Int. J. Production Economics 273 (2024) 109261

Contents lists available at ScienceDirect

International Journal of Production Economics

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

A budgeting resource allocation model for capacity expansion

Senay Solak [a], Zhuoxin Li [b,a], Mehmet Gumus [c], Yueran Zhuo [d,a,*]

[a] Isenberg School of Management, University of Massachusetts Amherst, USA
[b] Wisconsin School of Business, University of Wisconsin-Madison, USA
[c] Desautels Faculty of Management, McGill University, Canada
[d] College of Business, Mississippi State University, USA

ARTICLE INFO

Keywords:
Strategic planning
Capacity management
Dynamic programming
Resource allocation

ABSTRACT

The reconfiguration of the global supply chain has created opportunities for newly industrialized economies (NIEs) to strengthen and expand their production capacities. We develop a dynamic modeling framework to study a firm's resource allocation decisions while building capacity for a portfolio of production lines over multiple decision periods. Our model adds to the literature on versions of this problem in which distributions of market demands are incorporated into the investment–return function, and where no partial order fulfillment is possible. We consider two specific cases: one involving exogenously defined demand distributions, and one where the distributions of the parameters of the demand distribution are updated based on Bayesian inference. For each case, we first identify optimal capacity levels as a function of unit investment cost, holding cost, and the demand distribution of each production line. We then characterize the optimal path to reach these levels under different market conditions by deriving closed-form optimal resource allocation policies.

1. Introduction

Capacity management has emerged as a vital process in both manufacturing and service organizations, gaining heightened significance in the context of recent global trends in supply chain reconfiguration. This reconfiguration represents a strategic move by Western countries to mitigate risks associated with geopolitical tensions and to protect intellectual property rights. The result has been a relocation of numerous supply chains to newly industrialized economies (NIEs) such as Vietnam, Mexico, the Philippines, and India, seen as more geopolitically stable and friendly.

The shifting landscape has ushered in unparalleled opportunities for these NIEs to fortify and broaden their manufacturing and service capacities. As orders pour in, they are faced with the challenge of cementing long-term customer relationships through consistent, high-quality production. Successfully navigating this complex and turbulent period of global supply chain realignment could position these nations to emulate China's economic success and emerge as the new focal points of the world economy.

The on-going trend of supply base reduction, observed over the past few decades, has intensified competition among manufacturers in the newly industrialized economies (NIEs). This trend involves downstream buyers intentionally contracting with fewer suppliers to save costs and enhance coordination and communication. While supply base reduction offers many benefits to the buyer, such as improved efficiency, it also increases the risks of supply disruption for the buyer and heightens competition among potential suppliers. In this context, the importance of maintaining a stable and trustworthy level of supply capacity becomes paramount for suppliers. For instance, Apple mandates that many of its key suppliers keep two weeks of inventory within a mile of the company's assembly plants in Asia (Satariano and Burrows, 2011), while Volvo requires all potential suppliers to demonstrate capacity at each production stage through a formal assessment process (Volvo, 2010). Any potential supplier unable to demonstrate the required capacity when a market opportunity arises risks losing that opportunity. This underscores the significance of having a reliable supply capacity in a competitive market, particularly in light of global trends towards supply base reduction.

In this paper, we concentrate on scenarios where a supplier, acting as the decision-maker, manages a portfolio of production lines and strategically invests in capacity expansion, balancing the needs across multiple production lines. These production lines are designed to produce similar yet distinct products, a distinction that holds particular importance in newly industrialized economies. Here, manufacturers typically concentrate on a core product while diversifying into several subcategories to cater to a wider market demand. This variation in

* Correspondence to: 324 McCool Hall, Mississippi State, MS 39762, USA.
E-mail addresses: solak@isenberg.umass.edu (S. Solak), allen.li@wisc.edu (Z. Li), mehmet.gumus@mcgill.ca (M. Gumus), yzhuo@business.msstate.edu (Y. Zhuo).

https://doi.org/10.1016/j.ijpe.2024.109261
Received 8 December 2022; Received in revised form 5 April 2024; Accepted 30 April 2024
Available online 3 May 2024
0925-5273/

products indicates that, despite their similar profit margins, the demand dynamics and cost structures required to tailor these products to specific customer needs are significantly different.

An illustrative example is Murata, a Japan-based company manufacturing electronic components across 18 facilities in Malaysia, Taiwan, Thailand, and Vietnam. Murata's product portfolio includes modules used in electronic devices, telecom technology, mechatronics, and electrical sectors, serving major clients like Apple Inc and Samsung Electronics Co Ltd. During the global shortage of electronic parts, Murata faced the challenge of delicately balancing its supply capacity to meet orders from both Apple and Samsung, thus maintaining its market competitiveness. Similarly, PINACO, a leading battery manufacturer in Vietnam, offers a range of lead–acid batteries and lithium-ion batteries. The company has established strategic partnerships with top automobile and motorcycle manufacturers, navigating the complexities of supplying to both traditional and emerging markets, especially with the rise of electric vehicles.

As the world recovers from the pandemic, the global economy encounters a period marked by inflationary pressures and geopolitical risks, leading to a highly dynamic demand landscape for suppliers. The 3M Company exemplifies this shift in demand dynamics, with the pandemic increasing the demand for personal protective equipment (PPE) while decreasing the need for industrial materials. However, as the post-pandemic era unfolds and new norms are established, demand for automotive and construction industry products has surged. This fluctuating demand landscape complicates the decision-making process for capacity expansion. Suppliers must adeptly navigate these unpredictable market demands, understanding that capacity investment decisions have far-reaching implications over multiple time periods. This complexity suggests the necessity of a sophisticated modeling approach to budget allocation across production lines, cautioning against an oversimplified preference for ostensibly more efficient lines without considering the broader context of specific product demands and customization costs.

In this paper, we will explore the multifaceted decisions surrounding capacity expansion for manufacturing or service firms. Our investigation delves into *how a firm contends with stochastic demand, dynamically allocating resources within a portfolio of production lines for capacity expansion within the constraints of a given budget.* In constructing our stylized model, we adhere to certain simplifying assumptions to keep the framework tractable:

(1) We assume that capacity is reusable with proper maintenance and that the available capacity level determines whether a firm can bid for and win an order. Though our model is primarily geared towards manufacturing, it has potential extensions to service organizations such as consulting firms. In this context, capacity may be understood as the intellectual capability of a team to address certain client problems based on staffing levels.

(2) By "demand", we typically refer to a physical quantity that directly correlates with the capacity to supply products or services. However, this definition may be broadened to include nonphysical demand or other specific requirements tied to the firm's profit. An example of this would be when orders come with different quality requirements. Here, quality level can be treated as demand, and capacity as the ability to produce at a given quality level.

(3) We have normalized the lead time for the capacity expansion process to zero, since this time is generally predictable. With sufficient funding and a dedicated cross-functional team, companies can often streamline the process, reducing lead time further. Normalizing the lead time to zero in our model can be seen as an optimistic portrayal of these possibilities.

(4) Lastly, we assume that a customer order is lost if the firm's capacity is less than the amount of demand realized through a received order. This is a more complex scenario to model than the case with partial order fulfillment, which could be addressed through newsvendor-type models.

2. Review of relevant literature

The issue of resource allocation for capacity expansion primarily falls within the realm of strategic capacity management research. The typical modeling framework in this field involves making capacity investment decisions, such as sizes and timing, in the presence of stochastic demand information, and production decisions are made after observing demand realization (Van Mieghem, 2003). A common trend in the literature is to focus on two production lines or two decision periods, rather than considering a broader set of production lines. Examples of such studies, where the value of maintaining an optimal capacity portfolio is also discussed, include (Fine and Freund, 1990; Van Mieghem, 1998; Goyal and Netessine, 2007; Bassamboo et al., 2010). In these works, Fine and Freund (1990) outline the conditions for investing in flexible capacity, Van Mieghem (1998) develop a multi-dimensional newsvendor model, Goyal and Netessine (2007) introduce competition between firms, and Bassamboo et al. (2010) extend the problem to multiple production lines. These models offer valuable insights into short-term investment behaviors but are often limited to two-line and/or two-period frameworks.

In contrast, our study aims to identify optimal capacity investment policies for firms dealing with multiple production lines over various periods. This reflects the reality that capacity expansion often extends across a long planning horizon, with incremental additions in each period. Furthermore, our approach deviates from newsvendor-based models by assuming the loss of the entire order if firm capacity is less than the demand realized through a received order. This feature is especially relevant for strategic planning decisions among manufacturing firms, particularly those supplying large original equipment manufacturers. The resource allocation literature is highly relevant to the capacity investment issue, focusing on how to optimally distribute limited resources like capital, materials, and human resources to maximize benefits. Many such problems can be likened to different variations of the knapsack problem (Patriksson, 2008). Two areas of this literature particularly intersect with our research: the allocation of resources to multiple projects (Glasserman, 1996; Korhonen and Syrjanen, 2004) and dynamic allocation over several periods (Prastacos, 1981). Despite the relevance, few analytical models tackle both these aspects together, given their inherent complexity.

A significant exception is the work of Loch and Kavadias (2002), who constructed a multi-period model to explore new product development resource allocation. In this model, resources allocated are accumulated over time, with a fixed market size in each period and an increasing return function. Though insightful, this model is limited by its restrictive increasing return assumption, leaving questions about what might happen if this assumption were relaxed. Our paper extends this line of research by considering a return function defined according to the probability distribution of demand, encompassing both increasing and decreasing marginal returns. We also build upon the work of Loch and Kavadias (2002) by carrying over resources allocated to enhance the production line's current capacity level into the next period. We then formulate the return function based on demand distribution. Our model diverges further by demonstrating an optimal capacity level, a foundational element in deriving the best investment policies and budget-splitting rules under varying market conditions.

The sequential investment literature serves as another relevant research area, often following the general framework of option pricing models to address future uncertain demand (Grenadier and Weiss, 1997, Panayi and Trigeorgis, 1998, Kauffman and Li, 2005). Although these models focus on optimal investment strategy over time, they prove challenging to extend to a substantial portfolio of investment opportunities. A more flexible approach can be found in the multi-stage stochastic programming framework, allowing for more detailed considerations such as future budget uncertainty (Colvin and Maravelias, 2008, Solak et al., 2010). However, despite their ability to incorporate effects from multiple projects and decision periods, stochastic programming models are both analytically intractable and computationally demanding, particularly when dealing with continuous demand distributions.

In response to these challenges, our research identifies the existence of an optimal capacity level based on reasonable assumptions about demand distributions. By estimating this optimal capacity level, we define terminal capacity and return, thereby facilitating a tractable analytical study of the multi-production line, multi-period problem through backward induction. This approach otherwise would have been difficult, as it uncouples the terminal return from previous investment decisions. Additionally, we innovate a flexible approach to manage both known and unknown demand distributions, considering both stationary and non-stationary assumptions. To this end, we develop a dynamic programming based model to study resource allocation and capacity expansion decisions. Our analytical models illustrate the effects of resource competition among various capacity investment options across multiple decision periods. By explicitly modeling a *portfolio* of capacity investment projects, we can analyze the resource allocation problem and the cost-revenue trade-off. This yields a framework that supports optimization of a firm's investment opportunities and profitability (Cooper and Kleinschmidt, 1999, Beaujon et al., 2001). Furthermore, our attention to multiple decision periods enables a dynamic analysis of cost-revenue trade-offs across the planning horizon. This focus is vital in the context of capacity investment projects, which typically require significant capital and are implemented sequentially.

The remainder of this paper unfolds methodically, structured to guide readers through our progressive analyses and findings. In Section 3, we introduce our foundational model, elucidating the assumptions that underpin our research. Following this, Section 4 delves into the optimal capacity investment strategies, specifically addressing scenarios with known demand distributions. The analysis then expands in Section 5, where we consider the complexities of partially known demand distributions, adopting a Bayesian learning approach to provide nuanced insights. Further exploration occurs in Section 6, where we scrutinize the robustness of the Gamma Distribution-based solutions, examining their resilience to deviations from the Gamma Distribution assumption. Lastly, Section 7 succinctly summarizes our results, weaving together the various strands of our investigation, and outlines promising avenues for future research.

3. Base modeling framework setup for capacity expansion

We set up the model to capture three important characteristics for the capacity investment problem stated in Section 1: (1) A portfolio of n capacity investment projects, i.e. production lines, are competing for a limited resource; (2) demand on capacity for each production line is uncertain (initially stationary, and later extended to include nonstationarity); (3) capacity investment decisions are made sequentially over T decision periods. A complete nomenclature table is given in Online Supplemental Material Section A.

We denote a production line by $i = 1, 2, \ldots, n$ and a decision period by $t = 1, \ldots, T$. In this section, we consider a representative production line i. In period t, the accumulated capacity level s_{it} for production line i, which represents the state in (i.e. at the beginning of) period $t+1$, is defined as

$$s_{it} = s_{i0} + \sum_{\tau=1}^{t} x_{i\tau}, \tag{1}$$

where $x_{i\tau}$ is the amount of resource allocated to build production line i in period τ and s_{i0} is the initial capacity level. Note that resource is allocated at the beginning of each period but the newly invested capacity will be effective at the end of the period, i.e. there is a one-period lead time. The linear capacity accumulation function is a standard assumption in the capacity expansion literature (Luss, 1982, Campisi et al., 2001). While we do not specifically assume any units for invested resource amounts, it is natural to think of them in financial terms. However, it is possible to apply the analysis to any similar process as long as appropriate scaling is used for the corresponding parameters. A simplified event stream is illustrated in Fig. 1.

We assume that there is one order in each period, but the demand on capacity level imposed by the order is a random variable. This assumption captures several practical configurations. First, the case of one order per period can be directly applicable in some settings. For example, Cohen et al. (1999) note that the demand for a specific electronic testing equipment manufacturer is such that on average one customer order is received every three to five months, which may also be the frequency for capacity expansion decisions. Such a case might typically apply to industries with only a few dominant clients, or to firms providing products or services through certain government procurement processes. A second case of application is when demand and capacity respectively represent non-physical requirement and capability levels, and in each period there is a continuum of market opportunities each of which has a certain requirement level. In such a case, the demand distribution can be characterized based on the number of clients with a certain requirement level. The firm can only serve customers whose requirement level (demand) is below the capability (capacity) of the firm. Note that such a case can also be applicable to quantity based demand situations, if demand quantities are categorized into distinct levels, and the demand distribution is defined based on the number of clients with order sizes corresponding to each quantity level. A third, possibly less applicable case is when the firm faces multiple orders of the same amount, i.e. loss of one order due to lack of capacity would imply loss of all other orders.

