Inventory Management with Purchase Order Errors and Rework

Yan Jiang[*] , Nicole DeHoratius[†] , Diego Klabjan[‡]

August 10, 2012

Abstract

In a retail distribution center, fulfillment of purchase orders (POs) occasionally incurs errors retailers can correct through rework. PO fulfillment errors, or PO errors, include "ticket errors," whereby incorrect information (e.g., selling price) is put on tickets attached to some PO items. These errors are costly and different from random yield errors. This paper combines empirical and analytical methods to study the influence of PO errors on retailers' order policy and cost. Using data from a large retail chain, we determine key properties of these errors, in particular, how they vary with order complexity, including average order quantity per SKU and number of SKUs per PO. We propose one deterministic and one stochastic (Q,R) inventory model for POs with a single SKU that account for these correctable errors. We study the deterministic model analytically, and the stochastic model numerically, with parameters estimated from real retail data. We compare the performance of our adjusted order policy with standard policies that ignore correctable PO errors, and provide qualitative guidance in identifying SKUs more prone to these errors.

[*]E-mail: yanjiang2008@u.northwestern.edu; Department of Industrial Engineering and Management Sciences, Northwestern University

[†]E-mail: ndehoratius@zlc.edu.es; Zaragoza Logistics Center

[‡]E-mail: d-klabjan@northwestern.edu; Department of Industrial Engineering and Management Sciences, Northwestern University

1 Introduction

Retailers replenish inventory through supply networks, composed of vendors, retail distribution centers (DCs), and retail stores. Products ordered by retailers are usually shipped from vendors to DCs, which distribute them to retail stores. Inventory management at DCs is complicated and challenging not only because of the large variety of products and long lead times, but also due to vendor deliveries at variance with the order contract. Such vendor errors, which incur additional labor cost and complicate retail operations, are termed vendor non-compliance problems in the business literature.

Vendor non-compliance is penalized through chargebacks, which represent a significant cost for most vendors. The Credit Research Foundation reports that from 4% to 10% of all open items on accounts receivables are affected by chargeback deductions, which typically reduce overall vendor revenue by $2\% - 10\%$, according to the National Chargebacks Management Group (NCMG) of Charlotte, North Carolina (Zieger (2003)). Chargebacks also reflect, to some extent, the considerable cost to retailers occasioned by vendor non-compliance (Aron (1998)). Chargebacks, associated with vendor non-compliance problems, not only add unnecessary costs for retailers and vendors alike, but also make the inventory supply process less efficient. Retailers should thus work with vendors towards the elimination of non-compliance problems. Because, in practice, total elimination may not be feasible, retailers can adjust their inventory policies to account for vendor non-compliance. We examine how retailers might adjust their inventory policies in the presence of vendor non-compliance problems.

Vendor non-compliance problems are observed in approximately 5% of the roughly 4,000 purchase orders (POs) generated annually by Omega[1], the large retail chain that is the source of the data for the present study. Under vendor non-compliance, we include a range of PO fulfillment errors, such as quantity shortage errors and ticket errors. We henceforth refer to PO fulfillment errors as simply PO errors. A "ticket error" occurs when a vendor associates an incorrect stock keeping unit (SKU) number or selling price with items in a PO. Errors in fulfillment of POs at DCs results in additional labor cost, lead time, and lead time variability for retailers. Manufacturers incur millions of dollars in vendor chargebacks for these PO errors that add cost and complicate retail inventory management.

Only one type of PO error, random yield, has been extensively studied in the literature. Research on other types of PO errors in the retail setting is, to the best of our knowledge, limited. The substantial chargebacks imposed on vendors suggest that retailers have long realized the importance of dealing with PO errors. Yet, retailers have only partial knowledge of the real cost of PO errors, and do not know how to take these PO errors into account in their inventory policy. To help retailers understand the influence of PO errors on inventory policy and cost, our research attempts to identify (1) "PO error cost," defined as the additional cost a retailer incurs due to PO errors, (2) a new optimal order policy that anticipates PO errors that lead to rework, and (3) "potential cost savings," defined as the amount a retailer can save by adjusting order policies to account for PO errors. Because a retailer may need to manage hundreds of thousands of SKUs, it is useful, in practice,

[1] We disguise its name at the request of the retailer.

to identify SKUs with high potential cost savings, as there will be a per SKU cost to adjust and fine-tune an existing inventory system to account for PO errors. Under the emerging business model software-as-a-service, for example, inventory management software vendors like Predictix[2] charge retailers per SKU to adjust the standard inventory system. Retailer may thus prefer to incur this cost only for SKUs with high potential cost savings, using SKU characteristics (such as inventory cost parameters and demand estimates) to select such SKUs. It is also useful for retailers to identify SKUs for which PO errors are particularly costly without adjusting their existing inventory management systems. Retailer could then collaborate with their vendors to reduce the probability and magnitude of errors for those SKUs. To avoid paying inventory software vendors to calculate the cost of PO errors for each SKU, retailers need a procedure for choosing such SKUs based on their properties. Our research endeavors to provide such procedures together qualitative guidance in how to identify SKUs with high potential cost savings or PO error cost.

We consider retail inventory management of a single item under continuous review in the presence of PO errors eligible for rework at the retail DC. Properties of these errors, in particular, how they vary with order complexity, are obtained from data provided by our partner retailer. Order complexity consists of two components, average order quantity per SKU, and number of SKUs on a PO. We find the incidence and magnitude of PO errors to increase in both the average order quantity of SKUs and number of SKUs on a PO, and the properties of these correctable PO errors to differ from standard assumptions in the random yield literature. We further find that rework time for a PO increases in the error size of the PO. Using these new properties of PO errors, we first examine a deterministic inventory model analytically to obtain insights about the influence of PO errors. We show the optimal order quantity when accounting for PO errors to be less than the standard optimal order quantity, and both potential cost savings and PO error cost to decrease with fixed ordering cost, and increase in holding and backlogging cost as well as in demand rate. These properties provide qualitative guidance for identifying SKUs with high potential cost savings or PO error cost.

We then apply the single-item continuous time stochastic inventory model under a (Q, R) policy, with random yield and rework to reflect the properties of PO errors observed in the empirical data analysis. Due to the complexity of the stochastic model, we analyze the influence of PO errors via numerical experiments, the parameters for which are estimated using a proprietary data set from our partner retailer. The numerical experiments yield insights similar to those derived from the deterministic model. In addition, we find that PO errors cost our partner retailer tens to hundreds of dollars per transaction (i.e., per PO), and that whether our partner retailer under or over charges its vendors for these PO errors depends on the inventory cost parameters (including fixed ordering cost, holding cost, and backlogging penalty cost). We also observe substantial potential cost savings when retailers account for these PO errors in their inventory policies. Predictive models of potential cost savings and PO error cost, which are essential components of the procedures for identifying SKUs with higher potential cost savings or PO error cost, are generated through regression analysis using results from the

[2]Predictix Inc., a retail consulting firm headquartered in Atlanta, GA, serves companies in the retail, wholesale and consumer packaged goods industries. Service areas include inventory planning and allocation, assortment planning and shelf space allocation, and pricing and promotions planning, as well as demand forecasting and inventory replenishment.

numerical experiments.

The main contributions of our work are as follow. We are, to the best of our knowledge, the first to identify PO errors other than random yield, and empirically analyze the properties of such errors. Using a deterministic inventory model, we have obtained analytical insights about the influence of such PO errors on retail inventory cost and order policies. For the stochastic inventory model with uncertain demand and PO errors, we have conducted a thorough empirical study using real retail data to analyze the impact of PO errors. Lastly, we have proposed procedures for and provided qualitative guidance in identifying SKUs with high potential cost savings or PO error cost.

The rest of the paper is organized as follows. In Section 2, we review the related literature and position our work. An exploratory study of PO errors and rework time is presented in Section 3 to facilitate understanding of the problem. In Section 4, we first introduce a deterministic model that yields insights through an analytical study, and then a more complicated stochastic model that considers uncertainty in demand, PO errors, and rework time. We discuss the modeling and estimation of PO errors and rework time in Section 5. In Section 6, we report insightful observations gained from using the estimated parameters to perform numerical experiments on the stochastic inventory model, and propose procedures for identifying SKUs with high potential cost savings or PO error cost. In Section 7, we summarize the main findings of our study and suggest potential directions for future research.

2 Literature Review

Among existing papers that focus on random yield, ours is most closely related to those that use quality control measures (such as inspection and rework) to address the problem. These papers assume random yield to reflect production of defective items that can be detected through inspection and repaired through rework. Optimal inventory policy to account for quality control measures under these assumptions has been proposed by Porteus (1986), Peters et al. (1988), Zhang and Gerchak (1990), Lee (1992), and So and Tang (1995), among others. Our work is most similar to that of Peters et al. (1988), who employ an inventory model similar to ours except that they assume there to be at most one order outstanding at any given time, whereas we assume that orders do not cross in time. Our work differs from that of Peters et al. (1988) and that reported in other papers in the following ways as well. First, these other papers model random yield in the production setting. To the best of our knowledge, ours is the first paper to identify in the retail setting a set of PO errors different from random yield. The properties of these errors, in particular, how PO errors vary with order complexity (i.e., average order quantity per SKU and number of SKUs on a PO), differ from standard assumptions for random yield. They assume proportion of PO errors, defined as PO error size divided by order quantity, does not change with order quantity. Second, whereas existing papers rely solely on numerical experiments to study the influence of rework, we obtain some analytical properties from, and conduct numerical experiments with, real retail data.

Another closely-related paper in the random yield literature, that of Moinzadeh and Lee (1987), considers the case of an order that arrives in two shipments with constant lead time for both shipments. Upon arrival, the quantity of the first shipment is modeled as a random variable that is a function of order quantity. The exact cost function is derived, and an approximation provide based on the assumption that there is at most one order outstanding at any given time. A solution algorithm is provided for the approximate cost function under the assumption that the proportion of the first shipment decreases with order quantity. This assumption is satisfied under the binomial yield and random capacity models, but not for our PO errors. The second shipment in Moinzadeh and Lee (1989) can be thought as being obtained after rework at the vendor. For them, however, rework time does not depend on error size, as it does for our PO errors.

Our work is an extension of classical inventory management models under continuous time. Hadleyand Thomson M. Whitin (1963) and Zipkin (2000) are excellent references on classical inventory models. Because the PO errors in our problem depend on order quantity, so do rework and lead times. Papers that discuss how to find the optimal (Q, R) policy with lot-size dependent lead time include those by Kim and Benton (1995) and Hariga (1999). Kim and Benton (1995) also study how the dependence of lead time on lot-size influences the reorder point and order quantity. But whereas the dependence of lead time in their work is due to waiting and the production process at the manufacturer's plant, and modeled as a deterministic linear function of order quantity, lead time, in our study, depends on order quantity in a more complicated way through rework on PO errors at the DC, which cannot be adequately modeled as a deterministic linear function of order quantity.

3 Characterization of PO Errors and Rework Time

3.1 Data Description

We obtained two data sets from a single DC of Omega, a well-known 700-store discount retailer of apparel, electronics, and housewares (60% of its sales), as well as food items (40% of sales). Our focus is on the non-food items. One data set, generated by the DC's automated receiving system, records order quantity and order and arrival dates for each item on each PO for an entire year. The other, compiled from three months of paper audit reports, records errors and rework times as well as department and vendor for each PO. This data set includes 21 departments and 368 vendors.

3.2 Exploratory Analysis

Our partner retailer replenishes inventory via purchase orders (POs) that specify the quantity of each SKU ordered. A PO for a vendor such as Nike might, for example, be composed of 100 white and 200 black jackets and 50 white and 100 black pants. Figure 1 depicts the PO order process.

Our three months of audit data includes 995 POs and 1,217 incidences of errors of many types. One well-known error in the random-yield literature, the "short shipped" error, occurs when the quantity shipped by a vendor is less than that ordered by the retailer. Other error types observed include "ticket errors," whereby an

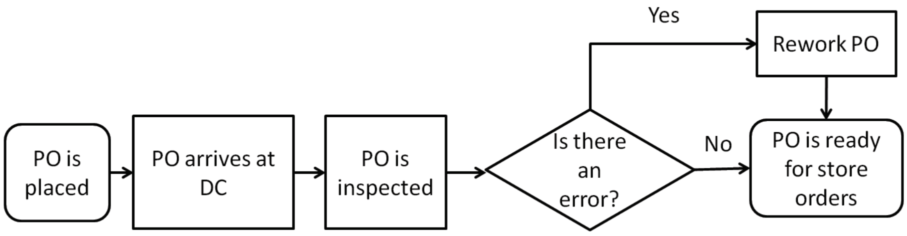

Figure 1: PO Process Flow Chart

incorrect SKU number or selling price is associated with a PO item, and the "Did not meet advanced shipping notice (ASN) requirements" error, which means that a vendor neglected to dispatch an ASN to the retailer in advance of the PO. Indeed, any breach of contractual requirements is a type of a PO fulfillment error. Figure 2 reports the PO error percentage breakdown by error type (specified in the labels on the left). The numbers to the right show the relative incidence of each error type.

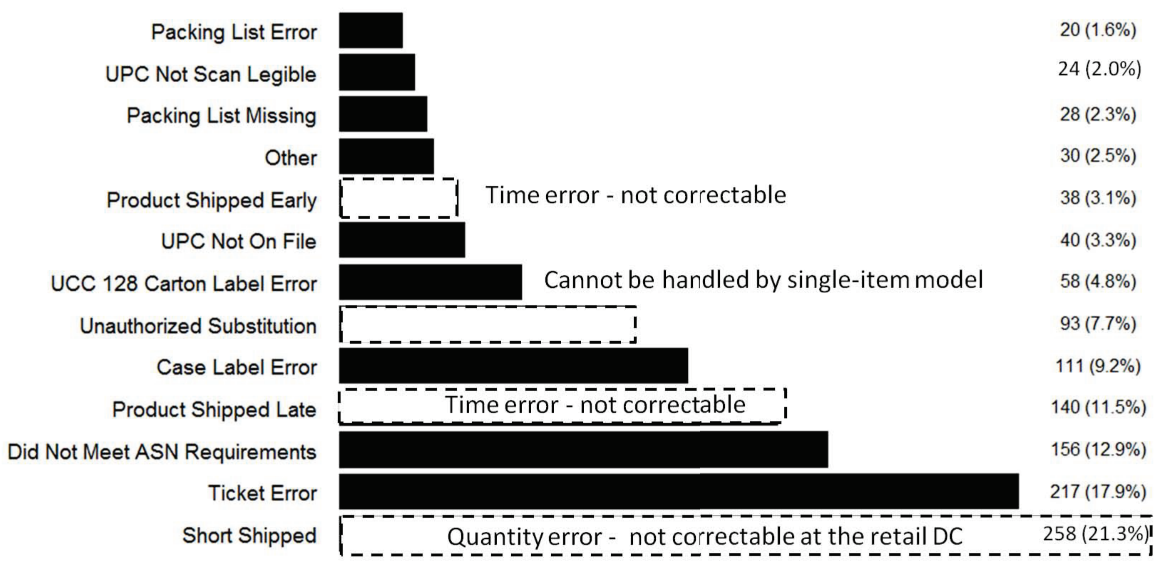

Figure 2: Incidence of PO Error Types

Error types and numbers vary greatly from department to department. Figure 3 is a heat map of the number of each type by department. The darker the color, the greater the number of errors. The 72 instances of "UCC 128 Carton Label Error" within "Children" is the largest number of errors. Other departments may not have this type of error. Shown in parentheses to the right of the heat map are the total number and percentage of errors across error types for each department. "Children," "Menswear/Luggage," "Ready To Wear," and "Domestics" have more error types and larger number of errors than other departments. Departments like

"Electrical & Plumbing" and "Paint and Décor" have only a few types and numbers of errors.

