Production, Manufacturing and Logistics

Quality and pricing decisions in production/inventory systems

Hamed Jalali [a,*], Raïsa Carmen [b], Inneke Van Nieuwenhuyse [b,c], Robert Boute [b,d]

[a] *NEOMA Business School, 1 Rue du Maréchal Juin, Mont-Saint-Aignan 76825, France*
[b] *Research Center for Operations Management, Department of Decision Sciences and Information Management, KU Leuven, Leuven 3000, Belgium*
[c] *Research Group Logistics, Hasselt University, Diepenbeek 3590, Belgium*
[d] *Technology and Operations Management, Vlerick Business School, Leuven, 3000, Belgium*

ARTICLE INFO

ABSTRACT

In this article, we consider the impact of finite production capacity on the optimal quality and pricing decisions of a make-to-stock manufacturer. Products are differentiated along a quality index; depending on the price and quality levels of the products offered, customers decide to either buy a given product, or not to buy at all. We show that, assuming fixed exogenous lead times and normally distributed product demands, the optimal solution has a simple structure (this is referred to as the *load-independent* system). Using numerical experiments, we show that with limited production capacity (which implies *load-dependent* lead times) the manufacturer may have an incentive to limit the quality offered to customers, and to decrease market coverage, especially in settings where higher product quality leads to higher congestion in production. Our findings reveal that the simple solution assuming load-independent lead times is suboptimal, resulting in a profit loss; yet, this profit loss can be mitigated by constraining the system utilization when deciding on quality and price levels. Our results highlight the importance of the relationship between marketing decisions and load-dependent production lead times.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

With increased competition and globalization of businesses, aligning marketing and production decisions have become essential to the profitability for manufacturing firms. Indeed, the marketing plan specifies among others the quality and price of the offered products, and the production facility in turn needs to fulfill the market demands resulting from the product offer. Consequently, taking production limits and constraints into account is essential to avoid an excellent marketing plan turning into a failure (Tang, 2010). In practice, however, the production limitations are often overlooked. For instance, Dr. Karl Kempf, Director of Decision Technologies at Intel reports that pricing decisions are very often taken without considering production capabilities (Pekgün, Griffin, & Keskinocak, 2008).

In this paper, we study the impact of finite production capacity on the optimal quality and pricing decisions of a make-to-stock manufacturer. We take the perspective of a centralized decision maker, who aims to maximize the expected profit (revenue from sales, minus material and inventory-related costs). The make-to-stock perspective is under-studied in the literature on quality and pricing decisions, see also Section 2. Yet, it adds another dimension to the production-marketing optimization problem, as the impact of the production lead times shows up in the inventory-related costs.

In a make-to-stock system, inventories of finished products are reviewed periodically, and replenishment orders are sent to the production facility. Similar to e.g., Dobson and Yano (2002) and Jayaswal, Jewkes, and Ray (2011), we assume that the different products *share* the same manufacturing facility. As price and quality decisions affect product demands, they naturally determine the load on the production system. This load, in turn, impacts replenishment lead times through the congestion effect in production. As a result, we have an integrated production/inventory (P/I) system (Benjaafar, Cooper, & Kim, 2005): replenishment lead times are endogenously generated by the finite-capacity production system and, consequently, are *load-dependent*. Yet, this production-marketing interaction is largely ignored in the current literature (see, e.g., Upasani & Uzsoy, 2008 and the literature review in Section 2).

The fact that different product types share the same operation system is not uncommon: particularly in settings with high fixed capacity investment costs, strong scale economies promote resource sharing as the bundling of product volumes leads to lower fixed cost per unit produced (Van Mieghem & Allon, 2015).

* Corresponding author.
  E-mail addresses: hamed.jalali@neoma-bs.fr (H. Jalali), raisa.carmen@kuleuven.be (R. Carmen), inneke.vannieuwenhuyse@kuleuven.be (I. Van Nieuwenhuyse), robert.boute@vlerick.com (R. Boute).

Examples of resource sharing can be found both in the high-tech and low-tech industries. Consider for instance Barco, a world leader in the development and production of high-tech screens. The company assembles different screen variants on the same finite-capacity line and has to decide on the quality of the offered screens (which is linked to product attributes such as number of megapixels and color intensity), their prices, and their stocking quantities (Boute, Van den Broeke, & Deneire, 2018). Another example is the bag manufacturer Timbuk2, which shares its San Francisco-based production capacity between both standard make-to-stock items and customized make-to-order items, despite the differences in variety required (Cattani, Dahan, & Schmidt, 2010). Finally, Dobson and Yano (2002) discuss a wheel supplier that dedicates each production line to a set of wheel types. Evidently, the range of product types that can be produced on a given resource depends on the resource's flexibility; our setting thus presumes that the production system *is* flexible. We make abstraction, though, of the actual configuration of the system. In reality, multiple types of configurations exist that offer flexibility (e.g., manufacturing cells, job shops, flexible machining centers, flexible assembly lines); as our goal is to study the impact of finite production capacity, we do not model the internal operation of the system in detail, but conceptualize it as a single-server capacitated resource. To gain maximal insight in the dynamics of this system, we limit the assortment to two product types and fix the quality of one of these. In calculating the expected profit, we assume that material and inventory-related costs per unit increase with the quality level of the product; the extent of the increase is case dependent. Analogously, the average and variance of the product processing time increase with product quality, reflecting the fact that higher quality products are often more complex to produce.

We study the following cases:

1. *load-independent optimization*: in this case, the decision maker simply ignores the impact of the finite capacity, and decides on prices and qualities assuming a fixed exogenous lead time. As discussed in Section 2, this assumption remains common in most of the production planning literature to date, and it might be reasonable in settings with abundant capacity (Gupta & Benjaafar, 2004). Even if capacity is limited, the complex relation between quality and price levels and production lead times is often overlooked in practice, and exogenous lead times are assumed.
2. *load-dependent optimization*: here, we determine optimal quality and price levels for the case in which the relationship between quality and price levels and production lead times can be exactly evaluated. Optimizing this system yields the maximum possible profit, given the current capacity of the system. Load-dependent optimization, however, can be difficult in practice given the complex relation between production congestion and demand.
3. *utilization-based optimization*: in this approach, the decision maker uses the expected processing times to constrain the utilization of the production system, and optimizes price and quality levels with this utilization constraint. As utilization-based optimization only requires information on the expected processing times as a function of quality levels, it might be more practical than load-dependent optimization in real-life settings.

We summarize our main contributions as follows:

1. We solve the load-independent optimization setting analytically, and show that the optimal solution follows a simple structure for normally distributed product demands. This yields several novel managerial insights for settings with

truly exogenous lead times (e.g., when inventory and production systems belong to separate firms or when capacity is abundant, Benjaafar et al., 2005; Gupta & Benjaafar, 2004).
2. We provide a Markov-based approach to analyze the load-dependent optimization of quality and pricing decisions. Using numerical experiments, we show that under load-dependent optimization, it is optimal for the manufacturer to limit the quality offered to customers, and to opt for a lower market coverage, especially in settings where higher product quality leads to higher system congestion. The simple solution assuming load-independent lead times is then clearly suboptimal, resulting in a profit loss. We demonstrate that this profit loss decreases when quality costs increase, and customers are more price sensitive and/or less quality sensitive.
3. We numerically show that utilization-based optimization can mitigate the profit loss incurred with load-independent optimization. Especially in settings where the dynamic lead time behavior cannot be evaluated (i.e., load-dependent optimization is difficult), this approach provides a pragmatic solution that outperforms load-independent optimization.

Our work is related to the literature on pricing and assortment optimization, and the research on production planning and scheduling with load-dependent lead times. Section 2 provides a brief review, highlighting the novelty of our work. Section 3 provides the details of our (stylized) model, and Section 4 discusses the analytical solution of the load-independent optimization approach. In Section 5, we consider the production/inventory system as a queueing system and study the load-dependent optimization. Numerical results are provided in Section 6, where we also discuss utilization-based optimization. We conclude in Section 7.

2. Literature review

Our quality and pricing problem relates to the pricing and assortment planning literature. Joint pricing and assortment planning, however, has not been studied in depth (Kök, Fisher, & Vaidyanathan, 2015; Shin, Park, Lee, & Benton, 2015). Many articles take a *retailer*'s perspective, and optimize prices and inventory levels for a given assortment. This is referred to as the joint inventory-pricing problem: see, e.g., the work by Federgruen and Heching (1999), Song and Xue (2007), Li and Huh (2011), and Yang and Zhang (2014) (for a recent review, see Chen & Simchi-Levi, 2012). Prices can be either *static*, meaning that they are kept unchanged over the planning horizon, or *dynamic*, so they can change from period to period (Zhu & Thonemann, 2009). Some papers also optimize the assortment carried by the retailer (see Kök et al., 2015 for a recent review). These articles mostly focus on *horizontally* differentiated products (i.e., items with equivalent quality levels, Pan & Honhon, 2012) and consider static pricing (see, e.g., Kök & Xu, 2011; Maddah & Bish, 2007, and Alptekinoğlu & Semple, 2016). Pan and Honhon (2012), on the other hand, consider products with different quality levels.

The optimal policies in all of the above articles are obtained under the assumption of negligible replenishment lead times. As noted by Yang and Zhang (2014), positive lead times render the joint inventory-pricing problem extremely difficult, especially in periodic review systems with dynamic pricing (see Pang, Chen, & Feng, 2012 for a first attempt to partially characterize the structure of the optimal solution to this problem, and Bernstein, Li, & Shang, 2015 for an effective heuristic).

The articles that are most closely related to our work are those that consider quality and pricing decisions in a *manufacturer* setting: unfortunately, many of these articles (see, for instance,

Netessine & Taylor, 2007; Tang & Yin, 2010, and Rong, Chen, & Shen, 2015) do not consider load-dependent lead times (Upasani & Uzsoy, 2008). Those that do, mostly consider *make-to-order* settings, and focus mainly on optimizing prices and lead time quotes for a single product (see, for instance, Jayaswal et al., 2011; Palaka, Erlebacher, & Kropp, 1998; Pekgün et al., 2008; Ray & Jewkes, 2004,; Hafızoğlu, Gel, & Keskinocak, 2016). An exception is Chayet, Kouvelis, and Yu (2011), who consider two products and also optimize product quality levels. For recent reviews of this literature, we refer to Tang (2010) and Upasani and Uzsoy (2008).

