Dynamic Capacity Expansion for a New Drug

Abstract

Due to demand and technical uncertainty and long lead times to build and add production capacity in compliance with government regulations, pharmaceutical manufacturers face potentially catastrophic capacity imbalance risks when introducing drugs. Motivated by the move towards partial outsourcing to mitigate such risks, this study investigates a pharmaceutical manufacturer's jointly optimal in-house capacity and outsourcing strategy for the life span of a new drug. In each period the manufacturer can invest in in-house capacity, which comes on line two periods after the investment, and also avail of outsourcing via a short-term agreement, which is modeled as a call option executable one period after the purchase date to get access to the reserved capacity for the period. The problem is formulated as a discrete-time, finite-period stochastic dynamic program, incorporating three uncertainties: clinical trial results, stochastic demand, and stochastic in-house capacity costs. We derive the optimal strategy in a three-period problem characterized by two state-dependent base-capacity levels and discuss the twofold advantage of partial outsourcing and managerial implications through a numerical study for a general N-period problem. Our model is most applicable to the industry with risky staged R&D, while it is generic, accommodating many R&D risk characterization, and can be applied to various cases where two types of capacity is available with non-trivial, heterogeneous investment lead times.

Keywords: capacity planning, partial outsourcing, Markov decision process, pharmaceutical manufacturing

1 Introduction

Capacity planning is one of the most challenging tasks drug manufacturers face in new drug introduction. The growing costs and complexity of drug R&D implies that a manufacturer could incur $1 million or more of lost revenues each day if capacity shortage occurs (Glasser 2002). Capacity overage is equally costly since a drug manufacturing plant can cost over $250 million (Grabowski 2008). Thus, balancing capacity is essential but difficult to achieve due to long lead times and demand and technical uncertainty. Building facilities and processes in compliance with FDA regulations requires multiple years. Our interviews with drug manufacturers reveal that they start capacity planning 2 years in advance of FDA approval, i.e., during the phase III (final stage) clinical trials. At this point there will be a 20% to 60% likelihood of not receiving FDA approval (DiMasi et al. 2010). Large uncertainty exists even after FDA approval. For example, Biogen Idec had to sell its plant built for Tysabri just a half year after its completion when the drug was linked to a potentially fatal brain disease and demand forecasts were revised significantly downwards (The Boston Globe 2005). On the other side, deferring capacity building can cause serious supply shortage. For instance, Immunex was acquired by Amgen after losing more than $250 million when capacity building could not catch up with the strong demand of Enbrel (The Seattle Times 2002).

Under these circumstances, outsourcing provides key flexibility in coping with demand uncertainty and long lead times, and has grown in importance in the face of changing technologies and market. Outsourcing can mitigate capacity overage risks attributable to drug failures and stochastic demand, and also provide investment time flexibility because contract manufacturing organizations (CMOs) have shorter lead times to bring capacity on line from their experience. Traditionally, manufacturers have kept high-margin drugs in-house and outsourced only off-patent drugs (e.g., see Singh (2005) for Eli Lilly's strategy). However, with many blockbuster drugs coming off-patent and struggling to develop new ones, manufacturers have reorganized manufacturing capacity and started outsourcing new drugs to CMOs to focus on R&D. Moreover, a shift to biopharmaceuticals is facilitating this trend, since biologic manufacturing is significantly more complex and expensive than small molecule drugs. In fact, most large-scale commercial biopharmaceutical production is partially outsourced to CMOs (High Tech Business Decisions 2007). For example, Bristol-Myers Squibb's Orencia has been manufactured in-house and at a CMO since its launch (Outsourcing Pharma 2007).

With this move towards partial outsourcing, a key question for a pharmaceutical manufacturer is how the capacity needs for in-house production and outsourcing evolve as technical and demand uncertainty evolves. To answer this question, various factors must be considered: capacity availability at the manufacturer and CMO, demand and clinical trial uncertainty, heterogeneous investment lead times, and stochastic costs evolving over a planning horizon. However, few studies have attempted to include all these factors into capacity planning.

Allowing for this concern, this study investigates a pharmaceutical manufacturer's jointly optimal in-house capacity and outsourcing policy for a new drug from the phase III clinical trial to patent expiration. The manufacturer can invest in in-house capacity, which comes on line two periods after the investment, and also reserve capacity at a CMO for the next period by purchasing options. This contract reflects the industry practice that most CMOs charge non-refundable upfront fees in the form of reservation fees, down payments or cancellation fees a year before the project starts in a large-scale project, which is typically 12 to 18 months in duration (High Tech Business Decisions 2006, 2007). We consider three uncertainties, drug failures and stochastic in-house capacity cost and demand. The problem is formulated as a discrete-time, finite-period stochastic dynamic program. The optimal policy in a 3-period problem is characterized by two state-dependent base-capacity levels. Managerial implications in an N-period problem are discussed in a numerical study.

The rest of the paper is organized as follows. In the next section, we review the relevant research. Section 3 details the model formulation and analytical results. The results of the numerical analysis are reported in section 4. Finally section 5 discusses final conclusions.

2 Literature Review

There are several streams of literature related to our study. The most relevant research is pharmaceutical project management. A key characteristic of capacity planning for drugs is that candidate drugs may fail during clinical trials. Scheduling problems for tasks that may succeed or fail was first studied by Schmidt and Grossman (1996), which is extended by many studies, including Jain and Grossmann (1999) and Levis and Papageorgiou (2004). Real options approach has also been used to manage risky projects. Rogers et al. (2002) apply binomial option pricing to drug R&D. Chambers et al. (2009) estimate the value of flexible capacity at a drug manufacturer. Flexible capacity is equivalent to a call option to defer building dedicated capacity until the drug receives FDA approval. Our study parallels to this except that we consider outsourcing. Overall literature on pharmaceutical project management employs a portfolio approach and focuses on project selection. When a manufacturer introduces a high demand drug, however, capacity planning dedicated to the drug becomes important, considering the huge risks illustrated by Tysabri and Enbrel. In addition, Myers and Howe (1997) estimate that the probability of a manufacturer introducing such a breakthrough drug is just 10% a year. Thus we take a product based approach.

Unlike these studies, Rajapakse et al. (2005) and George and Farid (2008) incorporate outsourcing into pharmaceutical project management. Rajapakse et al. (2005) simulate three scenarios, building in-house capacity early, late and outsourcing, and find that the building late option has the largest net present value if there are drug failure risks. George and Farid (2008) show in a simulation study that partnership and outsourcing are preferred to in-house operations if there is a budgetary constraint. Outsourcing in these studies is kept simple, while we consider more details, such as multiple timings to expand capacity, heterogeneous lead times, upfront payments and partial outsourcing. Partial outsourcing is especially important in our study, but operations management has paid little attention to it. Exceptions are Anderson and Parker (2002), Gray et al. (2009) and Wang et al. (2007), none of which examines partial outsourcing in capacity planning.

In supply chain contracting literature, option-like contracts, such as Barnes-Schuster et al. (2002), are similarity to our model in that they assume two procurement modes with different prices and lead times. Buyback and quantity flexibility contracts are also considered as option-like contractions. These studies are limited to at most two periods. Other studies consider capacity reservation combined with spot markets, some of which are in the multi-period setting. Serel (2007) assumes that the manufacturer can purchase inputs from its long-term supplier and a spot market. Chao et al. (2009) consider a spot market and a supply contract similar to our outsourcing contract. However, spot markets and instant delivery are not applicable to drugs. In fact, heterogeneous lead times are the key in our model, which have been studied in dual sourcing literature. For example, Veeraraghavan and

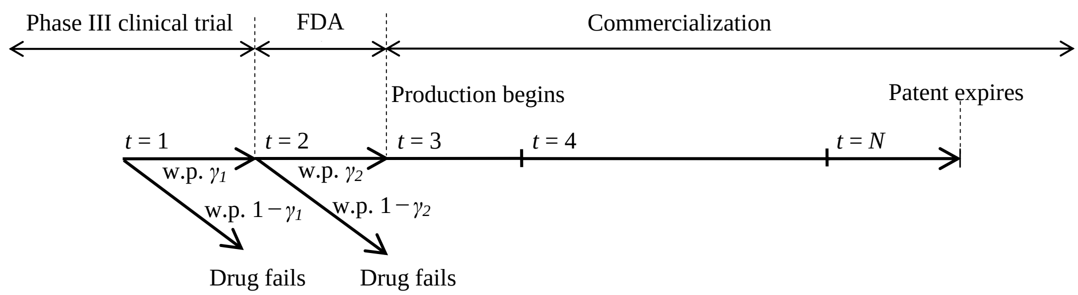

Figure 1: Planning Horizon

Scheller-Wolf (2008) study under general lead times a class of dual-index policies. Altogether, these models are studied in inventory management and the results are not transferred to our capacity planning problem, where capacity is assumed not to be dumped once installed.

3 The Model

Consider a pharmaceutical manufacturer with a potential blockbuster drug in its pipeline, facing capacity decisions for the drug. Figure 1 illustrates the planning horizon. At the beginning of the planning horizon, designated as $t=1$, the candidate drug is under the phase III clinical trial, which will last a single period (say, year), and the probability of passing this phase is γ_1. If alive, the drug will next undergo the FDA review process at the beginning of period 2, which also takes a single period in accordance with the Prescription Drug User Fee Act and the success rate is γ_2. If alive at the beginning of period 3, the drug will be produced in the next $N-2$ periods until its patent expires at the end of period N.