Whereas error types tend to spread across departments inasmuch as the sets of departments in which each error type occurs do not differ much from one another, individual vendors are usually associated with only one or two types of errors. From Figure 4, which shows the number of vendors with a certain number of error types or incidences, we see that 54.9% of vendors (202 of 368) had only one type of error, and 80.4% had no more than two types of errors. Numbers of errors, however, vary substantially among vendors, 42.9% (158 of 368) having only one error incidence, and only 35 vendors having more than six error incidences. A PO can have more than one type of error. From Figure 5, which shows the number of POs with a certain number of error types, we observe that 14.5% of the 995 POs in the audit data set have more than one type of error.

DC employees need to pull from the conveyor belt of an automatic receiving system and perform manual rework on PO order fulfillments that arrive with errors. For example, incorrect tickets need to be replaced with correct ones; the absence of an ASN necessitates going through each item to determine, for accounting purposes, which PO has arrived. The current rework system at our partner retailer is labor intensive. An erroneous PO is identified by attaching a hand-written slip of paper to its container; chargebacks are filed by completing, by hand, audit reports that provide evidence of errors (e.g., photographs of incorrect tickets attached to PO items). In terms of additional labor cost, lead time, and lead time variability incurred, for our partner retailer, rework time for a PO, defined as the time spent correcting errors, averages about 1.6 hours with a maximum of 3.6 days. The standard deviation of rework time is 5.2 hours. A PO that has to wait for an additional several hours may miss the shipping window from the DC to stores.[3]

Because PO errors not only add cost, but also complicate retail inventory management, our partner retailer penalizes PO errors by imposing a two-part charge on vendors, a fixed amount per erroneous PO and a variable amount proportional to rework time. Chargebacks for the three months of audit data totaled $302,675.50. Total chargebacks per vendor averaged $1,081 with a standard deviation of $1,897. Chargebacks can be as large as $16,960 and as small as $102.90. Figures 6 and 7 show total chargebacks in thousands of dollars by error type and department, respectively. "Ticket Error" accounts for 24.3% of total chargebacks, and the most common error type, "Short Shipped," 13.1%. "Children" and "Menswear/Luggage" are penalized more than other departments, accounting for 20.3% and 19.8% of total chargebacks.

4 Inventory Models

Retail inventory management is complicated by the additional lead time and lead time variability incurred by PO errors. In this section, we propose a deterministic and a stochastic inventory model to analyze the influence of PO errors on retailers' inventory systems. The main impact of PO errors on inventory systems being the extra lead time to perform rework, one might think standard inventory models with adjusted lead time that accounts for rework would suffice. But this is not the case. We found that inventory policies with adjusted lead

[3]We do not have data on waiting times for POs, which might be two to three days in the processing queue based on the retailer's knowledge.

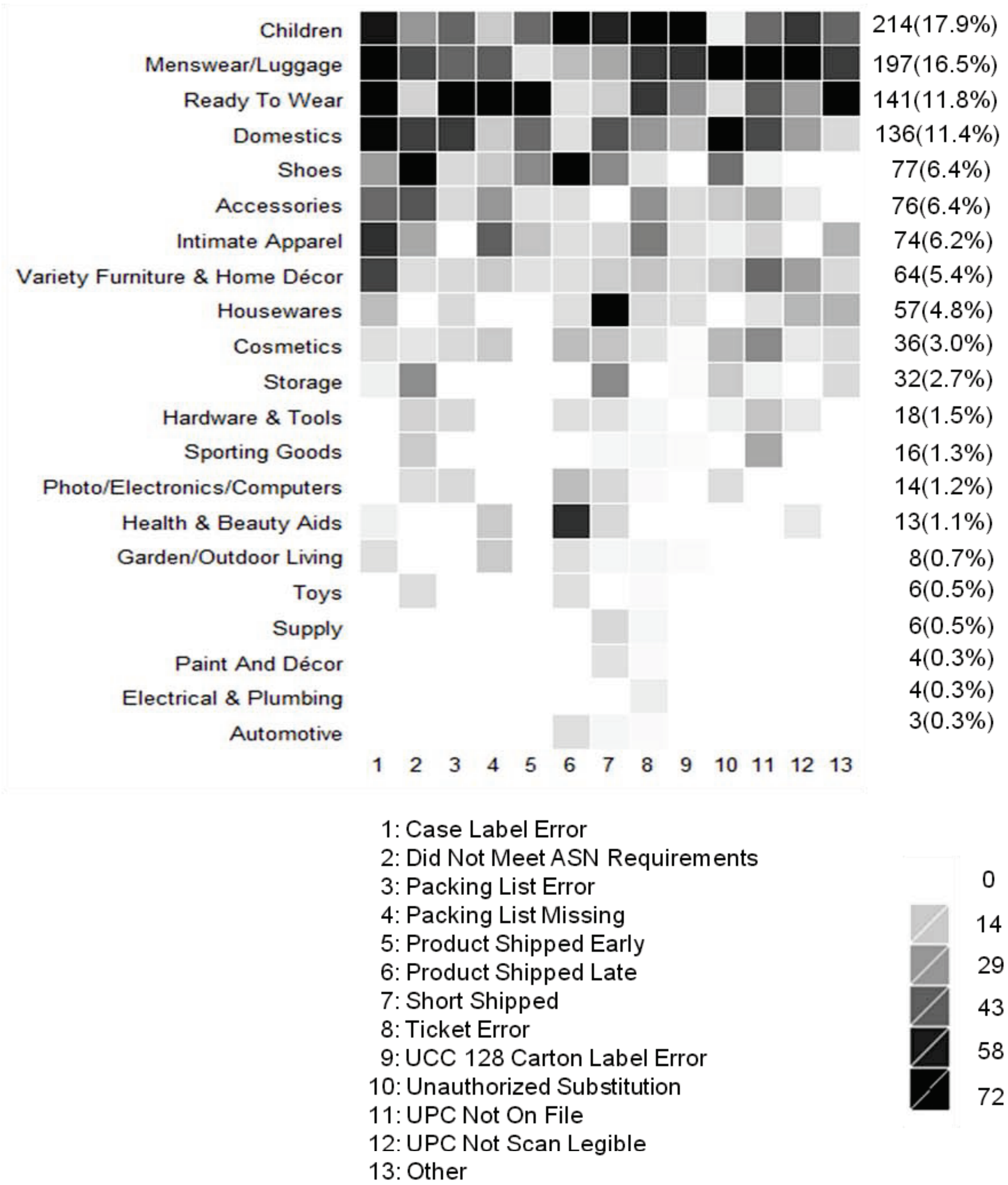

Figure 3: Heat Map for Error Incidences by Department and Error Type

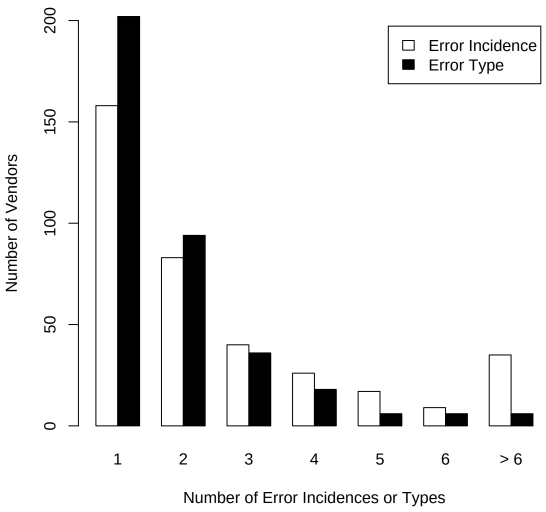

Figure 4: Error Incidences or Types for Vendors

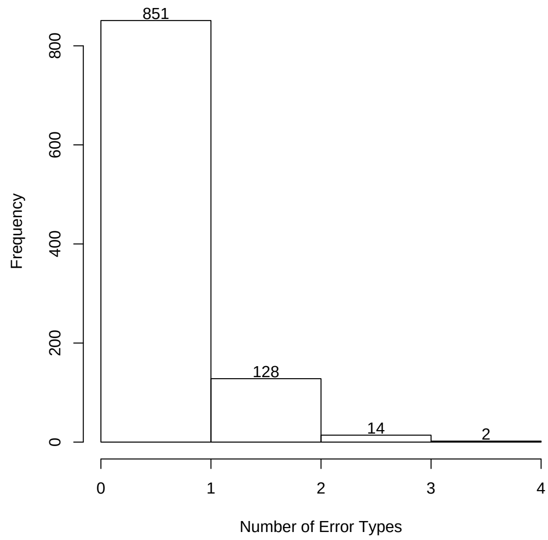

Figure 5: Error Types for POs

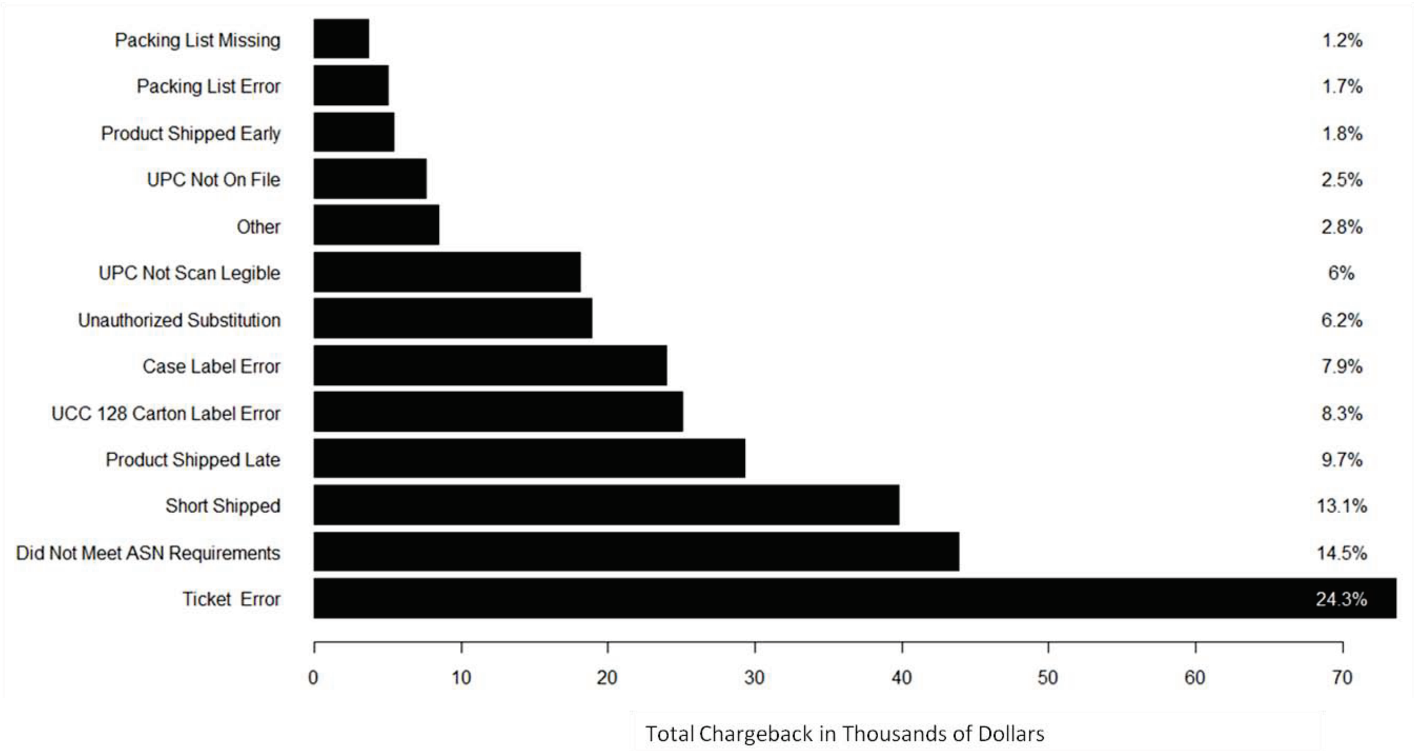

Figure 6: Chargeback by Error Type

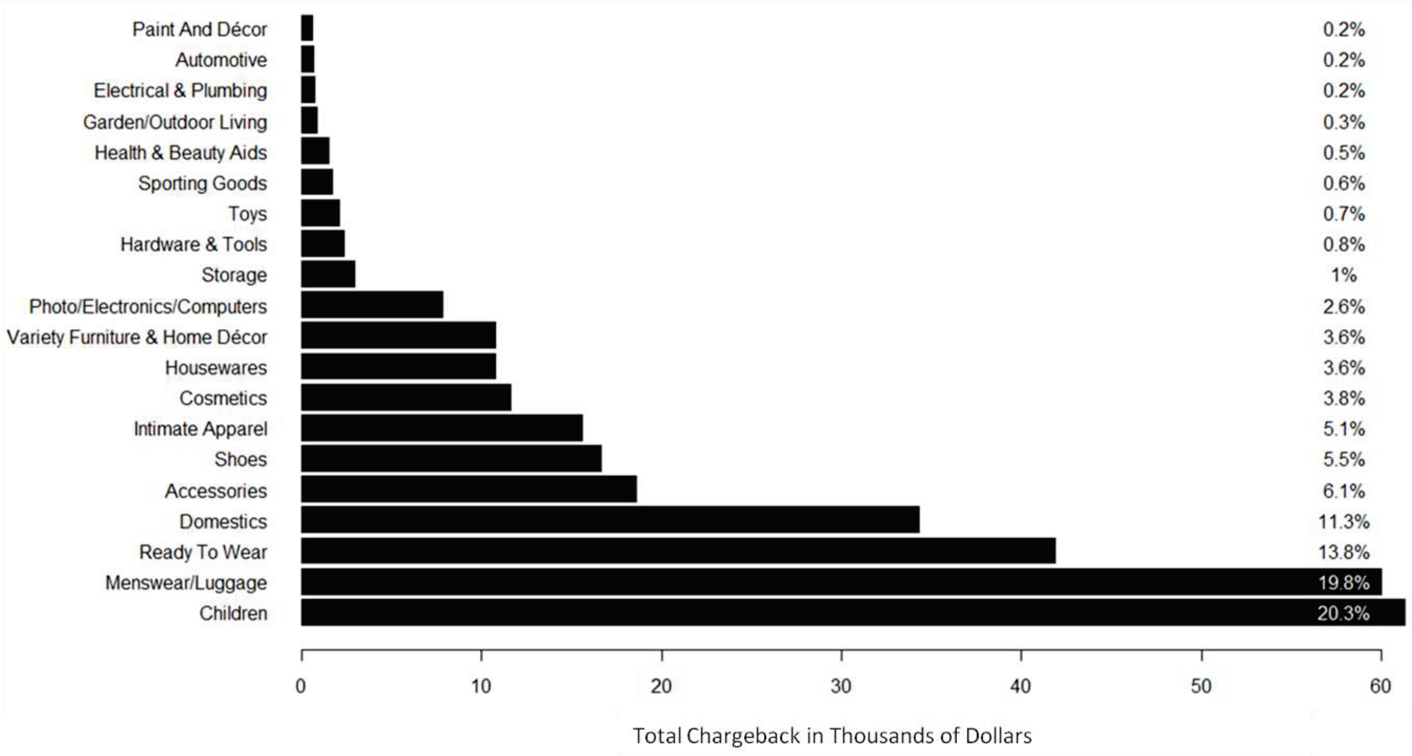

Figure 7: Chargeback by Department

time incur considerably more cost comparing to our proposed solution. Details of our findings are discussed in Section 6.3. We present here a deterministic and stochastic inventory model that capture the influence of rework on lead time.