To the best of our knowledge, our work is the first to consider pricing and quality decisions in *make-to-stock* settings with *load-dependent* lead times. Even in the area of production planning and scheduling, the body of work that considers load-dependent lead times remains relatively scarce. Usually, the common assumption of fixed and constant planned lead times at the aggregate level of production planning still prevails; yet, it is precisely this assumption that is blamed for a variety of problems occurring in current planning systems, such as MRP and MRPII (see, e.g., Hopp & Spearman, 2001; Selcuk, Fransoo, & De Kok, 2006; Van Nieuwenhuyse, De Boeck, Lambrecht, & Vandaele, 2011) As noted in the review by Pahl, Voß, and Woodruff (2007), "*although there is a large body of literature concerning queuing models for the analysis of the relationship between capacity utilization and lead times, and another body of work on control and order release policies that take lead times into consideration, there have been relatively few aggregate planning models that recognize the (nonlinear) relationship between the planned utilization of capacity and lead times*". The work so far mainly focuses on approaches to either (partially) avoid load-dependent lead times through job release policies that directly control work-in-process (such as KANBAN, CONWIP or POLCA; see, e.g., Krishnamurthy & Suri, 2009; Spearman, Hopp, & Woodruff, 1989; Spearman, Woodruff, & Hopp, 1990) or directly control workload (Land & Gaalman, 1996; Vandaele, Van Nieuwenhuyse, Claerhout, & Cremmery, 2008), or on attempts to integrate load-dependent lead times in aggregate production planning using queueing models (e.g., Lambrecht, Ivens, & Vandaele, 1998; Zijm & Buitenhek, 1996) or clearing functions (e.g., Asmundsson, Rardin, & Uzsoy, 2006). At this aggregate level, production lot sizing has been put forward as a means to minimize the weighted average lead time through production (see, e.g., Lambrecht et al., 1998; Van Nieuwenhuyse et al., 2011, and Vandaele N., 2000), in particular in settings with parallel and/or sequential process batching.

Only few articles so far study integrated production/inventory systems with load-dependent lead times. Alptekinoğlu and Corbett (2010) consider the possibility of make-to-stock production, in addition to make-to-order. They study a setting where all products have equal costs and production times (products are thus horizontally differentiated), and focus on the decision which products to produce to stock, versus to order. Wong and Naim (2011) study a similar setting, focusing on the benefits of postponement. Van Nyen, Bertrand, and Van Ooijen (2009) study cyclical production planning in a production/inventory system with job shop routing. Van Nieuwenhuyse, Vandaele, Rajaram, and Karmarkar (2007) and Van Nieuwenhuyse, Mahihenni, and Baudelocq (2014) study semi-process industry settings using a production/inventory model, in view of optimizing allocation and campaign sizing policies, and/or quantifying trade-offs between capacity, inventory and customer service. Noblesse, Boute, Lambrecht, and Van Houdt (2014) focus on optimizing the inventory parameters in a continuous review (s,S) inventory system. None of these articles considers the impact of product prices and attributes (such as quality) on system behavior.

3. Model formulation

3.1. Objective function

We consider a make-to-stock manufacturer who offers two products $i \in \{1, 2\}$ with qualities f_i. We assume f_1 to be fixed and we optimize f_2 within a discrete and finite set of potential values. Our methodology can technically handle a variable f_1, in both load-independent and load-dependent settings at the expense of extra computation time; this, however, does not yield additional insights. In each period, the firm earns revenue by offering the products with qualities f_i at prices p_i (see Section 3.2 for further details on the customer choice model). Inventories are periodically reviewed according to an order-up-to policy (see further details in Section 3.3).

The decision maker seeks to determine the optimal price levels of both products (p_i^*), the quality of the second product (f_2^*), and the base-stock inventory levels (S_i^*) to maximize the steady-state expected profit per period:

$$\Pi = \sum_{i=1}^{2} \left[(p_i - c_i)E(D_i) - h_i E((NS_i)^+) - b_i E((NS_i)^-) \right], \quad (1)$$

where D_i is the random variable denoting the demand of product i per period, and NS_i is a random variable referring to the steady-state net stock (on hand inventory minus backorder) of product i at the end of an arbitrary period (with $(X)^+ = \max(0, X)$ and $(X)^- = \max(0, -X)$).

The first term in Eq. (1) represents the expected gross profit per period: we assume that the unit material cost is quadratically related to the quality of the product: $c_i = m f_i^2$, where $m > 0$ is a constant. Such a relationship is commonly assumed in the literature (e.g., Bish & Chen, 2016; Chayet et al., 2011; Heese & Swaminathan, 2006; Kwark, Chen, & Raghunathan, 2017; Netessine & Taylor, 2007; Örsdemir, Kemahlıoğlu-Ziya, & Parlaktürk, 2014), and reflects the fact that it is increasingly more expensive to improve product quality; yet, this impact is mitigated for lower values of m (Bish & Chen, 2016).

The last two terms in Eq. (1) reflect the steady-state expected holding and backorder costs per period. The unit holding and backorder costs are assumed to be proportional to the material cost for each product i: $b_i = m_b f_i^2$ and $h_i = m_h f_i^2$, with $m_b, m_h > 0$. Equivalently, Eq. (1) can be written as:

$$\Pi = \sum_{i=1}^{2} \left[(p_i - c_i)E(D_i) - h_i E((S_i - IO_i)^+) - b_i E((IO_i - S_i)^+) \right], \quad (2)$$

where IO_i is the steady state distribution of inventory on-order of product i at the end of an arbitrary period.

3.2. Customer choice process

Depending on the price and quality levels, customers may decide to buy one of the products or may choose not to purchase at all. We use the multinomial logit (MNL) model to reflect this choice process. As noted by Alptekinoğlu and Semple (2016), this model is popular in the price and assortment optimization literature (e.g., Du, Cooper, & Wang, 2016; Sibdari & Pyke, 2010; Topaloglu, 2013; Van Ryzin & Mahajan, 1999), and has also been commonly used in practice (e.g., Kök & Fisher, 2007 use it to explain customer choice at the Dutch supermarket chain Albert Heijn, Vulcano, van Ryzin, & Chaar, 2010 successfully apply it in the airline industry, and Rusmevichientong, Shen, & Shmoys, 2010 use it
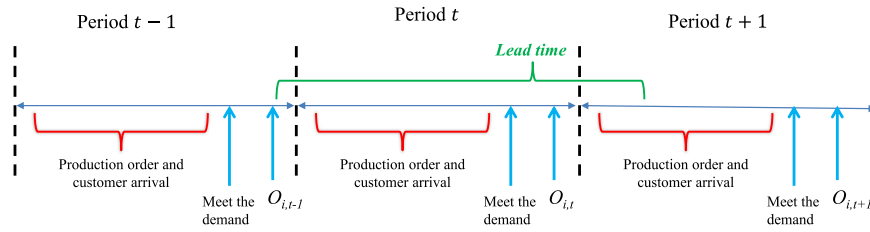
Fig. 1. Sequence of events in the periodic review system.

to explain DVD sales at a large online retailer). In this model, each customer opts for the choice that maximizes her utility. The *expected* utility of product i is denoted by μ_i: in our case $\mu_i = \varepsilon_f f_i + \varepsilon_p p_i$, where $\varepsilon_f > 0$ and $\varepsilon_p < 0$ represent the sensitivity of customers to quality and price, respectively. Increasing the price or decreasing the quality thus decreases the expected utility. We assume, without loss of generality, that the expected utility of the nonpurchase (i.e., not purchasing anything) is $\mu_0 = 0$. The MNL model captures both vertical and horizontal elements of customer choice; even if prices are equal, some customers still buy the low-quality product due to idiosyncratic product preferences which are unknown to the firm (Akcay, Natarajan, & Xu, 2010; Anderson, De Palma, & Thisse, 1992). Such a model has also been used in e.g., Dong, Kouvelis, and Tian (2009), Davis, Gallego, and Topaloglu (2013), and Du et al. (2016). The probability that a customer chooses product i (commonly referred to as the market share of product i) is then given by (Train, 2002):

$$q_i(\mathbf{p}, \mathbf{f}) = \frac{e^{\mu_i}}{1 + \sum_i e^{\mu_i}}, \tag{3}$$

and the nonpurchase probability is $q_0(\mathbf{p}, \mathbf{f}) = 1/(1 + \sum_i e^{\mu_i})$, where $\mathbf{p}$ and $\mathbf{f}$ represent the vector of prices and qualities of both products. According to this model, increasing the price or decreasing the quality of a product (while keeping everything else constant) reduces its market share, while increasing both the probability that customers buy the other product, and the nonpurchase probability.

We assume *static substitution*, meaning that if the chosen product is not available, customers do not switch to another product, but demand is backlogged. This is a reasonable assumption as we consider a high service level and stockouts are thus negligible (Kök & Fisher, 2007; Van Ryzin & Mahajan, 1999). Let the total number of customers per period be random, with mean λ and variance σ^2. The expected demand for product i in an arbitrary period then equals $E(D_i) = q_i \lambda$. Assuming multiplicative demand (Song & Xue, 2007), we have $Var(D_i) = q_i^2 \sigma^2$, which implies that the coefficient of variation of D_i is independent of q_i (and, thus, independent of price and quality levels).

3.3. Inventory control system

The manufacturer manages the inventory of finished products according to a periodic review base-stock policy with order-up-to level S_i for product i. This policy is well-studied, and has been proven optimal in settings with exogenous lead times, zero fixed ordering cost, and holding and shortage costs that are convex and proportional to the volume of on-hand inventory or shortage (Nahmias, 1997; Zipkin, 2000). In cases with endogenous (i.e., load-dependent) lead times, to the best of our knowledge, proofs about the optimal policy are not yet available.

Fig. 1 illustrates the sequence of events in our model. Customer demand observed during a period t is met at the end of the period. After meeting the demand (unfilled demands are backlogged), the inventory positions of both products are reviewed and orders of size $O_{1,t}$ and $O_{2,t}$ are placed to raise these up to S_1 and S_2,

respectively. The inventory position at the time of ordering thus equals $S_i - D_{i,t}$, where $D_{i,t}$ is the demand of item i in period t. Under i.i.d. demand, the order-up-to levels S_i are fixed and the order quantity of product i in period t equals the observed demand: $O_{i,t} = D_{i,t}$. Production orders delivered in period t are available to satisfy the demands of the same period.