If the drug receives FDA approval in period 2, demand D_3 is assumed to be normally distributed with mean $\tilde{\mu}_3 > 0$ and standard deviation σ_D, where $\tilde{\mu}_3 > 0 \gg \sigma_D$ so that the probability of $D_3 < 0$ is negligible. Given the demand observed in period t is $D_t = d_t$, the demand in period $t+n$, denoted by $D_{t+n}|D_t = d_t$, is normally distributed with mean $d_t+n\mu_D$ and standard deviation $\sqrt{n}\sigma_D$, where $\mu_D > 0$ represents a growth rate. The unconditional demand in the future period $3+n$, D_{3+n}, has the mean $\tilde{\mu}_3 + n\mu_D$ and standard deviation $\sqrt{n+1}\sigma_D$. We assume that $\tilde{\mu}_3 + n\mu_D \gg \sqrt{n+1}\sigma_D$ for any n so that the probability of negative demand is negligible. D is interpreted as the discrete version of an arithmetic Brownian motion (ABM) with drift μ_D and volatility σ_D. Empirically, the mean number of prescriptions grows linearly during the first decade and then slows down with patent expiry imminent (Lichtenberg and Duflos 2009). Since the expected growth of an ABM is linear, it gives a good approximation of the demand trends of ethical drugs. Note that our demand model can accommodate any trends by letting the drift be time dependent, i.e., $\mu_D(t)$.

The manufacturer can invest in two production modes: in-house capacity and outsourcing. Following previous studies (e.g., Pindyck 1993), we model the unit in-house capacity cost $K = \{K_t : t \geqslant 0\}$ as a geometric Brownian motion (GBM) satisfying the stochastic differential equation, $dK_t = \mu_K K_t dt + \sigma_K K_t dW_t$, where dW_t is the increment of a standard Brownian motion W_t with drift μ_K and volatility σ_K. The manufacturer observes the GBM at discrete times $t = 1, 2, \cdots, N$; if the realized cost at the beginning of period t is k_t, the cost at the beginning of period $t+n$, $n \geqslant 0$, is expressed as $K_{t+n} = k_t e^{(\mu_K - \frac{1}{2}\sigma_K^2)n + \epsilon\sqrt{n}\sigma_K}$, where ϵ is a standard normal random variable. In-house capacity costs include all the costs to build a manufacturing facility, such as costs of materials, equipments, labor and inspections. We assume a two-period investment lead time. Once installed, each unit of the in-house capacity can produce one unit of the drug per period until the end of the planning horizon. We model outsourcing as a European call option. If the manufacturer pays a premium c per unit capacity to a CMO in period t, the capacity is reserved for period $t+1$. Since an option is the right, but not an obligation, to engage in an agreed-upon transaction on an asset, the manufacturer may or may not use the reserved capacity, contingent on the realized demand. Pharmaceutical manufacturers usually form long term relationships with CMOs, particularly for commercial production, and often sign agreements for one-year forecast, in which unit prices are annually adjusted (47%) or not adjusted (53%) (High Tech Business Decisions 2006). Thus, our outsourcing agreement can also be interpreted as a long-term contract reviewed annually for capacity needs with no price adjustment.

The manufacturer incurs a per unit cost p for unsatisfied demand. We let $\lambda p \geqslant c$, where $0 < \lambda \leqslant 1$ is the discount factor, to avoid the trivial case in which no option will ever be purchased. A large p is justifiable because maintaining near a 100% service level is the primary goal at the manufacturer (Singh 2005). The manufacturer also incurs a unit cost h for idle in-house capacity in each period. Unsatisfied demand is lost because patients often must choose alternative drugs if no supply is available. The production time is assumed to be negligible, that is, the demand that can be met by the available capacity will be delivered in the same period. Table 1 summarizes the sets, parameters and variables used in our model.

In each period, the manufacturer makes decisions on the units of options to purchase and the units of new in-house capacity to order. Let a_t denote the *cumulative* in-house capacity available in period t, which is increasing in t as in-house capacity will not be abandoned. The in-house capacity ordered in period t is expressed as $a_{t+2} - a_{t+1}$. Also, let θ_t be the *total* capacity available in period t, that is, θ_t is the sum of the in-house and outsourcing capacity available in period t. The options purchased in period t satisfy $\theta_{t+1} - a_{t+1}$.

The manufacturer's capacity expansion problem is modeled as a discrete-time, finite-period stochastic dynamic program. We define the state of the system at the beginning of

Table 1: Sets, Parameters and Variables

$t = 1, \cdots, N$	periods
D	demand process, $D = \{D_t, t = 3, 4, \ldots, N\}$ is a discrete ABM; d_t is a realization of D_t
μ_D	drift parameter of D;
σ_D	volatility of D
K	per unit in-house capacity cost process, $K = \{K_t, t = 1, 2, \ldots, N\}$ is a discrete GBM; k_t is a realization of K_t
μ_K	drift parameter of K
σ_K	volatility of K
c	per unit option premium, $c > 0$
p	per unit lost sales cost, $p > 0$
h	per unit in-house capacity idleness cost, $h > 0$
γ_1	success rate in Phase III clinical trial, $0 < \gamma_1 \leqslant 1$
γ_2	success rate in FDA review process, $0 < \gamma_2 \leqslant 1$
λ	discount rate, $0 < \lambda \leqslant 1$
a_t	cumulative in-house capacity available in period t, $a_t \geqslant 0$
θ_t	total (in-house plus outsourcing) capacity available in period t, $\theta_t \geqslant 0$

period t by $(a_t, a_{t+1}, \theta_t, d_t, k_t)$, where a_t and θ_t are the in-house and total capacities available at the beginning of period t, a_{t+1} is the in-house capacity available at the beginning of period $t+1$, including the new capacity ordered in period $t-1$, and d_t and k_t are the demand and unit in-house capacity cost observed at the beginning of period t. We use notation $[x]^+$ to represent $\max\{x, 0\}$. At the beginning of period t, the manufacturer will first update (a_t, a_{t+1}, θ_t) and observe k_t and $d_{t.}$. The manufacturer then allocates d_t to available in-house capacity a_t and the reserved capacity $\theta_t - a_t$. The in-house capacity will be utilized first to avoid capacity idleness costs. As such, it will allocate $\min\{d_t, a_t\}$ units of demand to the in-house capacity and $\min\{\theta_t - a_t, [d_t - a_t]^+\}$ units of demand to the CMO. Unmet demand $[d_t - \theta_t]^+$ incurs the lost sales cost p per unit and unutilized in-house capacity $[a_t - d_t]^+$ incurs the capacity idleness cost h per unit. In addition, the manufacturer needs to decide its total capacity level θ_{t+1} in the next period by purchasing options $\theta_{t+1} - a_{t+1} \geqslant 0$ and its in-house capacity level in period $t+2$ by ordering new in-house capacity $a_{t+2} - a_{t+1} \geqslant 0$. Let $u_t(a_t, a_{t+1}, \theta_t, d_t, k_t)$ denote the minimum expected total cost from period t to the end of the planning horizon. For notation simplicity, we denote $K_{t+1}(k_t) = K_{t+1}|K_t = k_t$ and $D_{t+1}(d_t) = D_{t+1}|D_t = d_t$. The optimality equations can be expressed as

$$u_t(a_t, a_{t+1}, \theta_t, d_t, k_t) = p[d_t - \theta_t]^+ + h[a_t - d_t]^+ + \min_{a_{t+2} \geqslant a_{t+1}, \theta_{t+1} \geqslant a_{t+1}} \Big\{ k_t\,(a_{t+2} - a_{t+1}) + c(\theta_{t+1} - a_{t+1}) + \lambda E\left[u_{t+1}(a_{t+1}, a_{t+2}, \theta_{t+1}, D_{t+1}(d_t), K_{t+1}(k_t))\right] \Big\}, \quad (1)$$

where we restrict $a_t \leqslant \theta_t$. The last term within the minimum operator is the expected cost from the next period to the end of the planning horizon. The optimality equations in periods 1, 2 and $N-1$ are slightly different. In period 1, the manufacturer undergoes the phase III clinical trial with success rate γ_1. As the earliest possible time for demand is period 3, the manufacturer needs no option and the only decision is the in-house capacity level. Thus,

$$u_1(0, 0, 0, 0, k_1) = \min_{a_3 \geqslant 0} \left\{ k_1 a_3 + \gamma_1 \lambda E\left[u_2(0, a_3, 0, 0, K_2(k_1))\right] \right\}. \quad (2)$$

Period 2 involves the FDA review process. As demand D_3 occurs in the next period with probability γ_2, both the in-house capacity and total capacity levels must be chosen. Thus,

$$u_2(0, a_3, 0, 0, k_2) = \min_{a_4 \geqslant a_3,\, \theta_3 \geqslant a_3} \left\{ k_2(a_4 - a_3) + c(\theta_3 - a_3) + \gamma_2 \lambda E\left[u_3(a_3, a_4, \theta_3, D_3, K_3(k_2))\right] \right\}. \quad (3)$$

In period $N-1$, we must have $a_{N-1} = a_N$ because the last in-house capacity decision must be made in period $N-2$ due to the two-period lead time. The only decision is the total capacity level and thus we have

$$u_{N-1}(a_N, a_N, \theta_{N-1}, d_{N-1}, k_{N-1}) = p[d_{N-1} - \theta_{N-1}]^+ + h[a_N - d_{N-1}]^+ + \min_{\theta_N \geqslant a_N} \left\{ c(\theta_N - a_N) + \lambda E\left[u_N\left(a_N, 0, \theta_N, D_N(d_{N-1}), K_N(k_{N-1})\right)\right] \right\}. \quad (4)$$

Finally, the terminal value in period N is

$$u_N(a_N, 0, \theta_N, d_N, k_N) = p[d_N - \theta_N]^+ + h[a_N - d_N]^+, \tag{5}$$

subject to $\theta_N \geqslant a_N$.