Our inventory models are single SKU models in continuous time that incorporate important properties of PO errors based on the empirical analyses in Sections 5.1 and 5.2, namely, the dependence of PO errors and rework time on order quantities. Although in practice a PO often has more than one SKU, retailers do order SKUs independently, Chandran (2003), Levy and Weitz (2004), and van Donselaar et al. (2010) hinting that many retailers generate orders using single SKU replenishment systems. (Coordination, which occurs with trucking, is not captured by our models.) In regard to the choice of continuous review, our partner retailer employs continuous review for many items. In addition, based on Chandran (2003), Wal-Mart employs with suppliers like P&G an automated inventory replenishment system that reviews inventory continuously. Furthermore, periodic review with short periods behave as continuous review models; according to Axsater (2006), "Periodic review with a short review period T is very similar to continuous review."

We focus only on correctable PO errors at the DC, ignoring errors that cannot be corrected. The bars in solid boxes in Figure 2 represent errors, for all the error types at our partner retailer, that are correctable at the DC. The bars in dashed boxes correspond to errors that are not covered in this study, including time errors that cannot be corrected, substitution errors that cannot be handled by a single-item model, and quantity errors that cannot be corrected at the DC, and are anyway already covered in the random yield literature.

4.1 Deterministic Inventory Model

We first analyze the impact of PO errors using a deterministic model that ignores uncertainties in the inventory system. This deterministic model is simple enough to be studied analytically, yet captures some of the main effects of PO errors, including the additional lead time occasioned by rework and the dependence of lead time on order quantity. The main effects of PO errors are introduced in Section 5.

Assuming demand for a single SKU to arrive with constant rate λ, a retailer that places a PO for this SKU of size Q will receive the order after a constant lead time L. The order cannot be used until it finishes necessary rework, the rework time of a PO being a linear function of the error $W(Q)$, which is the number of erroneous units as a deterministic function of Q. If we denote per unit rework time by a, the rework time per order is $aW(Q)$. We make the following assumptions. First, a PO cannot be used until it finishes rework. In practice, retailers may occasionally use the correct portion of an order during rework. Often, though, retailers may prefer to clarify the error source before using the PO. Second, retailers do not pay holding costs for orders during rework (Zipkin (2000)).[4] Under these three assumptions, effective lead time is the actual lead time L plus rework time $aW(Q)$. Lastly, we assume that retailers recover all incorrect items of a PO through rework, which is true for many types of PO errors, including ticket errors. Figure 8 illustrates how inventory level changes over time.

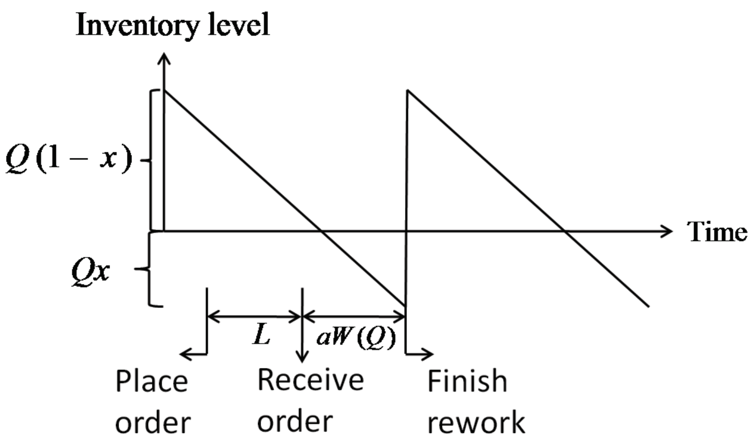

Figure 8: Deterministic Inventory Model

A retailer incurs fixed ordering cost K per order, holding cost h per unit per unit time, and rework cost c_r per unit time. Back orders are allowed, with backlogging cost π per unit per unit time. If x denotes the fraction of demand that is backlogged, then it is optimal for the retailer to place an order $L + aW(Q)$ before its inventory reaches $-xQ$, as shown in Figure 8. If we define $\rho(Q)$ as the error proportion of a PO, which is

[4] This assumption is to simplify the exposition. The model and relevant analysis do not change in substance when the retailer pays holding cost during rework.

$W(Q)/Q$, the retailer tries to minimize the total cost per unit time $M_r(Q,x)$.

$$M_r(Q,x) = \frac{\lambda K}{Q} + c_r a \lambda \rho(Q) + \frac{hQ(1-x)^2}{2} + \frac{\pi Q x^2}{2} \tag{1}$$

Here, $c_r a \lambda \rho(Q)$ is the rework cost per unit time, $\frac{hQ(1-x)^2}{2}$ the holding cost per unit time, and $\frac{\pi Q x^2}{2}$ the backlogging penalty cost per unit time.

The total cost per unit time $M_r(Q,x)$ is convex in the fraction of backlogged demand x. By setting the first derivative of $M_r(Q,x)$ with respect to x to 0, we find that the optimal x for any given Q is $h/(h+\pi)$. Substituting the optimal x back into $M_r(Q,x)$, we find the system cost as a function of Q to be:

$$C_r(Q) = \frac{\lambda K}{Q} + c_r a \lambda \rho(Q) + \frac{Q}{2}\frac{h\pi}{h+\pi}. \tag{2}$$

Let Q_r denote the optimal order quantity that minimizes $C_r(Q)$, and Q_s the EOQ order quantity with back orders, which is equal to $\sqrt{2K\lambda(h+\pi)/h\pi}$ (Axsater (2006)). Proposition 1 explains the influence of PO errors on optimal order quantity. Proofs for all propositions are provided in Appendix A.

Proposition 1. *If the error proportion $\rho(Q)$ decreases in Q (i.e., $\rho'(Q) < 0$), then the optimal order size in anticipation of PO errors Q_r is greater than the EOQ order quantity Q_s.*

We assume the error proportion decreases in Q because such a relationship is observed empirically.[5] Intuitively, rework cost per unit time increases with the error proportion, which decreases with order quantity, driving the optimal order quantity Q_r down from the EOQ order quantity Q_s.

We examine the influence of PO errors on system cost from two perspectives. We introduce two terms here. The first, "PO error cost," defined as equal to the difference between optimal total cost $C_r(Q_r)$ and optimal EOQ cost $C_s(Q_s)$, where $C_s(Q) = \frac{\lambda K}{Q} + \frac{Q}{2}\frac{h}{h+\pi}$. PO error cost specifies how much additional cost a retailer incurs due to the existence of PO errors, and is equal to $D_I + c_r a \lambda \rho(Q_r)$, where D_I is the difference between standard inventory costs $C_s(Q_r)$ and $C_s(Q_s)$. The second term, "potential cost savings," is defined as equal to the difference between the total cost with EOQ order quantity $C_r(Q_s)$ and the optimal total cost $C_r(Q_r)$. Potential cost savings represents how much a retailer can save by accounting for PO errors, and is equal to $D_r - D_I$, where D_r is the difference between rework costs $c_r a \lambda \rho(Q_s)$ and $c_r a \lambda \rho(Q_r)$. We obtain the following propositions regarding these two definitions. The assumption about $\rho(Q)$ in Proposition 3 is based on empirical data analysis of PO errors (see Section 5).

Proposition 2. *If the error proportion $\rho(Q)$ decreases in Q, then the PO error cost $C_r(Q_r) - C_s(Q_s)$ decreases in the fixed ordering cost per order K.*

Proposition 3. *If $\rho(Q) = cQ^{\beta-1}$ for a constant $c > 0$ and $0 < \beta < 1$, then the potential cost savings $C_r(Q_s) - C_r(Q_r)$ decreases in the fixed ordering cost K.*

[5] Further details of our empirical analysis can be found in Section 5.

As K increases, order quantities Q_r and Q_s increase, and the error proportion decreases, driving down rework costs $c_r a\lambda\rho(Q_r)$ and $c_r a\lambda\rho(Q_s)$. However, it is not clear how D_I changes as K increases. Although both $C_s(Q_r)$ and $C_s(Q_s)$ increase, Q_r also gets closer to Q_s. For PO error cost, which is equal to $D_I + c_r a\lambda\rho(Q_r)$, the decreasing effect is larger than the increasing effect of K, and therefore decreases in K. With regard to potential cost savings, although D_r decreases with K, it is not clear how D_I changes. We are nevertheless able to show that potential cost savings also decreases in K. Propositions 2 and 3 tell us that the cost impact of PO errors is greater for SKUs with smaller fixed ordering cost. Hence, retailers should, in the presence of PO errors, pay more attention to SKUs with smaller fixed ordering cost.

Proposition 4. *If the error proportion $\rho(Q)$ decreases in Q, then the PO error cost $C_r(Q_r) - C_s(Q_s)$ increases in the backlogging penalty cost π and the unit holding cost h.*

Proposition 5. *There exists a $\beta^* > 0$ such that for any $0 < \beta < 1$ and $\beta \leq \beta^*$, and $\rho(Q) = cQ^{\beta-1}$ for a constant $c > 0$, the potential cost savings $C_r(Q_s) - C_r(Q_r)$ increases in the backlogging penalty cost π and the unit holding cost h.*

As the backlogging penalty cost π or unit holding cost h increases, Q_r and Q_s decrease, and the error proportion thus increases. As a result, rework costs $c_r a\lambda\rho(Q_r)$ and $c_r a\lambda\rho(Q_s)$ increase, and both $C_s(Q_r)$ and $C_s(Q_s)$ increase in π and h. Although the directions of change for D_I and D_r are ambiguous, we are able to show that PO error cost and potential cost savings increase in the backlogging penalty cost π or unit holding cost h. We can interpret β as a measure of the degree of error. For a PO with a certain order quantity Q, a higher β corresponds to a larger error. For a PO with a small β, potential cost savings increases in the backlogging penalty cost π or unit holding cost h. Propositions 4 and 5 state that the cost impact of PO errors is greater for SKUs with higher purchase cost because both unit holding cost and backlogging penalty cost increase in the purchase cost.

Proposition 6. *If $\rho(Q) = cQ^{\beta-1}$ for a constant $c > 0$ and $0 < \beta < 1$, then the PO error cost $C_r(Q_r) - C_s(Q_s)$ increases in the demand rate λ. In addition, there exists some $\bar{\beta} > 0$ such that if for any $0 < \beta < 1$ and $\beta \leq \bar{\beta}$, and $\rho(Q) = cQ^{\beta-1}$ for a constant $c > 0$, the potential cost savings $C_r(Q_s) - C_r(Q_r)$ increases in the demand rate λ.*

As the demand rate λ increases, the error proportion decreases because Q_r and Q_s increase in λ. Therefore, rework costs $c_r a\lambda\rho(Q_r)$ and $c_r a\lambda\rho(Q_s)$ decrease. On the other hand, both $C_s(Q_r)$ and $C_s(Q_s)$ increase in λ. It is unclear, however, how D_I and D_r change. Putting these effects together, it turns out that both PO error cost and potential cost savings increase in the demand rate λ.

Propositions 3, 5 and 6 provide guidance in identifying SKUs with high potential cost savings in order to help retailers maximize the benefits and minimize the cost of adjusting and fine-tuning their existing inventory systems to account for PO errors. Propositions 2, 4 and 6 provide guidance in identifying SKUs with high PO error cost, to help retailers determine when to collaborate with vendors to reduce their error probability and magnitude. Detailed procedures for selecting SKUs with high potential cost savings or PO error cost are introduced for the stochastic inventory model in Section 6.

4.2 Stochastic Inventory Model

In practice, there are always uncertainties when managing retail inventory. To study the impact of PO errors in a more realistic setting, we propose a stochastic inventory model that includes three sources of uncertainty. First, the demand is uncertain. Unit demand (i.e., demand per unit of time) is commonly assumed to be normally distributed with mean μ and standard deviation σ. PO error size $\tilde{W}(Q)$, modeled by a random variable dependent on order quantity Q, is a second source of uncertainty. A third source is rework time $\tilde{t}_r(Q)$, which is modeled as a random variable dependent on error size and, hence, order quantity. Due to its simplicity, a (Q, R) inventory policy, whereby an order of size Q is placed when inventory position drops to R, is commonly used in practice and assumed herein. We utilize the assumptions from the deterministic model, and make the following additional assumptions. First, excessive demand is backlogged for the stochastic inventory model. Our study subject, a DC, serves demand from retail stores. In practice, this type of demand is usually backlogged, and the retailer incurs a backlogging penalty cost π per unit per unit time. One can also interpret backlogging cost as the imputed shortage cost of a service level. Second, orders do not cross during the effective lead time, which is equal to the sum of the actual lead time L and rework time $\tilde{t}_r(Q)$. Third, demand during the effective lead time is normally distributed with mean $\mu(Q) = \mu E(L + \tilde{t}_r(Q))$ and standard deviation $\sigma(Q) = (\sigma^2 E(L + \tilde{t}_r(Q)) + \mu^2 Var(L + \tilde{t}_r(Q)))^{\frac{1}{2}}$.[6] Explicit models of PO error size $\tilde{W}(Q)$ and rework time $\tilde{t}_r(Q)$ are obtained on the basis of the empirical data analysis in Section 5. Figure 9 shows how inventory level evolves over time in the stochastic inventory model.

