JOURNAL OF BUSINESS LOGISTICS
STRATEGIC SUPPLY CHAIN RESEARCH

Journal of Business Logistics, 2019, 1–13 doi: 10.1111/jbl.12227

Improving Order Fulfillment Performance through Integrated Inventory Management in a Multi-Item Finished Goods System

Liu Yang[1], Haitao Li[2], and James F. Campbell[2]

[1]*Sam Houston State University*
[2]*University of Missouri – St. Louis*

Effective inventory management is fundamental to order fulfillment excellence and supply chain success. In this paper, we develop a strategic inventory management decision tool that integrates inventory classification and inventory control policy decisions for maximizing order fulfillment performance, while accounting for a constraint on inventory budget and the profit expectation of a firm. This inventory solution tool provides critical enhancements to current inventory planning software, which is developed upon the traditional inventory classification scheme and where practitioners have to balance service levels and safety stock decisions through trial-and-error. The model allows firms to assess whether the current inventory performance is Pareto optimal, quantify the trade-offs between various performance measures, and identify the right inventory level according to the firms' strategic goals. In computational results, we demonstrate the trade-off and positive relationships between key item- and order-based inventory performance measures and short-term profitability under different levels of inventory budget in a multi-item finished goods inventory system.

Keywords: order fulfillment optimization; inventory classification; order fill rate; trade-off of profit and service level; mixed-integer programming

INTRODUCTION

A rich stream of interdisciplinary research on logistics and marketing has shown that order fulfillment performance is a significant determinant of customer satisfaction and loyalty, directly affecting customer future purchasing behavior and firm revenue (e.g., Daugherty et al. 1998; Stank et al. 2003; Davis-Sramek et al. 2010; Rao et al. 2011; Griffis et al. 2012). A fundamental element dictating order fulfillment performance is inventory control policies. While single-item inventory control has been extensively studied, there is little research available to guide companies on establishing inventory policies for order fulfillment improvement in a multi-item finished goods inventory system. The modern inventory planning software packages are developed upon the traditional inventory management approach and concern the service level of stock keeping unit (SKU) rather than customer orders. Typically, companies group the large variety of items into a few classes according to the Pareto principle or multicriteria inventory classification (MCIC) scheme and then assign each class a subjectively determined cycle service level (CSL; i.e., the expected probability of not having a stock-out during a replenishment cycle; Mohammaditabar et al., 2012; van Kampen et al. 2012; Teunter et al. 2017). For example, Microsoft prioritizes its hardware products into A, B, C, and D classes based on item revenue, life cycle status, profit contributions, and marketing factors and then sets CSL targets for each class (Neale and Willems 2009). With the help of inventory planning software, required inventory level is computed and reviewed by the management. If the inventory level is acceptable, the process is completed; otherwise, the inventory planner needs to adjust the CSL of each class until the inventory level is within the budget. Figure 1 depicts this process. It requires trial-and-error during inventory planning to balance service level and safety stock. Moreover, the entire procedure is performed independently of order fulfillment measures; it takes no consideration of order structure, and it is unclear how the inventory classification and the class-based service levels may impact order fulfillment. It has long been argued that the classification criterion, the number of classes, the determination of the cutoff value between classes, and the service level of each class rely more on managerial judgment than quantitative analysis (e.g., Viswanathan and Bhatnagar 2005; Stanford and Martin 2007; Teunter et al. 2010; Lajili et al. 2012). A review of the inventory literature has not identified any work that simultaneously optimizes inventory classification and service level (and the corresponding safety stock) decisions to maximize order fulfillment measures.

In addition, supply chain is a complex network, requiring companies to continuously weigh and balance key trade-offs in inventory management. The leading-edge inventory planning software offers trial-and-error exploration of trade-offs and alternatives, but is limited in two aspects (Figure 2). First, it is done at individual SKUs or SKU-class levels instead of a system-wide approach; second, only a small number of factors are considered in decision making, while the important business aspects such as inventory budget, short-term profit requirement, order criticality, and order-varying pricing are not accounted for. To fill these gaps, we develop a strategic inventory management decision tool in concert with a food product manufacturer to address the order-fulfillment-optimized inventory classification and safety stock decisions, while meeting the profit expectation of a firm and subject to an inventory budget constraint. The solution takes a holistic, systems approach to the inventory management and offers critical enhancements to the current inventory planning software packages.

Consider a particular order (order k) that consists of multiple items. The fill rate for this order, denoted as R_k, is defined as the fraction of items in order k that are filled from on-hand

Corresponding author:
Liu Yang, College of Business Administration, Sam Houston State University 1905 University Ave, Huntsville, TX 77340, USA; E-mail: LYang@shsu.edu

Figure 1: Inventory planning process.

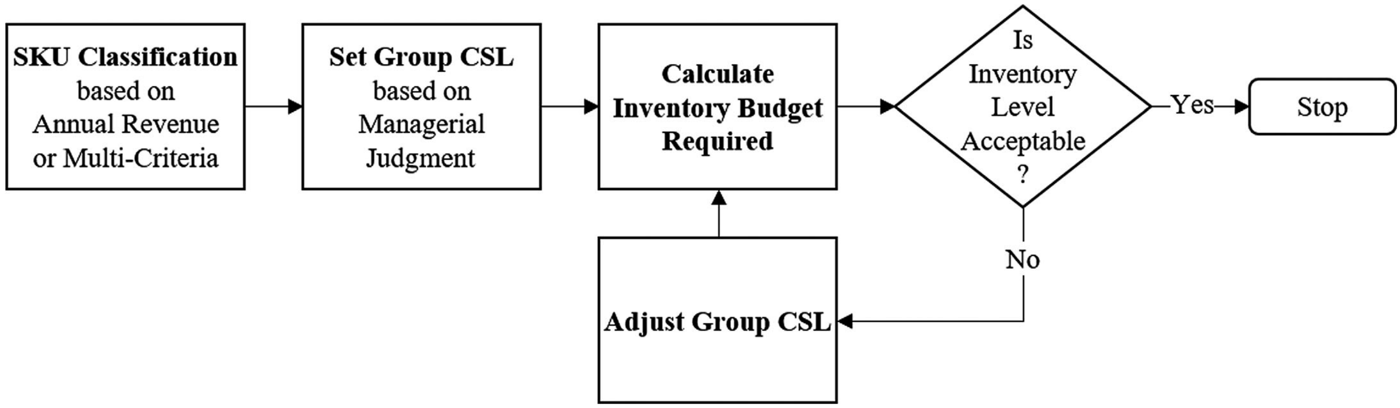
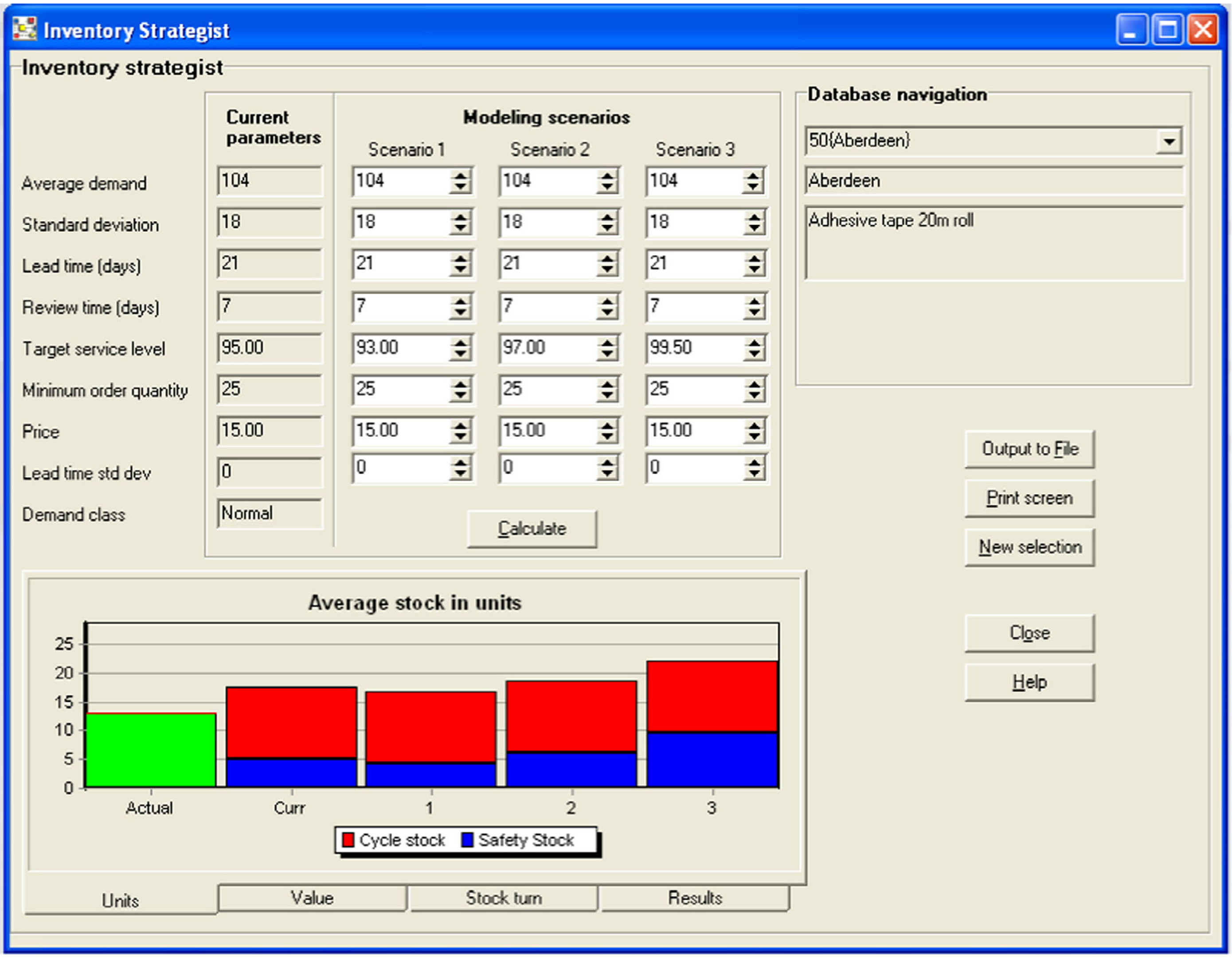


