigation. *Manuf. Serv. Op. Manag.* 9 (4), 518–534.
- DeHoratius, N., Raman, A., 2008. Inventory record inaccuracy: an empirical analysis. *Manag. Sci.* 54 (4), 627–641.
- DeHoratius, N., Mersereau, A.J., Schrage, L., 2008. Retail inventory management when records are inaccurate. *Manuf. Serv. Op. Manag.* 10 (2), 257–277.
- Eroglu, C., Williams, B.D., Waller, M.A., 2013. The backroom effect in retail operations. *Prod. Op. Manag.* 22 (4), 915–923.
- Fan, T., Chang, X., Gu, C., Yi, J., Deng, S., 2014. Benefits of RFID technology for reducing inventory shrinkage. *Int. J. Prod. Econ.* 147 (Part C), 659–665.
- Fiddaman, T., 2010. *MetaSD Model Library*. <http://models.metasd.com/pink-noise/>.
- Fisher, M.L., 2004. To me it's a store. To you it's a factory. *ECR J.* 4 (2), 9–18.
- Fisher, M.L., Krishnan, J., Netessine, S., 2006. Retail store execution: an empirical study. In: Working Paper. The Wharton School, University of Pennsylvania, Philadelphia.
- Fisher, M.L., Krishnan, J., Netessine, S., 2009. Are your staffing levels correct? *Int. Commer. Rev.* 8 (2–4), 110–115.
- Fleisch, E., Tellkamp, C., 2005. Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain. *Int. J. Prod. Econ.* 95 (3), 373–385.
- Forrester, J.W., 1958. Industrial dynamics – a major breakthrough for decision makers. *Harv. Bus. Rev.* 36 (4), 37–66.
- Forrester, J.W., Senge, P.M., 1980. Test for building confidence in system dynamics models. *TIMS Stud. Manag. Sci.* 14, 209–228.
- Gino, F., Pisano, G., 2008. Toward a theory of behavioral operations. *Manuf. Serv. Op. Manag.* 10 (4), 676–691.
- Hamada, M.S., Wilson, A., Reese, C.S., Martz, H., 2008. *Bayesian Reliability*. Springer, New York.
- Heese, H.S., 2007. Inventory record inaccuracy, double marginalization, and RFID adoption. *Prod. Op. Manag.* 16 (5), 542–553.
- Hoff, P.D., 2009. *A First Course in Bayesian Statistical Methods*. Springer, New York.
- Hollinger, R.C., Davis, J.L., 2003. 2002 National Retail Security Survey: Final Report. University of Florida, Gainesville, FL.
- Hollinger, R., 2009. National Retail Security Survey. University of Florida, Gainesville, FL.
- Ingrene, C.A., 1982. Labor productivity in retailing. *J. Mark.* 46 (4), 75–90.
- Jackman, S., 2009. *Bayesian Analysis for the Social Sciences*. Wiley, New York.
- Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S., Lohse, G.L., 2004. On the depth and dynamics of online search behavior. *Manag. Sci.* 50 (3), 299–308.
- Kang, Y., 2004. *Information Inaccuracy in Inventory Systems*. Unpublished PhD Dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- Kang, Y., Gershwin, S.B., 2005. Information inaccuracy in inventory systems: stock loss and stockout. *IIE Trans.* 37 (9), 843–859.
- Kapoor, G., Zhou, W., Piramuthu, S., 2009. Challenges associated with RFID tag implementations in supply chains. *Eur. J. Inf. Syst.* 18 (6), 185–205.
- Keating, E.K., Oliva, R., Repenning, N.P., Rockart, S.F., Sterman, J.D., 1999. Overcoming the improvement paradox. *Eur. Manag. J.* 17 (2), 120–134.
- Kesavan, S., Staats, B.R., Gilland, W.G., 2014. Labor-mix and flexible response to demand spikes: evidence from a retailer. *Manag. Sci.* 60 (8), 1884–1906.
- Khader, S., Rekik, Y., Botta-Genoulaz, V., Campagne, J., 2014. Inventory management subject to multiplicative inaccuracies. *Int. J. Prod. Res.* 52 (17), 5055–5069.
- Kok, A.G., Shang, K.H., 2007. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manuf. Serv. Op. Manag.* 9 (2), 185–205.
- Kok, A.G., Shang, K.H., 2014. Evaluation of cycle-count policies for supply chains with inventory inaccuracy and implications for RFID investments. *Eur. J. Operational Res.* 237 (1), 91–105.
- Lee, H.L., Ozer, O., 2007. Unlocking the value of RFID. *Prod. Op. Manag.* 16 (1), 40–64.
- Lyneis, J.M., Ford, D.N., 2007. System dynamics applied to project management: a survey, assessment, and directions for future research. *Syst. Dyn. Rev.* 23 (2–3), 157–189.
- Martinez-Moyano, I.J., 2012. Documentation for model transparency. *Syst. Dyn. Rev.* 28 (2), 199–208.
- McKee, M., West, E.G., 1984. Minimum wage effects on part-time employment. *Econ. Inq.* 22 (3), 421–428.
- Mersereau, A.J., 2013. Information-sensitive replenishment when inventory records are inaccurate. *Prod. Op. Manag.* 22 (4), 792–810.
- Morris, J.A., Gardner, M.J., 1988. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br. Med. J.* 296, 1313–1316.
- Mundlak, Y., 1978. Pooling of time-series and cross-section data. *Econometrica* 46 (1), 69–85.