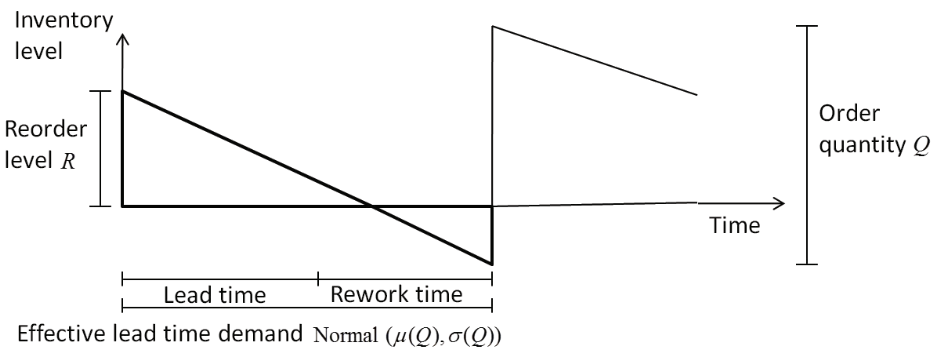

Figure 9: Stochastic Inventory Model

A retailer minimizes the total cost per unit time $C_r(Q, R)$ as follows:

$$(K + c_r E(\tilde{t}_r(Q)))\frac{\mu}{Q} + h(R + \frac{Q}{2} - \mu(Q)) + (h + \pi)\frac{\sigma^2(Q)}{Q}(H(\frac{R - \mu(Q)}{\sigma(Q)}) - H(\frac{R + Q - \mu(Q)}{\sigma(Q)})).$$

[6]Because we do not have actual demand data for the estimation of lead time demand, normal distribution is chosen, as in Axsater (2006) and many other works.

Here $H(x) = \frac{1}{2}((x^2+1)(1-\Phi(x))-x\varphi(x))$,[7] where $\Phi(x)$ and $\varphi(x)$ are the standard normal cumulative distribution and density functions, respectively. Because the stochastic model is more difficult than the deterministic model to analyze, we use mainly the numerical experiments introduced in Section 6, in which we find the optimal (Q, R) by solving a system of nonlinear first order equations using Newton's method. Starting value for Q is the optimal order quantity in the deterministic setting, and the starting value for R the expected effective lead time demand. The ordering policy (Q, R) found in this way is only a local optimal solution without proving unimodality of the total cost per unit of time $C_r(Q, R)$. Because we observed, in our numerical experiments, unimodality of $C_r(Q, R)$, we assume the total cost per unit of time $C_r(Q, R)$ to be unimodal, so that the locally optimal is also the globally optimal ordering policy.

Federgruen and Zheng (1992) and Rosling (1999) discuss efficient solution algorithms for finding the optimal Q and R for the standard stochastic inventory model with a (Q, R) policy. We use the iterative algorithm in Rosling (1999) to find optimal Q_s and R_s for the standard inventory model. A brief description of the algorithm is provided: Start with Q as the EOQ order quantity. Find the optimal reorder point R for a given Q by solving the equation generated by setting $\frac{\partial C_s(Q,R)}{\partial R} = 0$, where $C_s(Q, R)$ is the inventory cost function for the standard inventory model. Obtain the order quantity Q for the next iteration using an updating formula generated by setting $\frac{\partial C_s(Q,R)}{\partial Q} = 0$. Repeat this process until inventory cost $C_s(Q, R)$ converges. The solution algorithm used to find optimal (Q_r, R_r) for the stochastic inventory model that accounts for PO errors is as described in Section 4.2. Note that without proving the unimodality of system inventory cost with rework $C_r(Q, R)$, the optimal solution (Q_r, R_r) found using the previous algorithm is only locally optimal. Therefore, in the following numerical experiments, the potential cost savings calculated as $C_r(Q_s, R_s)-C_r(Q_r, R_r)$ is only a lower bound of the achievable potential cost savings. The PO error cost calculated as $C_r(Q_r, R_r)-C_s(Q_s, R_s)$ is an upper bound of the true PO error cost. But because we have observed unimodality of the cost function $C_r(Q, R)$ in our numerical experiments for various parameter settings, these bounds should reflect the actual potential cost savings and true PO error cost.

5 Estimation of PO Errors and Rework Time

To define appropriate parameters for our numerical experiments, we use data from a retail chain to model PO errors. Specifically, we evaluate whether, and to what extent, PO errors depend on the order quantity, and whether, and how, rework time depends on PO errors.

5.1 Regression Models and Estimation of PO Errors

We examine the PO characteristics that drive PO errors. Intuitively, the more complex a PO, the more likely it is to have an error and the larger the expected error size. We use two measures to characterize the complexity

[7] Here we define $H(x)$ in the same way as in Axsater (2006). See the expressions (5.65) and (6.10) there for the derivation of total cost per unit of time $C_r(Q, R)$.

of a PO, namely, average order quantity Q per SKU, and number of SKUs n on a PO. For a PO composed of 100 white jackets and 50 white pants, average order quantity is 75 and the number of SKUs two. PO errors depend on these two measures. From Section 3.2, we also need to account for variation in PO errors across vendors, departments, and error types. Because most vendors are associated with only one or two error types, as shown in Figure 4, we are unable to consider variation across vendors and error types at the same time. In practice, retailers are more interested in identifying vendors that need improvement than error types. We thus consider in the following analysis variation across vendors and departments but not error types.

We define error size as the sum of errors for all SKUs on a PO. Although a PO may have many different types errors, as shown in Figure 2, we do not differentiate among error types because we model the variation in PO errors across vendors. A PO error, being a generic error, can be of any type. The error size of a PO with an ASN error is always the total order quantity of SKUs on the PO, the error size of a PO with a ticket error usually the order quantity of one particular SKU. More than one type of error may be observed on a particular PO, as shown in Figure 5, in which case the error size is defined as the sum of errors of all types.

Because PO errors within the same vendor or department are expected to be similar to one another, we use cross-classified hierarchical linear modeling (HLM) which allows observations to be nested with two higher-level categories, namely, vendors and departments, to account for the multilevel structure of our data (Raudenbush and Bryk (2001)). To estimate these HLM models, we use the "lmer" function in R version 2.8.1 and employ a maximum likelihood estimation technique.[8]

We model PO error size $\tilde{W}_{ilj}$ as follows:

$$E(\tilde{W}_{lij}|Q_{lij}, n_{lij}) = P(\tilde{W}_{lij} > 0|Q_{lij}, n_{lij}) \times E(\tilde{W}_{lij}|\tilde{W}_{lij} > 0, Q_{lij}, n_{lij}). \tag{3}$$

Here, $E(\tilde{W}_{lij}|Q_{lij}, n_{lij})$ is the expected error size of PO l $(l = 1, ..., m_{ij})$ from vendor i $(i = 1, ..., 112)$ and department j $(j = 1, ..., 14)$ with average order quantity Q_{lij} and number of SKUs n_{lij}. The expression $P(\tilde{W}_{lij} > 0|Q_{lij}, n_{lij})$ represents the probability of an error for PO l from vendor i and department j with average order quantity Q_{lij} and number of SKUs n_{lij}. The expression $E(\tilde{W}_{lij}|\tilde{W}_{lij} > 0, Q_{lij}, n_{lij})$ is the expected error size, given that there is an error, of PO l from vendor i and department j with average order quantity Q_{lij} and number of SKUs n_{lij}.

With HLM models, we need to test for the existence of random effects using a null model before including our predictors and estimating the conditional model. A logistic model is used for the probability of an error. The null model for the probability of an error $P(\tilde{W}_{lij} > 0)$ is as follows:

$$\text{logit}(P(\tilde{W}_{lij} > 0)) = b_0 + c_i^p + d_j^p + e_{lij}^p, \tag{4}$$

where b_0 is a fixed intercept parameter and the superscript p stands for error probability. The random effect of vendor i is $c_i^p \sim N(0, \tau_c^p)$, the random effect of department j $d_j^p \sim N(0, \tau_d^p)$, and the random PO effect

[8]Consistent results are obtained using Proc Mixed in SAS v.9.1.

$e^p_{lij} \sim N(0, \sigma^2_p)$. Using a Chi-square test to test the significance of the random effects of vendor and department, we find both to be significant with p values of 0.02 and 0.04, respectively.

The conditional model for the probability of an error $P(\tilde{W}_{lij} > 0)$ with predictors Q_{lij} and n_{lij} is shown in equation (5). We include the predictors in their natural log forms to induce linearity, which we find to be the most effective way after analyzing the raw data. We employ the model

$$\text{logit}(P(\tilde{W}_{lij} > 0|Q_{lij}, n_{lij})) = b_0 + c^p_i + d^p_j + e^p_{lij} + b_1 \times log(Q_{lij}) + b_2 \times log(n_{lij}), \tag{5}$$

in which b_0, b_1, b_2 are fixed PO-level coefficients. We assume b_1 and b_2 to be fixed across vendors and departments, rather than randomly varying, because the number of POs under some vendor-department combinations is small.

To estimate $P(\tilde{W}_{lij} > 0)$, we filter out under the vendor-department combinations all POs with more than two error records. After filtering, we have $3,788$ POs. Table 1 presents the summary statistics for all of these POs for the regression variables. In the table, "StDev" stands for standard deviation.

Table 1: Summary Statistics for Error Likelihood Model

Variable	Min	Max	Median	Mean	StDev	Yes	No
$\tilde{W}_{lij} > 0$						536	3,252
Q_{lij}	1	125,100	810	2,253	5,650		
n_{lij}	1	592	12	27	46		

The null model for error size $\tilde{W}_{lij}$, given that there is an error, is

$$\tilde{W}_{lij}|\tilde{W}_{lij} > 0 = \alpha + c^s_i + d^s_j + e^s_{lij}, \tag{6}$$

where α is a fixed intercept parameter and the superscript s stands for error size. The random effect of vendor i is $c^s_i \sim N(0, \tau^s_c)$, the random effect of department j $d^s_j \sim N(0, \tau^s_d)$, and the random PO effect $e^s_{lij} \sim N(0, \sigma^2_s)$. Using a Chi-square test, we find the random effect of vendor to be significant, with p value smaller than 0.001, and the random effect of department to not be statistically significant. Hence, we do not include the random effect of department in the following conditional model. Instead, we account for the heterogeneity of POs under different departments by using fixed effects (e.g., dummy variables to represent each department).

We use a log-log model to determine error size, given that there is an error. Given that there is an error, the conditional model for the error size $\tilde{W}_{lij}$ with predictors Q_{lij} and n_{lij} is as follows:

$$\log(\tilde{W}_{lij}|\tilde{W}_{lij} > 0, Q_{lij}, n_{lij}) = \alpha_j + c^s_i + e^s_{lij} + \beta \times \log(Q_{lij}) + \gamma \times \log(n_{lij}). \tag{7}$$

Here, α_j, β, γ are fixed PO-level coefficients, and the constant terms α_j are again used to account for the heterogeneity of POs under different departments. We assume β and γ to be fixed across vendors and departments, rather than randomly varying, due to the small number of POs under some vendor-department combinations. Table 2 presents summary statistics for POs with positive error size for the dependent and independent variables of the error size model. In total, we have 536 POs with a nonzero error.

Table 2: Summary Statistics for Conditional Error Size Model

Variable	Min	Max	Median	Mean	StDev
$\tilde{W}_{lij}$	1	34,440	39	317	1,760
Q_{lij}	11	80,620	1,440	2,253	6,439
n_{lij}	1	592	12	27	63

Tables 3 and 4 show the estimation results for the likelihood of an error, and error size conditioning on the existence of an error respectively.[9] The deviance (i.e., negative twice the log-likelihood) of the fit in Table 3 is 2,899, and in Table 4, 2,231.[10] From the regression results in both tables, we can see that the likelihood and magnitude of an error increase in both average order quantity and the number of SKUs on a PO, confirming the intuition that the more complex a PO, the more likely there is to be an error. The estimates for $\tau_c^p(\tau_c^s)$, τ_d^p, and $\sigma_p^2(\sigma_s^2)$ are 0.450(0.354), 0.071, and 1(3.485), respectively.

Table 3: Error Likelihood Model Estimation Results ($n = 3,788$)

Regression Coefficients	Estimate	Standard Error	z value	p value
b_0	−3.344	0.306	−10.916	< 0.001
b_1 ($log(Q_{lij})$)	0.179	0.040	4.531	< 0.001
b_2 ($log(n_{lij})$)	0.369	0.055	6.659	< 0.001

Table 4: Conditional Error Size Model Estimation Results ($n = 536$)

Regression Coefficients	Estimate	Standard Error	t value	p value
β ($log(Q_{lij})$)	0.433	0.064	6.754	< 0.001
γ ($log(n_{lij})$)	0.239	0.092	2.589	0.010
α_1	1.278	0.594	2.151	0.032
α_2	1.159	0.492	2.353	0.020

5.2 Regression Model and Estimation for Rework Time

We now examine the PO characteristics that influence rework time. Intuitively, the larger the error size, the longer rework time. Rework time should also vary across vendors, departments, and error types. For reasons similar to those articulated in Section 5.1, variation across vendors and departments, but not error types, is accounted for in the rework time model. Rework time for POs with zero errors is zero. Thus, we model rework time only for POs with errors.

We first test for the existence of random effects using a null model. The null model for rework time $\tilde{t}_{lij}$

[9]The p values in Table 4 are calculated using the "pvals.fnc" function in the "languageR" package.

[10]Deviance is a measure of goodness of fit for the hierarchical linear models estimated using the maximum likelihood estimation technique. The larger the deviance, the worse the model's fit.

conditioning on there being an error is

$$\tilde{t}_{lij}|\tilde{W}_{lij} > 0 = \phi_0 + c_i^r + d_j^r + e_{lij}^r, \tag{8}$$

where ϕ_0 is a fixed intercept parameter, $\tilde{W}_{lij}$ denotes PO error size, and the superscript r stands for rework. The random effect of vendor i is $c_i^r \sim N(0, \tau_c^r)$, the random effect of department j $d_j^r \sim N(0, \tau_d^r)$, and the random PO effect $e_{lij}^r \sim N(0, \sigma_r^2)$. Finding through Chi-square tests, that only the random effect of vendor is significant with p value smaller than 0.001, we account for the heterogeneity of POs under different departments using fixed rather than random effects.

We use a linear model for rework time conditioning on the existence of an error, with per unit rework time $a_j + c_i^r$ varying by vendor i and department j. Intuitively, per unit rework time should also vary by error type. But because we have already accounted for the heterogeneity of per unit rework time for different vendor-department combinations, and certain vendor-department combinations are usually associated with certain types of errors, we are unable to consider the variability of per unit rework time by error type together with the variability across vendor-department combinations. A linear model is used to model rework time because we have tried the square and log transformation on the independent and/or dependent variables, and the linear model gives us the best fit (Kleinbaum et al. (1998)).