In the load-independent optimization model, the replenishment lead time is assumed to be exogenous and fixed (see Section 4). This can be reasonable when the inventory and production systems belong to separate firms, or when capacity is abundant. In settings with limited capacity, however, the replenishment lead times are endogenously generated by the production system and are load-dependent (Benjaafar et al., 2005; Gupta & Benjaafar, 2004; see the discussion in Section 5). Our numerical experiment in Section 6 reveals that ignoring this load-dependent lead times can result in suboptimal price and quality decisions.

4. Load-independent optimization

We first study the setting where the decision maker assumes a fixed, exogenous lead time of L periods. To get closed form expressions, we assume normally distributed product demands per period (as is common in literature, see e.g., Gaur & Honhon, 2006; Maddah & Bish, 2007; Mayorga, Ahn, & Aydin, 2013): $D_i \sim N(q_i\lambda, q_i^2\sigma^2)$. We first characterize the optimal order-up-to levels S_i^* for arbitrary prices and quality levels (Section 4.1); this result is then used to derive optimal prices p_i^* for any arbitrary f_2 (Section 4.2), and finally to derive the optimal quality level f_2^* (Section 4.3).

4.1. Optimizing order-up-to levels for arbitrary prices and quality levels

Assuming an exogenous lead time of L periods, the distribution of IO_i in Eq. (2) equals the distribution of the demand of product i during $L + 1$ periods (Zipkin, 2000). Thus,

$$\Pi = \sum_{i=1}^{2}\left[(p_i - c_i)E(D_i) - h_i E((S_i - D_i^{L+1})^+) - b_i E((D_i^{L+1} - S_i)^+)\right], \tag{4}$$

where $D_i^{L+1} \sim N((L+1)q_i\lambda, (L+1)q_i^2\sigma^2)$. This function is concave with respect to S_i; the optimal S_i can thus be obtained by the well-known newsvendor expression:

$$\Pr(D_i^{L+1} \leq S_i^*) = \frac{b_i}{b_i + h_i} = \frac{m_b}{m_b + m_h}. \tag{5}$$

Assuming normally distributed demands, we have:

$$S_i^* = (L+1)q_i\lambda + \Phi^{-1}\left(\frac{m_b}{m_b + m_h}\right)\sqrt{L+1}\, q_i\sigma, \tag{6}$$

where $\Phi^{-1}(.)$ is the standard normal inverse CDF. Eq. (6) reveals that the optimal order-up-to level of product i increases linearly with its market share q_i. Hence, it increases nonlinearly with its quality, and decreases nonlinearly in its price, see Eq. (3).

4.2. Optimizing prices for arbitrary quality levels

Assuming normal demand, and expressing c_i, b_i, and h_i in terms of f_i, we can expand profit function (4) as follows:

$$\Pi = \sum_{i=1}^{2} \Big[(p_i - m f_i^2) q_i \lambda - m_h f_i^2 (S_i - (L+1) q_i \lambda + \sqrt{L+1}\, q_i \sigma L(z_i))$$
$$- m_b f_i^2 \sqrt{L+1}\, q_i \sigma L(z_i) \Big], \tag{7}$$

where $z_i = (S_i - (L+1) q_i \lambda)/\sqrt{L+1} q_i \sigma$, and $L(z) = \phi(z) - z(1 - \Phi(z))$ is the standard normal loss function (ϕ and Φ are standard normal density and cumulative distribution functions). Using Eq. (6), the profit function (7) simplifies to (see Appendix A):

$$\Pi(p_1, p_2, f_2) = \sum_{i=1}^{2} \Big[(p_i - m f_i^2) q_i \lambda - (m_b + m_h) f_i^2 \sqrt{L+1} q_i \sigma \phi(z_c) \Big], \tag{8}$$

with $z_c = \Phi^{-1}(\frac{m_b}{m_b + m_h})$. For given quality levels, this profit function is in general not concave with respect to prices. Following the MNL model in Eq. (3), we can write prices in terms of the market shares:

$$p_i(q_1, q_2) = \frac{1}{\varepsilon_p} \big(\ln(q_i) - \varepsilon_f f_i - \ln(1 - q_1 - q_2) \big), \tag{9}$$

such that the expected profit (Eq. (8)) can be written as:

$$\Pi(q_1, q_2) = \sum_{i=1}^{2} \Big[\frac{1}{\varepsilon_p} \ln(q_i) q_i \lambda - \frac{\varepsilon_f}{\varepsilon_p} f_i q_i \lambda$$
$$- m f_i^2 q_i \lambda - (m_b + m_h) f_i^2 \sqrt{L+1}\, q_i \sigma \phi(z_c) \Big]$$
$$- \frac{1}{\varepsilon_p} \ln(1 - q_1 - q_2)(q_1 + q_2) \lambda. \tag{10}$$

Theorem 1. *The profit function $\Pi(q_1, q_2)$ in the load-independent lead time setting is jointly concave in q_1 and q_2.*

Proof. See Appendix B. □

Following Theorem 1, we optimize prices *indirectly* by optimizing the market shares (analogous to Song & Xue, 2007 and Dong et al., 2009). From Eqs. (9) and (10), we have:

$$\frac{\partial \Pi(q_1, q_2)}{\partial q_i} = \lambda p_i(q_1, q_2) + \frac{\lambda}{\varepsilon_p}$$
$$- m f_i^2 \lambda - (m_b + m_h) f_i^2 \sqrt{L+1} \sigma \phi(z_c)$$
$$+ \frac{\lambda(q_1 + q_2)}{\varepsilon_p(1 - q_1 - q_2)}, \tag{11}$$

which yields

$$p_i(q_1^*, q_2^*) = p_i^* = m f_i^2 + \frac{(m_b + m_h) f_i^2 \sqrt{L+1}\ \sigma \phi(z_c)}{\lambda}$$
$$- \frac{q_1^* + q_2^*}{\varepsilon_p(1 - q_1^* - q_2^*)} - \frac{1}{\varepsilon_p}. \tag{12}$$

This leads to the following insights:

Theorem 2. *In the setting with load-independent lead times, the optimal price of product i is always larger than its material cost (i.e., $p_i^* > m f_i^2$). The resulting optimal profit is always positive.*

Proof. . All the terms in Eq. (12) are positive since $\varepsilon_p < 0$ and $q_1 + q_2 \leq 1$ and, therefore, the optimal prices are always positive and larger than the corresponding material costs. Replacing the prices in Eq. (8) by (12), after simplification, we obtain:

$$\Pi(p_1^*, p_2^*) = \frac{-\lambda (q_1^* + q_2^*)}{\varepsilon_p(1 - q_1^* - q_2^*)} = \frac{-\lambda}{\varepsilon_p} (e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} + e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}), \tag{13}$$
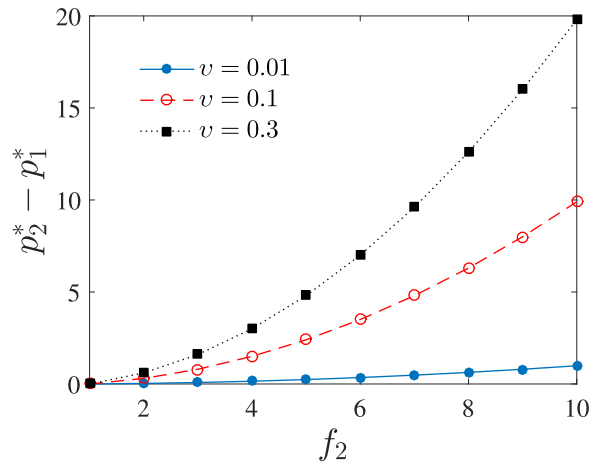

Fig. 2. Optimal price difference ($p_2^* - p_1^*$) in function of f_2 for different values of υ (with $f_1 = 1$).

which shows that the optimal profit for arbitrary quality levels is also always positive. □

Using Eqs. (13) and (12), we can write the optimal prices as follows:

$$p_i^* = \upsilon f_i^2 + \frac{\Pi(p_1^*, p_2^*)}{\lambda} - \frac{1}{\varepsilon_p}, \tag{14}$$

where

$$\upsilon = [m + (m_b + m_h) \sigma \phi(z_c) \sqrt{L+1}/\lambda] > 0. \tag{15}$$

The coefficient υ can be interpreted as a cost penalty for increasing quality, and is always strictly positive. Evidently, it increases when higher quality is more costly for the firm (higher m, m_b, and m_h). Note that the impact of the inventory-related costs is mediated by the demand coefficient of variation (σ/λ): as this gets larger, the inventory-related costs play a bigger role in υ (conversely, as σ/λ approaches zero, υ approaches the material cost m and backorder and holding costs become irrelevant). Lead time also affects the impact of inventory-related costs: the impact becomes stronger as lead time increases.

Combining (13) and (14), we obtain the following expression for the profit function:

$$\Pi(p_1^*, p_2^*) = \frac{-\lambda}{\varepsilon_p} \times \sum_{i=1}^{2} \exp \Big[\varepsilon_f f_i + \varepsilon_p \upsilon f_i^2 + \frac{\varepsilon_p \Pi(p_1^*, p_2^*)}{\lambda} - 1 \Big], \tag{16}$$

which can be solved numerically for the unknown $\Pi(p_1^*, p_2^*)$ for any arbitrary quality values. As the left-hand side is increasing in $\Pi(p_1^*, p_2^*)$ while the right-hand side is decreasing, this equation has a single solution. Once $\Pi(p_1^*, p_2^*)$ is obtained, we can find the optimal prices, for given quality levels, using Eq. (14).

As evident from Eq. (12), the difference between the optimal product prices is as follows:

$$p_2^* - p_1^* = \Big[m + \frac{(m_b + m_h) \sigma \phi(z_c) \sqrt{L+1}}{\lambda} \Big] (f_2^2 - f_1^2) = \upsilon(f_2^2 - f_1^2). \tag{17}$$

Fig. 2 illustrates this relationship, for different υ values and $f_1 = 1$. It is thus always optimal to sell a higher quality product at a higher price compared to the lower quality product; if the products have the same quality (and, thus, the same costs), their prices must be set equal. In settings where increasing the quality is more expensive (i.e., higher values of υ), the optimal price difference between

the products will be larger, which in turn reduces the demand of the high-quality product and thus lowers the total inventory-related costs. Also, as the coefficient of variation of demand decreases, the optimal price difference approaches the difference between material costs (i.e., $p_2^* - p_1^* = mf_2^2 - mf_1^2$, see Eq. (17)). In other words, in these cases, the optimal prices are characterized by "equal profit margins" ($p_2^* - c_2 = p_1^* - c_1$); a similar conclusion is obtained in Maddah and Bish (2007), among others.