- Nachtmann, H., Waller, M.A., Rieske, D.W., 2010. The impact of point-of-sale data inaccuracy and inventory record data errors. *J. Bus. Logist.* 31 (1), 149–158.
- Netessine, S., Fisher, M.L., Krishnan, J., 2010. Labor planning, execution, and retail store performance: an exploratory investigation. In: Working Paper. The Wharton School, University of Pennsylvania, Philadelphia.
- Oliva, R., 2001. Tradeoffs in responses to work pressure in the service industry. *Calif. Manag. Rev.* 43 (4), 26–43.
- Oliva, R., Sterman, J.D., 2001. Cutting corners and working overtime: quality erosion in the service industry. *Manag. Sci.* 47 (7), 894–914.
- Oliva, R., Sterman, J.D., 2010. Death spirals and virtuous cycles: Human resource dynamics in knowledge-based services. In: Maglio, P.P., Kieliszewski, C.A., Spohrer, J.C. (Eds.), *Handbook of Service Science*. Springer, New York, pp. 321–358.
- Oliva, R., Watson, N.H., 2011. Cross functional alignment in supply chain planning: A case study of Sales & Operations Planning. *J. Op. Manag.* 29 (5), 434–448.
- Oliva, R., Chuang, H.H.C., Kumar, S., 2015. Group-level information decay and item-level error distribution on retail inventory: empirical formulation with inspection optimization. In: Working Paper. Mays Business School. Texas A&M University, College Station, TX.
- Perdikaki, O., Kesavan, S., Swaminathan, J.M., 2012. Effect of traffic on sales and conversion rates of retail stores. *Manuf. Serv. Op. Manag.* 14 (1), 145–162.
- Rahmandad, H., Sterman, J.D., 2012. Reporting guidelines for simulation-based research in social sciences. *Syst. Dyn. Rev.* 28 (4), 396–411.
- Raman, A., 2000. Retail-data quality: evidence, causes, costs, and fixes. *Technol. Soc.* 22 (1), 97–109.
- Raman, A., DeHoratius, N., Ton, Z., 2001. Execution: the missing link in retail operations. *Calif. Manag. Rev.* 43 (3), 136–152.
- Redman, T., 1995. Improve data quality for competitive advantage. *Sloan Manag. Rev.* 36 (2), 99–107.
- Rekik, Y., Sahin, E., Dallery, Y., 2008. Analysis of the impact of the RFID technology on reducing product misplacement errors at retail stores. *Int. J. Prod. Econ.* 112 (1), 264–278.
- Rekik, Y., Sahin, E., Dallery, Y., 2009. Inventory inaccuracy in retail stores due to theft: an analysis of the benefit of RFID. *Int. J. Prod. Econ.* 118 (1), 189–198.
- Rice, J.A., 2006. *Mathematical Statistics and Data Analysis*. Thomson Higher Education, Belmont, CA.
- Richmond, B., 1993. Systems thinking: critical thinking skills for the 1990s and beyond. *Syst. Dyn. Rev.* 9 (2), 113–133.
- Sahin, E., Dallery, Y., 2009. Assessing the impact of inventory inaccuracies within a newsorder framework. *Eur. J. Operational Res.* 197 (3), 1108–1118.
- Sheppard, G.M., Brown, K.A., 1993. Predicting inventory record-keeping errors with discriminant analysis: a field experiment. *Int. J. Prod. Econ.* 32 (1), 39–51.
- Shin, S., Misra, S., Horsky, D., 2012. Disentangling preferences and learning in brand choice models. *Mark. Sci.* 31 (1), 115–137.
- Singhal, K., Singhal, J., 2012. Opportunities for developing the science of operations and supply-chain management. *J. Op. Manag.* 30 (3), 245–252.
- Schneiderman, A., 1988 April. Setting quality goals. *Qual. Prog.* 55–57.
- Sterman, J.D., 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill, Boston.
- Thiel, D., Hovelque, V., Le Hoa, V.T., 2010. Impact of inventory inaccuracy on service-level quality in (Q,R) continuous-review lost-sales inventory models. *Int. J. Prod. Econ.* 123 (2), 301–311.
- Ton, Z., 2009. The effect of labor on profitability: the role of quality. In: Working Paper. Harvard Business School, Boston.
- Ton, Z., 2012 January–February. Why good jobs are good for retailers? *Harv. Bus. Rev.* 90 (1–2), 124–131.
- Ton, Z., 2014. *The Good Jobs Strategy: How the Smartest Companies Invest in Employees to Lower Costs and Boost Profits*. New Harvest, New York.
- Ton, Z., Raman, A., 2010. The effect of product variety and inventory levels on retail store sales: a longitudinal study. *Prod. Op. Manag.* 19 (5), 546–560.
- Ventana Systems, Inc., 2010. *Vensim DSS Software 5.11a*. Ventana Systems Inc., Harvard, MA.
- Waller, M.A., Nachtmann, H., Hunter, J., 2006. Measuring the impact of inaccurate inventory information on a retail outlet. *Int. J. Logist. Manag.* 17 (3), 355–376.
- Wooldridge, J.M., 2001. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Boston.
- Ye, C., Liu, J., Ren, F., Okafo, N., 2000. Design of experiment and data analysis by JMP (SAS institute) in analytical method validation. *J. Pharm. Biomed. Analysis* 23 (2–3), 581–589.
- Zabriskie, N.B., Welch, J.L., 1978. Retail cashier accuracy: misreadings and some factors related to them. *J. Retail.* 54 (1), 43–50.