The regression model for rework time is as follows:

$$\tilde{t}_{lij}|\tilde{W}_{lij} > 0 = a_j \times \tilde{W}_{lij} + c_i^r \times \tilde{W}_{lij} + e_{lij}^r. \tag{9}$$

Note that we do not include an intercept term for the rework time model.[11] The intercept term can be interpreted as the set up time for rework. The rework needed for POs with errors does not require set up time. For example, the rework needed for the error type "Did not meet ASN requirements" is to go through each item to figure out which PO it is, a process that does not require set up. In addition, when testing multiple models with different functional forms (including a model with an intercept term), the no intercept model generates the best fit. Hence, we find that it appropriate to not include an intercept term. We have 465 POs with errors for the estimation of the rework time model. The following table presents summary statistics for the regression variables. The unit measure for rework time is minutes.

Table 5: Summary Statistics for Rework Time Model

Variable	Min	Max	Median	Mean	StDev
$\tilde{t}_{lij}$	0	5,235	30	96	314
$\tilde{W}_{lij}$	1	34,440	35	305	1,822

Table 6 shows the estimation results for rework time. We show a_j for only two departments, for purposes of illustration. The deviance of the fit is 6,155. The estimates for τ_c^r and σ_r^2 are, respectively, 1.762 and

[11] Note that e_{lij}^r has zero mean, and hence cannot act as an intercept.

18,969.465.

Table 6: Rework Time Model Estimation Results ($n = 465$)

Regression Coefficients	Estimate (min/unit)	Standard Error	t value	p value
a_1	0.738	0.251	2.940	0.003
a_2	0.316	0.325	0.973	0.331

5.3 Error Size and Rework Time for a PO of a Single SKU

We obtained forecast models for PO error size with average order quantity Q and number of SKUs n as predictors. Having single SKU inventory models, we use these forecasting models to derive expressions for the error mean $\mathrm{E}(\tilde{W}(Q))$ for a PO of a single SKU.

We use the error for a PO of a single SKU to estimate the error for that SKU. According to equation (3), for a PO of a single SKU with order quantity Q, the PO error mean $\mathrm{E}(\tilde{W}(Q))$ is $P(\tilde{W}(Q) > 0|Q, n = 1) \times E(\tilde{W}(Q)|\tilde{W}(Q) > 0, Q, n = 1)$. According to equation (5), the probability of an error for a PO of a single SKU $P(\tilde{W}(Q) > 0|Q, n = 1)$ is $\frac{e^{(b_0+b_1 log(Q))}}{1 + e^{(b_0+b_1 log(Q))}}$, henceforth denoted by $p(Q)$. Note that the number of SKUs for a PO of a single SKU is 1. According to equation (7), given that there is an error, the error size for a PO of a single SKU with order quantity Q has a log-normal distribution with mean equal to $e^{(\alpha+\sigma_s^2/2)}Q^\beta$. Hence, the PO error mean $\mathrm{E}(\tilde{W}(Q))$ is $p(Q)e^{(\alpha+\sigma_s^2/2)}Q^\beta$. The derivation of the PO error mean is given in Appendix B.

The error proportion $\rho(Q)$ for the deterministic model is considered to be the expected error proportion $E(\tilde{W}(Q))/Q$. The following proposition states a sufficient condition on the regression coefficient estimates of the PO error forecast models for $E(\tilde{W}(Q))/Q$ to decrease in the order quantity Q.

Proposition 7. *If $\beta + b_1 \le 1$, then $E(\tilde{W}(Q))/Q$ decreases in order quantity Q.*

From the regression results of the probability of an error and conditional error size, we can see the estimate for b_1 is 0.179 and the estimate for β 0.433. The sufficient condition in Proposition 7 is satisfied. Hence, the resulting error proportion $\rho(Q)$ decreases in order quantity Q, as assumed in Propositions 1 and 2. If we assume that the probability of an error $p(Q)$ is equal to a constant p, which does not depend on order quantity Q, then $\rho(Q)$ is equal to $pe^{(\alpha+\sigma_s^2/2)}Q^{\beta-1}$. Leting $c = pe^{(\alpha+\sigma_s^2/2)}$, we can see that $\rho(Q) = cQ^{\beta-1}$ for some constant $c > 0$ and $0 < \beta < 1$, as assumed in Propositions 3, 4, 5, and 6.

Having also obtained the forecast model for rework time conditioning on there being an error, we can derive the rework time mean $E(\tilde{t}_r(Q))$ and variance $Var(\tilde{t}_r(Q))$ for a PO of a single SKU with order quantity Q. The rework time mean, $E(\tilde{t}_r(Q))$, is $ap(Q)e^{(\alpha+\sigma_s^2/2)}Q^\beta$, and the rework time variance, $Var(\tilde{t}_r(Q))$, $a^2e^{2\alpha+\sigma_s^2}Q^{2\beta}(e^{\sigma_s^2} - p(Q))p(Q) + \sigma_r^2 p(Q)$. The derivations of the rework time mean and variance are given in Appendix B.

We can now derive expressions for the effective lead time demand mean $\mu(Q)$ and standard deviation $\sigma(Q)$. For a PO of a single SKU with order quantity Q, the effective lead time demand mean $\mu(Q)$ is given by the expression

$$\begin{aligned}\mu(Q) &= \mu E(L+\tilde{t}_r(Q))\\ &= \mu L + \mu E(\tilde{t}_r(Q))\\ &= \mu L + \mu a p(Q) e^{(\alpha+\sigma_s^2/2)} Q^{\beta},\end{aligned}$$

and the effective lead time demand standard deviation $\sigma(Q)$ by the expression

$$\begin{aligned}\sigma(Q) &= (\sigma^2 E(L+\tilde{t}_r(Q)) + \mu^2 Var(L+\tilde{t}_r(Q)))^{\frac{1}{2}}\\ &= (\sigma^2 L + \sigma^2 E(\tilde{t}_r(Q)) + \mu^2 Var(\tilde{t}_r(Q)))^{\frac{1}{2}}\\ &= (\sigma^2 L + \sigma^2 a p(Q) e^{(\alpha+\sigma_s^2/2)} Q^{\beta} + \mu^2 a^2 p(Q) e^{(2\alpha+\sigma_s^2)}(e^{\sigma_s^2} - p(Q)) Q^{2\beta} + \mu^2 \sigma_r^2 p(Q))^{\frac{1}{2}}.\end{aligned}$$

Note that POs from different vendor and department combinations have different estimates for the error likelihood parameter α, conditional error magnitude parameter b_0, and unit rework time parameter a.

6 Numerical Experiments

For the stochastic inventory model, we use numerical experiments to analyze the influence of PO errors on inventory policy and system cost. To perform the numerical experiments, we also need, in addition to the data on PO errors and rework time used for parameter estimation in Section 5, SKU demand data and inventory cost parameters. In the absence of sales or inventory data for estimating the demand for each SKU, we approximate demand for a particular month using the sum of order arrivals during that month. For seasonal products, months in which no orders of the SKU arrive are not included in the calculation of monthly demand mean and standard deviation. We spoke at length with contacts at our partner retailer about the inventory cost parameters. We were unable to obtain estimates of these values, but were provided with average purchase cost breakdowns by department, from which we estimated the holding cost h. The selected SKUs in the numerical experiments are from fourteen departments. We conducted experiments with the following cost parameter settings: fixed ordering cost ranges from \$10 to \$100 per order with an increment of \$10; holding cost per year ranges from 5% to 30% of the purchase cost with an increment of 5%; backlogging penalty cost per year ranges from 5 to 40 times the holding cost per year with an increment of 5. In the numerical experiments, we build for each SKU one standard (Q, R) inventory model and one inventory model with rework. There are approximately $13,500$ SKUs in our retail data set that provide sufficient data for estimation.[12]

[12] There are approximately 93,000 SKUs in the one year data set that records for each item within each PO, order quantity, order date, and receive date. There are approximately 14,000 SKUs in the three months of audit data that records error and rework time for each PO.

6.1 Influence of PO Errors

We ran 480 experiments for all the selected SKUs with inventory cost parameters, as described above. Figures 10 and 11 present, for some representative experiments, the yearly potential cost savings and PO error cost per SKU, averaged over all the selected SKUs. From Figure 10, we observe that, by accounting for PO errors that lead to rework, the retailer can achieve potential cost savings ranging from \$0.65 to \$19.63 per year per SKU, or about 9% to 260% of the average purchase cost per SKU. Potential cost savings vary with inventory cost parameters. More specifically, average yearly potential cost savings per SKU decrease in the fixed ordering cost, and increase in the holding cost percentage and the backlogging penalty cost. Average yearly potential cost savings per SKU are most sensitive to the holding cost percentage, and least sensitive to fixed ordering cost.[13] From Figure 11, we observe that PO errors cost retailers from \$11.19 to \$51.05 per year, per SKU, or about 150% to 670% of the average purchase cost per SKU. PO error cost varies with inventory cost parameters. More specifically, yearly PO error cost per SKU decreases in fixed ordering cost, and increases in holding cost percentage and backlogging penalty cost. Yearly PO error cost is most sensitive to holding cost percentage, and least sensitive to fixed ordering cost. Currently, the retailer charges vendors about \$25 per SKU for PO errors without considering SKU characteristics such as the inventory cost parameters. Based on our analysis, it might be undercharging or overcharging vendors depending on its inventory cost parameters. When holding cost is 15% of purchase cost, fixed ordering cost \$10 per order, and backlogging penalty cost 25 times holding cost, the retailer is undercharging its vendors for PO errors. When holding cost is 5% of purchase cost, fixed ordering cost \$10 per order, and backlogging penalty cost 40 times the holding cost, the retailer is overcharging vendors for PO errors. The foregoing analysis suggests that SKU characteristics should be taken into account when deciding chargebacks for PO errors.

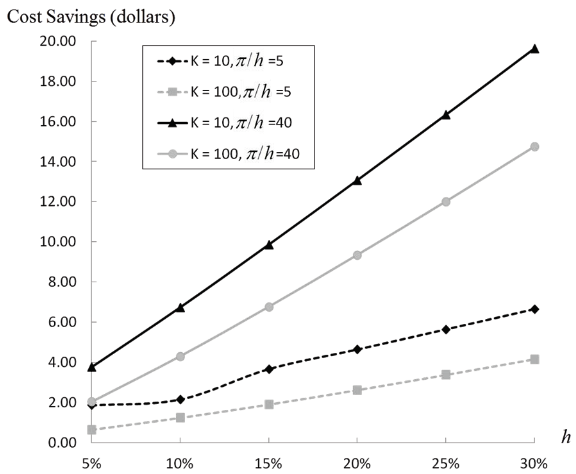

Figure 10: Yearly Cost Savings per SKU

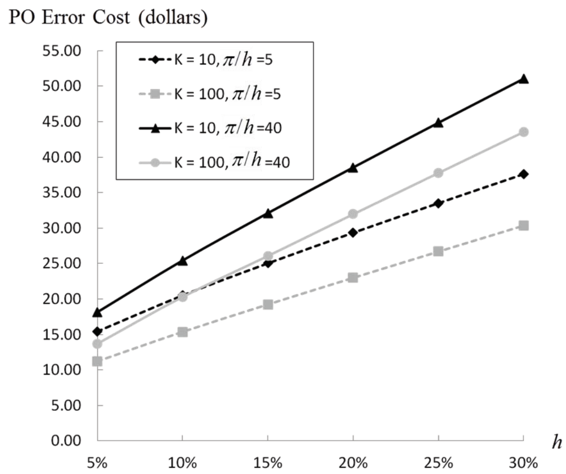

Figure 11: Yearly PO Error Cost per SKU

Optimal ordering policy is influenced by PO errors. In the deterministic model, we showed analytically

[13] It is most sensitive in the sense that $\frac{\Delta\text{potential cost savings}}{\text{potential cost savings}} / \frac{\Delta\text{cost parameter}}{\text{cost parameter}}$ is greatest for the holding cost percentage.

that the optimal order size Q_r, which accounts for PO errors, is greater than the optimal EOQ order quantity Q_s. In the stochastic case, we want to numerically evaluate whether a similar property holds, that is, whether the optimal order size that accounts for PO errors is greater than the optimal order quantity for the standard (Q, R) model. For all the selected SKUs in the previous 480 experiments, Q_r is always greater than Q_s. Because rework adds additional lead time and lead time variability, we expect the reorder point with rework to be higher than the standard one. For 99.36% of test cases, R_r is no less than R_s, though there are cases to the contrary.

As observed from Figures 10 and 11, both average yearly potential cost savings and PO error cost per SKU decrease in fixed ordering cost, and increase in holding cost percentage and backlogging penalty cost. In the deterministic model, we also showed analytically that potential cost savings and PO error cost decrease in fixed ordering cost K, and increase in holding cost h, and backlogging penalty cost π for each SKU, not just as an average value across all SKUs. Intuitively, the same property should hold in the stochastic case. We constructed several experiments to test this intuition.

First, to test whether potential cost savings decreases in the fixed ordering cost, we set the holding cost percentage at 30%, and the backlogging penalty cost at 20 times the holding cost, while varying fixed ordering cost from \$10 to \$100 per order, with an increment of \$10, and compared the potential cost savings of each SKU with the fixed ordering cost of $\$K$ and $\$(K + 10)$ per order for K from \$10 to \$90. We observed potential cost savings to be greater with lower fixed ordering cost per order K in 96.15% of test cases. We next verified, in the following way, that potential cost savings increases in the holding cost percentage. We fixed ordering cost at \$100 per order, and the backlogging penalty cost at 20 times the holding cost, with the holding cost percentage varying from 5% to 30% of purchase cost with an increment of 5%. Potential cost savings for each SKU with holding cost percentage h and $(h + 5\%)$ are compared for h from 5% to 25%. We observed potential cost savings to be greater with higher holding cost percentage in 98.17% of test cases. To determine whether potential cost savings increases in the backlogging penalty cost, we varied the backlogging penalty from 5 to 40 times the holding cost, with an increment of 5 times, while setting holding cost percentage at 30%, and fixed ordering cost at \$100 per order, and compared the potential cost savings for each SKU with backlogging penalty cost of π and $\pi + 5$ times the holding cost for π from 5 to 35. We observed that potential cost savings are greater with higher backlogging penalty cost in 98.82% of test cases. Similar observations can be made for PO error cost.