Consistent with Dieulle et al. (2003) and Hsu et al. (2008), we assume that market demand or capacity requirement r for product i, denoted by R_i, follows a gamma distribution with density function f_{R_i} with shape parameter α_i and rate parameter (also called inverse scale parameter) φ_i.[1] Note that this stationarity assumption implies that the demand is independent and identically distributed in each period, which may not be applicable for certain cases. We study this case as our basic model, and then extend our analysis to include the nonstationary case, where the values of demand distribution can change over time, i.e where dependency modeled through trending is captured. Moreover, further dependency issues are considered in Section 5 where partial demand information is used under a Bayesian setting. It is also important to note that the gamma family of distributions is appealing here since it is a flexible family on $[0, +\infty)$. The shape parameter and the rate parameter together can approximate a large number of demand distributions. Some well-known distributions such as the normal, exponential, chi-square and beta distributions can be seen as special cases of the gamma family and can be easily transformed to a gamma distribution.[2] In Section 6, we perform some numerical experiments to explore the robustness of the gamma distribution based solutions in our framework to departures from the gamma distribution.

We also assume that the value of received orders linearly correlates with their capacity requirements, denoted as r, and is characterized by a unit profit margin m_i for product line i. For simplification, the profit margin m_i is normalized to 1 across all production lines. This normalization reflects our assumption that the production lines manufacture similar yet distinct products to cater to different market segments. We note that a deterministic functional relationship between demand, i.e. the size of an order, and return is typical in many settings, as the firms will mostly have contracts specifying fixed cost-benefit structures per unit of product supplied or service provided. In the case where there is significant uncertainty due to negotiations, which may not be

[1] As shown in Lemma 1, the gamma distribution assumption can be relaxed for most cases, making our model applicable to a wider range of distributions.

[2] Leemis (1986) provides the transformations. For example, exponential distribution is a special case of the gamma distribution with $\alpha = 1$; $Normal(\mu, \sigma^2)$ is related to the gamma distribution by: $\mu = \frac{\alpha}{\varphi}, \sigma^2 = \frac{\alpha}{\varphi^2}$, where α is large.

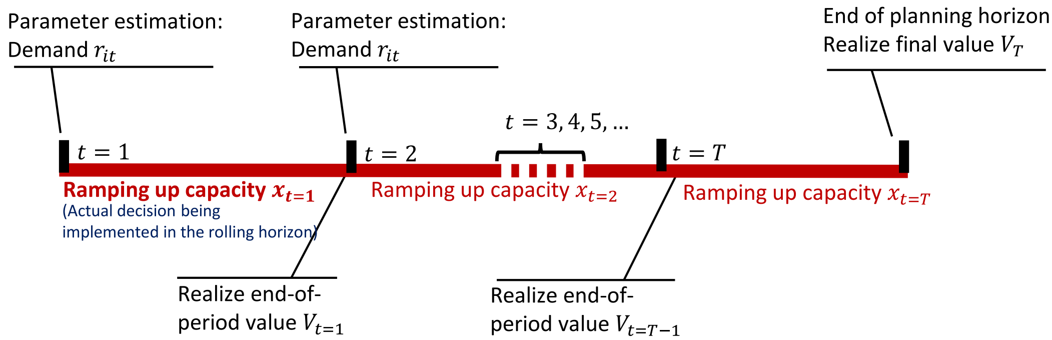


Fig. 1. Event stream of capacity expansion investment process Decision-makers utilize only the first period's decision, and the model is subsequently executed in a rolling horizon manner.

typical for most firms, this functional relationship between demand and return can be assumed to be based on expected rates. Under these assumptions, the expected gross return $g_{R_i}(s_i)$ from a product line i with a capacity level s_i is

$$g_{R_i}(s_i) = \int_0^{s_i} m_i r f_{R_i}(r)\, dr. \tag{2}$$

The return function states that if the capacity level required by a market opportunity, i.e. r, is lower than the capacity level the firm possesses, then the firm gets the order and receives the corresponding return r. Otherwise, the firm loses this market opportunity as a whole. As noted previously in the introduction section where we discuss practical settings for such an assumption, this framework also makes the model distinct from newsvendor-based models where only excess portion of demand is assumed to be lost. This distinction is important, because the latter case typically results in well-behaved concave return functions. On the other hand, the return functions in our model do not reveal simple concavity properties as shown in Lemma 1 below. We illustrate this difference in detail both analytically and numerically in Online Supplemental MateriL Section C.

Note that we use the subscript R_i for the probability density function f and the return function g when we do not assume any specific probability distribution for demand of product i. But when we refer to a specific distribution, we replace the subscript R_i by the parameter(s) of the distribution we are assuming. Using this convention, we state the properties of the gross return function for production line i as follows.

Lemma 1. *For production line i, the expected gross return $g_{R_i}(s_i)$ has the following properties:*

(1) For any probability distribution, $g_{R_i}(s_i)$ is increasing in the interval $[0, +\infty)$;

(2) Let θ_{i0} be defined such that $f_{R_i}(s_i)+sf'_{R_i}(s_i)$ is positive in the interval $[0, \theta_{i0}]$ and negative in the interval $[\theta_{i0}, +\infty)$. Then g_{R_i} is convex in the interval $[0, \theta_{i0}]$ and concave in the interval $[\theta_{i0}, +\infty)$; where for the gamma distribution with shape parameter α_i and rate parameter φ_i, $\theta_{i0} = \frac{\alpha_i}{\varphi_i}$ and for the exponential distribution with parameter λ_i, $\theta_{i0} = \frac{1}{\lambda_i}$.

Proof. Proof. All proofs are in the x. A. □

In Fig. 2 we illustrate two examples of the return and marginal return functions based on different probability distributions. In Fig. 2(a), corresponding functions for the $Gamma(2, 2)$ distribution are shown, where the probability density function f_{α_i,φ_i} is increasing in the interval $[0, 2]$ and decreasing in the interval $[2, +\infty)$. It can be observed that the return function g_{α_i,φ_i} is convex increasing in the interval $[0, 4]$ and concave increasing in the interval $[4, +\infty)$. Similar characteristics are displayed in Fig. 2(b) for the $Exponential(1)$ distribution, where the probability density function f_{λ_i} is decreasing in its support $[0, +\infty)$. In this case, the return function g_{λ_i} is convex increasing in the interval $[0, 1]$ and concave increasing in the interval $[1, +\infty)$.

In this problem, the decision variable is the capacity expansion level x_{it}. In each period $t \le T$, the decision maker observes the current capacity level $s_{i(t-1)}$, and identifies the expansion level x_{it} that maximizes the total expected value over the remaining periods. Here, $s_{i(t-1)}$ represents the state at period t, and the x_{it} corresponds to an action in that period. The state transition is defined by $s_{it} = s_{i(t-1)} + x_{it}$. The immediate reward for taking action x_{it} is given by $-c_i x_{it} - h_i s_{it} + \beta g_{R_i}(s_{it})$, where c_i, h_i, and β correspond to the unit investment cost, the unit holding cost and the discount rate, respectively. Investment cost is the one time acquisition cost for the resources. It consists of a variety of one time expenditure, such as facility acquisition and installation costs, workforce recruitment costs, and training costs. Consistent with the real options literature, we assume that resources are firm-specific and investment is irreversible (Dixit and Pindyck, 1994). Holding cost is the daily maintenance cost which includes maintenance costs associated with the technical systems, wages paid to the workforce, and other administrative costs. We assume that c_i and h_i do not change over time. Therefore, our model is more applicable to the cases where capacity acquisition and maintenance costs are stable. The discount rate β, $0 \le \beta \le 1$, measures the present value of a unit of future income. The β on the left of the return function g_{R_i} represents a one-period lag for a newly built capacity to be effective. The time lag might be the length of time required for a newly installed production system to operate under normal conditions, or might be the length of recruitment time plus the training time required for the workforce to acquire a new skill or enhance their current skill set. Given this problem setting, the expected total profit $V_{it}(s_{i(t-1)})$ for product i in period $t \le T$ can be expressed through the following dynamic programming recursion:

$$V_{it}(s_{i(t-1)}) = \max_{x_{it}\ge 0}\{-c_i x_{it} - h_i s_{it} + \beta g_{R_i}(s_{it}) + \beta V_{i(t+1)}(s_{it})\}. \tag{3}$$

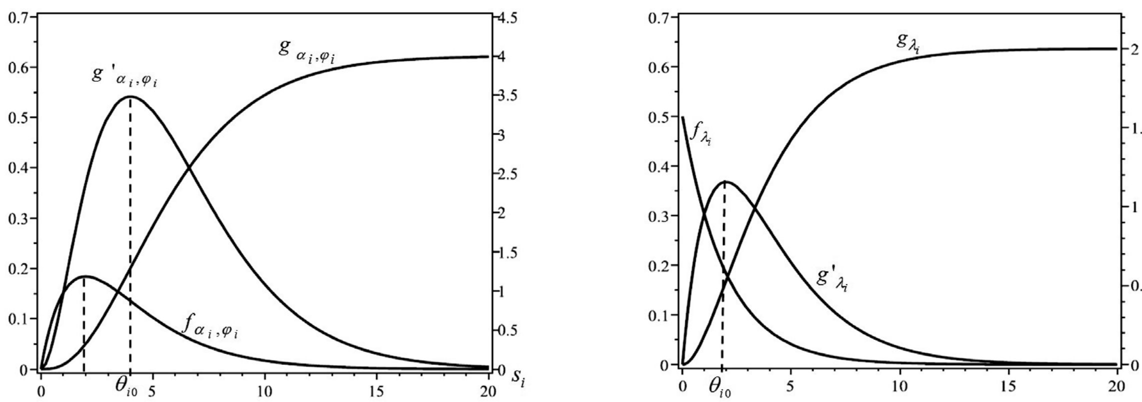
The term $\beta V_{i(t+1)}(s_{it})$ in Eq. (3) is the discounted value function for the next period $t+1$. The value function $V_{i(T+1)}(s_{iT})$ for the final capacity level is:

$$V_{i(T+1)}(s_{iT}) = \sum_{t=T+1}^{\infty} \beta^{t-1}\left[-h_i s_{iT} + \beta g_{R_i}(s_{iT})\right] = \frac{\beta^T\left[\beta g_{R_i}(s_{iT}) - h_i s_{iT}\right]}{1-\beta} \tag{4}$$

is based on the assumption that the firm continues to receive returns and pay holding costs over an infinite time horizon after reaching the terminal capacity level. From period $T+1$ on, there is no further investment and the capacity level remains stable. Thus, this term corresponds to the salvage value of the terminal capacity level s_{iT}.

Eventually, by combining the dynamic programming equations (3) and (4), we can obtain the following expression for the expected total profit at the beginning of the planning period when the available capacity is s_{i0}. We refer to this expression in several of the discussions and proofs in the paper.

$$V_{i1}(s_{i0}) = \max_{x_{i1},\ldots,x_{iT}\ge 0}\left\{\sum_{t=1}^{T}\beta^{t-1}\left[-c_i x_{it} - h_i s_{it} + \beta g_{R_i}(s_{it})\right]\right.$$



(a) Based on $Gamma(2,2)$ (b) Based on $Exponential(1)$

Fig. 2. Return and marginal return functions based on different distributions.

$$+\frac{\beta^T\left[\beta g_{R_i}(s_{iT})-h_i s_{iT}\right]}{1-\beta}\Bigg\} \tag{5}$$

The amount of resource allocated to capacity expansion for each production line is constrained by the total available budget. Based on the firm's budget allocation strategy, the budget constraints may have two distinctive forms: a pooled budget constraint for the whole planning horizon, or a periodic budget constraint for each decision period. We study these two cases separately. Note that the words *pooled* and *periodic* refer to the budget structure among different time periods, rather than among production lines in the portfolio, which always share the available budget.

Before we describe our optimal policy analyses under the two types of budget constraints, it is important to note the following: without a budget constraint, the optimal investment strategy for each production line i in the region with the convex increasing return function would be to invest as much as possible, at least to reach the critical point θ_{i0} where the marginal return is the maximum. This result is intuitive and consistent with basic results of the studies of convex increasing return functions (Loch and Kavadias, 2002). On the other hand, if the firm operates in the interval $[\theta_{i0}, +\infty)$, where the marginal return is strictly decreasing, there exists an optimal capacity level where the marginal cost equals the marginal return, as stated in the following lemma.

Lemma 2. *Suppose that $s_{i0} \geq \theta_{i0}$ for production line i. If no budget constraints exist, the optimal capacity level s_i^* for production line i is the unique solution to the following equation:*

$$s_i f_{R_i}(s_i) = \frac{(1-\beta)c_i + h_i}{\beta m_i}, \quad s_i \geq s_{i0}. \tag{6}$$

If there is no solution to Eq. (6), then $s_i^* = s_{i0}$, i.e. the firm should not invest in any additional capacity.

The existence of an optimal capacity level is not surprising. The marginal cost of a unit of investment is constant but the marginal return is strictly decreasing. Therefore, there exists a unique positive solution to the optimality Eq. (6) in the interval $[s_0^i, +\infty)$. The optimal capacity level is determined by the unit investment cost, the unit holding cost and the probability distribution of market demand. More specifically, it has the following relationships with the relevant parameters.

Corollary 1. *The optimal capacity level s_i^* is increasing in β, decreasing in c_i, and decreasing in h_i.*

If the unit investment cost is low, then it is optimal to invest more and reach a high capacity level which enables the firm to capture more market demand. Cost efficiency allows the firm to maintain a high level of capacity. The effect of the unit holding cost is similar. An increase in the discount factor, on the other hand, would result in an increase in the marginal returns, which may offset the impact of the decreasing marginal return structure in the interval $[s_{i0}, +\infty)$. We note that the existence of an optimal capacity level has significant value from an analysis perspective, as it allows for a backward induction based solution approach to the overall problem.