Figure 2: Inventory planning software interface. Adapted from Teunter et al. (2017).



inventory. In this study, we focus on two customer-order-based measures to gauge inventory performance: (1) the order fill rate (OFR), which is defined as the percentage of orders that are *completely* filled from on-hand inventory (e.g., Lu et al. 2003; Closs et al. 2010; Bowersox et al. 2012), and (2) the weighted average customer-order fill rate (WAFR), which is defined as a weighted average of the fill rates for the orders (i.e., the weighted R_k values). The weights are used to reflect the importance or criticality of individual orders or customers. When the same weight is applied across all customer orders, the WAFR becomes the average customer-order fill rate, a measure introduced by Larsen and Thorstenson (2014) in their investigation of order fulfillment performance in a single-item inventory system. The inventory literature addressing order fulfillment is limited and generally considers the OFR only (e.g., Song 1998; Song and Yao 2002; Lu et al. 2003; Closs et al. 2010; Hoen et al. 2011), which is a binary measure of a customer order—an order is either *completely* filled or *not completely* filled. The OFR is highly relevant to a firm's overall service performance in assemble-to-order (ATO) environments when all items are required to complete an activity, but in general cases this measure has two major drawbacks. First, it does not reflect the depth of demand satisfied by inventory on hand. An incomplete order disappoints the customer, but a 50%-filled order and a 98%-filled order can have drastically different effects on customer loyalty and incur different costs for suppliers. Second, it is difficult to account for order criticality when each order is measured by either completely or not completely filled. To the best of our knowledge, no prior research addressing the estimate of the OFR has taken into account order criticality. Hence, we use the WAFR to complement the OFR and thereby address both the depth of satisfied demand and the criticality of orders.

It is worth noting that companies typically use the item fill rate (IFR) to gauge inventory performance, where the IFR is defined as the percentage of total demand volume of a stock keeping unit (SKU) that can be filled from on-hand stock (e.g., Ballou and Burnetas 2003; Ballou 2005; Thomas 2005). The IFR (also known as the volume fill rate or the fill rate) is an item-based measure, but not (directly) an order-based measure like the OFR or the WAFR. The IFR can be directly translated to the CSL to guide inventory policy setting. However, it is important to recognize that the IFR is not a customer-centric indicator measuring the service received by customers; rather, it is a manufacturer/distributor/retailer-focused measure assessing internal operations (Anupindi and Tayur 1998; Zinn et al. 2002). The simulations conducted by Anupindi and Tayur (1998) demonstrate that the IFR cannot adequately indicate the OFR, and Song (1998) shows that the OFR may perform poorly, while the IFR is satisfactory in a multi-item inventory system. Finding an optimal inventory classification scheme and control policy with customer-oriented performance measures is a nontrivial task, especially for a myriad of SKUs. To this end, we develop a data-driven optimization approach based on mixed integer linear programming (MILP), with the objective of improving the OFR and the WAFR measures.

Our MILP model provides a vehicle for various sensitivity analysis on the impacts of varying problem parameters on the optimal solution; specifically, how the OFR, WAFR, and IFR are impacted when inventory is optimized for one performance measure under different levels of inventory capital, and with different profit targets. The results extend findings from single-item inventory research. For example, Larsen and Thorstenson (2014) show that the average customer-order fill rate always exceeds the IFR and OFR in a single-item inventory system. Our computational results show that in a multi-item finished goods inventory system, the IFR can be the highest among the three service measures even though the inventory is optimized for order fulfillment performance. We also document the significant impact that an inventory budget constraint has on the relationship between the IFR, OFR, and WAFR. The IFR may provide a good indication of the WAFR when inventory capital is relatively sufficient, but it can considerably overstate the WAFR when the budget is tight. On the other hand, the IFR can closely approximate the OFR only when the IFR is very high (99.9% or higher), consistent with the observation in a configure-to-order system by Closs et al. (2010). We also show that under a limited inventory budget, maximizing the OFR may result in much less immediate profit for a company, while maximizing profitability may lead to a significantly compromised OFR. Importantly, we show whether the company's inventory budget is most efficiently allocated, and when the company may have the opportunity to improve the OFR and (short-term) profitability simultaneously with additional inventory investment. Finally, by incorporating order criticality with the WAFR, we capture the ability to prioritize customer orders while optimizing inventory performance. Our model offers practitioners a tool to quantify the trade-off between customer service level and (short-term) profit, to identify the right level of inventory budget required to support the firm's strategic goals, and to evaluate the impacts of critical orders and customers.

The remainder of the paper is organized as follows. The next section provides a review of inventory literature concerning order-based performance measures and identifies the gaps motivating this study. The following section presents the modeling framework and formulation for an order-fulfillment-maximized inventory management problem. Then, comprehensive computational experiments are performed and the results are discussed, followed by conclusions and opportunities for future research. The model formulation for item-fill-rate-maximized inventory problem is provided in Appendix.

LITERATURE REVIEW

In this section, we focus on studies with order-based performance measures. Inventory classification and the importance of having an end-to-end, integrated inventory solution have been extensively discussed in Teunter et al. (2010), Lajili et al. (2012), and Yang et al. (2017). The present research extends the approach taken by Millstein et al. (2014) and Yang et al. (2017) that integrates SKU classification and inventory performance measures, presenting a strategic inventory optimization model that allows manufacturers and distributors with finished goods inventory to maximize order fulfillment performance under an inventory budget constraint, and subject to a minimum (short-term) profit requirement.

Much research on order fulfillment is related to assemble-to-order (ATO) or configure-to-order (CTO) systems, where the final product is made to order and components are controlled by a base-stock policy. For example, Song and Yao (2002) examine the trade-off between the OFR and inventory holding cost in a single-product assembly system where the customer demand follows a Poisson process. Lu et al. (2003) extend their research to a multi-item ATO system and develop approximations and bounds for the response-time-based OFR. Hoen et al. (2011) take a weighted average of the underestimated and overestimated OFRs to approximate the OFR in a lost-sales system, and demonstrate that the OFR is insensitive to the distribution of the component lead times. Closs et al. (2010) use simulation modeling to examine how the IFR and the OFR behave differently under various settings in a CTO environment, and find that a relatively high IFR does not necessarily lead to a high OFR unless the IFR is over 99%.

The distribution system setting for our research differs from the ATO/CTO system in two major aspects. First, unlike the ATO system with a fixed bill of materials (i.e., a fixed set of components with fixed quantity for each product), a customer order to a finished goods inventory system may have a fixed set of SKUs, but the demand for each SKU typically varies over time following certain probability distribution. Second, the OFR is most appropriate in an ATO environment since all the components of an end-product must be available before the product can be assembled and shipped. In a multi-item distribution system however, the items in a customer order are often not closely related and partial fulfillment is common in the business-to-business environment. A binary measure of order fulfillment (OFR) does not provide much insight on the service levels that individual customers receive.