6.2 Identify SKUs with High Potential Cost Savings or PO Error Cost

Both standard inventory cost $C_s(Q, R)$ and inventory cost with rework $C_r(Q, R)$ vary in the inventory cost parameters (K, h, and π), SKU unit time demand parameters (unit time demand mean μ and standard deviation σ), and SKU lead time L. Because inventory cost with rework $C_r(Q, R)$ also varies in the PO error related parameters, b_0 and α,[14] and SKU unit rework time a, we would expect both potential cost savings and PO error cost for any SKU to be functions of these parameters. An analysis of raw data suggests a nonlinear relationship between the outcome variable (potential cost savings or PO error cost) and each of these parameters. We transform the outcome variable and some of these parameters (including K, h, π, μ, and L) into their natural log form, which we we find to be most effective in inducing linearity. Because potential cost savings or PO error cost may be zero for some SKUs, we add 0.01 to it for the existence of its natural log form. We choose 0.01 because it is the smallest nonzero potential cost savings or PO error cost, and the smallest unit of money, corresponding to one cent. We find no substantial differences in the regression results when a number less than 0.01 is added. To capture the effect of the variation of SKU demand on potential cost savings or PO error cost, we use the SKU unit time demand coefficient of variation (defined as σ/μ) instead of standard deviation σ, the latter being highly positively correlated with demand mean μ. To capture the effect of the backlogging penalty cost on potential cost savings or PO error cost, we use the backlogging penalty ratio (defined as π/h) instead of the backlogging penalty cost per unit of time π, the latter being highly positively correlated with holding cost per unit per unit of time h.

We use the calculated potential cost savings or PO error cost and related parameters for the selected SKUs, as described in the previous numerical experiments, to fit the predictive models. Table 7 reports the regression results for the potential cost savings and PO error cost predictive models. The R-square values for these fits are 0.9675 and 0.9717. All the regression coefficients are statistically significant with p values less than 0.001. According to the signs of the regression coefficients, both potential cost savings and PO error cost decrease in the ordering cost K, and increase in holding cost h and backlogging penalty cost π, as can be shown for the deterministic inventory model and observed numerically for the stochastic model. Both potential cost savings and PO error cost increase in PO error related parameters b_0 and α, as the greater these parameters are, the greater the expected error size of a PO. Both potential cost savings and PO error cost increase in the unit time demand mean μ, as can be shown for the deterministic inventory model. Both potential cost savings and PO error cost increase in the unit time demand coefficient of variation σ/μ, as the more varied the demand, the greater are system inventory costs. Both potential cost savings and PO error cost decrease in lead time from vendor to retailer L, as the effect of PO errors is to increase lead time and lead time variability, and the greater lead time without rework L, the smaller this effect is relative to the lead time L.

Table 7: Regression Results of Predictive Models ($n = 1,076,320$)

	Intercept	$log(K)$	$log(h)$	$log(\pi/h)$	b_0	α	$log(a)$	$log(\mu)$	μ/σ	$log(L)$
Cost Savings	-3.12	-0.24	1.10	0.56	0.05	0.12	0.07	1.14	0.01	-0.48
Error Cost	-0.66	-0.14	0.68	0.16	0.39	0.57	0.45	0.95	0.70	-0.28

The performance of the potential cost savings and PO error cost predictive models is illustrated in Tables 8 and 9, in terms of values and ranks, respectively. In Table 8, "Absolute Error" is defined as the absolute

[14] Note that the remaining PO error-related parameters, β and b_1, are constant across all POs and, hence, all SKUs. Details can be found in Section 5.1.

difference between the predicted and actual values, "Absolute Relative Error" as the average absolute error divided by the actual value across SKUs with nonzero potential cost savings or PO error cost. From this table, we note that the absolute error is not significant for 50% of SKUs. To generate Table 9, we rank 1,076,320 SKUs according to their actual or predicted cost savings from highest to lowest. "Correlation" is calculated between the actual and predicted ranks; "Average Absolute Error" is the average absolute difference between the predicted and actual ranks (the percentages in parentheses are computed from the total number of SKUs to be ranked); and "Largest Absolute Error" is the largest absolute difference between the predicted and actual ranks. From this table, we observe that, on average, our predictive models perform very well in terms of ranking the SKUs.

Table 8: Predictive Performance in Value

	Absolute Error		Absolute Relative Error	
	Average	Median	Average	Median
Potential cost savings	0.26	0.03	19.95%	12.69%
PO error cost	1.09	0.17	17.65%	14.07%

Table 9: Predictive Performance in Rank

	Rank		
	Correlation	Average Absolute Error	Largest Absolute Error
Potential cost savings	0.99	28,690 (2.67%)	537,440 (49.93%)
PO error cost	0.98	37,160 (3.46%)	511,100 (47.48%)

A retailer can use the predictive models to select SKUs on the basis of potential cost savings or PO error cost, using only SKU characteristics. We propose the following procedure for this purpose.

- Step 1: Estimate related parameters for every SKU, as in Sections 5 and 6.
- Step 2: Predict potential cost savings or PO error cost for every SKU using the coefficients from the predictive models derived in Section 5.
- Step 3: Rank SKUs using the predicted cost savings or PO error cost, and choose those which rank highest.

Our methodology has limitations. Our procedure implicitly assumes regression coefficients for the potential cost savings or PO error cost predictive model to be the same for other retailers. The procedure provides a way to selectively adjust and fine tune existing inventory models or reduce error size based on properties of SKUs. This assumption is not very restrictive for retailers similar to our partner retailer. Our partner retailer's only fixed inputs to the foregoing regression analysis include the PO error related parameters b_1 and β, which are constant across all SKUs and will not influence their ranking. Although absolute value predictions may be off for other retailers, the ranking of SKUs should not be much influenced by using fixed inputs from our partner retailer. The procedure may not be suitable for retailers quite different from our partner retailer, but the regression analysis still provides valuable insights in terms of qualitative guidance. For example, PO errors are best addressed by focusing on SKUs with shorter lead time.

Retailers that can identify problematic vendors with a higher PO error cost can collaborate with them to reduce the incidence of such errors. Our procedure can be used to this end by accumulating the predicted PO error cost for SKUs from each vendor. We have used our procedure to predict PO error cost at the vendor level when fixed ordering cost is 100 dollars per order, holding cost is 20% of purchase cost, and backlogging penalty cost is 10 times the holding cost. The absolute relative error is, on average, 15.04% with a median of 12.05%. We can rank vendors according to actual or predicted PO error cost from highest to lowest. The correlation between the actual and predicted ranks is 0.98; the average absolute error between the predicted and actual ranks is 3.5, or 3.30% of the total number of vendors to rank (which is 106), and the largest absolute error between the predicted and actual ranks is 27, or 25.47% of the total number of vendors to rank.

6.3 Inventory Policy with Adjusted Lead Time

We now compare, with our optimal (Q_r, R_r) policy, the performance of an inventory policy with adjusted lead time that accounts for rework, which is the standard (Q, R) policy obtained by adding rework time $\tilde{t}_r$ to the lead time L. We refer to this policy as the adjusted inventory policy (Q_a, R_a). Figure 12 presents the yearly extra cost per SKU incurred by the adjusted inventory policy, averaged over the selected SKUs when the holding cost is 20% of the purchase cost. The extra cost is defined as the inventory cost difference between the adjusted policy and our optimal policy, which is equal to $C_r(Q_a, R_a) - C_r(Q_r, R_r)$. The rework time $\tilde{t}_r$ added to the lead time L averages 96 minutes, with a standard deviation of 314 minutes based on Section 5.2. From this figure, we observe that the adjusted inventory policy costs considerably more than our optimal inventory policy, ranging from \$6.49 to \$11.52 annually per SKU, or about 85.74% to 152.26% of the average purchase cost per SKU.

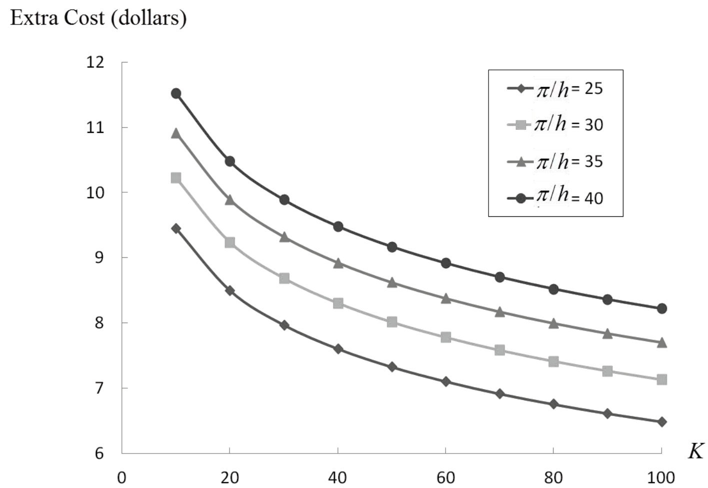

Figure 12: Yearly Extra Cost per SKU Incurred by the Adjusted Policy

7 Conclusions

Vendor noncompliance problems are quite common in retail practice, and retailers and vendors alike are interested in quantifying the impact of such problems on retail inventory management. Vendor noncompliance problems result in PO errors, only one type of which, "quantity errors," has been studied in the random yield literature. Many other types, such as "ticket errors," have received little attention from previous researchers. To the best of our knowledge, our paper is the first to identify the other types of PO errors, and empirically analyze real retail data to obtain their properties. We find the PO errors we studied to vary with order quantity in a way different from what is commonly assumed in the random yield literature. In order to analyze the influence of PO errors on retailers' ordering policy and inventory system cost, we propose deterministic and stochastic inventory models that account for these errors. We study the deterministic inventory model analytically, and the stochastic inventory model numerically, with parameters estimated from real retail data.

PO errors cost retailers tens of dollars per SKU per year, and whether vendors are undercharged or overcharged for these errors depends on the inventory cost parameters. PO error cost is higher for retailers with lower fixed ordering cost, higher holding cost, and higher backlogging penalty cost. Retailers and vendors have fiercely debated whether the chargebacks imposed for these problems are too low or too high. Our findings provide valuable insights into this unsettled debate.

Order size is almost always greater when PO errors are not accounted for than we they are not. Retailers can achieve considerable savings by adjusting their inventory policy to account for PO errors, the cost savings being highest for retailers with lower fixed ordering costs, and higher holding, and backlogging penalty costs. In practice, retailers with hundreds of thousands of SKUs to manage often purchase inventory management software from an outside vendor. We propose procedures for, and provide qualitative guidance in using SKU properties to identify SKUs for which adjusting and fine-tuning the inventory system or reducing error size would be particularly beneficial, thereby sparing retailers paying their software vendors to identify among all their SKUs those most prone to PO errors.

There are several possible directions for future research. First, one might want to relax the assumption that a PO can be used only after rework is finished. For some types of errors, the correct portion of a PO can be used while rework is performed. Second, one might study the joint replenishment problem for two SKUs while accounting for PO errors. One new feature in the two SKU case is that PO errors depend on the number of SKUs in a PO. Substitution errors, whereby a retailer orders SKU A but receives SKU B, can also be accounted for in the two SKU model. Third, it would be interesting to find ways to deal with uncorrectable errors not covered in this paper, time errors, for example. A PO with a time error either arrives earlier or later than stipulated in the contract. Finally, future work might want to examine ways to prevent and reduce PO errors at the source. Retailers might reduce PO errors by adjusting incentives to vendors, as by addressing PO quality in contracts.

Acknowledgement. *The authors thank the many Omega employees who participated in this study. This paper benefited from discussions with Nathan Craig, Ananth Raman, and Linus Schrage as well as seminar participants at Arizona State University, Babson College, Northwestern University, Ohio State University, University of Utah, Willamette University, and Vanderbilt University.*

Appendix A

Proof of Proposition 1:

For any $Q \leq Q_s$,

$$\begin{aligned} C_r'(Q) &= \frac{-\lambda K}{Q^2} + \frac{h\pi}{2(h+\pi)} + c_r a \lambda \rho'(Q) \\ &\leq c_r a \lambda \rho'(Q) < 0. \end{aligned} \tag{10}$$

The first inequality is true because $\frac{-\lambda K}{Q^2}$ increases in Q, and $\frac{-\lambda K}{Q_s^2} + \frac{h\pi}{2(h+\pi)} = 0$. The second inequality is true because $\rho(Q)$ decreases in Q. Hence, $C_r(Q)$ decreases in Q when $Q \leq Q_s$. Since Q_r minimizes $C_r(Q)$ and $C_r'(Q_s) < 0$, we must have $Q_r > Q_s$. □

Proof of Proposition 2:

From $C_s(Q_s) = \sqrt{\frac{2\lambda K h\pi}{h+\pi}}$, we know

$$\frac{dC_s(Q_s)}{dK} = \sqrt{\frac{\lambda h\pi}{2K(h+\pi)}}.$$

Recall that $C_r(Q;K) = \frac{\lambda K}{Q} + \frac{1}{2}\frac{h\pi}{h+\pi}Q + c_r a \lambda \rho(Q)$, and $Q_r = \arg\min_Q C_r(Q;K)$. According to the Envelope Theorem,

$$\frac{dC_r(Q_r)}{dK} = \frac{\lambda}{Q_r}.$$

Therefore we get

$$\frac{C_r(Q_r)}{dK} - \frac{C_s(Q_s)}{dK} = \frac{\lambda}{Q_r} - \sqrt{\frac{\lambda h\pi}{2K(h+\pi)}} \leq 0.$$

This inequality is due to the fact that by Theorem 1 $Q_r \geq Q_s = \sqrt{\frac{2\lambda K(h+\pi)}{h\pi}}$. Hence, $C_r(Q_r) - C_s(Q_s)$ decreases in the fixed ordering cost K. □

We need the following lemma to derive how potential cost savings $C_r(Q_s) - C_r(Q_r)$ changes with fixed ordering cost K.

Lemma 8. *If $\rho(Q) = cQ^{\beta-1}$ for a constant $\beta \in (0,1)$, as the fixed ordering cost $K \to \infty$, we have $\frac{Q_s}{Q_r} \to 1$.*

This lemma tells us that as the fixed ordering cost K goes to infinity, the optimal ordering quantity with rework Q_r gets closer to the optimal EOQ ordering quantity Q_s. Intuitively, as K goes to infinity, the relative influence of the rework cost diminishes. Therefore, the optimal ordering quantity is not much different with than without rework.

Proof of Lemma 8:

First, we show that as $K \to \infty$, $Q_r \to \infty$. Since $Q_r = \arg\min_Q C_r(Q)$, from (10) we know that

$$-\frac{\lambda K}{Q_r^2} + \frac{1}{2}\frac{h\pi}{h+\pi} - c_r ac\lambda(1-\beta)Q_r^{\beta-2} = 0. \tag{11}$$

To calculate $\frac{dQ_r}{dK}$, we take the derivative with regard to K on both sides. We get

$$\frac{2\lambda K}{Q_r^3}\frac{dQ_r}{dK} - \frac{\lambda}{Q_r^2} + c_r ac\lambda(1-\beta)(2-\beta)Q_r^{\beta-3}\frac{dQ_r}{dK} = 0,$$

which implies

$$\frac{dQ_r}{dK} = \frac{Q_r}{2K + c_r ac(1-\beta)(2-\beta)Q_r^{\beta}}. \tag{12}$$

Because $\beta < 1$, we have $\frac{dQ_r}{dK} \geq 0$, which means Q_r is nondecreasing in K. Therefore, as K increases from 0 to ∞, Q_r either converges to a constant or goes to ∞. From (11), we can see that Q_r has to go to ∞ because otherwise (11) does not hold when $K \to \infty$.