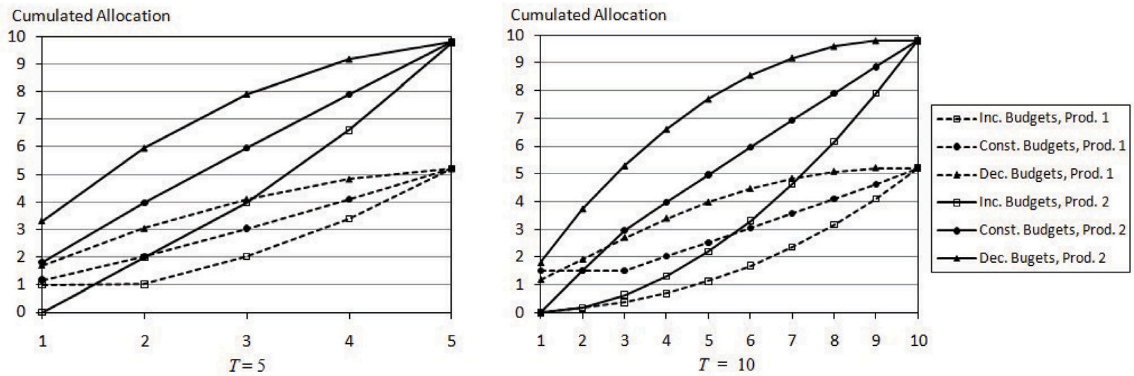
4. Capacity expansion with known demand distributions

This analysis assumes the capacity expansion budget allocation decisions are made under known demand distributions. The investment strategy is applied to a portfolio of production lines, where the firm has perfect knowledge about the probability distributions of demand over the planning horizon. We consider two cases regarding the status of the demands for all production lines: (1) stationary demand over time and (2) varying demand – i.e. non-stationary demand – for each time period. To this end, we adopt an analytical analysis approach aiming to derive structural properties for easy decision-making and to avoid the curse of dimensionality in computational solutions.

4.1. Stationary demand case

In our first model, we work under the premise of known and stationary demand distributions, implying that the demand in each period follows an identical distribution. Common products fitting this profile are baby and office/home supplies in stable markets. While this scenario may seem simplistic, it is anchored in prior literature, as referenced in works like Altintas et al. (2008). Shifting our gaze to global supply chain dynamics, we have observed that certain mature markets with consistent demand patterns are undergoing significant changes in their production bases. For instance, the weight training equipment industry's transition from China to Mexico or the shift in the shoe industry to regions like Vietnam and other Southeast Asian nations. Given these real-world shifts, we recognize the pertinence of capacity expansion in the context of stationary demand. This is particularly relevant for emerging suppliers in the Newly Industrialized Economies (NIE). Consequently, we have dedicated a section of our paper to shed light on this prevailing situation.

For the scope of this subsection, our analysis of stationary demand will be bifurcated into two scenarios. Initially, we delve into the periodic budget constraint, following which we aggregate the budget constraints across periods into a unified constraint. This approach not only provides insights into a specific case but also offers a perspective on flexible budgetary scenarios. By starting with the stationary demand scenario, we intend to elucidate the optimal policy for resource allocation among multiple production lines. This foundational understanding will later pave the way for more intricate analyses in subsequent sections of this paper.



(a) The increasing/constant/decreasing budget sequences are $\{B_t\} = \{1, 2, ..., 5\}$, $\{B_t\} = \{3, 3, ..., 3\}$, and $\{B_t\} = \{5, 4, ..., 1\}$, respectively

(b) The increasing/constant/decreasing budget sequences are $\{B_t\} = \{0, \frac{1}{3}, \frac{2}{3}, ..., \frac{9}{3}\}$, $\{B_t\} = \{\frac{3}{2}, \frac{3}{2}, ..., \frac{3}{2}\}$, and $\{B_t\} = \{\frac{9}{3}, \frac{8}{3}, ..., 0\}$, respectively

Fig. 3. Optimal splitting of period budgets under exponentially distributed market demand ($\lambda_1 = 1.0, \lambda_2 = 0.5$, and $\beta = 0.8$).

4.1.1. Resource allocation to a portfolio of production lines under periodic budget constraints

In the periodic budget constraints case, budget limitations exist for each period. For example, the firm annually reviews the ongoing projects and makes budget allocation decisions for period t based on the available budget B_t such that

$$\beta^{t-1} \sum_{i=1}^{n} c_i x_{it} \leq B_t, \forall t = 1, \dots, T \tag{7}$$

Note that the discounting term can be omitted in the constraint above by appropriately adjusting the budget B_t. We assume such an adjusted budget in the discussions below, and omit the discounting term. we state the optimal budget splitting policies for the T-period problem with two production lines as follows. Extension of the proposition to the general case with n production lines is through induction, and is included in Online Supplemental Material Section C.

Proposition 1. *Assume that $\sum_{t=1}^{T} B_t < \sum_{i=1}^{n} c_i(s_i^* - s_{i0})$. For the T-period 2-production line capacity investment problem with periodic budget constraints, it is optimal in every period to split the budget between two production lines according to their marginal profits, i.e. the optimal investment amount in each production line is determined by the following system of two equations:*

For period T:

$$-c_1 - \frac{h_1}{1-\beta} + \frac{\beta g'_{R_1}(s_{10} + \sum_{\tau=1}^{T} x_{1\tau})}{1-\beta} = -c_2 - \frac{h_2}{1-\beta} + \frac{\beta g'_{R_2}(s_{20} + \sum_{\tau=1}^{T} x_{2\tau})}{1-\beta}, \tag{8}$$

and for period $t, t = 1, 2, \dots, T-1$:

$$\begin{aligned} &-c_1 - \frac{h_1}{1-\beta} + \beta \sum_{\omega=0}^{T-t} \beta^{\omega} g'_{R_1}(s_{10} + \sum_{\tau=1}^{t+\omega} x_{1\tau}) + \frac{\beta^{T-t+1} g'_{R_1}(s_{10} + \sum_{\tau=1}^{T} x_{1\tau})}{1-\beta} \\ &= -c_2 - \frac{h_2}{1-\beta} + \beta \sum_{\omega=0}^{T-t} \beta^{\omega} g'_{R_2}(s_{20} + \sum_{\tau=1}^{t+\omega} x_{2\tau}) + \frac{\beta^{T-t+1} g'_{R_2}(s_{20} + \sum_{\tau=1}^{T} x_{2\tau})}{1-\beta} \end{aligned} \tag{9}$$

where $x_{2\tau} = \frac{B_\tau - c_1 x_{1\tau}}{c_2}$. If no solution to the above equations exist, then all the budget B_t is allocated to the production line corresponding to the dominant side.

Eq. (8) above determines the optimal splitting of the budget in the last period, given the splitting of budget in the previous periods. In each period $t, t = 1, 2, \dots, T-1$, budget is allocated to the two production lines according to their marginal profits which consider both the current period profit and the profit in the following periods from an additional unit of investment.

In Fig. 3, we illustrate Proposition 1 through a numerical example with the same cost and distribution parameters used in Fig. 4(a), and with two different lengths of planning horizon $T = 5$ and $T = 10$. We consider three budget sequences ${B_t^{Inc}}_{t=1}^{T}$, ${B_t^{Const}}_{t=1}^{T}$, and ${B_t^{Dec}}_{t=1}^{T}$ with equal sums such that $\sum_{t=1}^{T} B_t = 15$ for all three sequences. The numeric values for the three sequences are listed in the figure captions, and they correspond to increasing, constant and decreasing budget conditions, respectively.

Myopic behaviors are also observed in this numeric example. It can be seen in Fig. 3(a) that when the budget is initially small, i.e. for $B_1^{Inc} = 1$, production line 1, which has lower investment and holding costs, receives the whole budget ($x_{11}^* = 1$). On the other hand, with a relatively large budget, i.e. for $B_1^{Dec} = 5$, production line 2, which has higher gross return rate, receives more than half of the budget $x_{21}^* = 3.31$. Overall, although a biased split that is more favorable to higher return rates on average can be observed, the actual allocations in a given period may have the opposite structure. Interestingly, in both Figs. 3(a) and 3(b), the cumulated investment amounts for the two production lines are independent of the length of planning horizon and the budget arrangements, given that the total budget $\sum_{t=1}^{T} B_t$ is the same.

Not shown in Fig. 3 is the total expected value for each budget scenario. For $T = 5$, the firm's expected profits under increasing, constant, and decreasing budget sequences can be calculated as 8.32, 9.30, and 10.13, respectively. For $T = 10$, the corresponding profits are 5.82, 7.98, and 9.17. These examples demonstrate that under increasing budget sequences, the firm may not receive sufficient budget in the early periods of the planning horizon, which reduces the firm's expected profits. This negative impact of insufficient budget is more serious if the planning horizon is longer.

4.1.2. Resource allocation to a portfolio of production lines under a single pooled budget constraint

In the pooled budget case, there is a single budget constraint for the portfolio throughout the T-period planning horizon, which can be expressed as:

$$\sum_{t=1}^{T} \beta^{t-1} \sum_{i=1}^{n} c_i x_{it} \leq B \tag{10}$$

where the budget B is defined in terms of its present value. Under a typical pooled budget strategy, budget can be reserved for a medium or long-term capacity expansion project to guarantee that the project will not be suspended before completion due to inadequate funding, as all stages of the project cycle share a single budget pool B. This budget strategy gives managers the flexibility to optimally manage the budget throughout the project cycle.

As discussed in Section 3, the unit investment and holding costs are both assumed to be constant. Furthermore, for any production line i, the return function is convex increasing in the interval $[0, \theta_{i0}]$ and concave increasing in the interval $[\theta_{i0}, +\infty)$. Therefore, the marginal return g'_{R_i} is increasing in the interval $[0, \theta_{i0}]$ and decreasing in the interval $[\theta_{i0}, +\infty)$.

We demonstrate our analysis by initially considering two productions lines 1 and 2, and then by extending the results to cases with $n > 2$ production lines. Clearly, if the available budget exceeds the sum of the amount of budget that the two production lines need to reach their optimal capacity levels, there will be no competition for budget between the two production lines, and the optimal investment policy is trivial. We focus on the case where there is not enough budget to satisfy the whole requirement from any production line, i.e. $B < c_1(s_1^* - s_{10}) + c_2(s_2^* - s_{20})$, where s_i^* and s_{i0} are the optimal and initial capacity levels for production line i, respectively. The optimal policy for splitting the budget under this condition is as follows:

Proposition 2. *Assume that $B < c_1(s_1^* - s_{10}) + c_2(s_2^* - s_{20})$. For the T-period 2-production line capacity investment problem with a pooled budget constraint, there is positive investment only in the first period, i.e., $x_{1t}^* = x_{2t}^* = 0, \forall t = 2, 3, \dots, T$. In the first period, it is optimal to split the budget between the two production lines according to their marginal profits, i.e. x_{11}^* is the solution to the following equation:*

$$-c_1 - \frac{h_1}{1-\beta} + \frac{\beta g'_{R_1}(s_{10} + x_{11})}{1-\beta} = -c_2 - \frac{h_2}{1-\beta} + \frac{\beta g'_{R_2}(s_{20} + x_{21})}{1-\beta}, \quad (11)$$

where $x_{21} = \frac{B - c_1 x_{11}}{c_2}$. If no solution to the above equation exists, i.e. one side of the equation is always larger than the other for all $0 \le x_{it} \le \frac{B}{c_i}$, then all the budget is allocated to the production line corresponding to the dominant side.

Extension of Proposition 2 to n production lines is in the Supplemental Material. Given this structure, the following properties also hold for the optimal investment decisions.

Corollary 2. *For two production lines 1 and 2, assume that $B < c_1(s_1^* - s_{10}) + c_2(s_2^* - s_{20})$. For the T-period portfolio investment problem with a pooled budget constraint, the portion of budget allocated to production line 1 has the following properties:*

(1) x_{11}^* is increasing in c_2 and h_2, while decreasing in c_1 and h_1;

(2) If f_{R_1} and f_{R_2} are exponential distributions with parameters λ_1 and λ_2, then x_{11}^* is increasing in λ_2 while decreasing in λ_1.

The comparative statistics for x_{21}^* are similar.

In Fig. 4, we illustrate Proposition 2 and Corollary 2 with a numerical example using two exponential market demand distributions with rate parameters 1 and 0.5. For the parameter values listed in the caption of Fig. 4(a), the optimal capacity levels can be calculated using the results in Lemma 2 as $(s_1^*, s_2^*) = (7.62, 13.62)$. Similarly, for the parameter values listed in the caption of Fig. 4(c), the optimal capacity levels are $(s_1^*, s_2^*) = (7.62, 16.82)$. As shown in Figs. 4(a) and 4(c), the allocation to the production lines increase in the available budget B before the firm reaches the optimal capacity level, but the rate of increase for production line 2 is generally larger than that of production line 1. Note that the difference between the allocations to production lines 1 and 2, i.e. $x_{21}^* - x_{11}^*$ is increasing in B.

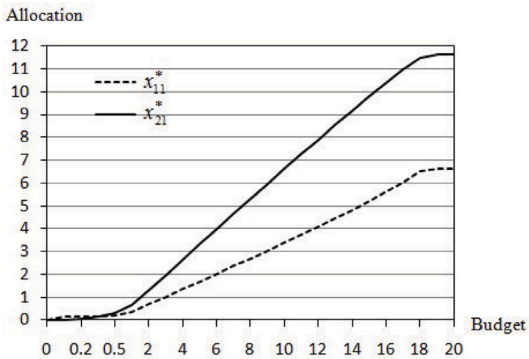
Moreover, Figs. 4(b) and 4(d) provide information about how an additional unit of budget is split. Cost advantage determines which production line has priority for budget allocation if the available budget is limited. However, marginal returns play the dominant role in the allocation of a relatively large budget. In Fig. 4(b), the investment and holding costs for production line 2 are twice as large as those for production line 1. If the available budget $B \le 0.15$, then all the budget is allocated to production line 1 due to its cost advantage. However, as the available budget increases, the amount of allocation to production line 2 increases faster than that of production line 1. This is due to the fact that as line 1's capacity level increases, its marginal return drops faster.

Results from Figs. 4(a) and 4(b) have relevant practical implications. Since production lines with higher marginal returns are typically associated with higher investment and holding costs, a firm with limited budget tends to invest only in the production lines with lower investment and holding costs, although their marginal returns are lower. However, a firm with relatively sufficient budget tends to invest more in production lines with higher marginal returns, although the investment and holding costs are much higher. Therefore, budget availability not only changes the absolute level of allocation to each production line in the portfolio, but also changes the structure, i.e. the relative levels of allocations, within the portfolio.

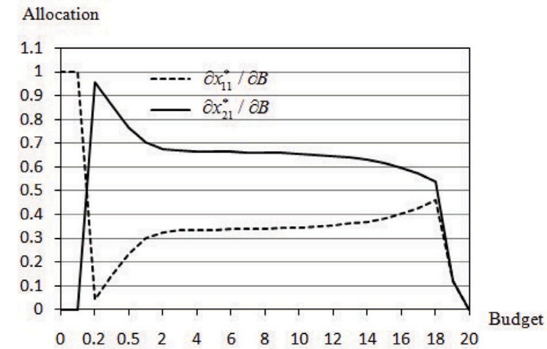
In Fig. 4(d), the investment and holding costs for production line 2 are half the costs for production line 1. Since the marginal profit of production line 2 is always higher than that of line 1 for any additional amount of budget, production line 2 will receive more than half of the budget. If the available budget $B \le 0.21$, then all the budget is allocated to production line 2 due to its cost advantage. As the available budget increases, the amount of allocation to production line 2 continues to increase faster than that of line 1, because the marginal return of the former drops at a slower rate.