Observe from Eq. (1) that function $u_t(a_t, a_{t+1}, \theta_t, d_t, k_t)$ depends on θ_t and a_t only for $p[d_t - \theta_t]^+$ and $h[a_t - d_t]^+$, which are sunk costs as θ_t and a_t are determined prior to time t. For convenience we define, for $1 \leqslant t \leqslant N$,

$$w_t(a_{t+1}, d_t, k_t) = u_t(a_t, a_{t+1}, \theta_t, d_t, k_t) - p[d_t - \theta_t]^+ - h[a_t - d_t]^+, \tag{6}$$

with $u_t \equiv w_t$ for $t = 1$ and 2 and $w_N \equiv 0$. The optimization problems (1) and (6) are equivalent and it is sufficient to focus on (6), which we can write as

$$\begin{aligned} w_t(a_{t+1}, d_t, k_t) &= \min_{a_{t+2} \geqslant a_{t+1}, \theta_{t+1} \geqslant a_{t+1}} \Big\{ k_t\ (a_{t+2} - a_{t+1}) + c(\theta_{t+1} - a_{t+1}) \\ &\quad + \lambda p E[D_{t+1}(d_t) - \theta_{t+1}]^+ + \lambda h E[a_{t+1} - D_{t+1}(d_t)]^+ \\ &\quad + \lambda E\left[w_{t+1}\left(a_{t+2}, D_{t+1}(d_t), K_{t+1}(k_t)\right)\right] \Big\}. \end{aligned} \tag{7}$$

Eq. (7) is separable, that is, total capacity decision, θ_{t+1}, and in-house capacity decision, a_{t+2}, can be made separately. This is intuitive because the options purchased in the current period can only be used in the next period and have no a long-term effect on the future cost. Therefore, for $3 \leqslant t \leqslant N-2$, we can decompose Eq. (7) into two optimization problems:

$$\begin{aligned} w_t(a_{t+1}, d_t, k_t) &= \lambda h E[a_{t+1} - D_{t+1}(d_t)]^+ - c a_{t+1} - k_t a_{t+1} \\ &\quad + \min_{\theta_{t+1} \geqslant a_{t+1}} \left\{ c\theta_{t+1} + \lambda p E[D_{t+1}(d_t) - \theta_{t+1}]^+ \right\} \qquad (8) \\ &\quad + \min_{a_{t+2} \geqslant a_{t+1}} \left\{ k_t a_{t+2} + \lambda E\left[w_{t+1}(a_{t+2}, D_{t+1}(d_t), K_{t+1}(k_t))\right]\right\}, \qquad (9) \end{aligned}$$

where we factor out the terms irrelevant to the current optimal decisions from the minimum operator. In the next two sections, we discuss the solutions to Eqs. (8) and (9).

3.1 Total Capacity Policy

Consider the total capacity decision θ_{t+1} first. The next proposition states that for each period t, there exists a total base-capacity level $\bar{\theta}_{t+1}(d_t)$, which depends only on the realized demand d_t, such that if the available in-house capacity in the next period is less than $\bar{\theta}_{t+1}(d_t)$, the manufacturer should make up the difference by purchasing options in the current period.

Proposition 1 *(Total base-capacity level): The optimal total capacity $\theta^*_{t+1}(d_t)$ is determined by the total base-capacity level $\bar{\theta}_{t+1}(d_t)$ as*

$$\theta^*_{t+1}(d_t) = \max\{\bar{\theta}_{t+1}(d_t), a_{t+1}\} \quad 2 \leqslant t \leqslant N-1, \tag{10}$$

where $\bar{\theta}_3(d_2) \equiv \bar{\theta}_3$. The total base-capacity level $\bar{\theta}_{t+1}(d_t)$ is given by

$$\bar{\theta}_{t+1}(d_t) = \begin{cases} d_t + \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right) & 3 \leqslant t \leqslant N-1 \\ \tilde{\mu}_3 + \sigma_D \Phi^{-1}\left(\frac{\gamma_2 \lambda p - c}{\gamma_2 \lambda p}\right) & t = 2 \end{cases} \tag{11}$$

where Φ is the standard normal cdf.

Proof: Consider $3 \leqslant t \leqslant N-1$. The unconstrained optimization problem of Eq. (8) is convex in θ_{t+1}, which is a newsvendor problem with the overage cost c and underage cost $\lambda p - c$. The optimal base-capacity level $\bar{\theta}_{t+1}(d_t)$ is determined by the fractile, $F_{D_{t+1}(d_t)}(\bar{\theta}_{t+1}) = \frac{\lambda p - c}{\lambda p}$. Since $D_{t+1}(d_t) \sim N(d_t + \mu_D, \sigma_D)$, the expression reduces to $\Phi\left(\frac{\bar{\theta}_{t+1}(d_t) - d_t - \mu_D}{\sigma_D}\right) = \frac{\lambda p - c}{\lambda p}$, leading to the first expression of Eq.(11). Imposing $\theta_{t+1} \geqslant a_{t+1}$, we obtain the optimal total capacity given in Eq. (10). The optimality equation for period 2 is solved similarly. □

Substituting the optimal solution $\bar{\theta}_{t+1}(d_t)$ into Eq. (8), we obtain, for $3 \leqslant t \leqslant N-2$,

$$\begin{aligned} w_t(a_{t+1}, d_t, k_t) &= \lambda h E[a_{t+1} - D_{t+1}(d_t)]^+ - k_t a_{t+1} + c[\bar{\theta}_{t+1}(d_t) - a_{t+1}]^+ \\ &\quad + \lambda p E[D_{t+1}(d_t) - \max\{a_{t+1}, \bar{\theta}_{t+1}(d_t)\}]^+ + \min_{a_{t+2} \geqslant a_{t+1}} \{g_t(a_{t+2}, d_t, k_t)\}, \end{aligned} \tag{12}$$

where

$$g_t(a_{t+2}, d_t, k_t) \quad = \quad k_t a_{t+2} + \lambda E\left[w_{t+1}(a_{t+2}, D_{t+1}(d_t), K_{t+1}(k_t))\right], \tag{13}$$

with $a_{N-1} \equiv a_N$. Next, we study the solution to the optimization problem (9).

3.2 In-house Capacity Policy: N=3

First we consider a special case with $N = 3$. To generalize the discussion, assume that the drug is already commercialized in period 1. We have a single in-house capacity decision in period 1 and wish to show that it is characterized by a base-capacity level. To this end, we show that the value functions $w_2(a_3, d_2, k_2)$ and $g_1(a_3, d_1, k_1)$ are quasi-convex in a_3, resorting to Proposition 3.1 in Cheng and Sethi (1999) (The result is due to Porteus 1971).

Lemma 1. If the density function $f(\xi)$ is log-concave and the function $h(a)$ is quasi-convex in a, then the convolution of functions f and h is quasi-convex. That is,

$$h * f(a) = \int_0^\infty h(a-\xi)f(\xi)d\xi$$

is quasi-convex in a.

Proposition 2 *1.* $w_2(a_3, d_2, k_2)$ *is a quasi-convex function of* a_3.

2. $g_1(a_3, d_1, k_1)$ *is a quasi-convex function of* a_3.

Proof of Part 1. Using Proposition 1, we can write Eq. (12) as

$$w_2(a_3, d_2) = \lambda h E[a_3 - D_3(d_2)]^+ + c[\bar{\theta}_3(d_2) - a_3]^+ + \lambda p E[D_3(d_2) - \max\{\bar{\theta}_3(d_2), a_3\}]^+$$

$$= \begin{cases} \lambda h E[a_3 - D_3(d_2)]^+ + c(\bar{\theta}_3(d_2) - a_3) + \lambda p E[D_3(d_2) - \bar{\theta}_3(d_2)]^+ & \bar{\theta}_3(d_2) \geqslant a_3 \\ \lambda h E[a_3 - D_3(d_2)]^+ + \lambda p E[D_3(d_2) - a_3]^+ & \bar{\theta}_3(d_2) < a_3 \end{cases} \tag{14}$$

where we omit the capacity cost k_2 in w_2 as no in-house capacity will be ordered in period 2 and also omit the last argument in Eq. (12) because no capacity decision will be made in period 3. We shall show that $w_2(a_3, d_2)$ is decreasing for $a_3 \leqslant \bar{\theta}_3(d_2)$ and convex for $a_3 \geqslant \bar{\theta}_3(d_2)$, that the two expressions meet when $a_3 = \bar{\theta}_3(d_2)$ and that $w_2(a_3, d_2)$ is quasi-convex in the entire domain $a_3 \in [0, \infty)$. We first consider each expression in its own domain.