Order fulfillment research on multi-item distribution systems is limited. Song (1998) develops heuristic bounds to estimate the OFR using item-based information, but the effectiveness of bounds varies by the order structure and the relative order size. Hausman et al. (1998) present optimization models to maximize the probability of filling all demands subject to an inventory budget constraint. The authors use the aggregated item demand; thus, their research only provides bounds on the fulfillment probability of individual orders. Anupindi and Tayur (1998) compare the IFR with order-based performance measures including order response times and the OFR for a multi-item cyclic production system, and demonstrate that the IFR is not a good indicator for order-based measures. Zinn et al. (2002) propose four order-based measures to assess the probabilities of no stock-out for a customer order and an item in a customer order. The authors present an analysis framework and an optimization model to minimize the safety stock investment subject to a service level constraint. The model is formulated for individual customers (customer orders), calculating the optimal safety stock factor for each order, which is then combined to determine the total safety stock required for an item. All the above works focus on the OFR or the OFR-equivalent measure. The only research that has been found to gauge the fill rate of an order (R_k) is Larsen and Thorstenson (2014), who investigate how the order-based and item-based performance measures are related in a single-product environment, but their observation on the relationship between the average customer-order fill rate, the OFR, and the IFR is not explored for a multiproduct distribution system.

In this study, we take a practice-oriented approach to the order-oriented inventory optimization problem and quantify the trade-offs between the OFR, WAFR, IFR, and short-term profit in a multi-item distribution system. We extend the framework of Zinn et al. (2002) to maximize the order fulfillment performance at the company level while meeting both a minimum profit requirement and an inventory budget constraint, in response to the fact that companies constantly need to balance profit, customer satisfaction, and investment.

ORDER-FULFILLMENT-BASED INVENTORY MODEL

In this section, an MILP model is developed to determine optimal inventory levels of individual SKUs for maximizing order fulfillment subject to inventory budget and (short-term) profit constraints.

Problem description and model assumptions

We consider a multi-item distribution system in a business-to-business environment where retailers or industrial customers repurchase the same variety of products from manufacturers/distributors over regular intervals. Hence, the demand pattern of a product requested in a particular order can be estimated from the customer's purchasing history as observed by Zinn et al. (2002). Let N denote the set of SKUs in the inventory and K be the set of customer orders. The set of SKUs required by customer order type $k \in K$ is known and denoted by N^k. Several customers may order the similar set of SKUs, but as the demand pattern likely differs by customers, we consider them as different types of customer orders in the model. Demand for SKU $i \in N$ in order $k \in K$ is assumed to follow a normal distribution with forecast mean d_{ik} units and standard deviation σ_{ik}. The assumption of normally distributed demand is widely adopted in inventory literature including Hadley and Whitin (1963), Johnson et al. (1995), Silver and Bischak (2011), and Disney et al. (2015) to name but a few. Experiments conducted by Porras and Dekker (2008) show that the models developed using the normal distribution generally perform well empirically, even for the demand process that does not follow a normal distribution. Because the actual demand is treated as a random variable, the expected fill rate for an order can never be 100% (unless the inventory level is infinite). Therefore, we consider an order to be completely filled if the expected fill rate for that order (R_k) is 99.99% or higher.

We assume that orders are independent of each other and an order occurs once per time unit, as manufacturers and distributors typically replenish the warehouses of downstream customers once per period (e.g., weekly) to take advantage of the economies of scale in transportation. In the cases where multiple orders occur regularly within a time period from the same customer, they can be treated as different orders in the model. Customer order k is prioritized through an assigned weight w_k ($w_k > 0$), and companies may set a minimum fill rate requirement Υ_k for an order (i.e., $R_k \geq \Upsilon_k$) if desired. We model a scenario where partially filling an order is acceptable if an SKU required in an order is in shortage. The model assumes no back-ordering and unmet demand is treated as lost sales, which is typical in a highly competitive market or if products have long lead time. The model can be easily adapted to accommodating penalty costs of stock-out.

The unit profit from SKU $i \in N^k$ is denoted by ρ_{ik}, an order-SKU-based parameter, considering that the company may quote different customers different prices based on the order quantity or the strategic position of a customer. The SKU unit cost, c_i, includes material cost and inbound shipping cost. Inventory holding cost h_i is measured as a percentage of the goods value, including capital cost, warehousing cost, physical handling, damage, and obsolescence cost (in relation to shelf life), and may differ by SKU. The proposed model takes into account the limited inventory budget ω, which sets the maximum dollar value of on-hand inventory in a time period. The inclusion of an inventory budget constraint is important in practical applications because it affects inventory turnover rate and the amount that a company can borrow is typically limited. Inventory budget is the decision of the company's management.

We model an inventory system under a periodic-review, order-up-to inventory policy with a common review period of one time unit. Our focus is on-hand inventory, which is affected by the demand patterns and the target service levels in a time period instead of replenishment lead times. Note that the "textbook" formula for the order-up-to inventory level, calculated as the sum of the expected demand and the safety stock over the lead time and review period, includes both pipeline and on-hand inventory. A company never needs to hold the amount of the stock beyond its review cycle in the warehouse. When inventory is reviewed and replenished every time unit, the inventory level on-hand is the sum of the regular stock and the safety stock of a period, regardless of lead times.

The safety stock levels of individual SKUs in an order are a function of the item CSL. Instead of using a piecewise linear approximation to address the nonlinear relationships between CSL, safety stock level, and order fulfillment, we take a similar approach adopted by Millstein et al. (2014) and Yang et al. (2017) to discretize the CSL and formulate the model as an MILP. The set of discrete CSL values can be customized for particular applications. Let J be the set of inventory classes to which an SKU can be assigned. Each inventory class j has a corresponding CSL l_j, a derived z-value z_j to determine safety stock, and a derived value e_j from the standardized loss function to calculate the expected lost sales. SKU i in order k can be assigned to at most one inventory class, but the same SKU can be assigned different classes in different orders. The goal is to maximize order fulfillment by optimizing inventory classification and CSL assignment decisions while satisfying the required profit threshold P and inventory budget ω.

MILP formulation

Table 1 summarizes the aforementioned notation as well as the decision variables.

Recall that the fill rate R_k is the fraction of total units required in an order that are filled directly from on-hand inventory. The approach to estimating R_k is similar to estimating the IFR, which examines the average number of units short over all orders (e.g., Silver et al. 1998; Chopra and Meindl 2015). The expected shortage in an order per replenishment cycle (*ESC*) is:

$$ESC = \int_{x=s}^{\infty} (x - S) f(x) dx \tag{1}$$

where S is the target amount of on-hand inventory, x is the demand in a replenishment cycle, and $f(x)$ is the density function of the demand distribution.

Table 1: Notation

N: set of SKUs in the inventory
K: set of customer order types
N^k: set of SKUs required by customer order type k, $\forall k \in K$
J: set of inventory classes to which an SKU can be assigned
d_{ik}: expected demand of SKU i in order k, $\forall i \in N, k \in K$
σ_{ik}: standard deviation of the forecast demand of SKU i in time period k, $\forall i \in N, k \in K$
w_k: assigned weight (priority) of customer order k, $w_k > 0$, $\forall k \in K$
ρ_{ik}: unit gross profit of SKU i in order k, $\forall i \in N, k \in K$
c_i: unit cost of SKU i, $\forall i \in N$
h_i: inventory holding cost (% of the cost of SKU i), $\forall i \in N$
l_j: CSL corresponded to class j, $\forall j \in J$
z_j: standard score (z-value) associated with CSL l_j, $\forall j \in J$
e_j: derived value from standardized loss function; that is, expected number of shortages as a fraction of the standard deviation for inventory class j, $\forall j \in J$
P: minimum profit expectation
ω: available inventory capital
r_k: minimum service-level requirement for order k, $\forall k \in K$
Decision Variables
x_{ikj} = 1 if SKU i in order k is assigned to inventory class j; 0 otherwise, $\forall i \in N^k, k \in K, j \in J$
y_k = 1 if order k is completely filled from on-hand inventory (i.e., if at least 99.99% of the units required in order k are expected to be filled immediately from on-hand inventory); 0 otherwise, $\forall k \in K$
$I_i \geq 0$: average on-hand inventory of SKU i, $\forall i \in N$

When demand is normally distributed with mean d_{ik} and standard deviation σ_{ik}, and the replenishment cycle is one time period, Equation (1) can be simplified to:

$$ESC = \sigma_{ik} \left[f_s(z_j) - z_j \left(1 - F_s(z_j) \right) \right] \tag{2}$$

where f_s is the standard normal density function, and F_s is the standard normal cumulative distribution function. $f_s(z_j) - z_j(1 - F_s(z_j))$ is the standardized loss function. For every CSL l_j, there are a corresponding z-value z_j and a value e_j returned from the standardized loss function. Equation (2) can thus be rewritten as:

$$ESC = \sigma_{ik} e_j \tag{3}$$

which is the expected shortage in an order per replenishment cycle when SKU i in order k is assigned to inventory class j. The expected fill rate of an individual customer order R_k is:

$$R_k = \frac{\sum_{i \in N^k} \sum_{j \in J} x_{ikj} \left(d_{ik} - \sigma_{ik} e_j \right)}{\sum_{i \in N^k} d_{ik}} \tag{4}$$

We adopt the approach by Zinn et al. (2002), using a pooled safety factor, denoted as $v_i \in \mathbb{R}$, to determine the SKU-level safety stock. The pooled safety stock factor is the average z-value of SKU i across all the orders, weighted by the standard deviation of the demand of SKU i in order k. Specifically:

$$v_i = (\sum_{k\in K}\sum_{j\in J} x_{ikj}z_j \times \sigma_{ik}) / \sum_{k\in K}\sigma_{ik} \tag{5}$$

The value of v_i may be negative (if z_j is negative).