Corollary 2 posits that demand and costs are pivotal in determining budget allocation among production lines. However, the situation becomes complex when a production line presents both high demand and elevated costs, leading to potentially conflicting preferences. To dissect this intricacy and elucidate the impact of high costs in the context of high demand, we conduct a sensitivity analysis. This analysis delineates the optimal capacity levels for a single production line – referencing Lemma 2 – across diverse demand and cost scenarios. Fig. 5 displays the outcomes, where the optimal capacity levels under a relaxed budget constraint are juxtaposed across various cost contours for different demand levels. Through this analysis, we identify specific conditions under which a high-demand production line may display a reduced optimal capacity due to escalated costs. For instance, point A – mirroring production line 1's configuration from Fig. 4 with parameters $c_1 = 0.01$, $h_1 = 0.001$, and $\lambda_1 = 1$ – achieves a higher unconstrained optimal capacity compared to point B. Point B, despite sharing the same demand parameter ($\lambda = 0.5$) as production line 2 from Fig. 4, incurs substantially higher costs ($c_2 = 0.3$, $h_2 = 0.03$), leading to a diminished capacity.

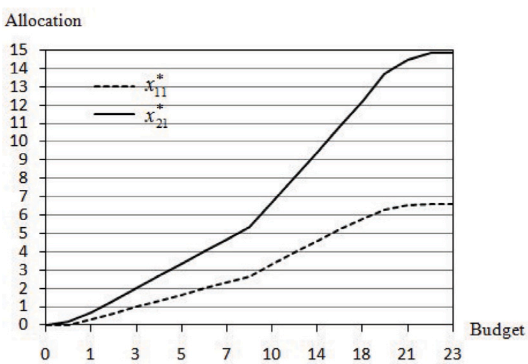
Fig. 6 suggests that within an unconstrained budget context, a high-demand production line facing significantly higher costs may become less preferred compared to its lower-demand, yet more cost-effective, counterpart. To further investigate, we examine two scenarios approximating the conditions at points A and B from Fig. 5, with configurations: $c_1 = 0.01, h_1 = 0.001, \lambda_1 = 1$ and $c_2 = 0.3, h_2 = 0.03, \lambda_2 = 0.5$. These scenarios illustrate the transition points where production line 1 begins to gain favorability under less stringent budget constraints. As depicted in Figs. 6(a) and 6(b), despite production line 1 attaining a higher capacity at the abundant end of the budget spectrum, production line 2 remains preferred under tighter budget constraints for the majority of the range. This preference also holds for a segment of budget values in Fig. 6(b), even when production line 2 becomes more expensive ($c_2 = 0.5, h_2 = 0.05$). From these observations, it is implied that the cost factor's role is intertwined with the influence of demand when overturning preferences between two production lines. Notably, under moderate budget restrictions, the impact of demand becomes more pronounced, whereas, in scenarios of relaxed budget constraints, the cost factor predominates in determining the favored production line for resource allocation.



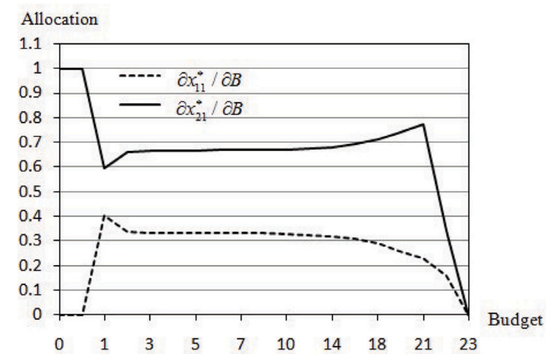
(a) $c_1 = 0.01, h_1 = 0.001, c_2 = 0.02, h_2 = 0.002$



(b) Split of an additional unit of budget for the example in Figure 4(a)



(c) $c_1 = 0.01, h_1 = 0.001, c_2 = 0.005, h_2 = 0.0005$



(d) Split of an additional unit of budget for the example in Figure 4(c)

Fig. 4. Optimal splitting of a pooled budget under exponentially distributed market demand ($\lambda_1 = 1.0, \lambda_2 = 0.5$, and $\beta = 0.8$).

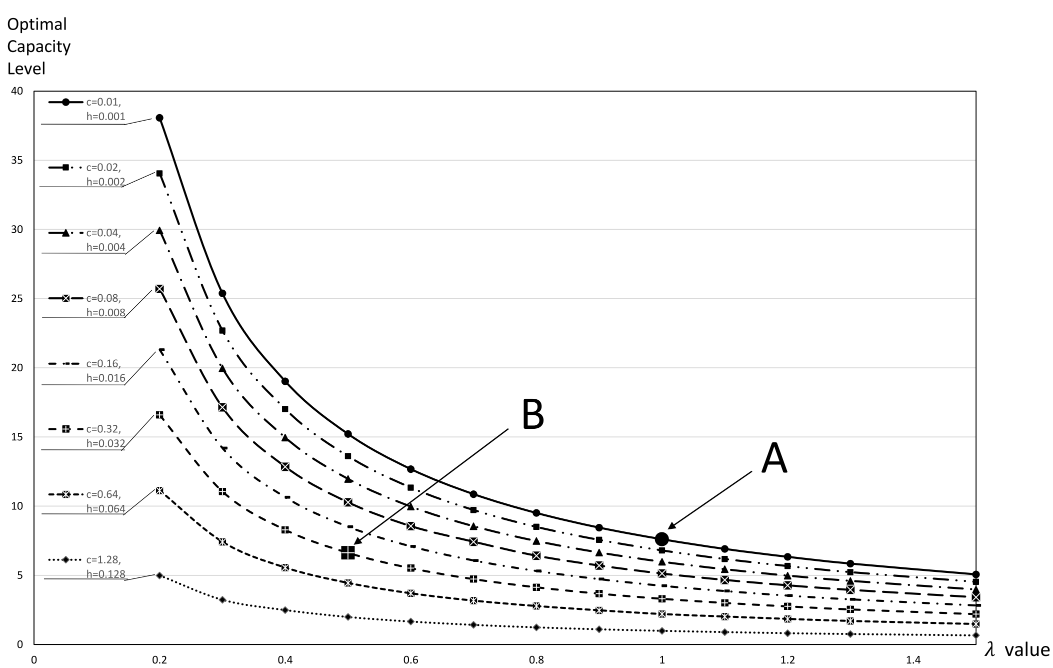
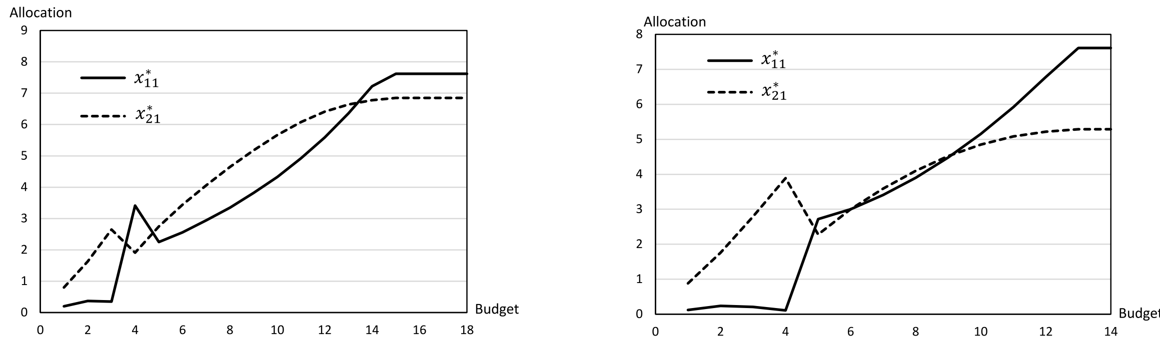


Fig. 5. Optimal capacity levels across various cost for different demand levels.

4.2. Nonstationary demand case

In this section we relax our assumption of stable markets with known identical demand distributions in each period, and model the case in which parameters of the demand distributions may change over time. Due to analytical tractability, we restrict our attention to cases where the parameters display a monotonically decreasing/increasing trend, corresponding to growing/shrinking market conditions. Within this framework, we assume that the demand distributions beyond the planning horizon, i.e. for $t > T$, are the same as those for period T. This



(a) $c_1 = 0.01, h_1 = 0.001, c_2 = 0.3, h_2 = 0.03$ (b) $c_1 = 0.01, h_1 = 0.001, c_2 = 0.5, h_2 = 0.052$

Fig. 6. Optimal splitting of a pooled budget under exponentially distributed market demand ($\lambda_1 = 1.0, \lambda_2 = 0.5$, and $\beta = 0.8$).

assumption is not very restrictive, as the planning horizon can simply be extended to capture all periods that demand distribution information is available.

For the nonstationary demand case, we only analyze the case of periodic budget constraints introduced in Section 4.1.1. This is because when demand distributions are nonstationary, there does not exist a single optimal capacity level throughout the planning horizon and a single pooled budget constraint is not practically as interesting. First, with a little abuse of notation, we denote $f_{R_{it}}$, $t = 1, 2, \dots, T$ as the probability density function of demand and $g_{R_{it}}$ as the return function in period t. Similar to the discussions above, we base our analysis on demand distributions that are in the gamma family. We assume that the shape parameters α_i are fixed, while the rate parameters φ_i can change over time, i.e. demand for production line i in period t is distributed as $Gamma(\alpha_i, \varphi_{it})$. This assumption is also not very restrictive, as the rate parameter impacts both the mean and variance of the gamma distribution. More specifically, the mean and variance are given as $\frac{\alpha_i}{\varphi_{it}}$ and $\frac{\alpha_i}{\varphi_{it}^2}$, respectively. For analytical exposition, suppose that α_i is a positive integer. Then the gamma function has a closed-form expression. Now let the rate parameter φ_{it} in each period be defined by a trend function such that $\varphi_{it} = \psi_i(t)$, where $\psi_i(t)$ characterizes how the rate parameters change over time. The function $\psi_i(t)$ can represent a monotonically decreasing/increasing trend as evidenced in a growing/shrinking market discussed above.

The structural properties of the above non-stationary demand setup suggests in an optimal capacity level for each product line in each period. Moreover, When the demand distribution is defined by $Gamma(\alpha_i, \psi_i(t))$, one sufficient condition for positive investment in period t is that the rate parameter of distribution of the period t is smaller than that of the period $t-1$, i.e. $\psi_i(t) < \psi_i(t-1)$. Hence, when the rate parameters $\psi_i(t)$ reveal a downward trend (notice that the mean of gamma distributed random variables is proportional to $\frac{1}{\psi_i(t)}$), there will be positive capacity expansion in each period.

With these properties, we can derive the optimal periodic budget splitting policies for two production lines under known nonstationary demand as follows.

Proposition 3. *Assume that $B_t < c_1(s^*_{1t} - s^*_{1(t-1)}) + c_2(s^*_{2t} - s^*_{2(t-1)})$, and $\psi_i(t) < \psi_i(t-1)$ for all $i = 1, 2$ and $t = 1, 2, \dots, T$. For the T-period 2-production line capacity investment problem with periodic budget constraints and known nonstationary demand, it is optimal in every period to split the budget between two production lines according to their marginal profits, i.e. the optimal investment amount to each production line is determined by the following system of two equations:*

For period T:

$$- c_1 - \frac{h_1}{1-\beta} + \frac{\beta m_1 f_{\alpha_1+1,\psi_1(T)}(s_{10} + \sum_{\tau=1}^{T} x_{1\tau})}{1-\beta}$$
$$= -c_2 - \frac{h_2}{1-\beta} + \frac{\beta m_2 f_{\alpha_2+1,\psi_2(t)}(s_{20} + \sum_{\tau=1}^{T} x_{2\tau})}{1-\beta}, \tag{12}$$

and for periods $t < T$:

$$-c_1 - \frac{h_1}{1-\beta} + \beta m_1 \sum_{\omega=0}^{T-t} \beta^{\omega} f_{\alpha_1+1,\psi_1(t+\omega)}(s_{10} + \sum_{\tau=1}^{t+\omega} x_{1\tau})$$
$$+ \frac{\beta^{T-t+1} m_1 f_{\alpha_1+1,\psi_1(T)}(s_{10} + \sum_{\tau=1}^{T} x_{1\tau})}{1-\beta}$$
$$= -c_2 - \frac{h_2}{1-\beta} + \beta m_2 \sum_{\omega=0}^{T-t} \beta^{\omega} f_{\alpha_2+1,\psi_2(t+\omega)}(s_{20} + \sum_{\tau=1}^{t+\omega} x_{2\tau})$$
$$+ \frac{\beta^{T-t+1} m_2 f_{\alpha_2+1,\psi_2(T)}(s_{20} + \sum_{\tau=1}^{T} x_{2\tau})}{1-\beta} \tag{13}$$

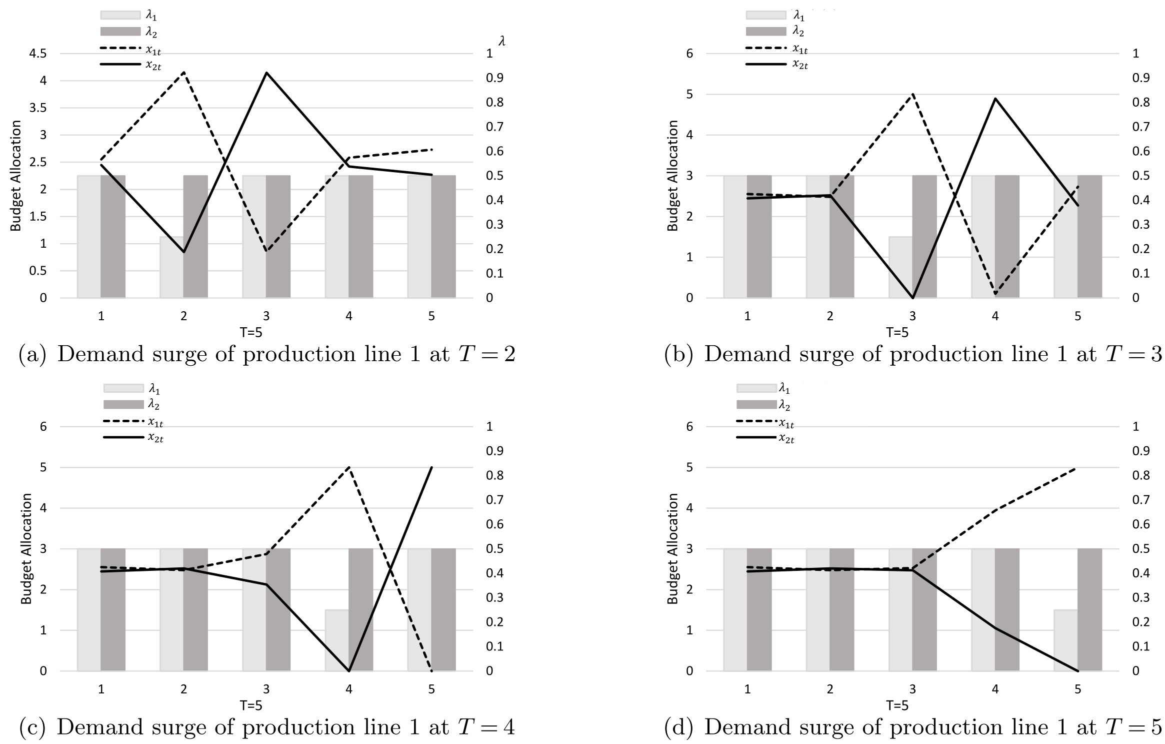
where $x_{2\tau} = \frac{B_\tau - c_1 x_{1\tau}}{c_2}$. If no solution to the above equations exist, then all the budget B_t is allocated to the production line corresponding to the dominant side.