Theorem 3. *The optimal price of the second product p_2^* increases with its quality level f_2; the optimal price of the first product p_1^* first increases with higher values of f_2 and then decreases, reaching a maximum at $f_2 = -\varepsilon_f/(2\upsilon\varepsilon_p)$.*

Proof. See Appendix C. □

4.3. Optimizing the quality of the second product

Now that we have derived the optimal price and order-up-to levels for both products, we optimize the quality level of the second product. From Eq. (16) we can derive the optimal quality of the second product (for a given fixed quality level f_1). This yields:

Theorem 4. *With load-independent lead times, the optimal quality of the second product is given by:*

$$f_2^* = \frac{-\varepsilon_f}{2\upsilon\varepsilon_p}. \tag{18}$$

Proof. See Appendix D. □

The optimal quality of the second product is thus independent of the quality of the first product; it is lower in environments with longer lead times, higher demand uncertainty (as measured by its coefficient of variation σ/λ), or when the cost of quality is higher (i.e., higher values of m, m_b, m_h). Conversely, f_2^* is higher as customers become more sensitive to quality (i.e., higher ε_f) and/or less sensitive to price (i.e., lower $|\varepsilon_p|$).

With load-independent lead times, the optimal solution thus has a simple structure. The optimal quality level f_2^* can be obtained from Eq. (18); we can then obtain the optimal profit by numerically solving Eq. (16). The optimal price levels can be readily calculated from Eq. (14), and Eq. (6) gives the optimal order-up-to levels.

Market coverage is defined as the portion of customers who purchase a product; it is equal to $1 - q_0$, where q_0 is the nonpurchase probability in Eq. (3) (Gaur & Honhon, 2006). For the load-independent setting, we find:

Theorem 5. *The market coverage with optimal price and quality levels is lower when customers are more sensitive to price (i.e., higher $|\varepsilon_p|$) or less sensitive to quality (i.e., lower ε_f), in environments with longer lead times, or higher demand uncertainty, and in high-cost environments, (i.e., high values of m, m_b, m_h). In these environments, the optimal profit levels will also be lower.*

Proof. See Appendix E. □

All of the conditions in Theorem 5 result in a lower f_2^* (see Eq. (18)). For instance, as price sensitivity increases, f_2^* decreases. The optimal price levels p_1^* and p_2^* also decrease (see Appendix E); yet, Theorem 5 implies that this decrease is insufficient to compensate customers for the lower quality: in the optimum, the firm will thus have lower market coverage. This confirms the observation by Li, Webster, Mason, and Kempf (2017), that market coverage and profit are often conflicting objectives. The other parameters in Theorem 5 have the same impact on f_2^*, and thus follow the same argument.

5. Load-dependent optimization

In this section, we explain how to optimize price and quality levels in the load-dependent optimization model, i.e., taking into account the impact of price and quality levels on the product demands and the resulting production load. We first introduce the key assumptions and notation (Section 5.1); next, we explain how to determine the optimal order-up-to levels in the resulting production/inventory (P/I) system (Section 5.2). Section 5.3 explains how to numerically optimize the quality levels and the corresponding product prices.

5.1. Assumptions and notation

Following the periodic review policy (Section 3.3), replenishment orders are sent to the production system with deterministic time intervals (equal to a single period or d time units). As both products have the same review period, each replenishment order consists of a number of items of *both* products; if production is busy, the order waits in queue (orders are processed FCFS without setup times). As mentioned in the introduction, we make abstraction of the actual configuration of the production system; to capture the impact of finite capacity on production lead times, we model the system as a single server. When the server becomes available, all items in the replenishment order are processed one by one. Once the whole order is finished, the items are sent to their respective inventories. For computational reasons (see Section 6.2 in Boute, Lambrecht, & Van Houdt, 2007), we model the production system as a *discrete* manufacturing system, having discrete unit production times M_i for product i. M_i follows a discrete phase-type (PH) distribution with a mean and standard deviation that increase quadratically with the product quality level: $E(M_i) = m_p f_i^2$ and $\sqrt{Var(M_i)} = CV m_p f_i^2$, with CV the coefficient of variation of the unit production times. Discrete PH distributions can approximate any non-negative discrete distribution arbitrarily closely and if the distribution is finite, the approximation is exact (Latouche & Ramaswami, 1999). The quadratic relation between the expected unit production time and quality presumes that it is increasingly more time-consuming to produce a higher quality product; yet, the impact is mitigated for lower values of m_p.

As the production system is a *discrete* manufacturing system, we also model the demand by a discrete distribution. More specifically, we use the outcomes of the customer choice model ($E(D_i)$ and $Var(D_i)$) to fit a discrete phase-type distribution to the demand of each individual product i, using the procedure in Section 6.1 of Boute et al. (2007).

5.2. Optimizing order-up-to levels for arbitrary prices and quality levels

To determine the expected profit in Eq. (2), we need to evaluate the distribution of IO_i (i.e., the number of units of product i queueing at the production facility, or being processed). This steady state distribution also determines the optimal order-up-to levels S_i^*. Taking the derivative of Eq. (2), we have:

$$G_{IO_i}(S_i^*) = b_i/(h_i + b_i), \tag{19}$$

where G_{IO_i} is the cumulative distribution function of IO_i.

As lead times are endogenous and load-dependent, IO_i is given by the amount of product i *in queue* prior to production, as well as the amount of product i *in service*, at the end of an arbitrary period t. Assume that the *age* of the order in service at time t is equal to k periods (which implies that the order was placed k periods ago; see Fig. 3 for an example), we then have (recall that

Age of order in service $= 2d \rightarrow 2$ orders in the queue
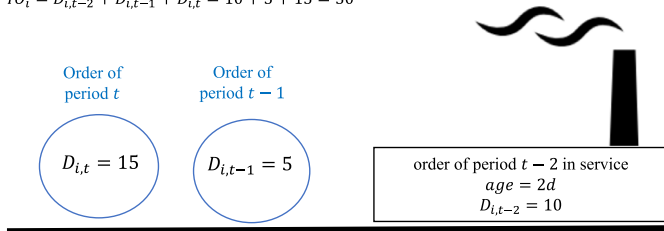$IO_i = D_{i,t-2} + D_{i,t-1} + D_{i,t} = 10 + 5 + 15 = 30$

Order of period t

Order of period $t-1$

$D_{i,t} = 15$

$D_{i,t-1} = 5$

order of period $t-2$ in service
$age = 2d$
$D_{i,t-2} = 10$

Fig. 3. An example of inventory on-order at the end of period t, immediately after sending O_t. Each period equals d time units.

$O_{i,t} = D_{i,t}$, see Section 3.3):

$$IO_{i,t} = D_{i,t-k} + \sum_{j=0}^{k-1} D_{i,t-j}, \tag{20}$$

with $D_{i,t-k}$ the amount of product i in service at time t, and $\sum_{j=0}^{k-1} D_{i,t-j}$ the amount of product i included in the k orders that are waiting in the production queue. The distribution of the amount in queue can be obtained from the k-fold convolution of the demand of product i. The distribution of the amount of products in service is more complex to determine: larger orders are more likely to have longer production times, implying that the age and the size of the order in production are correlated. We thus need to determine the joint probability that, at the end of an arbitrary period, the order in service has an age of k periods and contains a specific amount of product i. To that end, we model the production/inventory system as a discrete $D/PH/1$ queueing system, which can handle arbitrary processing time distributions. Boute et al. (2007) analyze such a queuing system but for a single product, we adapt their methodology for two products, and analyze the resulting Markov chain exactly using matrix-geometric methods (Neuts, 1981), see Online Appendix G for full details.

5.3. Optimizing prices and quality levels

In presence of load-dependent lead times, it is impossible to obtain closed form expressions for the optimal prices and the optimal quality of the second product. We thus resort to numerical techniques. Analogous to the exogenous lead time case, we optimize prices by first finding optimal market shares, and then applying Eq. (9). Although the expected profit function of the P/I system is not concave in market shares (we found examples showing that it might be non-concave), we prefer to optimize market shares since their feasible domain is clearly bounded ($0 < q_1 + q_2 < 1$). To avoid exhaustive search (which is very time-consuming), we use a neighborhood search to find the optimal market shares for each value of f_2.

The procedure starts from a point (q_1, q_2) and moves to the neighboring point with the highest improvement in expected profit (each point has 8 neighbors, the expected profit is calculated from Eq. (2), IO_i is obtained by solving the Markov chain). The search stops when no further improvement can be found. As the profit function might be non-concave, the neighborhood search may not converge to the global maximum. We mitigate this issue by using multiple starting points (this still does not guarantee global convergence, but reduces the chance of getting stuck in a local optimum). Based on some initial experiments and the computational time, we opted for 5 starting points.

6. Numerical experiments

This section compares the results of load-independent and load-dependent optimization. We demonstrate that the exogenous

lead time assumption (i.e., ignoring finite capacity) leads to suboptimal price and quality levels in settings where production capacity is limited, with lower profits as a result (Section 6.1). We also show how adding a utilization constraint based on expected processing times can mitigate this profit loss (Section 6.2). As our goal is to demonstrate the impact of congestion on optimal quality and pricing decisions, we focus on scenarios where the optimal solution of the load-independent optimization causes notable congestion in the production facility. As in these scenarios, the lead time L has only a minor effect on the optimal price and quality levels in the load-independent system (see Online Appendix H), without loss of generality, we set $L = 0$ periods (i.e., all orders arrive within one period after being placed, see Fig. 1). Finally, Section 6.3 discusses the generalizability of our observations.

Table 1 summarizes the parameter values of the base case scenario. We consider a period length of one day and use minutes as our time unit (i.e., $d = 1440$ minutes). In the MNL model, we set the quality sensitivity $\varepsilon_f = 1$ and the price sensitivity $\varepsilon_p = -0.8$, meaning that the customers are willing to pay up to $1/0.8 = 1.25$ units more for a unit increase in quality level f_i (Train, 2002). In practice, these sensitivities can be estimated by collecting data using software such as BIOGEME (Bierlaire, 2003).

We set $m_p = 5$, indicating that it takes on average 5 minutes to process a unit of product with quality level 1. The per unit average processing time of product i with quality f_i is thus $5f_i^2$ minutes. The material cost of product i is $c_i = mf_i^2$ where we set $m = 0.1$. The per unit holding cost of product i per *year* is 20%, implying a per unit holding cost per day of $h_i = 0.2c_i/365 = 0.000055f_i^2$. The service level equals $m_b/(m_b + m_h) = 0.995$. Unless otherwise mentioned, we set the quality of the first product to $f_1 = 1$ meaning that this product is always the lower quality product.