The expected (short-term) net profit in a time period, Π, is:

$$\Pi = \sum_{i\in N}\left[\sum_{k\in K}\sum_{j\in J} x_{ikj}\left(d_{ik} - \sigma_{ik}e_j\right)\rho_{ik} - I_i c_i h_i\right] \tag{6}$$

where the first term in the bracket calculates the expected gross profit and the second term computes the inventory holding cost.

The MILP formulation of integrated inventory classification and control for order fulfillment maximization problem can be written as:

$$\max\ \alpha\frac{1}{\sum_{k\in K} w_k}\sum_{k\in K} w_k R_k + \beta\frac{1}{|K|}\sum_{k\in K} y_k + \epsilon\Pi \tag{7}$$

subject to:

$$\sum\nolimits_{j\in J} x_{ikj} \le 1 \qquad \forall i\in N^k,\ k\in K \tag{8}$$

$$y_k \le R_k + 0.0001 \qquad \forall k\in K \tag{9}$$

$$\sum_{i\in N}\left(\sum_{k\in K}\sum_{j\in J} x_{ikj}d_{ik} + v_i\sqrt{\sum_{k\in K}\sigma_{ik}^2}\right)c_i \le \omega \tag{10}$$

$$\sum_{j\in J} x_{ikj}\left(d_{ik} + \sigma_{ik}z_j\right) \ge 0 \qquad \forall i\in N^k, k\in K \tag{11}$$

$$I_i \ge \sum_{k\in K}\sum_{j\in J} x_{ikj}d_{ik}/2 + v_i\sqrt{\sum_{k\in K}\sigma_{ik}^2} \tag{12}$$

$$I_i \ge \left(\sum_{k\in K}\sum_{j\in J} x_{ikj}d_{ik} + v_i\sqrt{\sum_{k'\in K}\sigma_{ik'}^2}\right)/2 \qquad \forall i\in N \tag{13}$$

$$R_k \ge \Upsilon_k \qquad \forall k\in K \tag{14}$$

$$\Pi \ge P \tag{15}$$

$$x_{ikj}, y_k \in \{0,1\}\forall i\in N^k, k\in K, j\in J;\quad I_i\ge 0\ \forall i\in N;\quad v_i\in\mathbb{R}\ \forall i\in N. \tag{16}$$

The objective function (7) maximizes order fulfillment performance, where parameters α and β are positive numbers reflecting the customer service strategy of a firm by allowing different emphasis on the WAFR and the OFR. The first term in the equation calculates the expected WAFR across $|K|$ customer orders and the second term computes the expected OFR. If the firm's strategy is toward adequately satisfying individual customer orders, especially the highly critical orders, α should be set significantly larger than β to focus on the WAFR. If the strategy is toward perfectly fulfilling as many orders as possible, α should be significantly smaller than β to focus on the OFR. The third term in the equation ensures that the solution is Pareto optimal; that is, the order fulfillment performance is maximized with the best profit possible. Parameter ϵ is a small number.

Constraint (8) restricts an SKU from being assigned more than one inventory class. Constraint (9) defines whether an order is expected to be completely filled (Recall that we use 99.99% for completely filled orders, as demand is a random variable so the expected fill rate for an order can never be 100%). Constraint (10) ensures that the total capital required for holding regular and safety stocks does not exceed the inventory budget; the first term $\sum_{k\in K}\sum_{j\in J} x_{ikj}d_{ik}$ calculates the regular stock, while the second term $v_i\sqrt{\sum_{k\in K}\sigma_{ik}^2}$ computes the total safety stock required for SKU i.

Constraint (11) restrains the expected inventory level to a non-negative value for any SKU in any order. A negative safety stock occurs when the CSL is less than 50%, and in the case where demand is also highly volatile with a high coefficient of variation (CV), the net inventory level may be negative, which is obviously not practical. In such case, Constraint (11) sets $\sum_{j=1}^{J} x_{ikj}$ to zero, implying no inventory is kept for the SKU in that order. Constraints (12) and (13) determine the average on-hand inventory. When the safety stock factor v_i is positive, average on-hand inventory is the sum of average cycle stock and safety stock, as described in Constraint (12). When the factor v_i is negative, inventory level is lower than cycle stock, and the average on-hand inventory is defined by Constraint (13). Constraints (12) and (13) become the same in the case of no safety stock. We remark that having a CSL less than 50% is not very common, though this may be desired for slow moving items that incur high inventory holding cost and/or have low profit margins. When the CSL is restricted to be no less than 50%, Constraints (11) and (13) can be removed.

Constraint (14) ensures the target or agreed minimum service levels for individual customers are met, and Constraint (15) guarantees that the expected net profit satisfies the minimum profit requirement. Constraint (16) sets the domains of all the decision variables. Note that the profit threshold P should not be higher than the maximum profit that can possibly be achieved given all the parameters. The maximum possible profit can be found by maximizing the expected net profit; that is:

$$max\ \Pi \tag{17}$$

subject to (5), (6), (8), (10–13), (16).

The model in (4)–(16) identifies the maximum order fulfillment performance that a firm can expect and also reveals the expected fill rate of individual customer orders through R_k. Such information is important to support the implementation of a firm's customer segmentation strategy, but has rarely been provided (if at all) in the existing literature. The proposed MILP model can be solved with commercial (or open source) MILP solvers such as CPLEX, Gurobi, among others.

COMPUTATIONAL STUDY

To illustrate how the proposed model can be implemented for improving order fulfillment performance and to gain insights into

Table 2: Summary of controlled parameters

Controlled parameters	Value/probability distribution
SKU cost	Normal (\$300, \$100)
SKU profit margin	Uniform (0.10, 0.30)
SKU mean demand	Pareto (100, 2)
Order-SKU mean demand	[Uniform (0, 1) / Sum of All Orders] X Pareto (100, 2)
Order-SKU CV of mean demand	Uniform (0.05, 0.85)
Inventory holding cost	35%

the relationships among various performance measures, as well as the impact of inventory budget and profit expectation, we develop a test problem based on a cheese company headquartered in Missouri that produces and distributes cheese products and specialty foods to retail grocers and wholesale brokers. The problem set represents a general multi-item finished goods inventory system.

Experimental design

We consider a Missouri-based cheese company that has 100 SKUs and generally receives 20 orders per time period. The total mean demand for an SKU across all the orders follows the Pareto distribution (Arnold, 2008). The number of SKUs in a customer order ranges from 21 to 35 with an average of 27. The monthly demand for most SKUs at this company falls within three standard deviations of the forecast means; to generalize our results to different settings, we consider the CV of the demand of an SKU in an order follows a uniform distribution of U (0.05, 0.85) instead. SKU unit cost is assumed to follow a normal distribution of N (\$300, \$100). The SKU profit margin follows a uniform distribution of U (0.10, 0.30) and is assumed to be the same for all orders in the experiment. Inventory holding cost is set at 35% for all SKUs. Table 2 summarizes all the controlled parameters.