1. $\bar{\theta}_3(d_2) \geqslant a_3$: From Leibniz integral rule, the derivative of $w_2(a_3, d_2)$ w.r.t. a_3 is

$$\frac{\partial w_2(a_3, d_2)}{\partial a_3} = \lambda h \frac{\partial}{\partial a_3} \int_0^\infty P(a_3 - D_3(d_2) \geqslant x) dx - c = \lambda h \Phi\left(\frac{a_3 - d_2 - \mu_D}{\sigma_D}\right) - c$$

   ϕ is the standard normal pdf. The first derivative is negative when $c > \lambda h$, the second derivative is $\frac{\lambda h}{\sigma_D} \phi\left(\frac{a_3 - d_2 - \mu_D}{\sigma_D}\right) > 0$, and thus $w_2(a_3, d_2)$ is decreasing convex for $a_3 \leqslant \bar{\theta}_3(d_2)$.

2. $\bar{\theta}_3(d_2) < a_3$: The first derivative of $w_2(a_3, d_2)$ w.r.t. a_3 is given by

$$\begin{aligned} \frac{\partial w_2(a_3, d_2)}{\partial a_3} &= \lambda h \frac{\partial}{\partial a_3} \int_0^\infty P(a_3 - D_3(d_2) \geqslant x) dx + \lambda p \frac{\partial}{\partial a_3} \int_0^\infty P(D_3(d_2) - a_3 \geqslant x) dx \\ &= \lambda (h + p) \Phi\left(\frac{a_3 - d_2 - \mu_D}{\sigma_D}\right) - \lambda p. \end{aligned}$$

   The second derivative is $\frac{\lambda(h+p)}{\sigma_D} \phi\left(\frac{a_3 - d_2 - \mu_D}{\sigma_D}\right) > 0$. Thus $w_2(a_3, d_2)$ is convex for $a_3 > \bar{\theta}_3(d_2)$.

Next, $w_2(a_3, d_2)$ is continuous and differentiable at $a_3 = \bar{\theta}_3(d_2)$, with the derivative equal to

$$\left.\frac{\partial w_2(a_3, d_2)}{\partial a_3}\right|_{a_3 = \bar{\theta}_3(d_2)} = \lambda h \left(\frac{\lambda p - c}{\lambda p}\right) - c \leqslant 0,$$

implying that $w_2(a_3, d_2)$ is decreasing at $a_3 = \bar{\theta}_3(d_2)$. As a decreasing function connected with a convex function must be quasi-convex, $w_2(a_3, d_2)$ is quasi-convex in a_3.

Proof of Part 2: Conditioning on $D_2(d_1) = \xi$ and noting $E[D_2(d_1)] = d_1 + \mu_D$, we have

$$\begin{aligned} g_1(a_3, d_1, k_1) &= k_1 a_3 + \lambda E\left[w_2(a_3, D_2(d_1))\right] \\ &= k_1(d_1 + \mu_D) + \int_0^\infty \left(k_1(a_3 - \xi) + \lambda w_2(a_3, \xi)\right) f_{D_2(d_1)}(\xi) d\xi, \end{aligned} \quad (15)$$

where $w_2(a_3, \xi)$ is defined by Eq. (14) with d_2 replaced by ξ. The integrand is written as

$$k_1(a_3-\xi)+\lambda w_2(a_3,\xi)=\begin{cases} k_1(a_3-\xi)+\lambda^2 hE[a_3-D_3(\xi)]^+ + \lambda c(\bar{\theta}_3(\xi)-a_3) & \\ +\lambda^2 pE[D_3(\xi)-\bar{\theta}_3(\xi)]^+ & \bar{\theta}_3(\xi)\geqslant a_3 \\ k_1(a_3-\xi)+\lambda^2 hE[a_3-D_3(\xi)]^+ + \lambda^2 pE[D_3(\xi)-a_3]^+ & \bar{\theta}_3(\xi)< a_3 \end{cases}$$

We can understand the above expression as follows. We condition on the demand in period 2 being ξ. If the in-house capacity a_3 is less than the total base-capacity level $\bar{\theta}_3(\xi)$, we purchase $\bar{\theta}_3(\xi)-a_3$ units of options in period 2 at cost $\lambda c(\bar{\theta}_3(\xi)-a_3)$, and incur the capacity idleness cost in period 3 if demand is less than a_3 and the lost sales cost if demand is more than $\bar{\theta}_3(\xi)$. On the other hand, if a_3 is more than $\bar{\theta}_3(\xi)$, we do not purchase any options in period 2, and incur either the capacity idleness cost or lost sales cost depending on whether the last period demand is less than the in-house capacity or not. Since $D_3(\xi)=\xi+\mu_D+\sigma_D Z$ and $\bar{\theta}_3(\xi)=\xi+\mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)$, where Z is a standard normal random variable, we can rewrite Eq. (16) as a function of $a_3-\xi$, as follows:

$$h_1(a_3-\xi)=k_1(a_3-\xi)+\lambda w_2(a_3,\xi)$$

$$=\begin{cases} k_1(a_3-\xi)+\lambda^2 hE[a_3-\xi-\mu_D-\sigma_D Z]^+ + \lambda c\left(\mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)-(a_3-\xi)\right) & \\ +\lambda^2 pE\left[\sigma_D Z-\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)\right]^+ & \mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)\geqslant a_3-\xi \\ k_1(a_3-\xi)+\lambda^2 hE[a_3-\xi-\mu_D-\sigma_D Z]^+ & \\ +\lambda^2 pE[\mu_D+\sigma_D Z-(a_3-\xi)]^+ & \mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)< a_3-\xi \end{cases}$$

where

$$h_1(a)=\begin{cases} k_1 a+\lambda^2 hE[a-(\mu_D+\sigma_D Z)]^+ + \lambda c\left(\mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)-a\right) & \\ +\lambda^2 pE\left[(\mu_D+\sigma_D Z)-\left(\mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)\right)\right]^+ & \mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)\geqslant a \\ k_1 a+\lambda^2 hE[a-(\mu_D+\sigma_D Z)]^+ & \\ +\lambda^2 pE[(\mu_D+\sigma_D Z)-a]^+ & \mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)< a \end{cases}$$

To show the quasi-convexity of $h_1(a)$, we again consider each expression in its own domain.

1. $\mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right) \geqslant a$: The first derivative of $h_1(a)$ is

$$\frac{\partial h_1(a)}{\partial a} = (k_1 - \lambda c) + \lambda^2 h \Phi\left(\frac{a - \mu_D}{\sigma_D}\right) \tag{16}$$

   As the second derivative is $\frac{\lambda^2 h}{\sigma_D}\phi\left(\frac{a-\mu_D}{\sigma_D}\right) \geqslant 0$, $h_1(a)$ is a convex function in this domain.

2. $\mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right) < a$: The first derivative of $h_1(a)$ is

$$\frac{\partial h_1(a)}{\partial a} = k_1 + \lambda^2 h \Phi\left(\frac{a - \mu_D}{\sigma_D}\right) - \lambda^2 p\left(1 - \Phi\left(\frac{a - \mu_D}{\sigma_D}\right)\right) \tag{17}$$

   As the second derivative is $\lambda^2(h+p)\phi\left(\frac{a-\mu_D}{\sigma_D}\right) \geqslant 0$, $h_1(a)$ is a convex in this domain.

Next, we argue that function $h_1(a)$ is a quasi-convex function of a. The function is continuous and differentiable at $a = \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$, with the first derivative satisfying

$$\left.\frac{\partial h_1(a)}{\partial a}\right|_{a=\mu_D+\sigma_D\Phi^{-1}\left(\frac{\lambda p-c}{\lambda p}\right)} = k_1 + \lambda^2(h+p)\frac{\lambda p - c}{\lambda p} - \lambda^2 p.$$

If the above expression is positive, $h_1(a)$ is convex for $a \leqslant \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$ and increasing and convex for $a \geqslant \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$, implying that $h_1(a)$ in the entire domain $a \in [0, \infty)$ is quasi-convex. If the above function is negative, $h_1(a)$ is decreasing and convex for $a \leqslant \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$ and convex for $a \geqslant \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$, implying again that $h_1(a)$ in the entire domain $a \in [0, \infty)$ is quasi-convex. From Eq. (15),

$$g_1(a_3, d_1, k_1) = k_1(d_1 + \mu_D) + \int_0^\infty h_1(a_3 - \xi) f_{D_2(d_1)}(\xi) d\xi. \tag{18}$$

The first term is a constant independent of a_3, the second term is a convolution of a quasi-convex function and a log-concave density function and hence is a quasi-convex function of a_3 by Lemma 1. Therefore, $g_1(a_3, d_1, k_1)$ is quasi-convex in a_3, which completes the proof. □