Extension of Proposition 3 to $n > 2$ production lines is very similar to the extension of Proposition 1 to $n > 2$ production lines, which is described in Online Supplementary Material Section C.

While our initial assumption considered demand to be monotonically decreasing, acknowledging the dynamic nature of markets – especially in Newly Industrialized Economies (NIEs) where trends of increasing demand are observed – is crucial. Moreover, the potential for demand fluctuations due to unforeseen events (e.g., the cancellation of significant orders or surges in demand from special events) underscores the need to examine the impact of such variability on decision-making. To this end, we conduct a numerical analysis to explore how these variations influence budget allocation across two production lines.

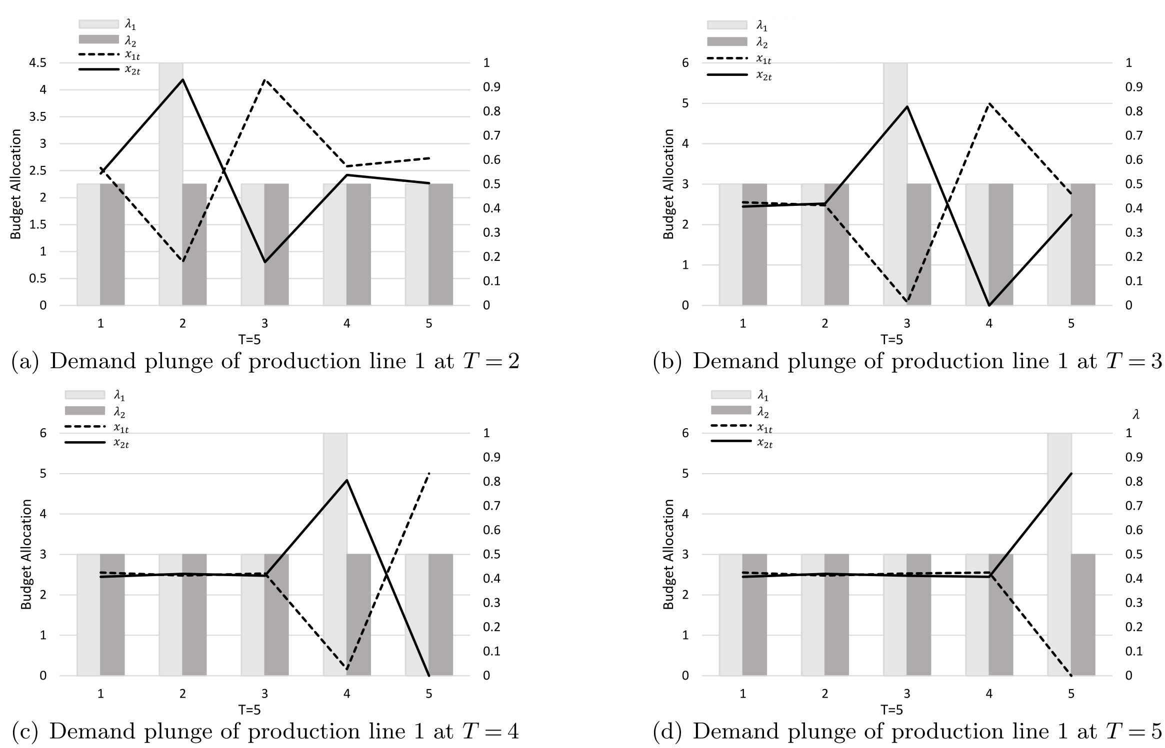
Fig. 7 presents the results of our numerical analysis, illustrating the budget allocation between two identical production lines under different demand fluctuation scenarios. We configure the parameters as follows: $c = 0.01$, $h = 0.001$, and $\beta = 0.8$, with the demand parameter λ displayed as bars. Production line 1 experiences constant demand with $\lambda = 0.5$ across five stages, except for a surge in demand at periods $T = 2$, $T = 3$, $T = 4$, and $T = 5$, as shown in Figs. 8(a) to 8(d), respectively. In contrast, production line 2 maintains constant demand throughout. The line plots in Fig. 7 depict the budget allocation, set to be periodic and constrained for each time period.

As observed in Figs. 8(a) to 8(d), the budget allocation favors production line 1 during periods of increased demand, disrupting the balance of an equal split typically seen when both production lines have identical demand. Notably, the timing of the demand surge significantly affects the budget allocation pattern. Early surges (at $T = 2$ and $T = 3$) lead to an immediate shift in budget towards production line 1, with a subsequent "revengeful" reversal favoring production line 2 in the following period. Conversely, when the surge occurs later (at $T = 4$ and $T = 5$), the budget allocation begins to shift towards production line 1 in advance of the anticipated demand increase — by one period for $T = 4$ and two periods for $T = 5$, with a gradual rather than immediate reallocation.



(a) Demand surge of production line 1 at $T = 2$
(b) Demand surge of production line 1 at $T = 3$
(c) Demand surge of production line 1 at $T = 4$
(d) Demand surge of production line 1 at $T = 5$

Fig. 7. Optimal splitting of periodic budget under fluctuating demand surge one production line.



(a) Demand plunge of production line 1 at $T = 2$
(b) Demand plunge of production line 1 at $T = 3$
(c) Demand plunge of production line 1 at $T = 4$
(d) Demand plunge of production line 1 at $T = 5$

Fig. 8. Optimal splitting of periodic budget under fluctuating demand plunge one production line.

Next, we extend our numerical analysis to scenarios depicted in Fig. 8, wherein the demand for production line 1 experiences significant drops at time periods $T = 2, T = 3, T = 4$, and $T = 5$, corresponding to an unusually high demand parameter value of $\lambda_1 = 1$. Meanwhile, demand for production line 2 remains constant at $\lambda_2 = 0.5$, as does the demand for production line 1 ($\lambda_1 = 0.5$) outside of these specific plunge periods.

As illustrated in Fig. 8 (Figs. 8(a) through 8(d)), it is observed that when demand for production line 1 drops, budget allocation shifts in favor of the unaffected production line 2. In the subsequent period, however, the budget allocation reverses, indicating a robust recovery for the demand resurgence in production line 1. Contrary to the demand surge scenarios presented in Fig. 7, the change in budget allocation in response to demand drops occurs simultaneously with the demand changes, without any anticipatory adjustment favoring the non-affected production line.

These findings affirm the expectation that the production line with higher demand will receive a larger budget allocation. In scenarios of demand surge, an interesting pattern emerges: optimal budget allocation tends to proactively adjust in anticipation of upcoming surges, especially when these surges are predicted further into the planning horizon. This anticipatory pattern is not observed in situations of demand drops. However, in both demand surge and drop scenarios, it is notable that the momentarily less-favored production line regains budgetary resources once the immediate impact of the surge or drop subsides. These insights can help practitioners capable of moderate-term demand forecasting, offering strategic guidance in budget allocation to adeptly navigate anticipated demand fluctuations.

5. Capacity expansion with partial demand distribution information

Our analyses in the last section rely on an important assumption that demand distribution functions are known to the decision maker. This assumption is valid for mature product markets where plenty of data are available to estimate the market demand. However, for newly introduced products, it may be challenging to estimate demand distributions based on limited sales and survey data (Berger, 1985). Moreover, for most products, seasonality and other exogenous effects may result in deviations from the assumed distributions of demand over time.

In this section we extend our analysis to the case where some partial information, in the form of a prior distribution for unknown demand distribution parameters, is available. We model this through a Bayesian inference approach. We note that in real practice, some data-driven models including machine learning (ML) methods can forecast the demand quite effectively. However, in this research we opt to utilize Bayesian inference for estimating the demand. While ML methods can indeed provide accurate demand forecasts, a stylized modeling approach using Bayesian inference allows the extraction of structural properties from analytical analysis. Such properties are invaluable to decision-makers as they provide overarching managerial insights, independent of technological dependencies. Moreover, Bayesian inference promotes transparency by elucidating the entire forecasting trajectory, enabling decision-makers to understand the underlying mechanisms and make informed choices based on a clear understanding of the forecasting process.

In the Bayesian framework, decision makers are assumed to have limited knowledge on some parameters of the distribution function, especially at the early stages of the planning horizon. However, they may have prior beliefs of the parameters, which they update through the Bayes' law as more information becomes available (Azoury, 1985, Lovejoy, 1990, Chen and Plambeck, 2008). When compared with the treatment of unknown distributions as known stochastic processes, the Bayesian learning approach may be appealing for strategic capacity planing with sequential investment decisions. We study the unknown demand structures with both stationarity and nonstationarity properties under our modeling framework.

Similar to our analysis for the known demand distribution case, we consider the highly flexible gamma family of distributions for demand, and note that the rate parameters determine both the mean and the variance in these distributions. We assume that the decision maker does not know the exact values for the rate parameters of the demand distributions, and that rate parameters themselves are assumed to be random variables following the gamma distribution. Hence, we incorporate the Bayesian inference approach into our model to update the decision maker's beliefs in the distributions of the rate parameters as market demand is revealed in each period.

5.1. Stationary demand case

We first assume that demand for product i is distributed according to $Gamma(\alpha_i, \varphi_i)$, where the shape parameter α_i is known and the rate parameter φ_i is unknown. Following the Bayesian inference approach, the prior belief in the distribution of the rate parameter φ_i for production line i can be expressed as:

$$f_{a_i,b_i}(\varphi_i) = \frac{b_i^{a_i}\varphi_i^{a_i-1}e^{-b_i\varphi_i}}{\Gamma(a_i)}, \quad \varphi_i \geq 0. \tag{14}$$

where a_i and b_i are hyperparameters, i.e. a_i is the shape parameter and b_i is the rate parameter for the posterior distribution of φ_i. This gamma prior is a conjugate prior for the likelihood function, i.e. the demand distribution function. Let the historical demand realizations for period t be $\{d_{i1}, d_{i2}, \ldots, d_{i(t-1)}\}$. Then the posterior belief of the rate parameter also follows a gamma distribution with updated parameters: $a_{it} = a_{i0} + (t-1)\alpha_i$ and $b_{it} = b_{i0} + \sum_{\tau=1}^{t-1} d_{i\tau}$.

We now modify the gross return function based on Bayesian inference. Letting $f_{a_{it},b_{it}}(\varphi_i)$ represent the distribution of the rate parameter for the demand distribution of product line i in period t, we get:

$$g_{a_{it},b_{it},\alpha_i}(s_{it}) = \int_0^{+\infty} f_{a_{it},b_{it}}(\varphi_i)\left(\int_0^{s_{it}} m_i r f_{\alpha_i,\varphi_i}(r)dr\right)d\varphi_i. \tag{15}$$

When this new gross return function is compared to the one in Eq. (2), the new gross return function has one more level of integration over all possible values of the unknown rate parameter. Although the gross return function has a complicated form, its first order derivative has a surprisingly simple and tractable structure:

$$g'_{a_{it},b_{it},\alpha_i}(s_{it}) = \frac{b_{it}^{a_{it}}\Gamma(a_{it}+\alpha_i)s_{it}^{\alpha_i}m_i}{\Gamma(a_{it})\Gamma(\alpha_i)(s_{it}+b_{it})^{a_{it}+\alpha_i}}. \tag{16}$$

We state the properties of the gross return function for production line i in Lemma 3, which leads to the optimal investment policy results in the next subsection.

Lemma 3. *For production line i, the expected gross return $g_{a_{it},b_{it},\alpha_i}(s_{it})$ is convex in the interval $[0, \frac{b_{it}\alpha_i}{a_{it}}]$ and concave in the interval $[\frac{b_{it}\alpha_i}{a_{it}}, +\infty)$.*

5.1.1. Resource allocation to a portfolio of production lines under Bayesian learning

In this section we use this revised modeling framework to analyze the optimal investment strategy under Bayesian learning for a firm with multiple production lines and budget constraints. We only analyze the case of periodic budget constraints introduced in Section 4.1.1. The benefit of a single pooled budget constraint is mostly to plan ahead, which is difficult when demand distributions are unknown and the investment policies are contingent on new demand realizations.