To ensure that load-independent and load-dependent systems are compared with the same conditions, we adopt the discrete PH distributions for product demands in the load-independent model as well; yet, this implies that the load-independent expected profit (Eq. (4)) no longer has a closed form, and is evaluated numerically.

6.1. Comparison of load-independent and load-dependent optimization

Observation 1. Load-independent optimization results in underpricing of the high-quality and (possibly) overpricing of the low-quality product compared to load-dependent optimization.

Fig. 4 compares the optimal product prices in the load-independent and load-dependent optimization models for different values of f_2: in both models $p_2^* > p_1^*$ (as product 2 is the higher quality product), but the optimal price difference between the products is markedly larger in the load-dependent case. The reason for this behavior is evident: as load-independent optimization disregards the impact of price and quality levels on the production load, it results in demands that are too high; in our case, demand can even exceed the production capacity, see Fig. 5. In practice, such overloaded settings are not sustainable in the long run: the company will either have to adjust capacity (for instance by outsourcing), or will have to adjust its product offer. In the load-dependent optimization, the system implicitly controls the congestion in production by shifting demand away from the high-quality product (which is most time-consuming to produce) to the low-quality product. This is achieved by increasing the price of the high-quality, while decreasing the price of the low-quality product.

Observation 2. The optimal quality level f_2^* resulting from load-dependent optimization is lower than that resulting from load-independent optimization; moreover, it decreases as the (fixed) quality of the first product (f_1) is higher.

Table 1
Parameter values in the base case scenario.

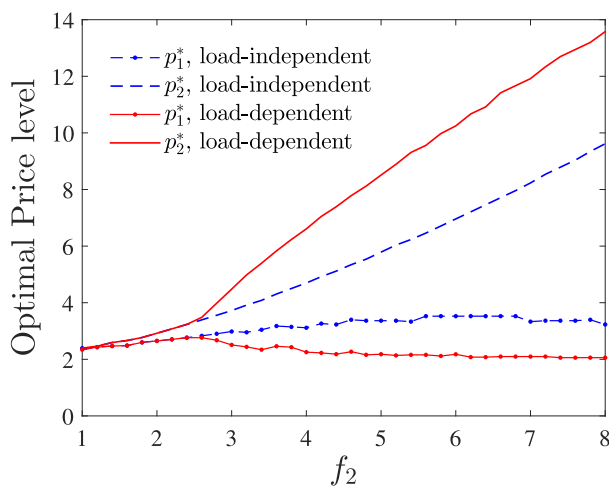|	Value
Parameters	
• Average and standard deviation of total demand per period (λ and σ)	100,10
• Quality of product 1 (f_1)	1
• Quality sensitivity (ε_f)	1
• Price sensitivity (ε_p)	−0.8
• Material, backorder, and holding cost coefficient (m, m_b, m_h)	0.1, 0.0109, 0.000055
• Single unit production time coefficient (m_p)	5
• Coefficient of variation of single unit production time (CV)	0.5
• Time units for one period (d)	1440 minutes
• Fixed lead time of load-independent system (L)	0
Feasible domain for decision variables	
• Quality of second product (f_2)	1: 0.2: 8
• Market share of product i (q_i)	0: 0.01: 0.999
• Feasible domain in load-independent setting	$q_1 + q_2 \leq 0.999$
• Feasible domain in load-dependent setting	$q_1 + q_2 \leq 0.999$ and $[(q_1\lambda)(m_p f_1^2) + (q_2\lambda)(m_p f_2^2)]/d < 0.98$

Fig. 4. Optimal product prices in the load-independent and load-dependent lead time optimization for different values of f_2 ($f_1 = 1$).
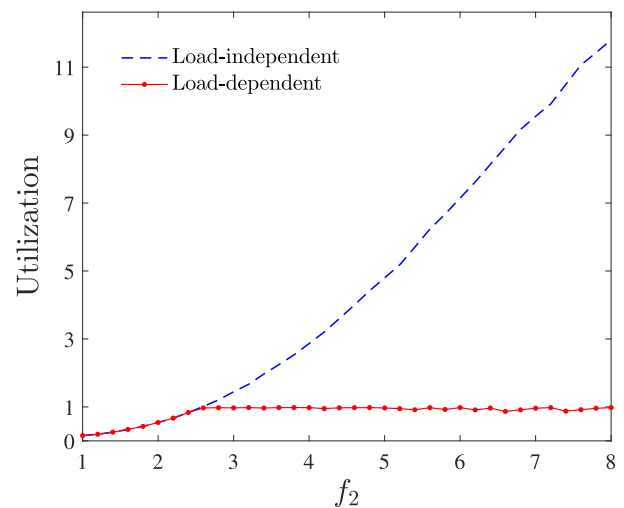

Fig. 5. Utilization in production corresponding to the optimal price configuration for different values of f_2 ($f_1 = 1$).

This is evident from Fig. 6, which shows the optimal expected profit for different values of f_2. With load-independent optimization, the optimal quality of the second product is $f_2^* = 6.2$, while with load-dependent optimization $f_2^* = 2.8$, for $f_1 = 1$ (see Fig. 6). For $f_1 = 2$, f_2^* remains the same with load-independent optimization (see Eq. (18)); in contrast, f_2^* reduces to 2.6 with load-dependent optimization. Again, this behavior can be explained by the system's attempt to control congestion: the (high) quality level resulting from load-independent optimization cannot be optimal in the load-dependent setting as it would require high price levels to control the congestion; such prices, in turn, result in low market coverage and profit loss. This is evident from Fig. 7, which shows the market coverage under optimized prices for different values of f_2. It is thus intuitive that the optimal quality of the load-dependent model cannot exceed the optimal quality of load-independent optimization, due to this price reaction. The impact on optimal market coverage follows immediately:

Observation 3. The market coverage resulting from load-dependent optimization will never exceed the market coverage resulting from load-independent optimization.

For load-independent optimization, higher average total demand (λ) decreases υ (i.e., the cost penalty for higher quality, see Eq. (15)), which in turn leads to higher f_2^* and higher market

coverage (see Theorems 4 and 5). This is no longer true with load-dependent optimization:

Observation 4. With load-dependent optimization, higher average total demand (λ) may lead to lower values of f_2^* and result in lower market coverage.

This contrasts with Theorems 4 and 5 in the case of load-independent optimization. Increasing average total demand in the base case from $\lambda = 100$ to $\lambda = 200$, for instance, results in $f_2^* = 2.2$ (versus $f_2^* = 2.8$ when $\lambda = 100$) and an optimal market coverage of 46% (down from 50%) with load-dependent optimization; for higher demand levels, the market coverage must indeed be decreased to control the congestion in production.

Observations 1–4 show that the optimal price and quality levels resulting from load-independent and load-dependent optimization differ; implementing the load-independent solution in settings with endogenous lead times may thus substantially hurt profits. The profit loss may be minor, though, in settings that are less prone to congestion.

As an illustration, Table 2 shows the profit loss for the base case scenario (see Table 1) and a number of variants (the difference with regard to the base case is shown in column 1). The second column reports the percentage profit loss when implementing the prices, quality, and order-up-to levels resulting from
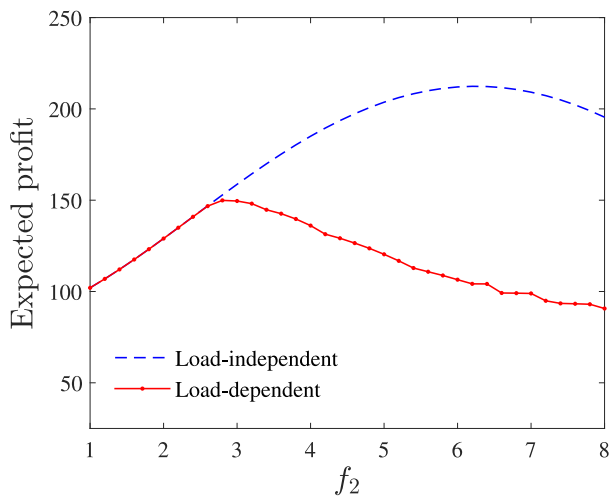
Fig. 6. Profit corresponding to the optimal prices in the load-independent and load-dependent optimization (Fig. 4) for different values of f_2.
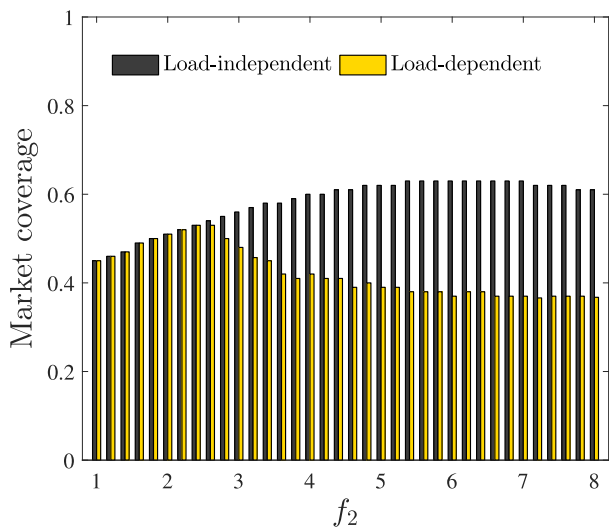

Fig. 7. Market coverage corresponding to the optimal price levels in the load-independent and load-dependent lead time settings for different values of f_2; $f_1 = 1$.

load-independent optimization in the corresponding P/I system. When this solution results in an unstable queue (as in the base case scenario), the profit loss is reported as ∞.

In settings with high cost coefficients, high price sensitivity or low quality sensitivity, the profit loss is minor: as Eq. (18) reveals, these conditions decrease the f_2^* in the load-independent optimization, such that the impact on utilization is minor. The same result is observed when the average and/or variance of the per-unit

processing time is reduced (i.e., lower m_p and CV), as the impact of f_2^* on system utilization is then automatically mitigated. This is summarized in the following observation:

Observation 5. The profit loss resulting from load-independent optimization is minor when average per unit processing times are short, processing time variability is low, cost of quality is high, and customers have high price sensitivity and/or low quality sensitivity.