Next we show that there exists a base-capacity level for period 3 so that the manufacturer brings its in-house capacity to this level as closely as possible. As period 1 is the decision period for in-house capacity and the total capacity in period 2 is prescribed by Proposition 1, we may treat Eq. (18) as the cost function of a newsvendor problem and determine the optimal in-house capacity level based on the tradeoff between the overage and underage costs. Suppose we bring the in-house capacity in period 3 to a_3. Then the expected overage costs include the unit in-house capacity cost k_1 and in-house capacity idleness cost $\lambda^2 h$, which occurs if the period 3 demand d_3 is less than a_3. Thus the expected overage cost is

$$\text{Expected overage cost} = k_1 + \lambda^2 h P(D_3 \leqslant a_3 | D_1 = d_1).$$

The expected underage costs include the option premium λc, which happens in period 2 if a_3 is less than the total base-capacity level $\bar{\theta}_3(D_2(d_1))$, prescribed by Proposition 1, and the lost sales cost $\lambda^2 p$, which incurs if a_3 is more than $\bar{\theta}_3(D_2(d_1))$ (so no option is purchased in period 2) and d_3 is more than a_3. Combining the two underage costs gives

$$\text{Expected underage cost} = \lambda c P(\bar{\theta}_3(D_2(d_1) > a_3) + \lambda^2 p P\left(D_3 > a_3, \bar{\theta}_3(D_2) \leqslant a_3 | D_1 = d_1\right).$$

The optimal in-house base-capacity level in period 3, denoted by $\bar{a}_3 \equiv \bar{a}_3(d_1, k_1)$, should be the value equalizing the expected overage and underage cost, that is,

$$\begin{aligned} & k_1 + \lambda^2 h P(D_3 \leqslant \bar{a}_3 | D_1 = d_1) \\ & = \lambda c P\left(\bar{\theta}_3(D_2(d_1)) > \bar{a}_3\right) + \lambda^2 p P\left(D_3 > \bar{a}_3, \bar{\theta}_3(D_2) \leqslant \bar{a}_3 | D_1 = d_1\right). \end{aligned} \tag{19}$$

Next, we formally state the above result and provide a rigorous proof.

Proposition 3 *(Optimal in-house base-capacity level:*$N = 3$*): The optimal in-house ca-*

pacity $a_3^*(a_2,d_1,k_1)$ in period 3 is determined by the in-house base-capacity level $\bar{a}_3(d_1,k_1)$ as

$$a_3^*(a_2,d_1,k_1) = \max\{\bar{a}_3(d_1,k_1), a_2\}, \tag{20}$$

where the in-house base-capacity level $\bar{a}_3(d_1,k_1)$ is the solution of Eq. (19).

Proof. We take the derivative of Eq. (18) wrt a_3 and set it to zero:

$$\frac{\partial g_1(a_3,d_1,k_1)}{\partial a_3} = \int_0^\infty \frac{\partial h_1(a_3-\xi)}{\partial a_3} f_{D_2(d_1)}(\xi)d\xi = 0.$$

Let $\bar{\theta}_3^{-1}(a_3) = a_3 - \mu_D - \sigma_D \Phi^{-1}(\frac{\lambda p - c}{\lambda p})$. From Eqs. (16) and (17), it works out that

$$\begin{aligned}
&\frac{\partial g_1(a_3,d_1,k_1)}{\partial a_3} = \int_{\bar{\theta}_3^{-1}(a_3)}^\infty \left((k_1-\lambda c) + \lambda^2 h \Phi\left(\frac{a_3-\xi-\mu_D}{\sigma_D}\right)\right) f_{D_2(d_1)}(\xi)d\xi \\
&+ \int_0^{\bar{\theta}_3^{-1}(a_3)} \left(k_1 + \lambda^2 h \Phi\left(\frac{a_3-\xi-\mu_D}{\sigma_D}\right) - \lambda^2 p \left(1-\Phi\left(\frac{a_3-\xi-\mu_D}{\sigma_D}\right)\right)\right) f_{D_2(d_1)}(\xi)d\xi \\
&= k_1 + \lambda^2 h P(D_3(d_1) \leqslant a_3) \\
&-\lambda c P(D_2(d_1) > \bar{\theta}_3^{-1}(a_3)) - \lambda^2 p P(D_3(d_1) > a_3, D_2(d_1) \leqslant \bar{\theta}_3^{-1}(a_3)).
\end{aligned} \tag{21}$$

Note that $P(D_2(d_1) > \bar{\theta}_3^{-1}(a_3)) = P\left(\bar{\theta}_3(D_2(d_1)) > \bar{\theta}_3(\bar{\theta}_3^{-1}(a_3))\right) = P\left(\bar{\theta}_3(D_2(d_1)) > a_3\right)$. Similarly, $P(D_3 > a_3, D_2 \leqslant \bar{\theta}_3^{-1}(a_3)|D_1 = d_1) = P\left(D_3 > a_3, \bar{\theta}_3(D_2) \leqslant a_3 | D_1 = d_1\right)$. Substituting these two expressions into Eq. (21), we derive Eq. (19). □

Hence, we have

$$\begin{aligned}
g_1^*(a_2,d_1,k_1) &= \min_{a_3 \geq a_2} \{g_1(a_3,d_1,k_1)\} \\
&= \begin{cases} k_1 \bar{a}_3(d_1,k_1) + \lambda E\left[w_2(\bar{a}_3(d_1,k_1), D_2(d_1))\right] & a_2 < \bar{a}_3(d_1,k_1) \\ k_1 a_2 + \lambda E\left[w_2(a_2, D_2(d_1))\right] & a_2 \geqslant \bar{a}_3(d_1,k_1) \end{cases}
\end{aligned}$$

Unfortunately, when $N > 3$, the value function (12) is ill-behaved and not convex nor quasi-convex in a_{t+1} (e.g. Figure 2). To proceed, we next discuss approximation methods.

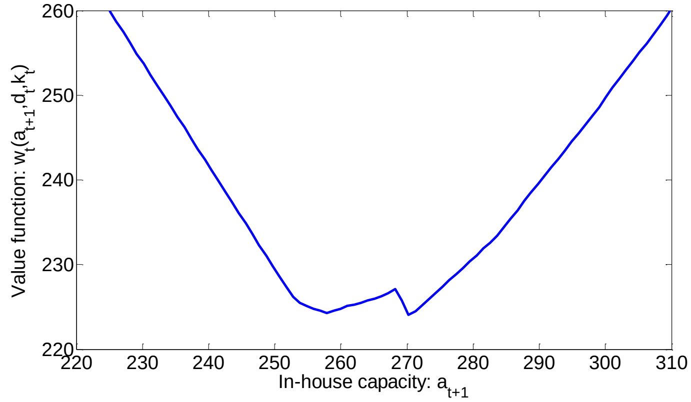

Figure 2: Value Function $w_t(a_{t+1}, d_t, k_t)$ in a General Case with $N > 3$

3.3 In-house capacity policy: $N > 3$

3.3.1 Myopic policy

We first use the previous result to approximate the optimal in-house capacity policy. Noting that the problem examined in (19) is equivalent to a general N-period problem with 3 periods remaining, we approximate the optimal in-house capacity, denoted by $\hat{a}^*_{t+2}(a_{t+1}, d_t, k_t)$, as

$$\hat{a}^*_{t+2}(a_{t+1}, d_t, k_t) = \max\{\hat{a}_{t+2}(d_t, k_t), a_{t+1}\} \quad 1 \leqslant t \leqslant N-2 \tag{22}$$

where the approximate in-house base-capacity level $\hat{a}_{t+2} \equiv \hat{a}_{t+2}(d_t, k_t)$ is the solution to

$$\begin{aligned} &\beta_t k_t + \lambda^2 h P(D_{t+2} \leqslant \hat{a}_{t+2} | D_t = d_t) \\ &= \lambda c P\left(\bar{\theta}_{t+2}(D_{t+1}(d_t)) > \hat{a}_{t+2}\right) + \lambda^2 p P\left(D_{t+2} > \hat{a}_{t+2}, \bar{\theta}_{t+2}(D_{t+1}) \leqslant \hat{a}_{t+2} | D_t = d_t\right). \end{aligned} \tag{23}$$

This solution is *myopic* in that in every period t, $1 \leqslant t \leqslant N-2$, we pretend that this is the last in-house capacity decision and determine the in-house base-capacity level by considering the costs in periods $t+1$ and $t+2$ but no cost in period $t+3$ onwards. β_t is a modifier to the unit in-house capacity cost k_t, which adjusts k_t to this short planning window so that it is comparable to the option premium c. For instance, we may use $\beta_t = \frac{1}{N-t-1}$. The intuition is that since the in-house capacity ordered in period t will be used for $N-t-1$ periods, we may allocate k_t evenly over the production periods. The myopic in-house capacity policy is easy to calculate owing to the closed form solution given in Eq. (23).

3.3.2 Optimal policy for the approximate value function

Next, we approximate the value function and derive the optimal solution for it. The non-convexity of the value function (12) comes from the lost sales cost, which is defined as

$$\Delta_t(a_{t+1}, d_t) \equiv \begin{cases} \lambda p E[D_{t+1}(d_t) - \max\{a_{t+1}, \bar{\theta}_{t+1}(d_t)\}]^+ & 3 \leqslant t \leqslant N-1 \\ \gamma_2 \lambda p E[D_3 - \max\{a_3, \bar{\theta}_3\}]^+ & t = 2 \end{cases} \tag{24}$$

with $\Delta_2(a_3, d_2) \equiv \Delta_2(a_3)$ and $\Delta_1 = \Delta_N \equiv 0$.