In our first set of experiments, we set the weights in the WAFR to all be equal to 1, reflecting the case where orders are equally important. We use a $20 \times 6 \times 2$ experimental design to examine model behavior and provide insights into the trade-off and positive relationships between four performance measures: OFR, WAFR, IFR, and profit. The design consists of twenty levels of inventory capital, six levels of required profit, and two weighting schemes in the objective (expression (7)). Specifically, inventory budget is varied from \$10 million to \$29 million, in increments of \$1 million, where \$29 million is the maximum inventory investment required to completely satisfy all the orders from inventory. The value of \$29 million is obtained by removing the inventory budget constraint ($\omega = \infty$), limiting the service level of individual orders to be 99.99% or higher ($F_k \geq 99.99\%$), and minimizing the total inventory value:

$$\sum_{i\in N}\left(\sum_{k\in K}\sum_{j\in J} x_{ikj}d_{ik} + v_i\sqrt{\sum_{k\in K}\sigma_{ik}^2}\right)c_i. \tag{18}$$

Note that the impact on the model behavior of the other parameters, such as the average demand for a SKU and the CV, is reflected through inventory budget. At a given inventory budget, a higher mean demand and a larger CV resemble the scenario where the inventory budget is tight, while a lower mean demand and a smaller CV mimic the scenario where the inventory budget is relatively sufficient. Six levels of required profit include the minimum profit when maximizing order fulfillment measures, the maximum profit when performing profit maximization, and another four levels of profits equally distributed in between. To examine the impact of one order-based measure on the other, parameters α and β in the model objective are set at two levels: $(1,10^4)$ and $(10^4,1)$. This allows the OFR and the WAFR to be maximized to two decimal places (i.e., .01%). We set ϵ at 10^{-12} to ensure that the weight of profit (in millions of dollars) is small enough (at 0.0001% for weighted profit) and does not impact the maximization of the OFR and WAFR. To allow the model to determine the best service level for each SKU order, we define 109 discrete CSLs, with integer percentage CSLs from 1% to 99% (1%, 2%, 3%, etc.), a more granular measure in tenths from 99.1% to 99.9% (99.1%, 99.2%, etc.), and a final maximum level of 99.99% (approximating a perfect CSL). This discretization scheme provides a wide range of service levels (from 0% to 99.99%) for the model to select from, yet allows computational tractability. It is also consistent with industry practice of using integer values between 50% and 99%, with a finer measure over 99% to define service levels.

Next, we use two order-weighting schemes (as shown in Table 3) to examine the impact of order criticality (compounded with profit constraint) on order fulfillment.

The proposed model was solved with the branch-and-cut method using GAMS/CPLEX 25.1.1 on a PC with 16GB RAM and Intel Core i7-6700 CPU 3.40GHz. The solution time is affected by the inventory budget and profit requirement constraints, with minimum solution times of less than 3 CPU seconds to maximum 131 CPU seconds (2.2 minutes). Note that in the experiment, we include 109 discrete CSL levels, while in

Table 3: Order-weighting scheme

Weighting scheme	Orders																			
	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12	O13	O14	O15	O16	O17	O18	O19	O20
W1	35	20	20	5	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
W2	1	1	1	1	1	35	30	10	9	1	1	1	1	1	1	1	1	1	1	1

Figure 3: Profit maximization vs WAFR maximization with different inventory budgets.

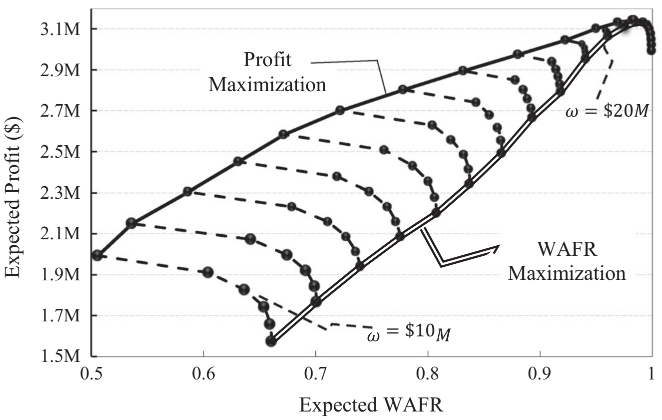
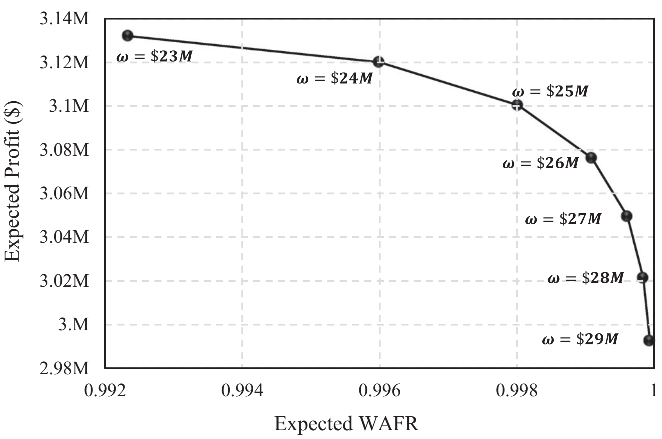


Figure 4: WAFR maximization with different inventory budgets (enlarged chart of upper-right tail in Figure 3).



practice, companies mostly desire a CSL above 50%, which significantly reduces the model size and possibly solution time. We remark that a large dataset may require hours to find the optimal solutions, but since the inventory control decisions are made periodically and generally do not require real-time processing, relatively long computation time is found acceptable to most companies.

RESULTS AND DISCUSSIONS

For each specified level of inventory budget, profit expectation, and objective function weighting, we solve the MILP in expressions (4)–(16) to obtain the expected maximum WAFR and OFR from the objective function (7). This provides 240 results (20 × 6 × 2) for the different combinations of factor levels. For each, we then calculate the corresponding IFR to examine its relationship with the two order fulfillment measures. The system-level IFR is:

$$\frac{\sum_{k\in K}\sum_{i\in N^k}\sum_{j\in J} x_{ikj}\left(d_{ik}-\sigma_{ik}e_j\right)}{\sum_{k\in K}\sum_{i\in N^k} d_{ik}}$$

Pareto optimality and impact of inventory budget

The expected maximum profit, the expected maximum WAFR, and their trade-off relationship under various levels of inventory budget are plotted in Figure 3. The bold solid line on the top describes the results of profit maximization as the inventory budget grows from \$10 million to \$29 million, while the double line shows the results of the WAFR maximization. Pareto efficiency frontiers for the expected WAFR and profit under different levels of inventory investment are approximated, as described by the dashed lines. The Pareto frontier represents a state of the firm's inventory capital being most efficiently allocated, and a situation falling below the frontier suggests the firm has the opportunity to increase profitability along with customer service level by rearranging the existing inventory capital.

Order fulfillment performance can be improved in two ways, with different implications to profit. For a company operating on the Pareto frontier, it can only improve its order fulfillment by reducing profit expectation. However, if increasing the inventory budget is an option, order fulfillment may be improved along with the increase in profit, as shown in Figure 3.

With our model, the inventory investment required to achieve the maximum profit (\$23 million of inventory in this case) is identified. Increasing the inventory budget improves both the WAFR and profit until reaching a threshold value (\$23 million), beyond which the inventory level becomes so high that the additional sales cannot compensate for the increase in inventory carrying costs, eventually causing a reduced profit, as shown in Figure 4 (an enlarged chart of the upper-right tail in Figure 3). At an inventory budget of \$29 million, all orders are expected to be filled from on-hand inventory, but the company will be 4.5% below the maximum level of expected profit.

The expected OFR and corresponding profit are plotted in Figure 5 for the 20 levels of the inventory budget. The percentages next to the line represent the expected OFR under different inventory budgets. While the overall relationships between the expected OFR, profit, and inventory budget appear to be positive, the specific impact of additional inventory investment hinges on where the firm stands today. For example, investing an additional \$1 million on the top of \$19 million inventory increases the expected OFR from 75% to 80%, but decreases

Figure 5: Maximize expected OFR with different levels of inventory budget.

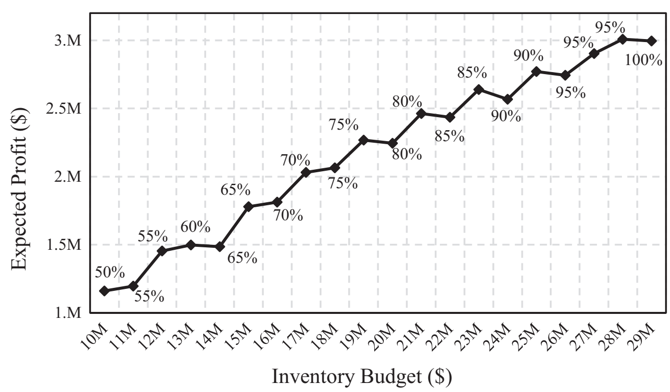


Figure 6: Expected fill rate for an order by varying profit expectation (OFR maximization).