Given that $Q_r = \arg\min_Q C_r(Q)$ and (10), we get

$$Q_r = \sqrt{\frac{\lambda K}{\frac{h\pi}{2(h+\pi)} + c_r a\lambda\rho'(Q_r)}}. \tag{13}$$

From (15) and (13), we get

$$\frac{Q_s}{Q_r} = \sqrt{\frac{\frac{h\pi}{2(h+\pi)} + c_r a\lambda\rho'(Q_r)}{\frac{h\pi}{2(h+\pi)}}}. \tag{14}$$

From $\rho(Q) = cQ^{\beta-1}$, we get that

$$\rho'(Q_r) = c(\beta-1)Q_r^{\beta-2}.$$

Because, as shown previously, as $K \to \infty$ so $Q_r \to \infty$, it follows that $\rho'(Q_r) \to 0$ as $K \to \infty$. Therefore, from (14), as $K \to \infty$, $\frac{Q_s}{Q_r} \to 1$. □

Proof of Proposition 3:

From $\rho(Q) = cQ^{\beta-1}$, we get

$$C_r(Q_s) = \sqrt{\frac{2\lambda K h\pi}{h+\pi}} + c_r ac\lambda Q_s^{\beta-1},$$

where

$$Q_s = \sqrt{\frac{2\lambda K(h+\pi)}{h\pi}}. \tag{15}$$

Therefore,

$$\frac{dC_r(Q_s)}{dK} = \sqrt{\frac{\lambda h\pi}{2K(h+\pi)}} - c_r ac\lambda(1-\beta)Q_s^{\beta-2}\sqrt{\frac{\lambda(h+\pi)}{2Kh\pi}}.$$

From the proof of Theorem 1, we know $\frac{dC_r(Q_r)}{dK} = \frac{\lambda}{Q_r}$. Therefore,

$$\begin{aligned}\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK} &= \sqrt{\frac{\lambda h\pi}{2K(h+\pi)}} - c_r ac\lambda(1-\beta)Q_s^{\beta-2}\sqrt{\frac{\lambda(h+\pi)}{2Kh\pi}} - \frac{\lambda}{Q_r} \\ &= \sqrt{\frac{\lambda h\pi}{2K(h+\pi)}} - \frac{\lambda}{Q_r} - c_r ac\lambda(1-\beta)Q_s^{\beta-2}\sqrt{\frac{\lambda(h+\pi)}{2Kh\pi}} \\ &= c_r a\lambda\sqrt{\frac{\lambda(h+\pi)}{2Kh\pi}}[\frac{\frac{h\pi}{h+\pi}(1-\frac{Q_s}{Q_r})}{c_r a\lambda} - c(1-\beta)Q_s^{\beta-2}].\end{aligned}$$

Let $f(K) = \frac{\frac{h\pi}{h+\pi}(1-\frac{Q_s}{Q_r})}{c_r a\lambda} - c(1-\beta)Q_s^{\beta-2}$. Then,

$$\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK} = c_r a\lambda\sqrt{\frac{\lambda(h+\pi)}{2Kh\pi}}f(K).$$

Next we show that $\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK} < 0$. First, as $K \to \infty$, $f(K) \to 0$ because, from Lemma 8, $\frac{Q_s}{Q_r} \to 1$, and $Q_s \to \infty$. Furthermore, $\frac{\lambda(h+\pi)}{2Kh\pi} \to 0$ as $K \to \infty$. Hence,

$$\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK} \to 0 \text{ as } K \to \infty.$$

Second, we show that the derivative of $\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK}$ as a function of K is greater than 0 by proving that $\frac{df(K)}{dK} > 0$. We derive that

$$\frac{df(K)}{dK} = \frac{-\frac{h\pi}{h+\pi}(Q_r\frac{dQ_s}{dK} - Q_s\frac{dQ_r}{dK})}{Q_r^2 c_r a\lambda} + c(1-\beta)(2-\beta)Q_s^{\beta-3}\frac{dQ_s}{dK}.$$

From (15), it follows that $\frac{dQ_s}{dK} = \frac{Q_s}{2K}$. We also know, from Lemma 8, that $\frac{dQ_r}{dK} = \frac{Q_r}{2K + c_r ac(1-\beta)(2-\beta)Q_r^{\beta}}$. Substituting $\frac{dQ_s}{dK}$ and $\frac{dQ_r}{dK}$ into $\frac{df(K)}{dK}$, we get

$$\frac{df(K)}{dK} = -\frac{\frac{h\pi}{h+\pi}(\frac{Q_rQ_s}{2K} - \frac{Q_rQ_s}{2K + c_r ac(1-\beta)(2-\beta)Q_r^{\beta}})}{Q_r^2 c_r a\lambda} + c(1-\beta)(2-\beta)Q_s^{\beta-3}\frac{Q_s}{2K}.$$

This implies

$$\frac{df(K)}{dK} = \frac{Q_sQ_r^{\beta-1}c(1-\beta)(2-\beta)}{2K(2K + c_r ac(1-\beta)(2-\beta)Q_r^{\beta})}g(K),$$

where

$$g(K) = -\frac{h\pi}{\lambda(h+\pi)} + \frac{2K(2K + c_r ac(1-\beta)(2-\beta)Q_r^\beta)}{2K} Q_r^{1-\beta} Q_s^{\beta-3}.$$

We know that

$$\begin{aligned} g(K) &> -\frac{h\pi}{\lambda(h+\pi)} + \frac{2K2K}{2K} Q_r^{1-\beta} Q_s^{\beta-3} \geq -\frac{h\pi}{\lambda(h+\pi)} + 2KQ_s^{-2} \\ &= -\frac{h\pi}{\lambda(h+\pi)} + 2K\left(\sqrt{\frac{h\pi}{2K\lambda(h+\pi)}}\right)^2 = 0. \end{aligned}$$

The first inequality holds because

$$c_r ac(1-\beta)(2-\beta)Q_r^\beta > 0$$

given that 0$< \beta < 1$ and $c > 0$. The second inequality is due to $Q_r \geq Q_s$. Therefore,

$$\frac{df(K)}{dK} = \frac{Q_s Q_r^{\beta-1} c(1-\beta)(2-\beta)}{2K(2K + c_r ac(1-\beta)(2-\beta)Q_r^\beta)} g(K) > 0.$$

As a result, the derivative of $\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK}$ as a function of K is greater than 0. We have shown previously that $\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK} \to 0$ as $K \to \infty$. Therefore, $\frac{dC_r(Q_s)}{dK} - \frac{dC_r(Q_r)}{dK} < 0$. Hence, the cost savings $C_r(Q_s) - C_r(Q_r)$ decreases in the fixed ordering cost K. □

Proof of Proposition 4:
From $C_s(Q_s) = \sqrt{\frac{2\lambda K h\pi}{h+\pi}}$, we get

$$\frac{dC_s(Q_s)}{d\pi} = \frac{1}{2}\sqrt{2K\lambda\frac{h+\pi}{h\pi}}\frac{h^2}{(h+\pi)^2} = \frac{Q_s}{2}\frac{h^2}{(h+\pi)^2}.$$

The second equation is due to the fact that $Q_s = \sqrt{\frac{2K\lambda(h+\pi)}{h\pi}}$. Because $Q_r = \arg\min_Q C_r(Q;\pi)$, according to the Envelope Theorem, we get

$$\frac{dC_r(Q_r)}{d\pi} = \frac{Q_r}{2}\frac{h^2}{(h+\pi)^2}.$$

Therefore, given that $Q_r \geq Q_s$, it holds that

$$\frac{dC_r(Q_r)}{d\pi} - \frac{dC_s(Q_s)}{d\pi} = \frac{Q_r}{2}\frac{h^2}{(h+\pi)^2} - \frac{Q_s}{2}\frac{h^2}{(h+\pi)^2} \geq 0.$$

Hence, $C_r(Q_r) - C_s(Q_s)$ increases in the backlogging penalty cost π. The property that PO error cost $C_r(Q_r) - C_s(Q_s)$ increases in the unit holding cost can be proved similarly by replacing π with h in the previous derivations. □

Proof of Proposition 5:

We now have $\rho(Q) = cQ^{\beta-1}$. Therefore,

$$C_r(Q_s) = \sqrt{\frac{2\lambda K h\pi}{h+\pi}} + c_r a c\lambda Q_s^{\beta-1},$$

where $Q_s = \sqrt{\frac{2K\lambda(h+\pi)}{h\pi}}$. Hence, because $\frac{dQ_s}{d\pi} = -\frac{Q_s}{2}\frac{h}{\pi(h+\pi)}$, we can derive that

$$\begin{aligned}\frac{dC_r(Q_s)}{d\pi} &= \frac{Q_s}{2}\frac{h^2}{(h+\pi)^2} - c_r a c\lambda(1-\beta)Q_s^{\beta-2}\frac{dQ_s}{d\pi} \\ &= \frac{Q_s}{2}\frac{h^2}{(h+\pi)^2} + c_r a c\lambda(1-\beta)Q_s^{\beta-1}\frac{Q_s}{2}\frac{h}{h+\pi}.\end{aligned}$$

Given that, according to the Envelope Theorem, $Q_r = \arg\min_Q C_r(Q;\pi)$, we get $\frac{dC_r(Q_r)}{d\pi} = \frac{Q_r}{2}\frac{h^2}{(h+\pi)^2}$, and

$$\begin{aligned}\frac{dC_r(Q_s)}{d\pi} - \frac{dC(Q_r)}{d\pi} &= \frac{Q_s}{2}\frac{h^2}{(h+\pi)^2} + c_r a c\lambda(1-\beta)Q_s^{\beta-1}\frac{Q_s}{2}\frac{h}{h+\pi} - \frac{Q_r}{2}\frac{h^2}{(h+\pi)^2} \\ &= \frac{h}{2(h+\pi)}\left(\frac{Q_s h}{h+\pi} - \frac{Q_r h}{h+\pi} + c_r a\lambda c(1-\beta)Q_s^{\beta-1}\frac{1}{\pi}\right).\end{aligned}$$

Let $f(\pi) = \frac{Q_s h}{h+\pi} - \frac{Q_r h}{h+\pi} + c_r a\lambda c(1-\beta)Q_s^{\beta-1}\frac{1}{\pi}$. Therefore,

$$\frac{dC_r(Q_s)}{d\pi} - \frac{dC_r(Q_r)}{d\pi} = \frac{h}{2(h+\pi)}f(\pi).$$

Next, we show that there exists some $\beta^* > 0$ such that $f(\pi) \geq 0$ when $\beta \leq \beta^*$. By multiplying (10) by Q_r and using the definition of $\rho(Q)$, we obtain

$$\frac{1}{2}\frac{h\pi}{h+\pi}Q_r - \frac{K\lambda}{Q_r} - (1-\beta)c_r a\rho(Q_r)\lambda = 0. \tag{16}$$

From (15), we obtain

$$\frac{1}{2}\frac{h\pi}{h+\pi}Q_s - \frac{K\lambda}{Q_s} = 0. \tag{17}$$

Subtracting (17) from (16), we get

$$(Q_r - Q_s)\left(\frac{1}{2}\frac{h\pi}{h+\pi} + \frac{K\lambda}{Q_s Q_r}\right) = c_r a(1-\beta)cQ_r^{\beta-1}\lambda,$$

which implies

$$Q_r - Q_s = \frac{c_r a(1-\beta)cQ_r^{\beta-1}\lambda}{\left(\frac{1}{2}\frac{h\pi}{h+\pi} + \frac{K\lambda}{Q_s Q_r}\right)}. \tag{18}$$

Plugging (18) into $f(\pi)$, it follows

$$
\begin{aligned}
f(\pi) &= -\frac{h}{h+\pi}\frac{c_r a(1-\beta)cQ_r^{\beta-1}\lambda}{\left(\frac{1}{2}\frac{h\pi}{h+\pi}+\frac{K\lambda}{Q_sQ_r}\right)} + c_r a\lambda c(1-\beta)Q_s^{\beta-1}\frac{1}{\pi} \\
&= c_r a(1-\beta)c\lambda\left(-\frac{h}{h+\pi}\frac{Q_r^{\beta-1}}{\frac{1}{2}\frac{h\pi}{h+\pi}+\frac{K\lambda}{Q_sQ_r}}+\frac{1}{\pi}Q_s^{\beta-1}\right) \\
&= \frac{c_r a(1-\beta)c\lambda}{\pi}\left(-\frac{Q_r^{\beta}}{\frac{1}{2}(Q_r+Q_s)}+Q_s^{\beta-1}\right) \\
&= \frac{c_r a(1-\beta)c\lambda}{\pi}\left(\frac{-Q_r^{\beta}Q_s^{1-\beta}+\frac{1}{2}(Q_s+Q_r)}{\frac{1}{2}(Q_r+Q_s)Q_s^{1-\beta}}\right)
\end{aligned}
$$

Let $g(\beta) = -Q_r^{\beta}Q_s^{1-\beta}+\frac{1}{2}(Q_s+Q_r)$. We can see that $g(0) > 0$ because, from Theorem 1, $Q_r > Q_s$. Given that $g(\beta)$ is a continuous function of β, there must exist some $\beta^* > 0$ such that, for any $\beta \leq \beta^*$, it follows that $g(\beta) > 0$ and, thus, $f(\beta) \geq 0$. Therefore, potential cost saving $C_r(Q_s) - C_r(Q_r)$ increases in the backlogging penalty cost π if $\beta \leq \beta^*$. The property that potential cost savings $C_r(Q_s) - C_r(Q_r)$ increases in the unit holding cost can be proved similarly by replacing π with h in the previous derivations. $\square$

Proof of Proposition 6:

First, we show PO error cost $C_r(Q_r) - C_s(Q_s)$ to increase in demand rate λ. From $C_s(Q_s) = \sqrt{\frac{2\lambda Kh\pi}{h+\pi}}$, we get

$$\frac{dC_s(Q_s)}{d\lambda} = \sqrt{\frac{2Kh\pi}{h+\pi}}\frac{1}{2\sqrt{\lambda}}.$$

Given that, according to the Envelope Theorem, $Q_r = \arg\min_Q C_r(Q;\lambda)$, we get

$$\frac{dC_r(Q_r)}{d\lambda} = \frac{K}{Q_r} + c_r a\rho(Q_r).$$

Therefore,

$$\frac{dC_r(Q_r)}{d\lambda} - \frac{dC_s(Q_s)}{d\lambda} = \frac{K}{Q_r} + c_r a\rho(Q_r) - \sqrt{\frac{2Kh\pi}{h+\pi}}\frac{1}{2\sqrt{\lambda}}.$$

After multiplying (10) by Q_r, noting that $\rho'(Q) = (\beta-1)\frac{\rho(Q)}{Q}$ when $\rho(Q) = cQ^{\beta-1}$, we have

$$\frac{1}{2}\frac{h\pi}{h+\pi}Q_r - \frac{K\lambda}{Q_r} - (1-\beta)c_r a\rho(Q_r)\lambda = 0,$$

and, in turn,

$$\frac{K}{Q_r} + c_r a\rho(Q_r) > \frac{K}{Q_r} + (1-\beta)c_r a\rho(Q_r)$$

$$= \frac{1}{2}\frac{h\pi}{h+\pi}\frac{Q_r}{\lambda} > \sqrt{\frac{2Kh\pi}{h+\pi}}\frac{1}{2\sqrt{\lambda}}.$$

The first inequality is because $\beta < 1$; the second follows from Theorem 1, $Q_r > Q_s = \sqrt{\frac{2K\lambda(h+\pi)}{h\pi}}$ and, thus, $\frac{dC_r(Q_r)}{d\lambda} - \frac{dC_s(Q_s)}{d\lambda} > 0$. That is, $C_r(Q_r) - C_s(Q_s)$ increases in the demand rate λ.