Proposition 4 *The lost sales cost in the value function* (12), *defined in Eq.*(24), *satisfies*

$$\lim_{p\to\infty} \Delta_t(a_{t+1}, d_t) = 0 \quad 2 \leqslant t \leqslant N-1.$$

Proof. For $3 \leqslant t \leqslant N-1$, we have the following:

$$\begin{aligned} \lim_{p\to\infty} \Delta_t(a_{t+1}, d_t) &= \lim_{p\to\infty} \lambda \sigma_D p E\left[Z - \bar{\Phi}^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)\right]^+ & (25) \\ &\leftrightarrow \lim_{K\to\infty} \frac{\sigma_D c}{1-\Phi(K)} E[Z-K]^+ & (26) \\ &= \lim_{K\to\infty} \sigma_D c \left[\frac{\phi(K) - K(1-\Phi(K))}{1-\Phi(K)}\right] & (27) \end{aligned}$$

Eq. (25) is obtained by substituting $\bar{\theta}_{t+1}(d_t) = d_t + \mu_D + \sigma_D \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$ from Proposition 1 and $D_{t+1}(d_t) = d_t + \mu_D + \sigma_D Z$ into Eq. (24) and assuming sufficiently large p, meaning $\max\{a_{t+1}, \bar{\theta}_{t+1}(d_t)\} = \bar{\theta}_{t+1}(d_t)$. For notation simplicity, we let $K = \Phi^{-1}\left(\frac{\lambda p - c}{\lambda p}\right)$, which can be expressed as $p = \frac{c}{\lambda(1-\Phi(K))}$ after some algebra. By substituting K and p into Eq. (25)

and noting that K approaches to infinity as p approaches to infinity, we get Eq. (26). Eq. (27) follows because $E[Z-K]^{+}$is a standard normal linear loss function. Clearly, the denominator of Eq. (27) approaches to zero when K approaches to infinity. The numerator also approaches to zero when K goes to infinity because $\lim_{K\to\infty}\phi(K)=0$ and

$$\lim_{K\to\infty} K\left(1-\Phi(K)\right)=\lim_{K\to\infty}\frac{K}{\frac{1}{1-\Phi(K)}}=\lim_{K\to\infty}\frac{1}{\frac{\phi(K)}{(1-\Phi(K))^2}}=\lim_{K\to\infty}\frac{2\phi(K)\left(1-\Phi(K)\right)}{K\phi(K)}=0,$$

where we apply L'Hôpital's rule to indeterminate forms repeatedly and use the fact, $\phi'(K)=-K\phi(K)$. Therefore, Eq. (27) is also an indeterminate form. After applying L'Hôpital's rule several times, Eq. (27) become zero. A similar argument holds for $t=2$. $\square$

Proposition 4 is intuitive because as p gets larger, the optimal total base-capacity level $\bar{\theta}_{t+1}(d_t)$ becomes higher and thus the probability diminishes that the demand exceeds $\bar{\theta}_{t+1}(d_t)$. Based on this, we propose the approximate value function $\tilde{w}_t(a_{t+1},d_t,k_t)$ defined as follows:

$$\tilde{w}_t(a_{t+1},d_t,k_t)\equiv w_t(a_{t+1},d_t,k_t)-\Delta_t(a_{t+1},d_t)\quad 1\leqslant t\leqslant N-1. \tag{28}$$

Accordingly we substitute $w_t(a_{t+1},d_t,k_t)$ in Eq. (13) with $\tilde{w}_t(a_{t+1},d_t,k_t)$ and define the function as $\tilde{g}_t(a_{t+2},d_t,k_t)$. Proposition 5 says that $\tilde{w}_t(a_{t+1},d_t,k_t)$ has a base-capacity policy.

Proposition 5: *(Optimal in-house base-capacity level for approximate value function):The optimal policy $\tilde{a}^*_{t+2}(a_{t+1},d_t,k_t)$ for the approximate value function (28) is determined by the in-house base-capacity level $\tilde{a}_{t+2}(d_t,k_t)$ as*

$$\tilde{a}^*_{t+2}(a_{t+1},d_t,k_t)=\max\{\tilde{a}_{t+2}(d_t,k_t),a_{t+1}\}\qquad 1\leqslant t\leqslant N-2, \tag{29}$$

where $\tilde{a}_{t+2}(d_t,k_t)$ is the minimizer of the approximate value function (28).

Proof: It suffices to show the convexity of the approximate value function (28). Employing backward induction, we first consider the last decision period $t=N-2$.

$\tilde{g}_{N-2}(a_N, d_{N-2}, k_{N-2}) = k_{N-2}a_N + \lambda E\left[\tilde{w}_{N-1}\left(a_N, D_{N-1}(d_{N-2})\right)\right]$ is convex in a_N because $\tilde{w}_{N-1}(a_N, d_{N-1}) = \lambda h E\left[a_N - D_N(d_{N-1})\right]^+ + c\left[\bar{\theta}_N(d_{N-1}) - a_N\right]^+$ is convex in a_N, and expectation and summation preserve convexity. Next, we assume that $\tilde{w}_{t+1}(a_{t+2}, d_{t+1}, k_{t+1})$ and $\tilde{g}_t(a_{t+2}, d_t, k_t)$ are convex in a_{t+2} and shall show that $\tilde{w}_t(a_{t+1}, d_t, k_t)$ and $\tilde{g}_{t-1}(a_{t+1}, d_{t-1}, k_{t-1})$ are convex in a_{t+1}. From the induction hypothesis, $\tilde{g}_t(a_{t+2}, d_t, k_t)$ has a global minimizer, denoted by $\tilde{a}_{t+2}(d_t, k_t)$, and its constrained optimization problem is solved as

$$\min_{a_{t+2} \geqslant a_{t+1}} \{\tilde{g}_t(a_{t+2}, d_t, k_t)\} = \begin{cases} \tilde{g}_t(\tilde{a}_{t+2}(d_t, k_t), d_t, k_t) & a_{t+1} \leqslant \tilde{a}_{t+2}(d_t, k_t) \\ \tilde{g}_t(a_{t+1}, d_t, k_t) & a_{t+1} > \tilde{a}_{t+2}(d_t, k_t) \end{cases}$$

The above expression is constant for $a_{t+1} \leqslant \tilde{a}_{t+2}(d_t, k_t)$, increasing and convex for $a_{t+1} > \tilde{a}_{t+2}(d_t, k_t)$ and thus convex in the entire domain, $a_{t+1} \in [0, \infty)$. Then, $\tilde{w}_t(a_{t+1}, d_t, k_t) = \lambda h E[a_{t+1} - D_{t+1}(d_t)]^+ - k_t a_{t+1} + c[\bar{\theta}_{t+1}(d_t) - a_{t+1}]^+ + \min_{a_{t+2} \geqslant a_{t+1}} \{\tilde{g}_t(a_{t+2}, d_t, k_t)\}$ is also convex in a_{t+1} because all the terms are convex in a_{t+1}. Finally, it is clear that $\tilde{g}_{t-1}(a_{t+1}, d_{t-1}, k_{t-1}) = k_{t-1}a_{t+1} + \lambda E\left[\tilde{w}_t\left(a_{t+1}, D_t(d_{t-1}), K_t(k_{t-1})\right)\right]$ is convex in a_{t+1}. Hence, $\tilde{w}_t(a_{t+1}, d_t, k_t)$ and $\tilde{g}_{t-1}(a_{t+1}, d_{t-1}, k_{t-1})$ are convex in a_{t+1} and this concludes the proof. □

The approximate value function is convex and we can find the optimal solution. The accuracy of these two approximation methods are discussed in the next section.

4 Numerical analysis

4.1 Methodology

We employed Monte Carlo simulation and interpolation to calculate the optimal policies. For reduced-size problems, we calculated the exact solution using Monte Carlo simulation;

Table 2: Interpolation vs. Exact Solution: Absolute Percentage Errors

			Initial in-house capacity cost (k_1): 12		14		16	
			In-house capacity cost volatility (σ_K)		In-house capacity cost volatility (σ_K)		In-house capacity cost volatility (σ_K)	
			0.05	0.1	0.05	0.1	0.05	0.1
Customer service level 95%	Demand volatility (σ_D)	15	0.03%	0.04%	0.08%	0.68%	0.00%	0.00%
		30	0.05%	0.04%	0.07%	0.12%	0.00%	0.00%
Customer service level 99%	Demand volatility (σ_D)	15	0.05%	0.05%	0.07%	0.65%	0.00%	0.01%
		30	0.28%	0.24%	0.07%	0.11%	0.00%	0.00%

we generated a tree of sample paths of a demand and unit in-house capacity cost pair, (D_t, K_t), with a branching number 30, and calculated the optimal policy and corresponding cost for each sample path. This method is limited to small problems as the tree grows exponentially. To overcome this problem, we next employed interpolation to estimate value functions. To examine the accuracy of the interpolation, we compared the optimal costs of the policies estimated by interpolation to the optimal costs of exact policies calculated from Monte Carlo simulation in 5-period problems for various parameter sets (Table 2). The same tree of sample paths with a branching number 30 was used in both methods. The absolute percentage errors are less than 1% in all cases, indicating no large errors from interpolation in larger problems as well. Thus, we employed interpolation in the rest of the analysis.