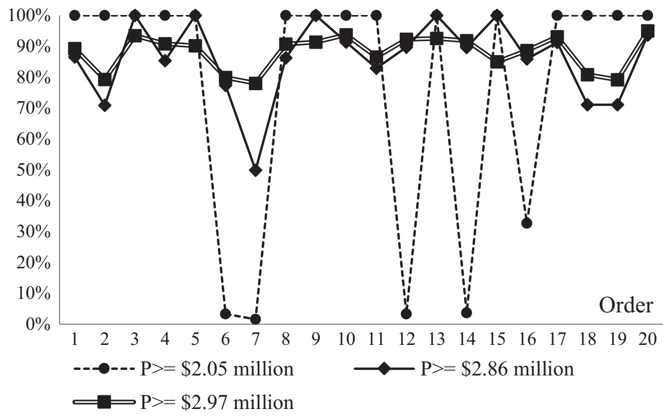
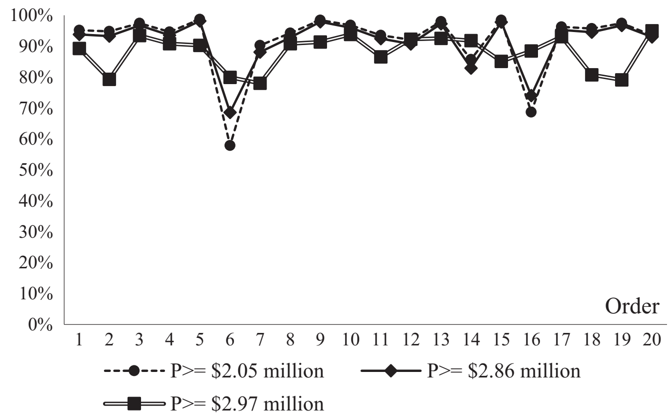


Figure 7: Expected fill rate for an order by varying profit expectation (WAFR maximization).



profit by a marginal 1%, as resources are moved away from more profitable but unlikely to be fully filled orders to maximize the overall OFR. However, if the firm can secure another $1 million to maintain a $21 million inventory, a profit increase in 8.6% can be expected. To achieve the optimal return on the additional inventory investment, it is well known that two criteria must be met: the right level of additional investment and the appropriate allocation of new resources. Both can be attained by solving the model for different levels of inventory investment.

Impact of short-term profit requirement

The expected fill rates for individual orders (R_k) as a result of the OFR and the WAFR maximization at an inventory budget of $18 million are reported in Figures 6 and 7, respectively. From Figure 6, at a minimum profit requirement of $2.05 million, 15 orders are expected to be completely filled, while 4 orders have an expected fill rate less than 4%, indicating resources being concentrated on orders most likely to be filled (e.g., because inventory requirement is relatively low). As the profit expectation increases, a more balanced order management approach has to be implemented. With a minimum profit requirement of $2.86 million, 4 orders are expected to be met in full, and the expected fill rates for the rest of the orders are 50% or higher. When the minimum profit expectation increases to $2.97 million, none of the orders are expected to be filled in complete, but the range of R_k is significantly reduced (78%–95%), showing the resources being more evenly distributed.

In comparison, with a WAFR-focused approach, the impact of profit constraint on the expected fill rates for individual orders is much less drastic, as shown in Figure 7. In fact, the profit requirement of $2.05 million is not a binding constraint for an inventory system that already operates at Pareto optimality. With a profit constraint of $2.97 million, the expected fill rate for individual orders under the WAFR maximization is largely the same as with the OFR maximization. This is due to the fact that such profit expectation is already near the maximum profit that a firm may achieve for a given inventory budget of $18 million.

Balance WAFR and IFR

Maximizing the WAFR and maximizing the OFR have different implications for profit. We illustrate this by plotting the results of profit, the WAFR, and the OFR maximizations on one chart in Figure 8. The dashed line on the top provides a benchmark of the maximum profit that can be achieved at each level of inventory budget, while the double line and the solid line demonstrate the profit consequences from the WAFR maximization and the OFR maximization, respectively. The results show that maximizing the OFR generally yields much less expected profit than the WAFR approach when an SKU's profit contribution does not vary across orders. With an inventory budget less than $22 million, a WAFR-focused process generates 20%–35% more profit than an OFR-oriented system. The two approaches converge only when the inventory capital is large enough (over $28 million in this experiment) to support perfect fulfillment for almost all orders.

Relationship with IFR

The expected WAFR and IFR as a result of the WAFR maximization under different inventory budgets are presented in Figure 9, and the expected OFR and IFR resulting from the OFR maximization are shown in Figure 10. The IFR remains high

Figure 8: Effects of OFR and WAFR maximization on profit.

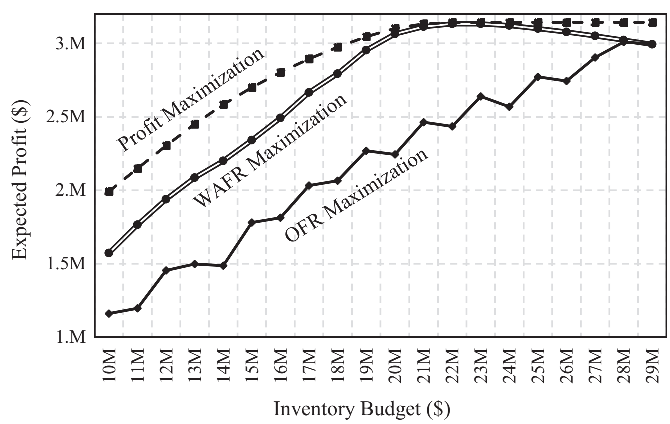


Figure 9: Expected WAFR and IFR from WAFR maximization.

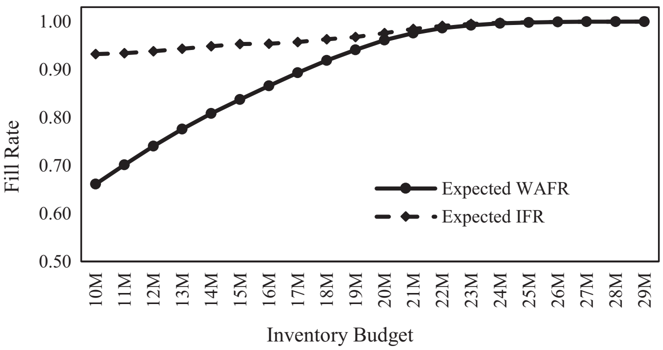
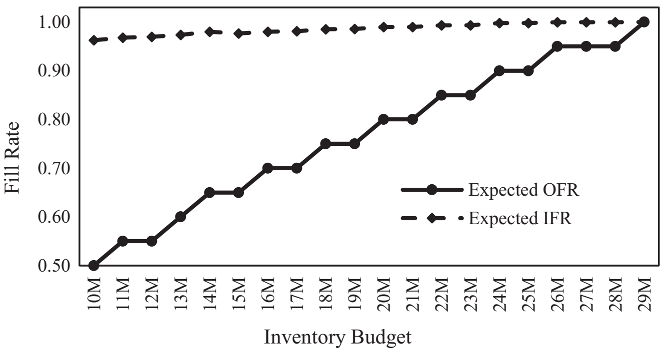


Figure 10: Expected OFR and IFR from OFR maximization.



(over 90%) in all circumstances, even when the OFR is expected to be only 50% or the WAFR only 66% (when the inventory budget is $10 million). The WAFR measures the average distance between shipment and order quantity across orders. With a tight inventory budget, the expected WAFR differs significantly from the IFR as shown in Figure 9. However, as the inventory budget increases, the gap narrows, and when the inventory investment reaches $21 million or higher, the difference is less than 1%, suggesting that the IFR may provide a good indication of the WAFR when the inventory budget is large enough. In comparison, from Figure 10, the IFR cannot adequately approximate OFR unless the inventory level reaches $29 million where a perfect OFR is expected.

Impact of order criticality

We test two order-weighting schemes at an inventory budget of $18 million, with a binding (P = $2.9 million) and a nonbinding (P = $2.0 million) profit expectation, respectively. Results are reported in Figures 11 and 12, and Table 4. When the weight is set equal for all orders, the optimal solution to achieve a high WAFR without a binding profit constraint is to assign orders #6 and #16 a relatively low expected fill rate (58% and 69%), and give the first 5 orders relatively high fill rates (above 94%). Order-weighting scheme 1 (W1) focuses on the first 5 orders, three of which have significantly higher weights. From Figure 11, the expected fill rates for orders #1 to #5 are indeed improved

Figure 11: Expected fill rate for an order by varying order weights (P = $2.0 million).