It is assumed that at the beginning of period t, all available information is summarized as a state vector $(a_{it}, b_{it}, s_{i(t-1)})$, where a_{it} and b_{it} contain the updated beliefs of the demand distribution function and $s_{i(t-1)}$ is the capacity level in the last period. For the moment, we first ignore the budget constraint and consider a single production line. Under the Bayesian learning assumption, the value function for product i in period t is:

$$V_{it}(s_{i(t-1)}, a_{it}, b_{it}) = \max_{s_{it}\geq 0}\{c_i s_{i(t-1)} - (c_i+h_i)s_{it} + \beta g_{R_{it}}(s_{it}) + \beta\mathbb{E}\left[V_{i(t+1)}(s_{it}, a_{it}+\alpha_i, b_{it}+d_{it})\right]\} \tag{17}$$

where R_{it} is a vector of the distribution parameters $(a_{it}, b_{it}, \alpha_i)$ and the state vector is $(s_{i(t-1)}, a_{it}, b_{it})$, with state transition relationships given as $s_{it} = s_{i(t-1)} + x_{it}$, $a_{it} = a_i^o + (t-1)\alpha_i$, and $b_{it} = b_i^o + \sum_{\tau=1}^{t-1} d_{i\tau}$. Similar to Eq. (4), the value function $V_{i(T+1)}(s_{iT}, a_{i(T+1)}, b_{i(T+1)})$ for the final capacity level is:

$$V_{i(T+1)}(s_{iT}, a_{i(T+1)}, b_{i(T+1)}) = \sum_{t=T+1}^{\infty} \beta^{t-1}\left[-h_i s_{iT} + \beta g_{R_{i(T+1)}}(s_{iT})\right]$$

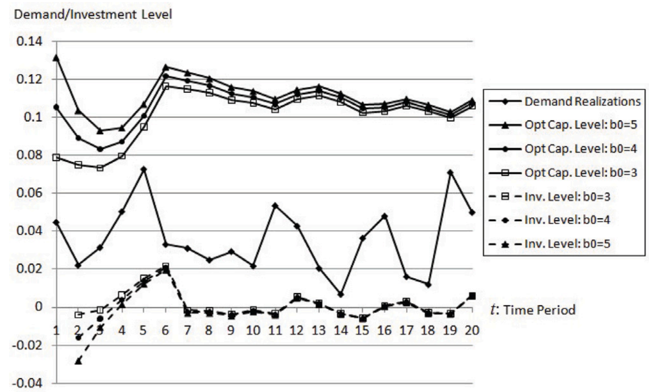


Fig. 9. Prior beliefs, demand realizations, and myopic capacity levels in each period.

$$= \frac{\beta^T \left[\beta g_{R_{i(T+1)}}(s_{iT}) - h_i s_{iT}\right]}{1-\beta} \tag{18}$$

Based on this, the following conclusions can be reached about the optimal capacity level and investment amount in period t for a production line i:

Lemma 4. *Suppose that $s_{i0} \geq \frac{b_{i0}\alpha_i}{a_{i0}}$. Under the Bayesian learning assumption for a single production line, if the state vector in period t is given as $(s_{i(t-1)}, a_{it}, b_{it})$, then*

1. $V_{it}(s_{i(t-1)}, a_{it}, b_{it})$ is constant in $s_{i(t-1)}$ when $s_{i(t-1)} \leq \frac{b_{it}\alpha_i}{a_{it}}$, and concave-decreasing in $s_{i(t-1)}$ when $s_{i(t-1)} \geq \frac{b_{it}\alpha_i}{a_{it}}$.

2. There exists an optimal capacity level s^*_{it} for each period t where $s^*_{it} \geq \frac{b_{it}\alpha_i}{a_{it}}$ such that if the initial capacity in period t, i.e. $s_{i(t-1)}$, is less than s^*_{it}, then it is optimal to raise the capacity up to s^*_{it}, otherwise, it is optimal not to invest in capacity in that period.

Given that the analytical characterization of s^*_{it} is not possible, we describe below a heuristic policy and numerically assess the difference between this policy and the optimal solution to the overall problem.

5.1.2. A myopic resource allocation policy

Our proposed heuristic is a myopic resource allocation policy which we define as follows:

Lemma 5. *Under the Bayesian learning assumption for a single production line, if the state vector in period t is given as $(s_{i(t-1)}, a_{it}, b_{it})$ with $s_{i(t-1)} \geq \frac{b_{it}\alpha_i}{a_{it}}$, then the optimal capacity level s^*_{it} can be approximated through a myopic resource allocation policy as the unique solution $\tilde{s}^*_{it}$ to the following equation:*

$$\frac{s_{it}^{\alpha_i}}{(s_{it}+b_{it})^{a_{it}+\alpha_i}} = \frac{\Gamma(a_{it})\Gamma(\alpha_i)[(1-\beta)c_i + h_i]}{\beta b_{it}^{a_{it}}\Gamma(a_{it}+\alpha_i)m_i}. \tag{19}$$

If there is no solution to Eq. (19), then $\tilde{s}^*_{it} = s_{i0}$. Assuming irreversible investment, the myopic investment level without any budget constraints in period t is $\tilde{x}^*_{it} = \max\{0, \tilde{s}^*_{it} - s_{i(t-1)}\}$.

The right hand side of Eq. (19) is constant given the state vector, and thus the capacity level $\tilde{s}^*_{it}$ and investment amount $\tilde{x}^*_{it}$ can easily be obtained for $s_{i(t-1)} \geq \frac{b_{it}\alpha_i}{a_{it}}$. Fig. 9 illustrates the application of Lemma 5 to a production line with Bayesian updating. The horizontal axis represents decision period t and the vertical axis represents the heuristic capacity levels and investment amounts in each period. Demand realizations are shown as the middle line. We calculate the heuristic capacity levels and investment amounts based on demand realizations and the three different priors with $a_i^o = 2$ and $b_i^o = 3, 4, 5$, respectively (higher value of b_i^o means that the decision maker holds a more optimistic prior belief).

Fig. 9 shows that the myopic capacity levels and investment amounts fluctuate with demand realizations with one period lag. A high demand realization in one period is a signal that the demand distribution might have been underestimated, while a low demand realization in another period shows that the demand distribution might have been overestimated. Therefore, the decision maker updates his/her beliefs of the demand distribution and myopic capacity level according to Eq. (14) and Eq. (19), respectively. For example, in period 2, the demand realization is 0.022, which is lower than the decision maker's expectation in period 1. The decision maker incorporates this new demand realization to update his/her belief of the demand distribution. Under the updated belief, the myopic capacity level is smaller. Thus, capacity is reduced (i.e., the investment level is negative). Assuming irreversibility of capacity, this would mean that no capacity expansion is made. In period 5, the demand realization is high. The decision maker invests a positive amount to reach a higher capacity level. As more demand realizations are available, posterior beliefs are updated and the impact of (inaccurate) prior beliefs become less significant. The heuristic capacity levels and investment amounts stay relatively stable after period $t = 7$ when there is sufficient data to estimate the rate parameter φ_i. As seen from the top three lines and bottom three lines, the capacity levels and investment amounts based on three different prior beliefs converge to the same levels. It is also important to note that demand realizations between periods 2 and 8 reveal an S-shaped product life cycle model, and the myopic capacity levels in these periods also reveal the same shape. However, a potential drawback of the Bayesian model is that there might be a lag between demand realization and capacity adjustment.

We now consider the portfolio budget allocation problem with multiple production lines and budget constraints under Bayesian learning, and derive a myopic policy based on Lemma 5. Similar to previous discussions, we initially demonstrate our analysis using two production lines 1 and 2.

Proposition 4. *Assume that $B_t < c_1(s^*_{1t} - s_{1(t-1)}) + c_2(s^*_{2t} - s_{2(t-1)})$. If $s_{i(t-1)} \geq \frac{b_{it}\alpha_i}{a_{it}}$ for $i = 1, 2$, in period t, a myopic policy is to split the budget between the two production lines according to their marginal profits, i.e. s_{1t} and s_{2t} solve the following equations:*

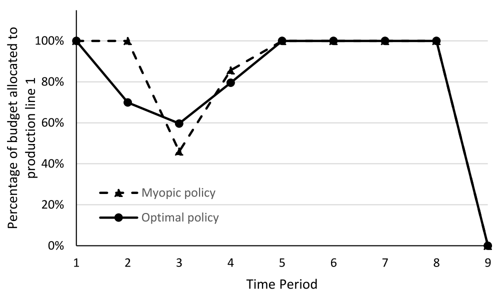
$$-[(1-\beta)c_1 + h_1] + \frac{\beta b_{1t}^{a_{1t}}\Gamma(a_{1t}+\alpha_1)s_{1t}^{\alpha_1}m_1}{\Gamma(a_{1t})\Gamma(\alpha_1)(s_{1t}+b_{1t})^{a_{1t}+\alpha_1}} = -[(1-\beta)c_2 + h_2] + \frac{\beta b_{2t}^{a_{2t}}\Gamma(a_{2t}+\alpha_2)s_{2t}^{\alpha_2}m_2}{\Gamma(a_{2t})\Gamma(\alpha_2)(s_{2t}+b_{2t})^{a_{2t}+\alpha_2}}, \tag{20}$$

and $c_1(s_{1t} - s_{1(t-1)}) + c_2(s_{2t} - s_{2(t-1)}) = B_t$.

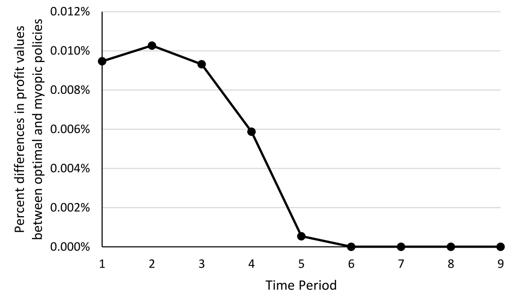
If $s_{i(t-1)} < \frac{b_{it}\alpha_i}{a_{it}}$ for $i = 1, 2$, then the policy for that product line is either to invest nothing or to invest B_t, depending on the maximizer for $\beta g_{R_{it}}(s_{it}) - c_i(s_{it} - s_{i(t-1)}) - h_i s_{it}$.

Extension of Proposition 4 to n production lines case is in the Online Supplemental Material. We note that one issue in modeling and solving multiple-period resource allocation problems involves computational complexity, even for numerical implementations. The concavity property of the gross return function in our model, which is stated in Lemma 3, is appealing from that perspective, especially when the decision horizon T and the number of production lines n is large. Most nonlinear optimization packages can efficiently solve the system of first order conditions for this multiple-period portfolio optimization problem.

We illustrate Proposition 4 through some numerical examples, and compare the results of the proposed myopic policy with the optimal policy in these numerical cases. To emphasize the impacts of prior beliefs and demand realizations, we control the effect of various costs. To this end, we let $c_1 = c_2 = 1, h_1 = h_2 = 0.001$. Moreover, the hyperparameters are such that $a_{1(t-1)} = a_{2(t-1)} = 2, b_{1(t-1)} = 1$, and $b_{2(t-1)} = 2$. The demand of the two lines in the example are fixed in each time period as $d_{1(t-1)} = 2$ and $d_{2(t-1)} = 3$. We consider the budget



(a) Percentage of budget allocated to production line 1 under optimal and myopic policies



(b) Percent difference in profit values V_t between optimal and myopic policies

Fig. 10. Differences in profit values and the split of budget in the optimal and myopic heuristic policies ($a_{1(t-1)} = a_{2(t-1)} = 2, b_{1(t-1)} = 3, b_{2(t-1)} = 4, B_t = 0.2$, and $d_{2(t-1)} = 6$).

allocation suggested by myopic policy and optimal policy over the two production lines in a multi-period model, as the myopic policy and optimal policy would suggest the same budget allocation split for a two-period model over the two production lines. In an extended multi-period model, there exist some differences between the myopic policy and optimal policy. In Fig. 2 we show the trend of split of the budget ($B_t = 1$) between production line 1 and production line 2 in a planning horizon through time period 1 to time period 9.

From Fig. 10(a), the myopic policy behaves in a similar fashion to the optimal policy, but it is generally more aggressive in cutting budget on line 1, which may indicate a greater focus on immediate or short-term returns. Despite this difference, the discrepancy in profit values between the two policies is quite minimal, with the largest difference being only 0.01% in month 2, where the optimal policy outperforms the myopic policy.

From Fig. 10(b), it appears the myopic policy demonstrates an impressive performance by closely mirroring the profit outcomes of the optimal policy with negligible differences. While the optimal policy does marginally outperform the myopic policy, this difference is remarkably small and may not significantly impact the overall profitability in a real-world scenario. More importantly, the myopic policy holds a considerable advantage in terms of its practical implementation. This policy is derived from a straightforward closed-form formula, which is far easier and quicker to compute. Therefore, the myopic policy can be applied as a suitable alternative to the optimal policy in practical situations. By doing so, organizations can streamline decision-making processes and still achieve near-optimal profitability

5.1.3. Using demand data from related product markets

Estimating the parameters of probability distributions for market demand is challenging, because most of the time there is only limited historical data available, especially for new products. The firm observes a realization of market demand in each period and has to make investment decisions based on a few observations. For firms operating multiple production lines, managers can exploit the relationships among demand distributions for the portfolio of products the manufacturer produces.

In some special cases, the demands for two products (product i and product j) can be represented by $d_{it} = \rho_{ij} d_{jt}$, where ρ_{ij} is the correlation coefficient between product i and j. Such a linear relationship is typically observed when two products are complements, i.e. when consumers buy one more unit of product j, they tend to also buy ρ_{ij} units of product i. The value of ρ_{ij} can be derived from theoretical constraints of product characteristics, consumer purchase patterns, and/or estimated using historical data. In this paper, we assume that the values for these parameters are given.

The dependency between the demands for two related products offer a way to deal with inadequate data. When estimating the parameters of the distribution function for newly introduced product i, managers can also use data from the established product j which the managers have better knowledge of its demand. This methodology is especially useful for the estimation of the demand distribution for new products that the managers have limited market data about. We can demonstrate how data from the related product j can be used to infer the distribution of new product i as follows.

Suppose in period t, the historical demand realizations for product i are $\{d_{i1}, d_{i2}, \ldots, d_{i(t-1)}\}$ and the available realizations for product j are $\{d_{j(-T')}, d_{j(-T'+1)}, \ldots, d_{j0}, d_{j1}, d_{j2}, \ldots, d_{j(t-1)}\}$ (product j has been on the market $T' + 1$ periods before the introduction of product i). The posterior belief of the rate parameter of the demand distribution can be updated as: $a_{it} = a_i^o + (2(t-1) + T' + 1)\alpha_i$ and $b_{it} = b_i^o + \sum_{\tau=1}^{t-1} d_{it} + \sum_{t=-T'}^{t-1} \rho_{ij} d_{jt}$.

In Fig. 11, we calculate the optimal capacity levels and investment levels for production line i with additional demand data from production line j. In period 1, the decision maker already has demand data from product j ($T' = 9$). Moreover, from period 1 on, the decision maker has two sets of demand realization in each period. Fig. 11 shows that even when at the beginning of the planning horizon, the optimal capacity levels and investment amounts are much more stable (converge quickly) than when there is no additional demand data from product j. Additional data from product j helps estimate the parameter of the demand distribution more accurately and in a timely manner.

5.2. Nonstationary demand case

In this section, we consider the capacity management problem under partially known nonstationary demand distributions, which has been referred to as a challenging but important problem in the literature (Van Mieghem, 2003, Patriksson, 2008).