Next, we show that there exists some $\bar{\beta} > 0$ such that if $\beta \leq \bar{\beta}$, then cost savings $C_r(Q_s) - C_r(Q_r)$ increases in demand rate λ. From $C_r(Q_s) = \sqrt{\frac{2\lambda Kh\pi}{h+\pi}} + c_r a\rho(Q_s)\lambda$, we know that

$$\begin{aligned}\frac{dC_r(Q_s)}{d\lambda} &= \sqrt{\frac{2Kh\pi}{h+\pi}}\frac{1}{2\sqrt{\lambda}} + c_r a\rho(Q_s) + c_r a\lambda\rho'(Q_s)\frac{dQ_s}{d\lambda} \\ &= \sqrt{\frac{2Kh\pi}{h+\pi}}\frac{1}{2\sqrt{\lambda}} + \frac{\beta+1}{2}c_r a\rho(Q_s)\end{aligned}$$

because $\frac{dQ_s}{d\lambda} = \frac{Q_s}{2\lambda}$ and $\rho(Q) = cQ^{\beta-1}$. Therefore,

$$\begin{aligned}\frac{dC_r(Q_s)}{d\lambda} - \frac{dC_r(Q_r)}{d\lambda} &= \sqrt{\frac{2Kh\pi}{h+\pi}}\frac{1}{2\sqrt{\lambda}} + \frac{\beta+1}{2}c_r a\rho(Q_s) - \frac{K}{Q_r} - c_r a\rho(Q_r) \\ &= \frac{K}{Q_s} + \frac{\beta+1}{2}c_r a\rho(Q_s) - \frac{K}{Q_r} - c_r a\rho(Q_r). \qquad (19)\end{aligned}$$

Subtracting (17) from (16), we get

$$\frac{K\lambda}{Q_s} - \frac{K\lambda}{Q_r} = -\frac{1}{2\lambda}\frac{h\pi(Q_r - Q_s)}{h+\pi} + (1-\beta)c_r a\rho(Q_r). \qquad (20)$$

Therefore, it follows from (19) and (20) that

$$\frac{dC_r(Q_s)}{d\lambda} - \frac{dC_r(Q_r)}{d\lambda} = -\frac{1}{2\lambda}\frac{h\pi(Q_r - Q_s)}{h+\pi} + \frac{\beta+1}{2}c_r a\rho(Q_s) - \beta c_r a\rho(Q_r).$$

Plugging from (18), we get

$$\begin{aligned}\frac{dC_r(Q_s)}{d\lambda} - \frac{dC_r(Q_r)}{d\lambda} &= -\frac{1}{2\lambda}\frac{h\pi}{h+\pi}\frac{c_r a(1-\beta)cQ_r^{\beta-1}\lambda}{\frac{1}{2}\frac{h\pi}{h+\pi} + \frac{K\lambda}{Q_r Q_s}} + \frac{\beta+1}{2}c_r a\rho(Q_s) - \beta c_r a\rho(Q_r) \\ &= -\frac{c_r a(1-\beta)cQ_r^{\beta}}{Q_r + Q_s} + \frac{\beta+1}{2}c_r a\rho(Q_s) - \beta c_r a\rho(Q_r) \\ &= \frac{c_r ac}{Q_r + Q_s}\left(-Q_r^{\beta} - \beta Q_r^{\beta-1}Q_s + \frac{\beta+1}{2}Q_s^{\beta} + \frac{\beta+1}{2}Q_s^{\beta-1}Q_r\right).\end{aligned}$$

When $\beta = 0$, we have

$$\frac{dC_r(Q_s)}{d\lambda} - \frac{dC_r(Q_r)}{d\lambda} = -\frac{1}{2} + \frac{1}{2}\frac{Q_r}{Q_s}. \qquad (21)$$

Because, according to Theorem 1 and by (21), $Q_r > Q_s$, it follows that $\frac{dC_r(Q_s)}{d\lambda} - \frac{dC_r(Q_r)}{d\lambda} > 0$ when

$\beta = 0$. Given that $\dfrac{dC_r(Q_s)}{d\lambda} - \dfrac{dC_r(Q_r)}{d\lambda}$ is a continuous function of β, there exists some $\bar{\beta} > 0$ such that $\dfrac{dC_r(Q_s)}{d\lambda} - \dfrac{dC_r(Q_r)}{d\lambda} \geq 0$ if $\beta \leq \bar{\beta}$. Therefore, potential cost savings $C_r(Q_s) - C_r(Q_r)$ increases in demand rate λ if $\beta \leq \bar{\beta}$. $\square$

Proof of Proposition 7:

We know that $E(\tilde{W}(Q))/Q = cp(Q)Q^{\beta-1}$, where c is defined as $e^{(\alpha+\sigma_s^2/2)}$. The probability that there is an error has the following form:

$$p(Q) = \frac{e^{b_0+b_1 logQ}}{1+e^{b_0+b_1 logQ}} = \frac{dQ^{b_1}}{1+dQ^{b_1}},$$

where d is defined as e^{b_0}. Therefore,

$$\frac{dp(Q)}{dQ} = \frac{db_1Q^{b_1-1}}{(1+dQ^{b_1})^2} = p(Q)\frac{b_1}{Q(1+dQ^{b_1})}.$$

Based on this, we have

$$\begin{aligned}
\frac{dE(\tilde{W}(Q))/Q}{dQ} &= (\beta-1)cp(Q)Q^{\beta-2} + cQ^{\beta-1}\frac{dp(Q)}{dQ} \\
&= (\beta-1)cp(Q)Q^{\beta-2} + cQ^{\beta-1}p(Q)\frac{b_1}{Q(1+dQ^{b_1})} \\
&= cp(Q)Q^{\beta-2}((\beta-1) + \frac{b_1}{(1+dQ^{b_1})}) \\
&\leq cp(Q)Q^{\beta-2}((\beta-1)+b_1).
\end{aligned}$$

The last inequality holds because $dQ^{b_1} \geq 0$ for any $Q \geq 0$. Hence, if $\beta + b_1 \leq 1$, then $\dfrac{dE(\tilde{W}(Q))/Q}{dQ} \leq 0$. As a result, $E(\tilde{W}(Q))/Q$ decreases in Q. $\square$

Appendix B

We first derive the mean of the error size for a PO for a single SKU with order quantity Q, given that there is an error. We know that, given $\tilde{W}(Q) > 0$,

$$log(\tilde{W}(Q)) = \alpha + \beta log(Q) + \gamma log(n) + e_s$$

with number of SKUs n equal to 1 and e_s normally distributed with mean 0 and variance σ_s^2. Hence, given $\tilde{W}(Q) > 0$, we have $\tilde{W}(Q) = e^{\alpha}Q^{\beta}\exp(e_s)$, where $\exp(e_s)$ has a log-normal distribution with mean $e^{\sigma_s^2/2}$. Therefore,

$$E(\tilde{W}(Q)|\tilde{W}(Q) > 0) = e^{\alpha+\sigma_s^2/2}Q^{\beta}.$$

In addition, $E(\tilde{W}(Q)) = E(\tilde{W}(Q)|\tilde{W}(Q) > 0)P(\tilde{W}(Q) > 0) = p(Q)e^{\alpha+\sigma_s^2/2}Q^{\beta}$.

Next, we derive the rework time mean $E(\tilde{t}_r(Q))$ and variance $Var(\tilde{t}_r(Q))$. Recalling that e_s is normally distributed with mean 0 and variance σ_s^2, and $exp(e_s)$ has a log-normal distribution with mean $e^{\sigma_s^2/2}$ and variance $(e^{\sigma_s^2}-1)e^{\sigma_s^2}$, we get

$$\begin{aligned}
E(\tilde{t}_r(Q)) &= E(\tilde{t}_r(Q)|\tilde{W}(Q)>0)P(\tilde{W}(Q)>0) + E(\tilde{t}_r(Q)|\tilde{W}(Q)=0)P(\tilde{W}(Q)=0) \\
&= E(a\tilde{W}(Q)+e^r|\tilde{W}(Q)>0)p(Q) + 0\times P(\tilde{W}(Q)=0) \\
&= aE(\tilde{W}(Q)|\tilde{W}(Q)>0)p(Q) \\
&= ap(Q)e^{(\alpha+\sigma_s^2/2)}Q^\beta,
\end{aligned}$$

and

$$\begin{aligned}
Var(\tilde{t}_r(Q)) &= E(\tilde{t}_r(Q)^2) - E^2(\tilde{t}_r(Q)) \\
&= E(\tilde{t}_r(Q)^2|\tilde{W}(Q)>0)P(\tilde{W}(Q)>0) + E(\tilde{t}_r(Q)^2|\tilde{W}(Q)=0)P(\tilde{W}(Q)=0) - E^2(\tilde{t}_r(Q)) \\
&= (Var(\tilde{t}_r(Q)|\tilde{W}(Q)>0) + E^2(\tilde{t}_r(Q)|\tilde{W}(Q)>0))p(Q) + 0\times P(\tilde{W}(Q)=0) - E^2(\tilde{t}_r(Q)) \\
&= (Var(a\tilde{W}(Q)+e^r|\tilde{W}(Q)>0) + E^2(a\tilde{W}(Q)+e^r|\tilde{W}(Q)>0))p(Q) - E^2(\tilde{t}_r(Q)) \\
&= [a^2Var(\tilde{W}(Q)|\tilde{W}(Q)>0) + \sigma_r^2 + (ae^{(\alpha+\sigma_s^2/2)}Q^\beta)^2]p(Q) - (ap(Q)e^{(\alpha+\sigma_s^2/2)}Q^\beta)^2 \\
&= [a^2Var(e^\alpha Q^\beta exp(e_s)) + \sigma_r^2 + (ae^{(\alpha+\sigma_s^2/2)}Q^\beta)^2]p(Q) - (ap(Q)e^{(\alpha+\sigma_s^2/2)}Q^\beta)^2 \\
&= [(ae^\alpha Q^\beta)^2(e^{\sigma_s^2}-1)e^{\sigma_s^2}) + \sigma_r^2 + (ae^{(\alpha+\sigma_s^2/2)}Q^\beta)^2]p(Q) - (ap(Q)e^{(\alpha+\sigma_s^2/2)}Q^\beta)^2 \\
&= a^2e^{2\alpha+\sigma_s^2}Q^{2\beta}(e^{\sigma_s^2}-p(Q))p(Q) + \sigma_r^2 p(Q).
\end{aligned}$$

References

Aron, L. J. (1998). More carrot, less stick for vendors. *Chain Store Age*, 74(12):232.

Axsater, S. (2006). *Inventory Control*. Springer Science + Business Media, LLC, second edition.

Chandran, P. M. (2003). Wal-mart's supply chain management practices. *ICFAI Center for Management Research*.

Federgruen, A. and Zheng, Y. (1992). An efficient algorithm for computing an optimal (Q,R) policy in continuous review stochastic inventory systems. *Operations Research*, 40(4):808–813.

Hadleyand Thomson M. Whitin, G. (1963). *Analysis of inventory systems*. Prentice Hall.

Hariga, M. A. (1999). A stochastic inventory model with lead time and lot size interaction. *Production Planning and Control*, 10(5):434–438.

Kim, J. and Benton, W. (1995). Lot size dependent lead times in a (Q,R) inventory system. *International Journal of Production Research*, 33(1):41–58.

Kleinbaum, D. G., Kupper, L. L., and Muller, K. E. (1998). *Applied Regression Analysis and Other Multivariable Methods*. Duxbury Press, Pacific Grove, CA.

Lee, H. L. (1992). Lot sizing to reduce capacity utilization in a production process with defective items, process corrections and rework. *Management Science*, 38(9):1315–1328.

Levy, M. and Weitz, B. A. (2004). *Retailing Management*, chapter 14 Buying Systems. McGraw-Hill/Irwin, fifth edition.

Moinzadeh, K. and Lee, H. L. (1987). A continuous-review inventory model with constant resupply time and defective items. *Naval Research Logistics*, 34(4):457–467.

Moinzadeh, K. and Lee, H. L. (1989). Approximation order quantities and reorder points for inventory systems where orders arrive in two shipments. *Operations Research*, 37(2):277–287.

Peters, M. H., Schneider, H., and Tang, K. (1988). Joint determination of optimal inventory and quality control policy. *Management Science*, 34(8):991–1004.

Porteus, E. L. (1986). Optimal lot sizing, process quality improvement and setup cost reduction. *Operations Research*, 34(1):137–144.

Raudenbush, S. W. and Bryk, A. S. (2001). *Hierachical Linear Models: Applications and Data Analysis Methods*. Sage Publications, second edition.

Rosling, K. (1999). An efficient algorithm for computing an optimal (Q,R) policy in continuous review stochastic inventory systems. *Working Paper*.

So, K. C. and Tang, C. S. (1995). Optimal batch sizing and repair stages for operations with repairable jobs. *Management Science*, 41(5):894–908.

van Donselaar, K. H., Gaur, V., van Woensel, T., Broekmeulen, R. A. C., and Fransoo, J. C. (2010). Ordering behavior in retail stores and implications for automated replenishement. *Management Science*, 56(5):766–784.

Zhang, X. and Gerchak, Y. (1990). Joint lot sizing and inspection policy in an EOQ model with random yield. *IIE Transactions*, 22(1):41–47.

Zieger, A. (2003). Retailer chargebacks: Is there an upside? Retailer compliance initiatives can lead to efficiency. *Frontline Solutions*.

Zipkin, P. H. (2000). *Foundations of Inventory Management*. McGraw-Hill.