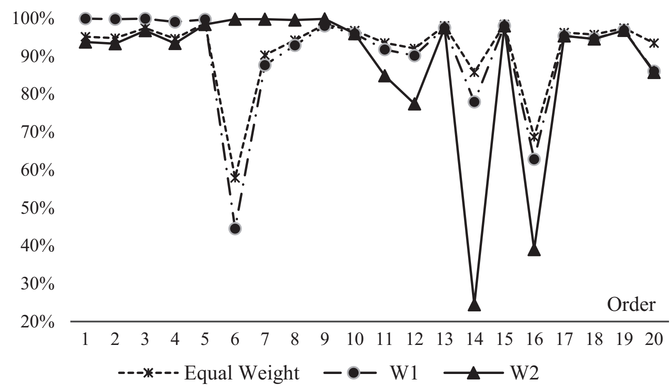
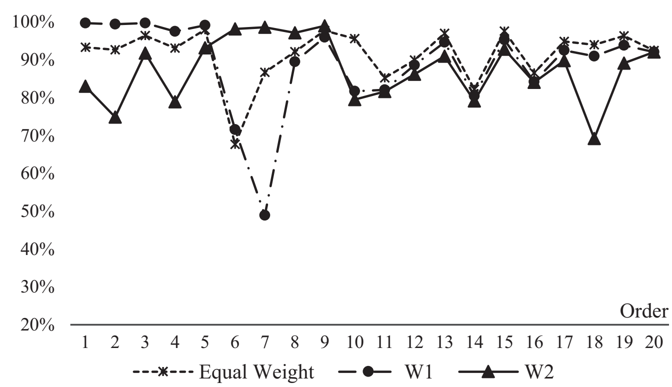


Figure 12: Expected fill rate for an order by varying order weights (P = $2.9 million).



(above 99%) by taking away a small portion of resources from each of the other orders (line W1). Weighting scheme 2 (W2) moves the focus to order #6–#9. Recall that #6 only has an expected fill rate of 58% if all orders are treated equal. With a very high priority, 99.8% of the demand in order #6 is now expected to be filled (line W2). The resource allocation structure under weighting scheme 2 obviously differs from an unweighted or scheme 1 system.

The profit constraint changes how each order is expected to be filled (Figure 12). An interesting phenomenon is that the range of R_k is narrowed across all three weighting structure (unweighted, W1, and W2; Table 4), showing that a higher profit expectation pushes resources to be more evenly distributed across orders. And, the results once again demonstrate that the WAFR-focused optimization under a limited inventory budget generally does not support concentrating resources on a small number of orders for an expected "perfect" fill rate, even if the orders are prioritized, because a "perfect" fill rate often requires a significantly higher inventory investment. However, if the expected "perfect" fill rate is desired, firms can achieve this by adjusting the value of parameters α and β, or simply specifying γ_k at 99.99%, which will guarantee the critical orders an expected "perfect" fill rate.

Table 4: Range of R_k with the WAFR maximization

	P = \$2.0 M		P = \$2.9 M	
	Min R_k	Max R_k	Min R_k	Max R_k
Equal weight	57.9%	98.7%	67.7%	98.0%
W1	44.5%	99.9%	48.9%	99.7%
W2	24.5%	99.9%	69.2%	99.0%

CONCLUSIONS

This paper addresses a customer-order-fulfillment-oriented inventory management problem. Our work contributes to the inventory literature and practice along several dimensions. First, it provides practitioners a decision-support tool that integrates inventory classification and inventory control policy decisions for maximizing order fulfillment performance, while accounting for various real-world complexities, including inventory budget, short-term profit expectation, order-varying pricing, demand variability, and order criticality. The proposed modeling framework is a novel enhancement to the current inventory planning software, which is developed upon the traditional inventory classification scheme with many critical business aspects not accounted for, and in which practitioners can only balance trade-offs through trial-and-error. Our model assesses whether the current inventory performance is Pareto optimal, quantifies the trade-off between customer service level and short-term profit, identifies the right inventory level according to the company's strategic goals, and evaluates the impacts of critical orders and customers.

Second, we adopt an alternative measure, the WAFR, to gauge the fulfillment performance of individual customer orders in a multi-item finished goods inventory system. This complements the traditional OFR measure by offering insight into the depth of demand satisfied in a customer order and accounting for order criticality. The combination of the WAFR and OFR gives companies a more complete picture of customer service performance and a balanced approach to inventory management.

Moreover, this research sheds new light on the trade-off and positive relationships between the OFR, WAFR, IFR, and short-term profit, and the implications of inventory budget and profit requirement. Our results show that the IFR may adequately indicate the WAFR when the inventory budget is sufficiently large, but may not effectively describe the OFR unless the IFR is 99.9% or higher. Different from the single-item inventory system where the average customer-order fill rate always exceeds the IFR and OFR (Larsen and Thorstenson 2014), in a multi-item finished goods inventory system, the IFR can be the highest among the three service measures even though the inventory is optimized for order fulfillment.

Results show that with a limited inventory budget, a strategy focusing on maximizing the OFR is less profitable than focusing on the WAFR. Profit constraint affects the expected fill rate of individual orders. A higher profit expectation tends to push resources more evenly distributed across orders, while distinct service levels are observed at a lower profit requirement when the OFR is measured. We also demonstrate that the company has the opportunity to improve customer service level along with profitability with additional inventory investment, but must be aware of the threshold value where the maximum expected profit is reached.

This study, while providing a practical solution to customer-focused inventory management, has a number of limitations that may be addressed in future research. First, similar to many inventory models, the performance of the proposed model relies on the quality of demand forecasting, especially order-level forecasting. Many manufacturers and distributors today still use shipment data for demand forecasting, which can never truly capture the real demand unless the ordered quantity is 100% shipped. Research addressing the issue of effectively tracking customer orders and forecasting at order-SKU level would represent a significant contribution to both the marketing and the inventory literature. Second, further studies on the relationship between the item fill rate and the average customer-order fill rate based on different demand processes with the consideration of product returns would provide further insights and comparison with the results of this work. Third, customer orders might be correlated due to the macroeconomic factors; hence, extending the model to include coupling factors that reflect order correlations would be useful. Fourth, our approach considers the short-term profit resulting from fulfilling a set of orders based on random demands. Longer-term profit issues such as from changes in demand patterns driven by changing service levels would be an interesting future research. In addition, the impact of the system condition, such as order structure, on the relationships between WAFR, OFR, and IFR would deserve further investigation. Moreover, there might be occasions where SKUs offered by the company are substitutable with each other, and thus, extending the model to take into account competing or substitutable relationships among SKUs would be of value. Finally, our work is based on a periodic-review, order-up-to inventory system in which all SKUs are replenished in the same period. In a perpetual review system, however, SKUs may be replenished across time periods, the effect of which on order fulfillment is unclear. Further, firms with a perpetual review system would benefit from the joint optimization of the reorder point and the order quantity. A comparison of the results of the two inventory review system would be a valuable contribution.

REFERENCES

Anupindi, R., and Tayur, S. 1998. "Managing Stochastic Multiproduct Systems: Model, Measures, and Analysis." *Operations Research* 46(3):S98–S111.

Arnold, B.2008. "Pareto and Generalized Pareto Distributions." In *Modeling Income Distributions and Lorenz Curves*, edited by D. Chotikapanich, 119–145. New York, NY: Springer.

Ballou, R.H. 2005. "Expressing Inventory Control Policy in the Turnover Curve." *Journal of Business Logistics* 26(2):143–164.

Ballou, R.H., and Burnetas, A. 2003. "Planning Multiple Location Inventories." *Journal of Business Logistics* 24 (2):65–90.

Bowersox, D.J., Closs, D.J., and Cooper, M.B. 2012. *Supply Chain Logistics Management*. 4th ed. New York, NY: McGraw-Hill.

Chopra, S., and Meindl, P. 2015. *Supply chain Management: Strategy, Planning, and Operation*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Closs, D.J., Nyaga, G.N., and Voss, M.D. 2010. "The Differential Impact of Product Complexity, Inventory Level, and Configuration Capacity on Unit and Order Fill Rate Performance." *Journal of Operations Management* 28(1):47–57.

Daugherty, P.J., Stank, T.P., and Ellinger, A.E. 1998. "Leveraging Logistics/Distribution Capabilities: The Effect of Logistics Service on Market Share." *Journal of Business Logistics* 19(2):35–51.

Davis-Sramek, B., Germain, R., and Stank, T.P. 2010. "The Impact of Order Fulfillment Service on Retailer Merchandising Decisions in the Consumer Durables Industry." *Journal of Business Logistics* 31(2):215–230.

Disney, S.M., Gaalman, G.J.C., Hedenstierna, C.P.T., and Hosoda, T. 2015. "Fill Rate in a Periodic Review Order-Up-To Policy under Auto-correlated Normally Distributed, Possibly Negative, Demand." *International Journal of Production Economics* 170:501–512.

Griffis, S.E., Rao, S., Goldsby, T.J., Voorhees, C.M., and Iyengar, D. 2012. "Linking Order Fulfillment Performance to Referrals in Online Retailing: An Empirical Analysis." *Journal of Business Logistics* 33(4):279–294.