6.2. Utilization-based optimization

The question remains how to mitigate the profit loss in settings where it *is* substantial. Estimating the impact of marketing decisions on replenishment lead times may be difficult in practice, as it requires to explicitly account for congestion and variability effects. One approach is to optimize decisions subject to an explicit utilization constraint:

$$[q_1\lambda E(M_1) + q_2\lambda E(M_2)]/d < \delta. \tag{21}$$

where we set $\delta = \min(0.99,$ utilization of the optimal load-independent solution). The third column in Table 2 reports the percentage profit loss resulting from this *utilization-based* optimization, along with the value of δ. Evidently, this approach is most effective when load-independent optimization results in high congestion (as in the base case).

The advantage of this approach is that, contrary to load-dependent optimization, it only requires insight into the relationship between quality levels and the per unit average processing times (i.e., $E(M_i) = m_p f_i^2$). We acknowledge that that this approach can be further fine-tuned, as the choice for the utilization threshold (δ) may be improved. We hope our work may provide a stepping stone for researchers to help develop other pragmatic approaches for optimizing decisions in P/I systems with endogenous lead times.

6.3. Discussion

Although some of our insights are based on numerical observations, we are confident that they are not specific to the values used and can be generalized, as the underlying "mechanics" are ubiquitous in practice: with scarce capacity, the firm needs to adjust the price and quality levels to control the congestion in the load-dependent system. Increasing the price of the high-quality product dampens its demand and, consequently, decreases congestion (as the high-quality product is more time-consuming to produce). Lowering its quality level also decreases its demand, and additionally relieves congestion by shortening the average per-unit production time. It also lowers the per-unit backorder and holding costs, making the impact of congestion less costly. As we have shown, it may thus be optimal to both increase the price and decrease the quality of the high-quality product in P/I systems with load-dependent lead times.

Although our observations rely on several functional assumptions, these are not restrictive for our main insights. For instance,

Table 2
Percentage profit loss with load-independent and utilization-based optimization (δ denotes the utilization threshold for utilization-based optimization).

Variant	Percentage profit loss under load-independent optimization	Percentage profit loss under utilization-based optimization
0) Base case ($\lambda = 100, \sigma = 10, m = 0.1, m_b = 0.0109, m_h = 0.000055, m_p = 5, CV = 0.5, \varepsilon_p = -0.8, \varepsilon_f = 1$)	∞	1.17 ($\delta = 0.99$)
1) High cost environment ($m = 0.22, m_b = 0.024, m_h = 0.00012$)	1.01	0.71 ($\delta = 0.9$)
2) High price sensitivity ($\varepsilon_p = -1.8$)	1.05	0.7 ($\delta = 0.9$)
3) Low quality sensitivity ($\varepsilon_f = 0.5$)	0.79	0.42 ($\delta = 0.91$)
4) Low average processing time per unit ($m_p = 0.6, CV = 0.5$)	2.6	2.06 ($\delta = 0.91$)
5) Low average processing time per unit with high uncertainty ($m_p = 0.6, CV = 0.8$)	3.14	2.61 ($\delta = 0.91$)

we opted for the Multinomial Logit model to reflect customer choice, as it conveniently allows for modeling customers' reaction to price and quality. In fact, regardless of the choice model used, ignoring congestion-related costs naturally makes decision makers aim for higher quality levels at lower prices, as they believe that the increase in market coverage will lead to higher expected profit. The quadratic relation between quality and material cost ($c = mf^\alpha$ with $\alpha = 2$) is convex, as is commonly observed in practice (Jerath, Kim, & Swinney, 2017). In fact, our insights hold for any $\alpha \geq 1$, as the actual value of α only impacts the extent to which quality impacts costs, and has no direct impact on the congestion-related insights.

We assumed that the average per unit production time increases quadratically in the quality level ($E(M_i) = m_p f_i^\beta$ with $\beta = 2$). In practice, this production time might increase at a slower rate. Even when $\beta = 1$, load-independent optimization may still overestimate the quality of the high-quality product as an excessive quality level may lead to high demands and high per-unit production costs (see Online Appendix I for an example).

7. Conclusions

This paper has shown that the optimal prices and qualities of the products offered by a finite-capacity make-to-stock manufacturer are impacted by the endogeneity of the (load-dependent) production lead times. While the assumption of fixed, exogenous lead times is common in current research, and in many real-life planning systems, it results in a product offer that aims for an excessive market coverage, resulting in profit loss due to the increased congestion in the production system. Only in case of abundant capacity, or when the system is not prone to congestion (e.g., low unit processing times in production, low quality sensitivity in the market), assuming load-independent and fixed lead times is justified. In that case, closed form results can be obtained for the optimal price and quality levels. Though our insights are partly based on numerical results, we are confident that they are generalizable, as the main drivers of system behavior (e.g., negative price sensitivity, positive quality sensitivity, unit processing times that are increasing in product quality level, endogenous production lead times) are ubiquitous in practice.

Acknowledgments

Appendix A. Calculations for Eq. (8)

Replacing the order-up-to levels with their optimal expressions $S_i^* = (L+1)q_i\lambda + \Phi^{-1}(\frac{m_b}{m_b+m_h})\sqrt{L+1}\ q_i\sigma$, the $z_i = (S_i - (L+1)q_i\lambda)/\sqrt{L+1}q_i\sigma$ becomes $z_c = \Phi^{-1}(\frac{m_b}{m_b+m_h})$. Therefore, instead of $L(z_i)$, we obtain:

$$L(z_c) = \phi(z_c) - z_c(1 - \Phi(z_c)) = \phi(z_c) - z_c\left(\frac{m_h}{m_b+m_h}\right),$$

and we can re-write the profit function (7):

$$\Pi = \sum_{i=1}^{2}\left[(p_i - mf_i^2)q_i\lambda - m_h f_i^2(S_i^* - (L+1)q_i\lambda + \sqrt{L+1}\ q_i\sigma L(z_c))\right.$$
$$\left. - m_b f_i^2\sqrt{L+1}\ q_i\sigma L(z_c)\right]. \tag{22}$$

Expanding S_i^*, we have:

$$\Pi(p_1, p_2) = \sum_{i=1}^{2}\left[(p_i - mf_i^2)q_i\lambda - (m_h + m_b)f_i^2\sqrt{L+1}q_i\sigma L(z_c)\right.$$
$$\left. - m_h\sqrt{L+1}f_i^2 q_i\sigma z_c\right]. \tag{23}$$

Expanding $L(z_c)$ in the second term, we obtain expression (8).

Appendix B. Proof of Theorem 1

All the terms inside the summation sign in Eq. (10) are jointly concave in q_1 and q_2. The last term is also jointly concave because its Hessian is negative semi-definite. Indeed,

$$Hessian = \begin{bmatrix} T & T \\ T & T \end{bmatrix}, \quad T = \frac{2\lambda}{\varepsilon_p(1 - q_1 - q_2)}$$
$$+ \frac{(q_1 + q_2)\lambda}{\varepsilon_p(1 - q_1 - q_2)^2} < 0 \text{ as } \varepsilon_p < 0,$$

such that, for any real numbers x and y, we obtain

$$\begin{pmatrix} x & y \end{pmatrix}\begin{bmatrix} T & T \\ T & T \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = T(x + y)^2 \leq 0.$$

As the sum of concave terms is concave, this proves Theorem 1.

Appendix C. Proof of Theorem 3

In this appendix we prove that $\partial p_2^*/\partial f_2 > 0$. Using Eq. (13) in (14), we have:

$$p_i^* = \upsilon f_i^2 - \frac{1}{\varepsilon_p}(e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} + e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}) - \frac{1}{\varepsilon_p}. \tag{24}$$

Using implicit differentiation, we obtain:

$$\frac{\partial p_2^*}{\partial f_2} = 2\upsilon f_2 - \frac{1}{\varepsilon_p}\left(\varepsilon_p\frac{\partial p_1^*}{\partial f_2}\right)e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} - \frac{1}{\varepsilon_p}\left(\varepsilon_f + \varepsilon_p\frac{\partial p_2^*}{\partial f_2}\right)e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}. \tag{25}$$

After simplification, we have:

$$\left(1 + e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}\right)\frac{\partial p_2^*}{\partial f_2} = 2\upsilon f_2 - \frac{\partial p_1^*}{\partial f_2}\ e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} - \frac{\varepsilon_f}{\varepsilon_p}\ e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}. \tag{26}$$

Analogously,

$$\frac{\partial p_1^*}{\partial f_2} = -\frac{1}{\varepsilon_p}\left(\varepsilon_p\frac{\partial p_1^*}{\partial f_2}\right)e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} - \frac{1}{\varepsilon_p}\left(\varepsilon_f + \varepsilon_p\frac{\partial p_2^*}{\partial f_2}\right)e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}. \tag{27}$$

From Eqs. (25) and (27), it follows that

$$\frac{\partial p_1^*}{\partial f_2} = \frac{\partial p_2^*}{\partial f_2} - 2\upsilon f_2, \tag{28}$$

such that Eq. (26) reduces to

$$\left(1 + e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}\right)\frac{\partial p_2^*}{\partial f_2} = 2\upsilon f_2 - \left(\frac{\partial p_2^*}{\partial f_2} - 2\upsilon f_2\right)e^{\varepsilon_f f_1 + \varepsilon_p p_1^*}$$
$$- \frac{\varepsilon_f}{\varepsilon_p}\ e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}, \tag{29}$$

and thus

$$\frac{\partial p_2^*}{\partial f_2} = \frac{2\upsilon f_2\left(1 + e^{\varepsilon_f f_1 + \varepsilon_p p_1^*}\right) - \frac{\varepsilon_f}{\varepsilon_p}e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}}{1 + e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} + e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}} > 0. \tag{30}$$

The above expression is always positive (recall that $\upsilon > 0$ and $\varepsilon_p < 0$), proving that the optimal price of the second product increases as its quality increases. The relationship between p_1^* and f_2

is not monotone. From Eqs. (30) and (28), it follows that:

$$\frac{\partial p_1^*}{\partial f_2} = \frac{\left(-\frac{\varepsilon_f}{\varepsilon_p} - 2\upsilon f_2\right)e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}}{1 + e^{\varepsilon_f f_1 + \varepsilon_p p_1^*} + e^{\varepsilon_f f_2 + \varepsilon_p p_2^*}} \begin{cases} < 0 & \text{if } f_2 > -\varepsilon_f/(2\upsilon\varepsilon_p) \\ > 0 & \text{if } f_2 < -\varepsilon_f/(2\upsilon\varepsilon_p), \end{cases}$$

(31)

such that the sign of $\partial p_1^*/\partial f_2$ depends on f_2.