The parameters are set as follows. A planning horizon is $N = 15$ periods, reflecting the average duration from the late stage clinical trail to patent expiration. Demand follows an ABM with drift $\mu_D = 25$ and volatility $\sigma_D = \{15, 30\}$, with initial demand $D_3 \sim N(100, \sigma_D)$. The drift is approximated as the industry average of high demand drugs. The option premium is $c = 10$, whereas the unit in-house capacity cost K follows a GBM with drift $\mu_K = 0.05$ and volatility $\sigma_K = \{0.05, 0.1\}$, with initial cost $k_1 = \{20, 40, 60\}$. The drift matches the discount rate $\lambda = 0.95\%$. The unit lost sales cost p satisfies customer service levels $CSL = \{95\%, 99\%\}$ and the unit in-house capacity idleness cost h is 5% of k_1. Finally, the success rates in the phase III clinical trial and FDA review are $\gamma_1 = 60\%$ and $\gamma_2 = 85\%$, respectively, based on previous literature (DiMasi et al. 2010). This leads to 24 combinations of parameters.

For each parameter combination, we generated 10,000 sample paths of a pair $(D_t, K_t), 1 \leqslant t \leqslant N$, and calculated the optimal policy and corresponding cost for each sample path by interpolation. Then an average was taken over all the sample paths. In order to see the effect of a particular parameter on the optimal policy, we grouped the combinations by the specific parameter values and averaged their optimal policies and costs.

4.2 The optimal capacity strategies

We discuss the main findings of our numerical analysis regarding the manufacturer's optimal in-house capacity and outsourcing strategies. The overall strategy is summarized as follows. The manufacturer defers building in-house capacity during the phase III clinical trial and uses the CMO as the sole source of capacity in all cases. Once the drug passes the phase III clinical trial, the manufacturer starts building in-house capacity, but a significant part of the demand is still met by the CMO during the FDA review process. Thus, under clinical trials uncertainty, the CMO plays a major role in mitigating large capacity overage risks. When the drug finally receives FDA approval, large in-house capacity is added immediately to satisfy early demands, followed by gradual expansion in the next several periods. Meanwhile,

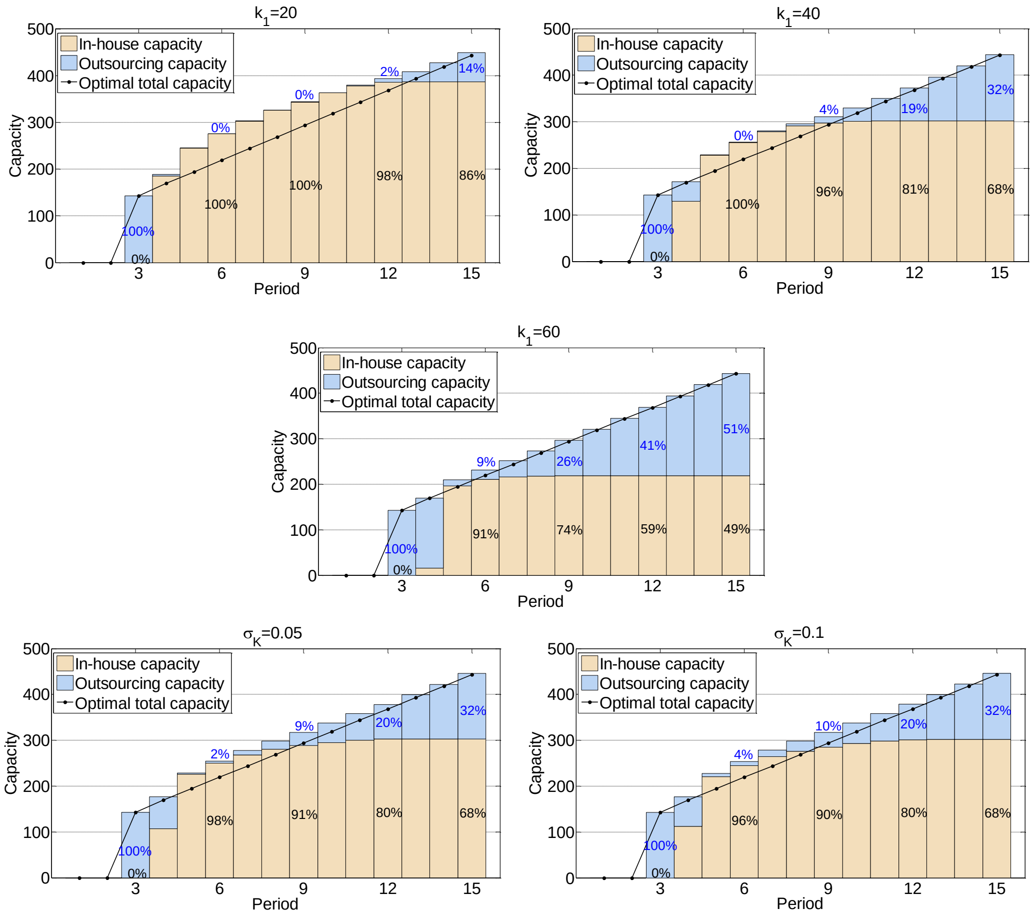

Figure 3: Optimal policies for different initial unit in-house capacity costs and unit in-house capacity cost volatilities: (From top left) $k_1 = 20$, $k_1 = 40$, $k_1 = 60$, $\sigma_K = 0.05$ and $\sigma_K = 0.1$.

the manufacturer also pays the CMO for short-term capacity imbalance. As the remaining periods diminishes, in-house capacity investments become relatively expensive and the manufacturer becomes more dependent on the CMO to expand the total capacity. That is, the CMO has the important function of reducing capacity overage risks during these periods.

Figure 3 shows that the initial unit in-house capacity cost k_1 has significant impacts on the manufacturer's optimal capacity decisions. When k_1 is favorable, the manufacturer builds huge in-house capacity early on, which is more than sufficient to satisfy early demands. This overbuilding tendency diminishes as k_1 increases. On the other hand, the volatility of

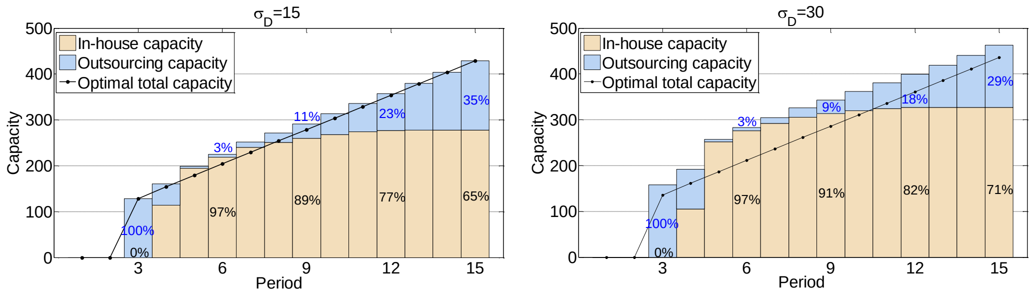

Figure 4: Optimal policies for different demand volatilities: 15% (left) and 30% (right).

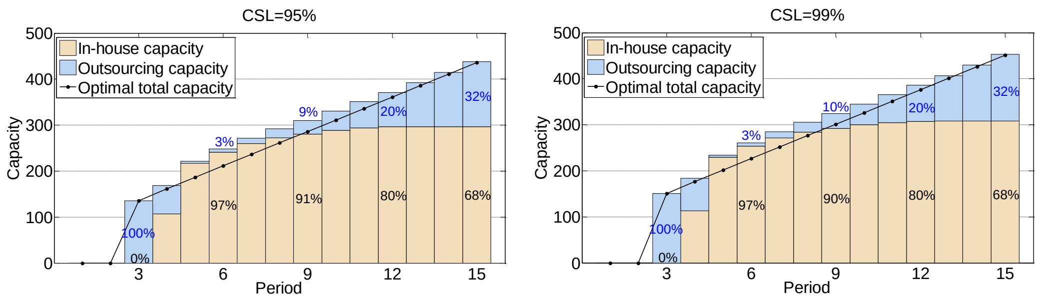

Figure 5: Optimal policies for different customer service levels: 95% (left) and 99% (right).

the unit in-house capacity σ_K has little effect on the manufacturer's optimal strategies. The absolute percentage difference in the optimal costs with $\sigma_K = 0.05$ and $\sigma_K = 0.1$ is 0.63%. The result is desirable because the manufacturer need not consider the potentially high or low future costs and can make reasonable decisions based on the observed cost in period 1.