Hadley, G., and Whitin, T. 1963. *Analysis of Inventory Systems*. Englewood Cliffs, NJ: Prentice-Hall.

Hausman, W.H., Lee, H.L., and Zhang, A.X. 1998. "Joint Demand Fulfillment Probability in a Multi-item Inventory System with Independent Order-Up-To Policies." *European Journal of Operational Research* 109(3):646–659.

Hoen, K.M., Güllü, R., Van Houtum, G.J., and Vliegen, I.M. 2011. "A Simple and Accurate Approximation for the Order Fill Rates in Lost-Sales Assemble-To-Order Systems." *International Journal of Production Economics* 133(1):95–104.

Johnson, M.E., Lee, H.L., Davis, T., and Hall, R. 1995. "Expressions for Item Fill Rates in Periodic Inventory Systems." *Naval Research Logistics* 42:57–80.

Lajili, I., Babai, M.Z., and Ladhari, T.2012. "Inventory Performance of Multi-Criteria Classification Methods: An empirical Investigation." Proceedings of 9th international conference of modeling, optimization and simulation–MOSIM'12.

Larsen, C., and Thorstenson, A. 2014. "The Order and Volume Fill Rates in Inventory Control Systems." *International Journal of Production Economics* 147:13–19.

Lu, Y., Song, J.S., and Yao, D.D. 2003. "Order Fill Rate, Leadtime Variability, and Advance Demand Information in an Assemble-To-Order System." *Operations Research* 51 (2):292–308.

Millstein, M.A., Yang, L., and Li, H. 2014. "Optimizing ABC Inventory Grouping Decisions." *International Journal of Production Economics* 148:71–80.

Mohammaditabar, D., Ghodsypour, S.H., and O'Brien, C. 2012. "Inventory Control System Design by Integrating Inventory Classification and Policy Selection." *International Journal of Production Economics* 140(2):655–659.

Neale, J.J., and Willems, S.P. 2009. "Managing Inventory in Supply Chains with Nonstationary Demand." *Interfaces* 39 (5):388–399.

Porras, E., and Dekker, R. 2008. "An Inventory Control System for Spare Parts at a Refinery: An Empirical Comparison of Different Re-Order Point Methods." *European Journal of Operational Research* 184(1):101–132.

Rao, S., Griffis, S.E., and Goldsby, T.J. 2011. "Failure to Deliver? Linking Online Order Fulfillment Glitches with Future Purchase Behavior." *Journal of Operations Management* 29(7):692–703.

Silver, E.A., and Bischak, D.P. 2011. "The Exact Fill Rate in a Periodic Review Base Stock System under Normally Distributed Demand." *Omega: The International Journal of Management Science* 39:346–349.

Silver, E.A., Pyke, D.F., and Peterson, R. 1998. *Inventory Management and Production Planning and Scheduling*. 3rd ed. New York, NY: John Wiley & Sons Inc.

Song, J.S. 1998. "On the Order Fill Rate in a Multi-Item, Base-Stock Inventory System." *Operations Research* 46(6):831–845.

Song, J.S., and Yao, D.D. 2002. "Performance Analysis and Optimization of Assemble-To-Order Systems with Random Lead Times." *Operations Research* 50(5):889–903.

Stanford, R.E., and Martin, W. 2007. "Towards a Normative Model for Inventory Cost Management in a Generalized ABC Classification System." *Journal of the Operational Research Society* 58(7):922–928.

Stank, T.P., Goldsby, T.J., Vickery, S.K., and Savitskie, K. 2003. "Logistics Service Performance: Estimating its Influence on Market Share." *Journal of Business Logistics* 24 (1):27–55.

Teunter, R.H., Babai, M.Z., and Syntetos, A.A. 2010. "ABC Classification: Service Levels and Inventory Costs." *Production and Operations Management* 19(3):343–352.

Teunter, R.H., Syntetos, A.A., and Babai, M.Z. 2017. "Stock Keeping Unit Fill Rate Specification." *European Journal of Operational Research* 259:917–925.

Thomas, D.J. 2005. "Measuring Item Fill-Rate Performance in a Finite Horizon." *Manufacturing & Service Operations Management* 7(1):74–81.

van Kampen, T.J., Akkerman, R., and van Donk, D.P. 2012. "SKU Classification: A Literature Review and Conceptual Framework." *International Journal of Operations & Production Management* 32(7):850–876.

Viswanathan, S., and Bhatnagar, R. 2005. "The Application of ABC Analysis in Production and Logistics: An Explanation for the Apparent Contradiction." *International Journal of Services and Operations Management* 1(3):257–267.

Yang, L., Li, H., Campbell, J.F., and Sweeney, D.C. 2017. "Integrated Multi-Period Dynamic Inventory Classification and Control." *International Journal of Production Economics* 189:86–96.

Zinn, W., Mentzer, J.T., and Croxton, K.L. 2002. "Customer-Based Measures of Inventory Availability." *Journal of Business Logistics* 23(2):19–43.

APPENDIX

EXTENSION TO AN IFR-FOCUSED INVENTORY PROBLEM

In this section, we present an MILP model that optimizes inventory control policies for improving the IFR for critical items. The IFR for an SKU i is:

$$IFR_i = \frac{\sum_{j \in J} x_{ij}\left(d_i - \sigma_i e_j\right)}{d_i} \tag{19}$$

where d_i and σ_i are the expected demand and standard deviation of SKU $i \in N$, and x_{ij} is a binary decision variable describing whether SKU i is assigned to inventory class j.

The objective function of maximizing the demand-weighted IFR is:

$$\max \frac{\sum_{i \in N} d_i IFR_i}{\sum_{i \in N} d_i} \tag{20}$$

subject to:

$$\sum_{j \in J} x_{ij} \leq 1 \qquad \forall i \in N \tag{21}$$

$$\sum_{i \in N} \sum_{j \in J} x_{ij}(d_i + \sigma_i z_j) c_i \leq \omega \tag{22}$$

$$\sum_{j \in J} x_{ij}\left(d_i + \sigma_i z_j\right) \geq 0 \qquad \forall i \in N \tag{23}$$

$$I_i \geq \sum_{j \in J} x_{ij}(d_i/2 + z_j \sigma_i) \qquad \forall i \in N \tag{24}$$

$$I_i \geq \Sigma_{j \in J} x_{ij}(d_i + z_j \sigma_{ii})/2 \qquad \forall i \in N \tag{25}$$

$$IFR_i \geq \Upsilon_i \qquad \forall i \in I \tag{26}$$

$$\sum_{i \in N} \left[\sum_{j \in J} x_{ij}\left(d_i - \sigma_i e_j\right)\pi_i - I_i c_i h_i \right] \geq p \tag{27}$$

$$x_{ij} \in \{0,\ 1\}\ \forall i \in N, j \in J;\ I_i \geq 0\ \forall i \in N. \tag{28}$$

Constraints (21)–(28) are equivalent to aforementioned Constraints (8), and (10)–(16), after pooling order-level demand and variability for an SKU. The IFR maximization focuses on SKU-level fill rates and does not guarantee order-level performance. Moreover, using the IFR to estimate the order fulfillment performance is challenging; the existing research (e.g., Song 1998; Hausman et al. 1998; Hoen et al. 2011) only provides bounds on the order fill rate, the effectiveness of which varies by the order structure. We solved the model in (20)–(28) for the same problem set. As expected, the IFR-oriented inventory system generates the highest expected profit among three service measures since inventory capital can now focus on the SKUs with high profit contribution. Practitioners shall implement the IFR-maximization and order-fulfillment-maximization models individually according to the respective needs.

SHORT BIOGRAPHIES

Liu Yang is Assistant Professor of Supply Chain Management in the College of Business Administration at Sam Houston State University. Dr. Yang focuses her effort on use-inspired, applied research motivated by practical problems in industry where optimal decisions and innovative approaches are sought. Her current research includes inventory management, healthcare supply chain, and distribution network design. She was the principal investigator of a number of industry-funded research projects.

Haitao Li is Associate Professor in the Supply Chain & Analytics Department, College of Business Administration at University of Missouri—St Louis. Dr. Li's research interests focus on optimization modeling and algorithm design with applications in scheduling, workforce optimization, and supply chain configuration. His funded projects include those sponsored by U.S. Army Research Office, Department of Transportation, HP Labs, and APICS, among others.

James F. Campbell is Professor of Supply Chain & Analytics in the College of Business Administration at the University of Missouri—St. Louis. Dr. Campbell's research centers around modeling and optimization of transportation, logistics, and supply chain systems, including applied problems in drone delivery, hub location and network design, and continuous approximation modeling. He is Associate Editor of *Transportation Research Part B: Methodological* and *Transportation Science.*