Appendix D. Proof of Theorem 4

Eq. (16) can be rewritten as follows:

$$\frac{-\varepsilon_p \Pi(p_1^*, p_2^*)}{\lambda} \, e^{\frac{-\varepsilon_p \Pi(p_1^*, p_2^*)}{\lambda}} = e^{\varepsilon_f f_1 + \upsilon\varepsilon_p f_1^2 - 1} + e^{\varepsilon_f f_2 + \upsilon\varepsilon_p f_2^2 - 1}.$$

The left-hand side is nondecreasing in $\Pi(p_1^*, p_2^*)$; consequently, the larger the right-hand side, the larger the value of $\Pi(p_1^*, p_2^*)$ that satisfies the equation. The quality of the first product (f_1) is fixed; thus, the right-hand side reaches a maximum when $\varepsilon_f f_2 + \upsilon\varepsilon_p f_2^2 - 1$ is at its maximum, which happens at $f_2 = -\varepsilon_f/(2\upsilon\varepsilon_p)$.

Appendix E. Proof of Theorem 5

The probability that an arriving customer doesn't purchase any of the products (i.e., nonpurchase probability) is given by the MNL model (Eq. (3)):

$$q_0 = \frac{1}{1 + \exp(\varepsilon_f f_1 + \varepsilon_p p_1) + \exp(\varepsilon_f f_2 + \varepsilon_p p_2)}.$$

(32)

Therefore, the expected size of the market that remains uncovered is $q_0\lambda$. For the optimal quality of the second product ($-\varepsilon_f/(2\upsilon\varepsilon_p)$), the optimal prices in Eq. (12) can be written as:

$$p_1^* = \upsilon f_1^2 - \frac{1}{\varepsilon_p q_0^{**}},$$

$$p_2^* = \upsilon\left(\frac{\varepsilon_f^2}{4\upsilon^2\varepsilon_p^2}\right) - \frac{1}{\varepsilon_p q_0^{**}},$$

(33)

where q_0^{**} denotes the nonpurchase probability when p_1, p_2, and f_2 are chosen optimally. Therefore, from Eq. (32), we have:

$$q_0^{**} = \frac{1}{1 + \exp(\varepsilon_f f_1 + \varepsilon_p\upsilon f_1^2 - 1/q_0^{**}) + \exp(-\varepsilon_f^2/(2\upsilon\varepsilon_p) + \varepsilon_f^2/(4\upsilon\varepsilon_p) - 1/q_0^{**})}$$

$$= \frac{1}{1 + \exp(\varepsilon_f f_1 + \varepsilon_p\upsilon f_1^2 - 1/q_0^{**}) + \exp(-\varepsilon_f^2/(4\upsilon\varepsilon_p) - 1/q_0^{**})}.$$

(34)

Using implicit differentiation, we calculate the derivative of q_0^{**} with respect to υ, ε_f, and ε_p and show how the optimal market coverage $(1 - q_0^{**})$ changes with these parameters. To simplify the notations, let $\exp(\mu_1^{**}) = \exp(\varepsilon_f f_1 + \varepsilon_p\upsilon f_1^2 - 1/q_0^{**})$ and $\exp(\mu_2^{**}) = \exp(-\varepsilon_f^2/(4\upsilon\varepsilon_p) - 1/q_0^{**})$:

$$\frac{\partial q_0^{**}}{\partial \varepsilon_p} = \frac{-\left[\left(\upsilon f_1^2 + \frac{1}{(q_0^{**})^2} \times \frac{\partial q_0^{**}}{\partial \varepsilon_p}\right)\exp(\mu_1^{**}) + \left(\frac{\varepsilon_f^2}{4\upsilon\varepsilon_p^2} + \frac{1}{(q_0^{**})^2} \times \frac{\partial q_0^{**}}{\partial \varepsilon_p}\right)\exp(\mu_2^{**})\right]}{[1 + \exp(\mu_1^{**}) + \exp(\mu_2^{**})]^2}$$

$$\rightarrow \frac{\partial q_0^{**}}{\partial \varepsilon_p} = \frac{-\upsilon f_1^2 \exp(\mu_1^{**}) - \frac{\varepsilon_f^2}{4\upsilon\varepsilon_p^2}\exp(\mu_2^{**})}{[1 + \exp(\mu_1^{**}) + \exp(\mu_2^{**})]^2 + \exp(\mu_1^{**})/(q_0^{**})^2 + \exp(\mu_2^{**})/(q_0^{**})^2} < 0.$$

(35)

Thus, as ε_p decreases (higher price sensitivity), optimal market coverage decreases (higher q_0^{**}). Analogously:

$$\frac{\partial q_0^{**}}{\partial \varepsilon_f} = \frac{-\left[\left(f_1 + \frac{1}{(q_0^{**})^2} \times \frac{\partial q_0^{**}}{\partial \varepsilon_f}\right)\exp(\mu_1^{**}) + \left(\frac{-2\varepsilon_f}{4\upsilon\varepsilon_p} + \frac{1}{(q_0^{**})^2} \times \frac{\partial q_0^{**}}{\partial \varepsilon_f}\right)\exp(\mu_2^{**})\right]}{[1 + \exp(\mu_1^{**}) + \exp(\mu_2^{**})]^2}$$

$$\rightarrow \frac{\partial q_0^{**}}{\partial \varepsilon_f} = \frac{-f_1 \exp(\mu_1^{**}) + \frac{\varepsilon_f}{2\upsilon\varepsilon_p}\exp(\mu_2^{**})}{[1 + \exp(\mu_1^{**}) + \exp(\mu_2^{**})]^2 + \exp(\mu_1^{**})/(q_0^{**})^2 + \exp(\mu_2^{**})/(q_0^{**})^2} < 0,$$

(36)

which shows that optimal market coverage decreases as customers become less sensitive to quality. Finally, we have:

$$\frac{\partial q_0^{**}}{\partial \upsilon} = \frac{-\left[\left(\varepsilon_p f_1^2 + \frac{1}{(q_0^{**})^2} \times \frac{\partial q_0^{**}}{\partial \upsilon}\right)\exp(\mu_1^{**}) + \left(\frac{\varepsilon_f^2}{4\upsilon^2\varepsilon_p} + \frac{1}{(q_0^{**})^2} \times \frac{\partial q_0^{**}}{\partial \upsilon}\right)\exp(\mu_2^{**})\right]}{[1 + \exp(\mu_1^{**}) + \exp(\mu_2^{**})]^2}$$

$$\rightarrow \frac{\partial q_0^{**}}{\partial \upsilon} = \frac{-\varepsilon_p f_1^2 \exp(\mu_1^{**}) - \frac{\varepsilon_f^2}{4\upsilon^2\varepsilon_p}\exp(\mu_2^{**})}{[1 + \exp(\mu_1^{**}) + \exp(\mu_2^{**})]^2 + \exp(\mu_1^{**})/(q_0^{**})^2 + \exp(\mu_2^{**})/(q_0^{**})^2} > 0,$$

(37)

Thus, the optimal market coverage decreases as υ increases (which occurs when lead time increases, cost coefficients increase, or the coefficient of variation of demand increases).

Based on the above results and Eq. (33), we can readily show that p_1^* and p_2^* decrease with higher $|\varepsilon_p|$ and lower ε_f. As υ increases, p_2^* decreases but p_1^* may increase or decrease depending on the fixed value of f_1.

It remains to be shown that the resulting profit also decreases. From Eq. (13), we have:

$$\Pi^* = -\frac{\lambda}{\varepsilon_p}\left(\frac{1}{q_0^{**}} - 1\right),$$

(38)

and, thus,

$$\frac{\partial \Pi^*}{\partial \varepsilon_p} = \frac{\lambda}{\varepsilon_p^2}\left(\frac{1}{q_0^{**}} - 1\right) - \frac{\lambda}{\varepsilon_p} \times \frac{-\frac{\partial q_0^{**}}{\partial \varepsilon_p}}{(q_0^{**})^2} > 0,$$

$$\frac{\partial \Pi^*}{\partial \varepsilon_f} = -\frac{\lambda}{\varepsilon_p} \times \frac{-\frac{\partial q_0^{**}}{\partial \varepsilon_f}}{(q_0^{**})^2} > 0,$$

$$\frac{\partial \Pi^*}{\partial \upsilon} = -\frac{\lambda}{\varepsilon_p} \times \frac{-\frac{\partial q_0^{**}}{\partial \upsilon}}{(q_0^{**})^2} < 0.$$

(39)

Consequently, as $|\varepsilon_p|$ increases, ε_f decreases, or υ increases, the optimal profit drops.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2018.06.013.

References

Akcay, Y., Natarajan, H. P., & Xu, S. H. (2010). Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science, 56*(8), 1345–1361.

Alptekinoğlu, A., & Corbett, C. J. (2010). Leadtime-variety tradeoff in product differentiation. *Manufacturing & Service Operations Management, 12*(4), 569–582.

Alptekinoğlu, A., & Semple, J. H. (2016). The exponomial choice model: A new alternative for assortment and price optimization. *Operations Research, 64*(1), 79–93.

Anderson, S. P., De Palma, A., & Thisse, J. F. (1992). Discrete choice theory of product differentiation. MIT Press.

Asmundsson, J., Rardin, R. L., & Uzsoy, R. (2006). Tractable nonlinear production planning models for semiconductor wafer fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing, 19*(1), 95–111.

Benjaafar, S., Cooper, W. L., & Kim, J.-S. (2005). On the benefits of pooling in production-inventory systems. *Management Science, 51*(4), 548–565.

Bernstein, F., Li, Y., & Shang, K. (2015). A simple heuristic for joint inventory and pricing models with lead time and backorders. *Management Science, 62*(8), 2358–2373.

Bierlaire, M. (2003). Biogeme: A free package for the estimation of discrete choice models. In *Proceedings of the Swiss transport research conference*. TRANSP-OR–CONF-2006-048.

Bish, E. K., & Chen, W. (2016). The optimal resource portfolio under consumer choice and demand risk for vertically differentiated products. *Journal of the Operational Research Society, 67*(1), 87–97.

Boute, R. N., Van den Broeke, M. M., & Deneire, K. A. (2018). Barco implements platform-based product development in its healthcare division. *Interfaces, 48*(1), 35–44.

Boute, R. N., Lambrecht, M. R., & Van Houdt, B. (2007). Performance evaluation of a production/inventory system with periodic review and endogenous lead times. *Naval Research Logistics (NRL), 54*(4), 462–473.

Cattani, K. D., Dahan, E., & Schmidt, G. M. (2010). Lowest cost may not lower total cost: Using spackling to smooth mass-customized production. *Production and Operations Management, 19*(5), 531–545.