The influence of demand volatility σ_D is illustrated in Figure 4. On the whole, the higher demand volatility requires more capacity to satisfy potentially large demand and leads to a higher cost. The absolute percentage difference in the optimal costs with $\sigma_D = 15$ and $\sigma_D = 30$ is 16.88%. Before receiving FDA approval, lost sales risks are mitigated by securing extra capacity at the CMO. Once the drug is approved, however, the potentially high demand justifies large investments in in-house capacity and the manufacture builds huge in-house capacity up front. Although demand could also potentially be small, overbuilding

Table 3: Myopic Policy vs. Original Policy: Absolute Percentage Errors

			Initial in-house capacity cost (k_1)					
			20		40		60	
			In-house capacity cost volatility (σ_K)		In-house capacity cost volatility (σ_K)		In-house capacity cost volatility (σ_K)	
			0.05	0.1	0.05	0.1	0.05	0.1
Customer service level 95%	Demand volatility (σ_D)	15	1.08%	1.02%	1.06%	0.94%	1.93%	1.84%
		30	0.09%	0.16%	0.10%	0.25%	1.41%	1.71%
Customer service level 99%	Demand volatility (σ_D)	15	1.31%	1.19%	1.06%	0.97%	1.96%	1.88%
		30	0.06%	0.00%	0.07%	0.04%	1.59%	1.94%

capacity is optimal because holding capacity is cheaper than losing sales.

Finally, Figure 5 represents the optimal policies for different customer service levels. A higher customer service level increases the total capacity and cost, but the proportion of partial outsourcing remains the same. The absolute percentage difference in the optimal costs with CSL=95% and CSL=99% is 5.02%. In other words, the manufacturer can increase the customer service level from 95% to 99% by increasing the investments by 5%.

4.3 Accuracy of approximate policies

We discuss the accuracy of the two approximation methods derived before. In Table 3, we compared the costs of the myopic policy defined in Eqs. (22) and (23), using $\beta_t = \frac{1}{N-t-1}$, to the costs of the original policy. The absolute percentage errors are within 2% in all cases, ranging from 0.002% to 1.96%. Similarly, Table 4 compares the costs of the optimal policies of the approximate value functions defined in Proposition 5 to the costs of the original policy.

Table 4: Approximated Value Function vs. Original Policy: Absolute Percentage Errors

			Initial in-house capacity cost (k_1)					
			20		40		60	
			In-house capacity cost volatility (σ_K)		In-house capacity cost volatility (σ_K)		In-house capacity cost volatility (σ_K)	
			0.05	0.1	0.05	0.1	0.05	0.1
Customer service level 95%	Demand volatility (σ_D)	15	0.11%	0.14%	0.02%	0.04%	0.04%	0.03%
		30	0.87%	0.92%	0.60%	0.63%	0.34%	0.31%
Customer service level 99%	Demand volatility (σ_D)	15	0.08%	0.10%	0.02%	0.03%	0.02%	0.03%
		30	0.74%	0.77%	0.47%	0.49%	0.27%	0.25%

The absolute percentage errors are less than 1%, between 0.02% to 0.92%, in all cases. Note that this approximation works better for larger lost sales cost, p, that is, the approximation is more accurate when the manufacturer sets a higher customer service level.

5 Discussion and conclusion

This study examined the pharmaceutical manufacturer's jointly optimal in-house capacity and short-term outsourcing policy for a new drug in a multi-period problem. The main contribution is the inclusion of partial outsourcing, though our model can accommodate total in-house/outsourcing as well. More firms are using a hybrid of in-house production and outsourcing for commercial production, but few studies have been done to examine such a strategy. We also considered heterogeneous investment lead times, since a CMO's shorter lead time is one of the keys to gain flexibility. The hybrid strategy is characterized by two base-capacity levels and we derived a closed form solution in a 3-period problem. For a general N-period problem, two approximation methods were proposed.

Partial outsourcing has two advantages: (1) the strategic use to hedge against capacity overage at the beginning, and capacity underage at the end of the planning horizon; and (2) the tactical use to mitigate temporal capacity imbalance during the planning horizon. Managerial insights are summarized as follows. First, the manufacturer should always defer investments in in-house capacity during clinical trials. We used optimistic success rates in clinical trials, so even a moderate risk is sufficient to postpone the construction of in-house manufacturing plants during clinical trials. Hence, securing capacity at a CMO is crucial during the early stage. Second, the manufacturer may benefit from building large in-house capacity early on only if the initial in-house capacity cost is favorable over the CMO price or the demand volatility is high; Otherwise overbuilding in-house capacity is not recommended. The overbuilding tendency has been observed for small molecule drugs, but firms are gradually moving to partial or total outsourcing for biopharmaceutical production. Our results suggest that this trend may be attributable to the high costs to build biologic manufacturing plants. Thus, for biopharmaceutical manufacturers, building huge in-house capacity early on is justifiable only when they suspect demand to be potentially very high.

References

[1] Anderson, E. G., Parker, G. G. 2002. The effect of learning on the make/buy decision. Production and Opr. Management, 11(3), 313-339.

[2] Barnes-Schuster, D., Bassok, Y., Anupindi, R. 2002. Coordination and flexibility in supply contracts with options. Manufacturing and Service Opr. Management, 4(3), 171-207.

[3] Chambers, C. G., Snir, E. M., Ata, A. 2009. The use of flexible manufacturing capacity in pharmaceutical product introductions. Decision Sci. 40(2), 243-267.

[4] Chao, X., Chen, H., Zheng, S. 2009. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty. Opr. Res. 57(1), 82-93.

[5] Cheng, F., Sethi, S. P. 1999. Optimality of state-dependent (s, S) policies in inventory models with Markov-modulated demand and lost sales. Production and Opr. Management, 8(2), 183-192.

[6] DiMasi, J. A., Feldman, L., Seckler, A., Wilson, A. 2010. Trends in risks associated with new drug development: success rates for investigational drugs. Clinical Pharmacology and Therapeutics, 87(3), 272-277.

[7] George, E. D., Farid, S. S. 2008. Strategic biopharmaceutical portfolio development: an analysis of constraint-induced implications. Biotechnology Progress, 24(3), 698-713.

[8] Glasser, P. 2002, October 10. Case Study: Eli Lilly and Using IT to Accelerate Research. CIO Insight. http://www.cioinsight.com/c/a/Past-News/Case-Study-Eli-Lilly-and-Using-IT-to-Accelerate-Research/

[9] Grabowski, H. 2008. Follow-on biologics: data exclusivity and the balance between innovation and competition. Nature Rev. Drug Discovery, 7, 479-488.

[10] Gray, J. V., Tomlin, B., Roth, A. V. 2009. Outsourcing to a powerful contract manufacturer: The effect of learning-by-doing. Production and Opr. Management, 18(5), 487-505.

[11] High Tech Business Decisions. 2007. Biopharmaceutical contract manufacturing 2007: Quality, capacity and emerging technologies. San Jose: High Tech Business Decisions.

[12] High Tech Business Decisions. 2006. Biopharmaceutical contract manufacturing: best practices pricing study 2006. Moraga: High Tech Business Decisions.

[13] Jain, V., Grossmann, I. E. 1999. Resource-constrained scheduling of tests in new product development. Ind. Eng. Chem. Res. 38, 3013-3026.

[14] Levis, A. A., Papageorgious, L. G. 2004. A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry. Computers and Chemical Engineering, 28(5), 707-725.

[15] Lichtenberg, F. R., Duflos, G. 2009. Time Release: The effect of patent expiration on U.S. drug prices, marketing, and utilization by the public. New York: Manhattan Institute.

[16] Myers, S. C., Howe, C. D. 1997. A life cycle financial model of pharmaceutical R& D. Working paper # 41-97. Cambridge: Sloan School of Management.

[17] Outsourcing Pharma. 2007, March 13. BMS extends large Lonza deal. http://www.outsourcing-pharma.com/Contract-Manufacturing/BMS-extends-large-Lonza-deal

[18] Pindyck, R. S. 1993. Investments of uncertain cost. J. Financial Economics, 34(1), 53.

[19] Porteus, E. L. 1971. On the Optimality of Generalized (s, S) Policies. Management Sci. 17(7), 411-426.

[20] Rajapakse, A., Titchener-Hooker, N. J., Farid, S. S. (2005). Modelling of the cbiopharmaceutical drug development pathway and portfolio management. Comp. and Chem. Eng. 29, 1357-1368.

[21] Rogers, M. J., Gupta, A., Maranas, C. D. 2002. Real options based analysis of optimal pharmaceutical research and developement portfolio. Ind. Eng. Chem. Res. 41, 6607-6620.

[22] Schmidt, C. W., Grossmann, I. E. 1996. Optimization models for the scheduling of testing tasks in new product development. Ind. Eng. Chem.Res. 35, 3498-3510.

[23] Serel, D. A. 2007. Capacity reservation under supply uncertainty. Comp. and Opr. Res. 34, 1192-1220.

[24] Singh, M. (2005). The pharmaceutical supply chain: a diagnosis of the state-of-the-art. ME thesis, MIT.

[25] The Boston Globe. 2005, June 17. To cut costs, Biogen to sell facility for $408m: Firm planned to make Tysabri at Calif. plant. http://www.boston.com/business/articles/2005/06/17/to_cut_costs_biogen_to_sell_facility_for_408m/

[26] The Seattle Times. 2002, March 23. Immunex warns of temporary shortage of Enbrel. http://community.seattletimes.nwsource.com/archive/?date=20020323&slug=immunex23

[27] Veeraraghavan, S., Scheller-Wolf, A. 2008. Now or later: a simple policy for effective dual sourcing in capacitated systems. Oper. Res. 56(4), 850-864.

[28] Wang, L. M., Liu, L. W., Wang, Y. J. 2007. Capacity decisions and supply price games under flexibility of backward integration. Int. J. Production Economics, 110, 85-96.