We first describe how the Bayesian framework can be used to analyze the capacity management problem under these assumptions. Let $\{\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{iT}\}$ refer to the rate parameters of the demand distributions for period $t = 1, 2, \ldots, T$. Further assume that $\psi(t; \mathbf{w_i})$ represents the nonstationarity in demand distributions, where $\mathbf{w_i}$ is a vector of structural parameters which characterize the nature of the time-dependency among the distribution parameters across periods. Note that these structural parameters themselves may be unknown and can be treated as parameters to be jointly estimated. We discuss this case later in this subsection. Suppose the prior distribution of the parameters is summarized by $f(\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{iT}, \mathbf{w_i})$. At the end of period t, the decision maker can use demand realizations $\{d_{i1}, d_{i2}, \ldots, d_{it}\}$ to update the posterior distribution of the parameters according to Bayes' theorem:

$$
\begin{aligned}
&f(\varphi_{i1}, \ldots, \varphi_{iT}, \mathbf{w_i} | d_{i1}, \ldots, d_{it}) \\
&\quad = \frac{f(\varphi_{i1}, \ldots, \varphi_{iT}, \mathbf{w_i}) f(\varphi_{i1}, \ldots, \varphi_{iT}, \mathbf{w_i} | d_{i1}, \ldots, d_{it})}{f(d_{i1}, \ldots, d_{it})}
\end{aligned}
$$

where $f(d_{i1}, \ldots, d_{it}) = \int \ldots \int f(\varphi_{i1}, \ldots, \varphi_{iT}, \mathbf{w_i}) f(d_{i1}, \ldots, d_{it} | \varphi_{i1}, \ldots, \varphi_{iT}, \mathbf{w_i}) \mathrm{d}\varphi_{i1} \ldots \mathrm{d}\varphi_{iT} \mathrm{d}\mathbf{w_i}$.

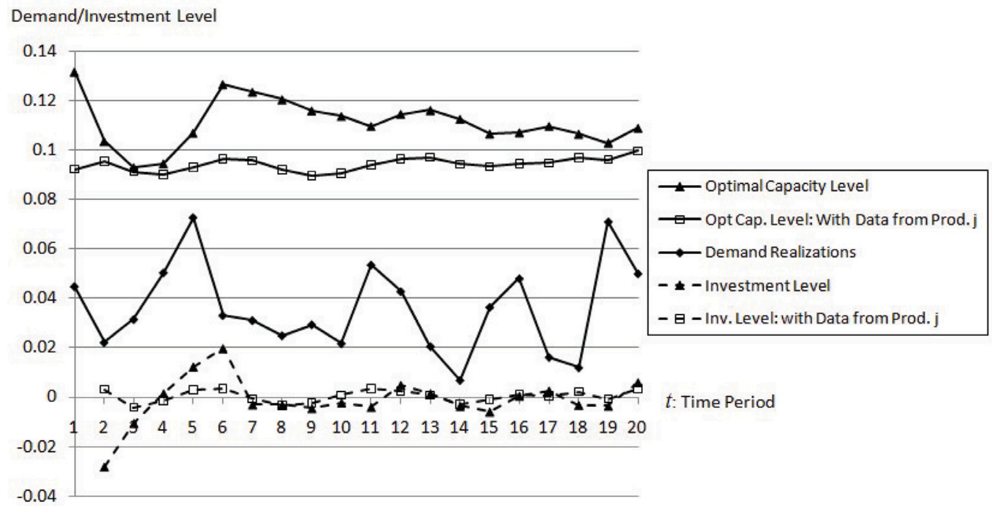


Fig. 11. Comparison of investment behaviors with and without additional demand data.

Based on the standard assumptions on conditional independency,

$$f(d_{i1},\dots,d_{it}|\varphi_{i1},\dots,\varphi_{iT},\mathbf{w_i}) = \prod_{\tau=1}^{t} f(d_{i\tau}|\varphi_{i1},\dots,\varphi_{iT},\mathbf{w_i})$$

$$= \prod_{\tau=1}^{t} f(d_{i\tau}|\varphi_{i\tau},\mathbf{w_i})$$

It follows as described in Winkler (1975) that:

$$f(\varphi_{i1},\dots,\varphi_{iT},\mathbf{w_i}|d_{i1},\dots,d_{it})$$
$$= f(\varphi_{i1},\dots,\varphi_{it},\mathbf{w_i}|d_{i1},\dots,d_{it})f(\varphi_{i(t+1)}...,\varphi_{iT},|\varphi_{i1},\dots,\varphi_{it},\mathbf{w_i}).$$

A common theoretical issue for nonstationary demand distributions is the treatment of the salvage value of the terminal capacity level. For stationary demand distributions, this is not an issue because the firm's undiscounted profit in any period beyond period T is the same as that in period T as seen in the last term of Eq. (5). As in the known distribution case, we assume that the demand distribution beyond the planning horizon, i.e. for $t > T$, is the same as that of period T since the estimates for last period are the latest estimates which incorporate all available information. Alternatively, we can extend the planning horizon such that it is somewhat reasonable to assume that the salvage value of the terminal capacity is zero. However, this alternative approach might not be as realistic because firms may rarely do such long-term capacity planning.

In the following subsections, we look into the cases when the demand shifts over time are deterministically and stochastically unknown within the general Bayesian inference and decision framework.

5.2.1. Unknown distribution parameters with deterministic trend

Consider the general deterministic case $\varphi_{it} = \psi(t;\mathbf{w_i})$, where $\mathbf{w_i}$ is a vector of parameters characterizing how the rate parameters change over time. In the first period, i.e. $t = 1$, the prior distribution of the rate parameter φ_{i1} can be characterized as $\psi(t = 1;\mathbf{w_i}) \sim Gamma(\mathbf{a_i^o},\mathbf{b_i^o})$, where $\mathbf{a_i^o}$ and $\mathbf{b_i^o}$ are the two vectors of hyperparameters which have the same dimension as $\mathbf{w_i}$. The procedure to update $\mathbf{w_i}$, $\mathbf{a_{it}}$ and $\mathbf{b_{it}}$ are the same as that for the case with stationary demand discussed above.

We now look at a special case where the rate parameters of the demand distributions for product line i reveal a linear time-dependent trend, i.e., $\varphi_{it} = \omega_i t$, where ω_i is the unknown linear trend factor that links the prior distribution of each period (Popovic, 1987, Kamath and Pakkala, 2002). In the first period, the prior distribution of the rate parameter is $\varphi_{i1} \equiv \omega_i \sim Gamma(a_i^o, b_i^o)$. After observing the demand realization d_{i1}, the posterior distribution is $\omega_i|d_{i1} \sim Gamma(a_i^o+\alpha_i, b_i^o + d_{i1})$. Notice that for a gamma distribution, if $\omega \sim Gamma(a,b)$, then $\omega t \sim Gamma(a, \frac{b}{t})$. Hence, the prior distribution of the rate parameter in the second period is

$$\varphi_{i2}|d_{i1} \equiv 2\omega_i|d_{i1} \sim Gamma(a_i^o+\alpha_i^o, \frac{b_i^o+d_{i1}}{2}).$$

After observing the demand realization d_{i2}, the posterior distribution of the rate parameter in the second period follows to be

$$\varphi_{i2}|d_{i1},d_{i2} \sim Gamma\left(a_i^o+2\alpha_i, \frac{b_i^o+d_{i1}}{2}+d_{i2}\right),$$

while the posterior distribution of ω_i in the second period is

$$\omega_i|d_{i1},d_{i2} \equiv \frac{\varphi_{i2}}{t}|d_{i1},d_{i2} \sim Gamma\left(a_i^o+2\alpha_i, 2(\frac{b_i^o+d_{i1}}{2}+d_{i2})\right).$$

Following the same procedure, we can derive by induction that the prior distribution of the rate parameter in period t is

$$\varphi_{it}|d_{i1},\dots,d_{i(t-1)}$$
$$\equiv \omega_i t|d_{i1},\dots,d_{i(t-1)} \sim Gamma\left(a_i^o+(t-1)\alpha_i, \frac{b_i^o+\sum_{\tau=1}^{t-1}\tau d_{i\tau}}{t}\right). \quad (21)$$

Hence, the Bayesian updating framework for the unknown but constant parameter still applies for the case of unknown and time-dependent demand distributions. Our analysis and results in Section 5.1 remain unchanged except that we replace b_{it} by $\frac{b_i^o+\sum_{\tau=1}^{t-1}\tau d_{i\tau}}{t}$. Also different from the case of the unknown and constant rate parameter, expression (21) shows that demand realizations in recent periods have higher weights and thus have a large impact on the estimation of φ_{it}. The reason is that demand realizations in previous periods ($\tau = 1,2,\dots,t$) impacts the estimation of b_{it} indirectly through ω_{it}, which is used to estimate b_{it}. However, the demand realization in the current period is directly used to come up with the posterior distribution.

We can extend the analysis of the linear time-dependent trend to capture the general case $\varphi_{it} = \psi(t;\mathbf{w_i})$. The procedure to update $\mathbf{w_i}$, $\mathbf{a_{it}}$ and $\mathbf{b_{it}}$ are the same as the linear trend case discussed above. However, there might not exist a closed-form result for a general deterministic trend function and numerical approaches may need to be used.

5.2.2. Unknown distribution parameters with stochastic trend

Stochasticity can also be included into the trend relationship $\psi(\cdot)$. Suppose shifts in the rate parameter from one period to the next follow a random walk, i.e. $\varphi_i = \varphi_{i(t-1)} + \varepsilon_{it}$, where ε_{it} are independently and identically distributed according to a normal distribution with mean μ_i and variance σ_i^2, i.e. $\varphi_{it} \sim Normal(\varphi_{i(t-1)}+\mu_i, \sigma_i^2)$. If the distribution parameters (μ_i, σ_i^2) are known, we can easily extend our analysis for the deterministic trend to deal with stochastic trend. Specifically, the posterior distribution of $\varphi_{i(t-1)}$ in period $t-1$ after the decision maker observes d_{it} is $\varphi_{i(t-1)}|d_{i(t-1)} \sim Gamma(\alpha_{i(t-1)}, b_{i(t-1)})$ where $\alpha_{i(t-1)} = a^o_{i(t-1)} + \alpha_i$ and $b_{i(t-1)} = b^o_{i(t-1)} + d_{i(t-1)}$. This posterior distribution can then be used to form an updated belief on the distribution of φ_{it} as follows:

$$f(\varphi_{it}) = \int f(\varphi_{it}|\varphi_{i(t-1)}, \mu_i, \sigma_i^2) f(\varphi_{i(t-1)}|d_{i(t-1)}) \mathrm{d}\varphi_{i(t-1)}$$

$$= \int \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{\varphi_{it}-(\mu_i+\varphi_{i(t-1)})^2}{2\sigma_i^2}} \frac{b_{i(t-1)}^{a_{i(t-1)}} \varphi_{i(t-1)}^{a_{i(t-1)}-1} e^{-b_{i(t-1)}\varphi_i}}{\Gamma(a_i)} \mathrm{d}\varphi_{i(t-1)}$$

We can use this updated distribution to replace $f_{a_{it},b_{it}}(\varphi_i)$ in Eq. (15).

6. Exploration of the robustness of the gamma distribution based solutions to departures from the gamma distribution assumption

In this section, we expand our numerical analyses and explore the robustness of the gamma distribution based solutions to departures from the gamma distribution assumption for both known and partially known demand distributions. These analyses consist of various cases of comparisons involving the normal distribution as described in detail below. In general, we conclude that the flexibility and tractability of the gamma distribution, which allow for accurate approximations of several common distributions, results in it being a good and practically reasonable choice for modeling normal or other demand distribution. On the other hand, in line with the empirical findings of Swan (2002), we observe that the quality of approximations for normally distributed demand deteriorate in the relatively less common cases where the mean of the demand distribution is very low.

6.1. Known stationary demand distribution

We perform our analysis for the known stationary demand distribution case using two examples with different structures. In the first example, demand distributions for two product lines are given as $Normal(10, 10)$ and $Normal(20, 40)$. In the second example, which is based on the instance described in Fig. 4(a), the demand distributions are $Normal(1, 1)$ and $Normal(0.5, 0.25)$, corresponding to an instance with lower parameter values. For both examples, we compare the optimal policies obtained through the direct use of the normal distribution versus their gamma approximations. We assume that unit investment and holding costs are given as $c_1 = 0.01, h_1 = 0.001, c_2 = 0.005$, and $h_2 = 0.0005$.

For the first example, the corresponding gamma approximations for $Normal(10, 10)$ and $Normal(20, 40)$ are $Gamma(10, 1)$ and $Gamma(10, 0.5)$, respectively. In the interest of space, these results are illustrated in the online supplemental materials. In Figure C.3(a) and Figure C.3(c) we show how the budget is split under different budget levels for the actual distributions and their gamma approximations, while in Figure C.3(b) and Figure C.3(d) the same comparison is performed based on the split of an additional unit of budget in the two cases. A clear similarity in the corresponding plots within Fig. 4(a) is observed, indicating the accuracy of the approximation through the gamma distribution.

As an additional analysis, we also consider the case with periodic budgets in each period and compare the results obtained under the normal distribution and its gamma approximation. In Online Supplemental Figure C.4, we show the budget allocations for increasing, constant, and decreasing budgets over a five period planning horizon, where the budget levels are the same as those described in Section 4.1.1. Again, the accuracy of the approximations are observed, which demonstrates the quality of the approximations in the periodic budget case for distributions with large means.

For the second example, the demand distributions are $Normal(1, 1)$ and $Normal(0.5, 0.25)$, and the corresponding gamma approximations are $Gamma(1, 1)$ and $Gamma(1, 2)$ respectively, which are the distributions used in the first example in Section 4.1.2. We note that the gamma approximation of the normal distribution with small means as in this example is known to be not very accurate, and thus we consider such an example for exploration purposes. The allocation of a given budget among the two product lines under the demand distributions of $Gamma(1, 1)$ and $Gamma(1, 2)$ is shown in Figs. 4(a) and 4(b). When we use the normal distributions directly in our modeling framework, the corresponding budget allocation decisions for capacity expansion of the two product lines are calculated to be as shown in Online Supplemental Figures C.5(a) and Figure C.5(b).

Comparing these two plots with Figures 12b 4(a) and 12d 4(b), it can be observed that while the optimal allocations are pretty much the same for budgets levels less than 9 units, at larger budget levels the gamma approximation results deviate from the normal distribution case. More specifically, there is an overestimation of the optimal capacity levels when the available budget gets high. Overall, we observe that the quality and robustness of the gamma approximation of the normal distribution are maintained for demand distributions with means that are not too low, but the quality deteriorates when the mean of the normally distributed demand is very low.