Chayet, S., Kouvelis, P., & Yu, D. Z. (2011). Product variety and capacity investments in congested production systems. *Manufacturing & Service Operations Management, 13*(3), 390–403.

Chen, X., & Simchi-Levi, D. (2012). Pricing and inventory management. In *The handbook of pricing management* (pp. 784–822). Oxford University Press: Oxford, UK.

Davis, J., Gallego, G., & Topaloglu, H. (2013). Assortment planning under the multinomial logit model with totally unimodular constraint structures. Working paper. Ithaca, NY: Cornell University.

Dobson, G., & Yano, C. A. (2002). Product offering, pricing, and make-to-stock/make-to-order decisions with shared capacity. *Production and Operations Management, 11*(3), 293–312.

Dong, L., Kouvelis, P., & Tian, Z. (2009). Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management, 11*(2), 317–339.

Du, C., Cooper, W. L., & Wang, Z. (2016). Optimal pricing for a multinomial logit choice model with network effects. *Operations Research, 64*(2), 441–455.

Federgruen, A., & Heching, A. (1999). Combined pricing and inventory control under uncertainty. *Operations Research, 47*(3), 454–475.

Gaur, V., & Honhon, D. (2006). Assortment planning and inventory decisions under a locational choice model. *Management Science, 52*(10), 1528–1543.

Gupta, D., & Benjaafar, S. (2004). Make-to-order, make-to-stock, or delay product differentiation? A common framework for modeling and analysis. *IIE transactions, 36*(6), 529–546.

Hafızoğlu, A. B., Gel, E. S., & Keskinocak, P. (2016). Price and lead time quotation for contract and spot customers. *Operations Research, 64*(2), 406–415.

Heese, H. S., & Swaminathan, J. M. (2006). Product line design with component commonality and cost-reduction effort. *Manufacturing & Service Operations Management, 8*(2), 206–219.

Hopp, W. J., & Spearman, M. L. (2001). Factory physics: Foundations of manufacturing management. Irwin/McGraw-Hill.

Jayaswal, S., Jewkes, E., & Ray, S. (2011). Product differentiation and operations strategy in a capacitated environment. *European Journal of Operational Research, 210*(3), 716–728.

Jerath, K., Kim, S.-H., & Swinney, R. (2017). Product quality in a distribution channel with inventory risk. *Marketing Science, 36*(5), 747–761.

Kök, A. G., & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research, 55*(6), 1001–1021.

Kök, A. G., Fisher, M. L., & Vaidyanathan, R. (2015). Assortment planning: Review of literature and industry practice. In *Retail supply chain management* (pp. 175–236). Springer.

Kök, A. G., & Xu, Y. (2011). Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Science, 57*(9), 1546–1563.

Krishnamurthy, A., & Suri, R. (2009). Planning and implementing polca: A card-based control system for high variety or custom engineered products. *Production planning and control, 20*(7), 596–610.

Kwark, Y., Chen, J., & Raghunathan, S. (2017). User-generated content and competing firms product design. *Management Science*. doi:10.1287/mnsc.2017.2839.

Lambrecht, M. R., Ivens, P. L., & Vandaele, N. J. (1998). Aclips: A capacity and lead time integrated procedure for scheduling. *Management Science, 44*(11-part-1), 1548–1561.

Land, M., & Gaalman, G. (1996). Workload control concepts in job shops a critical assessment. *International journal of production economics, 46*, 535–548.

Latouche, G., & Ramaswami, V. (1999). Introduction to matrix analytic methods in stochastic modeling: Vol. 5. Philadelphia, Siam.

Li, H., & Huh, W. T. (2011). Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing & Service Operations Management, 13*(4), 549–563.

Li, H., Webster, S., Mason, N., & Kempf, K. (2017). Winner-2017 M&SOM practice-based research competition product-line pricing under discrete mixed multinomial logit demand. *Manufacturing & Service Operations Management*. doi:10.1287/msom.2017.0675.

Maddah, B., & Bish, E. K. (2007). Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics (NRL), 54*(3), 315–330.

Mayorga, M. E., Ahn, H.-S., & Aydin, G. (2013). Assortment and inventory decisions with multiple quality levels. *Annals of Operations Research, 211*(1), 301–331.

Nahmias, S. (1997). Production and operation analysis (3rd ed.). McGraw-Hill, New York.

Netessine, S., & Taylor, T. A. (2007). Product line design and production technology. *Marketing Science, 26*(1), 101–117.

Neuts, M. F. (1981). Matrix-geometric solutions in stochastic models: an algorithmic approach. Baltimore, MD: John Hopkins University Press.

Noblesse, A. M., Boute, R. N., Lambrecht, M. R., & Van Houdt, B. (2014). Lot sizing and lead time decisions in production/inventory systems. *International Journal of Production Economics, 155*, 351–360.

Örsdemir, A., Kemahlıoğlu-Ziya, E., & Parlaktürk, A. K. (2014). Competitive quality choice and remanufacturing. *Production and Operations Management, 23*(1), 48–64.

Pahl, J., Voß, S., & Woodruff, D. L. (2007). Production planning with load dependent lead times: an update of research. *Annals of Operations Research, 153*(1), 297–345.

Palaka, K., Erlebacher, S., & Kropp, D. H. (1998). Lead-time setting, capacity utilization, and pricing decisions under lead-time dependent demand. *IIE Transactions, 30*(2), 151–163.

Pan, X. A., & Honhon, D. (2012). Assortment planning for vertically differentiated products. *Production and Operations Management, 21*(2), 253–275.

Pang, Z., Chen, F. Y., & Feng, Y. (2012). Technical note-a note on the structure of joint inventory-pricing control with leadtimes. *Operations Research, 60*(3), 581–587.

Pekgün, P., Griffin, P. M., & Keskinocak, P. (2008). Coordination of marketing and production for price and leadtime decisions. *IIE Transactions, 40*(1), 12–30.

Ray, S., & Jewkes, E. M. (2004). Customer lead time management when both demand and price are lead time sensitive. *European Journal of operational research, 153*(3), 769–781.

Rong, Y., Chen, Y.-J., & Shen, Z.-J. M. (2015). The impact of demand uncertainty on product line design under endogenous substitution. *Naval Research Logistics (NRL), 62*(2), 143–157.

Rusmevichientong, P., Shen, Z.-J. M., & Shmoys, D. B. (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research, 58*(6), 1666–1680.

Selcuk, B., Fransoo, J. C., & De Kok, A. (2006). The effect of updating lead times on the performance of hierarchical planning systems. *International Journal of Production Economics, 104*(2), 427–440.

Shin, H., Park, S., Lee, E., & Benton, W. (2015). A classification of the literature on the planning of substitutable products. *European Journal of Operational Research, 246*(3), 686–699.

Sibdari, S., & Pyke, D. F. (2010). A competitive dynamic pricing model when demand is interdependent over time. *European Journal of Operational Research, 207*(1), 330–338.

Song, J.-S., & Xue, Z. (2007). *Demand management and inventory control for substitutable products*. Durham, NC: The Fuqua School of Business, Duke University.

Spearman, M. L., Hopp, W. J., & Woodruff, D. L. (1989). A hierarchical control architecture for constant work-in-process (conwip) production systems. *Journal of Manufacturing and Operations Management, 2*(3), 147–171.

Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). Conwip: A pull alternative to kanban. *The International Journal of Production Research, 28*(5), 879–894.

Tang, C. S. (2010). A review of marketing–operations interface models: From co-existence to coordination and collaboration. *International Journal of Production Economics, 125*(1), 22–40.

Tang, C. S., & Yin, R. (2010). The implications of costs, capacity, and competition on product line selection. *European Journal of Operational Research, 200*(2), 439–450.

Topaloglu, H. (2013). Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management, 22*(5), 1182–1199.

Train, K. E. (2002). Discrete choice methods with simulation. Cambridge University Press.

Upasani, A., & Uzsoy, R. (2008). Incorporating manufacturing lead times in joint production-marketing models: A review and some future directions. *Annals of Operations Research, 161*(1), 171–188.

Van Mieghem, J., & Allon, G. (2015). Operations Strategy: Principles and Practice. Dynamic Ideas.

Van Nieuwenhuyse, I., De Boeck, L., Lambrecht, M., & Vandaele, N. J. (2011). Advanced resource planning as a decision support module for erp. *Computers in Industry, 62*(1), 1–8.

Van Nieuwenhuyse, I., Mahihenni, M. H., & Baudelocq, S. (2014). Trading off inventory, capacity and customer service in semi-process industries: a case study at seppic. *European Journal of Industrial Engineering, 8*(1), 90–111.

Van Nieuwenhuyse, I., Vandaele, N., Rajaram, K., & Karmarkar, U. S. (2007). Buffer sizing in multi-product multi-reactor batch processes: Impact of allocation and campaign sizing policies. *European Journal of Operational Research, 179*(2), 424–443.

Van Nyen, P. L., Bertrand, J. W. M., & Van Ooijen, H. P. (2009). A computational comparison of cyclical and non-cyclical control for stochastic production-inventory systems. *International Journal of Production Research, 47*(16), 4609–4627.

Van Ryzin, G., & Mahajan, S. (1999). On the relationship between inventory costs and variety benefits in retail assortments. *Management Science, 45*(11), 1496–1509.

Vandaele, N., Van Nieuwenhuyse, I., Claerhout, D., & Cremmery, R. (2008). Load-based polca: An integrated material control system for multiproduct, multimachine job shops. *Manufacturing & Service Operations Management, 10*(2), 181–197.

Vandaele N., D. S. N. C. R., & Lambrecht, M. (2000). Spicer off-highway products division - brugge improves its lead-time and scheduling performance. *Interfaces, 30*, 83–95.

Vulcano, G., van Ryzin, G., & Chaar, W. (2010). Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management, 12*(3), 371–392.

Wong, H., & Naim, M. (2011). Re-examining postponement benefits: An integrated production-inventory and marketing perspective. In *Mass customization* (pp. 305–330). Springer.

Yang, N., & Zhang, R. (2014). Dynamic pricing and inventory management under inventory-dependent demand. *Operations Research, 62*(5), 1077–1094.

Zhu, K., & Thonemann, U. W. (2009). Coordination of pricing and inventory control across products. *Naval Research Logistics (NRL), 56*(2), 175–190.

Zijm, W. H., & Buitenhek, R. (1996). Capacity planning and lead time management. *International Journal of Production Economics, 46*, 165–179.

Zipkin, P. H. (2000). Foundations of inventory management. New York: McGraw-Hill.