6.2. Partially known stationary demand distribution

We also study the robustness of the gamma approximation for the partially known demand distributions case, which we address through a Bayesian approach in our modeling framework. However, we note that under Bayesian learning it is not possible to have an exact comparison between the normally distributed demand case and its gamma approximation. This is because as the parameters of the posterior of the unknown parameter φ_i are updated in $Gamma(\alpha_i, \varphi_i)$, it has an impact on both the mean and the variance of the distribution. On the other hand, if either the mean or the variance is unknown in $Normal(\mu_i, \sigma_i^2)$, the updates on the parameters of the posterior of that unknown parameter affect only the mean or the variance of the distribution, but not both. Given this issue, we consider two cases or normally distributed demand: one with known mean but unknown variance; and the other with unknown mean but known variance. We derive the corresponding optimal allocation equations separately for each case and then compare the results with the gamma approximation using numerical examples.

6.2.1. Normally distributed demand with known mean and unknown variance

Assume that demand for product i follows a normal distribution $Normal(\mu_i, \sigma_i^2)$, for which the decision maker knows the mean but does not know the variance of the distribution. The decision maker updates the belief in the distribution of the variance as market demand is realized in each period. We derive the optimal capacity allocation policy for this case under the Bayesian framework as a counterpart to the equation in Proposition 4 the following equation to be solved for optimal allocation of budget for two product lines:

$$\begin{aligned}-[(1-\beta)c_1 + h_1] + &\frac{s_{1t} b_{1t}^{a_{1t}} \Gamma(a_{1t} + \frac{1}{2}) m_1}{\sqrt{2\pi}\Gamma(a_{1t})[\frac{1}{2}(s_{1t} - \mu_1)^2 + b_{1t}]^{a_{1t}+\frac{1}{2}}} \\ &= -[(1-\beta)c_2 + h_2] \\ &\quad + \frac{s_{2t} b_{2t}^{a_{2t}} \Gamma(a_{2t} + \frac{1}{2}) m_2}{\sqrt{2\pi}\Gamma(a_{2t})[\frac{1}{2}(s_{2t} - \mu_2)^2 + b_{2t}]^{a_{2t}+\frac{1}{2}}},\end{aligned} \tag{22}$$

The comparison of a Normal distribution example and a Gamma distribution example are shown in Online Supplemental Figure C.6 and Figure C.7. If a normal distribution with known mean and unknown variance is assumed, the resulting budget allocations for capacity expansion would be as shown in Figure C.6. The corresponding policy when a gamma approximation is used for the similar set-up is shown in Figure C.7. Although an exact comparison is not possible due to the reasons mentioned above, when we consider the results in Figure C.6 and Figure C.7, there is a clear difference. This implies that if demand is distributed normally with known mean and unknown variance, an approximation based on a gamma distribution with known shape parameter and unknown rate parameter is likely not to be an accurate representation. We show below that the approximation is much better when the mean of the normal distribution is unknown and the variance is known.

6.2.2. Normally distributed demand with unknown mean and known variance

Similar to the setup above, in this section we assume that demand distribution for each product line i is again $Normal(\mu_i, \sigma_i^2)$, but the variance of demand is known and the mean μ_i is unknown by the decision maker. The conjugate prior for this case is also a normal distribution with corresponding mean a_i and variance b_i as follows:

$$f_{a_i,b_i}(\mu_i) = \sqrt{\frac{b_i}{2\pi}} e^{-\frac{b_i(\mu_i - a_i)^2}{2}}$$

The counterpart of the equation in Proposition 4 is:

$$-[(1-\beta)c_1 + h_1] + \beta m_1 s_{1t} \sqrt{\frac{b_{1t}\sigma_1^2}{2\pi(\sigma_1^2 + b_{it})}} e^{\frac{\frac{(s_{1t}\sigma_1^2 + a_{1t}b_{1t})^2}{\sigma_1^2 + b_{1t}} - \sigma_1^2 s_{1t}^2 - b_{1t}a_{1t}^2}{2}}$$
$$= -[(1-\beta)c_2 + h_2] + \beta m_2 s_{2t} \sqrt{\frac{b_{2t}\sigma_2^2}{2\pi(\sigma_2^2 + b_{2t})}} e^{\frac{\frac{(s_{2t}\sigma_2^2 + a_{2t}b_{2t})^2}{\sigma_2^2 + b_{2t}} - \sigma_2^2 s_{2t}^2 - b_{2t}a_{2t}^2}{2}} \quad (23)$$

By observing Online Supplemental Figure C.8, we get the optimal budget allocations shown for the two product lines when a normal distribution with unknown mean and known variance is assumed for the demand of each product line. It can be shown that these results are very similar to those in Figure C.7, where a gamma approximation is used. Hence, it can be noted that gamma approximation is likely to work well for cases involving normally distributed demands with unknown means and known variances.

7. Conclusions

In this paper, we derive closed-form optimal resource allocation policies for capacity expansion, which address the two dimensional trade-offs in this problem: the split of budget among the portfolio of capacity investment projects and the split of budget among multiple decision periods. Complementary and distinct from other capacity expansion models, our analysis considers cases where an order is lost completely if the firm does not have sufficient capacity to fulfill the entire demand associated with the order. Moreover, instead of using an exogenously given distribution of returns as a function of realized demand and deriving results based on assumptions about the properties of that return function, we formulate the return function directly based on a stochastic distribution of market demand. As part of the analysis, we identify the existence of an optimal capacity level for each production line and derive the optimal path to reach this level both under known demand distributions and unknown demand distributions with Bayesian updating. We consider both stationary and nonstationary demand cases for multi-production line multi-period versions of the problem, and provide some numerical examples.

Our study has led to several managerial insights that can have direct implications for decision-makers responsible for production line capacity decisions including the following:

1. Our analysis demonstrates the existence of an optimal capacity level for each production line, fundamentally influenced by the unit investment cost, the unit holding cost, and the market demand distribution. This insight suggests the critical balance between cost efficiency and market responsiveness, advocating for a strategic approach to capacity planning that aligns investment and holding costs with demand dynamics (Section 3).
2. The allocation of budget across two or multiple production lines hinges on the marginal profit of each line, which in turn is shaped by cost structures and demand patterns. Our findings elucidate that:
   (a) In contexts of extremely limited budget availability, cost advantages dictate budget prioritization towards more cost-saving production lines.
   (b) Conversely, when the budget limitation becomes more relaxed, the allocation decisions are predominantly influenced by the marginal returns, which is determined by the interplay between cost and demand (Section 4.1.2).
3. In the face of non-stationary demand – characteristic of many NIE manufacturers experiencing growth – strategic budget allocation over multiple periods becomes imperative. Our study highlights:
   (a) The proactive adjustment of budget allocation in anticipation of demand surges, suggesting that forward-looking planning can significantly benefit firms in volatile markets.
   (b) The resilience of temporarily less-favored production lines, which are capable of reclaiming budgetary resources following the resolution of demand fluctuations.
   (c) The value of moderate-term demand forecasting as a tool for strategic budget allocation, enabling firms to effectively navigate anticipated market changes (Section 4.2).
4. Our exploration of myopic budget allocation strategies reveals their potential as viable, simplified alternatives to more complex optimal strategies, particularly in dynamic demand environments. Despite their simplicity, myopic strategies closely align with the outcomes of optimal policies. This adaptability suggests that myopic strategies can serve as practical, efficient approaches for managing budget allocation in real-time, reducing computational burdens while maintaining strategic effectiveness (Section 5.1.2).

By integrating these insights into strategic planning, managers can enhance their capacity expansion decisions, placing them more closely with market realities and optimizing their budget allocation to maximize profitability and competitiveness.

Meanwhile, we concur that the dynamic capacity expansion problems may involve several additional complexities that are not explicitly captured by the modeling assumptions in this paper. One such extension is the deterioration or growth of capacity due to learning effects. This can be incorporated into the model by slightly changing the capacity accumulation Eq. (1) through the addition of a multiplier to represent deterioration or growth effects. In addition, we have not analyzed the impact of decreasing unit investment costs and economies of scale on capacity expansion. While decreasing investment costs can be modeled by replacing the costs c_i by a function $c_{it} = c(t)$, incorporating economies of scale into the model would impact the concavity of the value function and would result in limitations in the analysis.

In conclusion, our research represents a unique angle in understanding optimal resource allocation policies for capacity expansion. Our model addresses trade-offs in budget allocation among multiple production lines and multiple decision periods. The closed-form solutions derived provide a rigorous, comprehensive framework for managing capacity investment, with unique considerations for unfulfilled orders and a novel formulation of the return function. Our study provides managerial insights, guiding strategic decisions tied to cost structures, market demand, budget availability, and dynamic demand scenarios. It also opens the door to future research, particularly in areas such as capacity deterioration or growth due to learning effects, and the implications of decreasing unit investment costs and economies of scale. Ultimately, this paper contributes to the broader understanding of capacity expansion, enhancing both efficiency and responsiveness in the market environment in the era of global supply chain reconfiguration.

CRediT authorship contribution statement

Senay Solak: Conceptualization, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Zhuoxin Li:** Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Mehmet Gumus:** Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Yueran Zhuo:** Conceptualization, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used GPT-4 in order to analyze and rectify grammatical inaccuracies. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijpe.2024.109261.

References

Altintas, N., Erhun, F., Tayur, S., 2008. Quantity discounts under demand uncertainty. Manage. Sci. 54 (4), 777–792.
Azoury, K.S., 1985. Bayes solution to dynamic inventory models under unknown demand distribution. Manage. Sci. 31 (9), 1150–1160.
Bassamboo, A., Randhawa, R.S., Mieghem, J.A.V., 2010. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. Manage. Sci. 56 (8), 1285–1303.
Beaujon, G.J., Marin, S.P., McDonald, G.C., 2001. Balancing and optimizing a portfolio of R&D projects. Naval Res. Logist. 48 (1), 18–40.
Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.
Campisi, D., Mancuso, P., Nastasi, A., 2001. R&D competition, absorptive capacity, and market shares. J. Econ. 73 (1), 57–80.
Chen, L., Plambeck, E.L., 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. Manuf. Serv. Oper. Manag. 10 (2), 236–256.
Cohen, M.A., Zheng, Y., Wang, Y., 1999. Identifying opportunities for improving Teradyne'S service-parts logistics system. Interfaces 29 (4), 1–18.
Colvin, M., Maravelias, C.T., 2008. A stochastic programming approach for clinical trial planning in new drug development. Comput. Chem. Eng. 32 (11), 2626–2642.
Cooper, R.G.S.J.E., Kleinschmidt, E.J., 1999. New product portfolio management: Practices and performance. J. Prod. Innov. Manage. 16 (4), 333–351.
Dieulle, L., Berenguer, C., Grall, A., Roussignol, M., 2003. Sequential condition-based maintenance scheduling for a deteriorating system. European J. Oper. Res. 150 (2), 451–461.
Dixit, A., Pindyck, R., 1994. Investment Under Uncertainty. Princeton University Press, Princeton, NJ.
Fine, C.H., Freund, R.M., 1990. Optimal investment in product-flexible manufacturing capacity. Manage. Sci. 36 (4), 449–466.
Glasserman, P., 1996. Allocating production capacity among multiple products. Oper. Res. 44 (5), 724–734.
Goyal, M., Netessine, S., 2007. Strategic technology choice and capacity investment under demand uncertainty. Manage. Sci. 53 (2), 192–207.
Grenadier, S.R., Weiss, A., 1997. Investment in technological innovations: An option pricing approach. J. Financ. Econ. 44 (3), 397–416.
Hsu, Y.C., Pearn, W.L., Wu, P.C., 2008. Capability adjustment for gamma processes with mean shift consideration in implementing Six Sigma program. European J. Oper. Res. 191 (2), 517–529.
Kamath, K., Pakkala, T., 2002. A Bayesian approach to a dynamic inventory model under an unknown demand distribution. Comput. Oper. Res. 29 (4), 403–422.
Kauffman, R.J., Li, X., 2005. Technology competition and optimal investment timing: A real options perspective. IEEE Trans. Eng. Manage. 52 (1), 15–29.
Korhonen, P., Syrjanen, M., 2004. Resource allocation based on efficiency analysis. Manage. Sci. 50 (8), 1134–1144.
Leemis, L.M., 1986. Relationships among common univariable distributions. Am. Stat. 40 (2), 143–146.
Loch, C.H., Kavadias, S., 2002. Dynamic portfolio selection of NPD programs using marginal return. Manage. Sci. 48 (10), 1227–1241.
Lovejoy, W.S., 1990. Myopic policies for some inventory models with uncertain demand distributions. Manage. Sci. 36 (6), 724–738.
Luss, H., 1982. Operations research and capacity expansion problems: A survey. Oper. Res. 30 (5), 907–947.
Panayi, S., Trigeorgis, L., 1998. Multi-stage real options: The cases of information technology infrastructure and international bank expansion. Q. Rev. Econ. Finance 38 (3), 675–692.
Patriksson, M., 2008. A survey on the continuous nonlinear resource allocation problem. European J. Oper. Res. 185 (1), 1–46.
Popovic, J., 1987. Decision making on stock levels in cases of uncertain demand rate. European J. Oper. Res. 32 (2), 276–290.
Prastacos, G.P., 1981. Optimal sequential investment decisions under conditions of uncertainty. Manage. Sci. 29 (1), 118–134.
Satariano, A., Burrows, P., 2011. Apple's supply-chain secret? Hoard lasers. Bloomberg Businessweek.
Solak, S., Clarke, J.-P.B., Johnson, E.L., Barnes, E.R., 2010. Optimization of R&D project portfolios under endogenous uncertainty. European J. Oper. Res. 207 (1), 420–433.
Swan, W., 2002. Airline demand distributions: passenger revenue management and spill. Transp. Res. E 38 (3), 253–263.
Van Mieghem, J.A., 1998. Investment strategies for flexible resources. Manage. Sci. 44 (8), 1071–1078.
Van Mieghem, J.A., 2003. Capacity management, investment and hedging: Review and recent developments. Manuf. Serv. Oper. Manag. 5 (4), 269–302.
Volvo, 2010. Initial capacity assessment sheet: Communication kit for suppliers. Initial Capacity Assessment Sheet: Communication Kit for Suppliers. URL http://www.volvogroup.com/suppliers/global/en-gb/supplierselection/ourrequirements/keyelements/relateddocuments/Pages/related_documents.aspx.
Winkler, R., 1975. A Bayesian approach to nonstationary processes. In: Tri, M.P.H., Ponssard, J. (Eds.), Theorie de la Decision et Applications. Foundation Nationale pour 1'Enseignement de la Gestion, Paris, France, pp. 1–13.