IIE Transactions (2008) **40**, 221–237

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170701488052

Tactical capacity management under capacity flexibility

OSMAN ALP[1] and TARKAN TAN[2,*]

[1]*Industrial Engineering Department, Bilkent University, 06800 Ankara, Turkey*
E-mail: osmanalp@bilkent.edu.tr
[2]*Department of Technology Management, Technische Universiteit Eindhoven, P.O. Box 513, 5600MB Eindhoven, The Netherlands*
E-mail: t.tan@tue.nl

Received February 2006 and accepted November 2006

In many production systems a certain level of flexibility in the production capacity is either inherent or can be acquired. In that case, system costs may be decreased by managing the capacity and inventory in a joint fashion. In this paper we consider such a make-to-stock production environment with flexible capacity subject to periodic review under non-stationary stochastic demand, where we allow for positive fixed costs both for initiating production and for acquiring external capacity. Our focus is on tactical-level capacity management which refers to the determination of in-house production capacity while the operational-level integrated capacity and inventory management is executed in an optimal manner. We first develop a simple model to represent this relatively complicated problem. Then we elaborate on the characteristics of the general problem and provide the solution to some special cases. Finally, we develop several useful managerial insights as to the optimal capacity level, the effect of operating at a suboptimal capacity level and the value of utilizing flexible capacity.

Keywords: Inventory, production, temporary workforce, overtime production, fixed costs, flexible capacity, capacity management

1. Introduction and related literature

The issue of capacity management is of vital importance in most production systems, especially under demand volatility. In a make-to-stock system with fixed capacity and volatile demand, elevated levels of inventory and/or significant underutilization of capacity is unavoidable in order to be able to meet demand in a timely fashion. Nevertheless, in many production systems a certain level of flexibility in the production capacity is either inherent or can be acquired. In that case, system costs may be decreased by managing the capacity and inventory in a joint fashion. In this paper we consider such a make-to-stock production environment with flexible capacity subject to periodic review under non-stationary stochastic demand, where our focus is on tactical-level capacity management.

Capacity can be defined as the total productive capability of all the utilized productive resources including workforce and machinery. These productive resources can be permanent or contingent. We define permanent capacity as the maximum amount of production possible in regular work time by utilizing the internal resources of a company such as existing workforce level on the steady payroll or the machinery owned or leased by the company. Total capacity can be increased temporarily by acquiring contingent resources, which can be internal or external, such as hiring temporary workers from external labor supply agencies, subcontracting, overtime production, renting work stations, and so on. We refer to this additional temporarily acquired capacity as the contingent capacity. Capacity flexibility refers to the ability to adjust the total production capacity in any period with the option of utilizing contingent resources in addition to permanent resources.

The capacity decisions can be made in all decision-making hierarchies: strategic, tactical and operational. Examples of these decisions include determining how many production facilities to operate, determining the permanent capacity of a facility and making contingent capacity adjustments, respectively. Our focus is on the tactical level. In particular, we consider the problem of determining the permanent capacity of a facility, while the operational level integrated capacity and inventory management is executed in an optimal manner. For ease of exposition, we refer to the workforce capacity setting in some parts, but this does not mean that our analysis is solely confined to that environment. A possible application area of the problem we consider is an environment in which the production is mainly determined by the workforce size. This workforce size is flexible, in the sense that temporary (contingent) workers can be hired in any period in addition to the permanent workforce that is fixed through the planning horizon. Contingent

*Corresponding author

workers are paid only for the periods they work, whereas permanent workers are on a payroll. The firm wishes to find the optimal permanent workforce size as well as their optimal operating policies. We assume that the lead time to acquire contingent labor is zero. Indeed, it takes as little as 1 or 2 days to acquire temporary workers from external labor supply agencies for jobs that do not require high skill levels, according to our experience. In some cases temporary labor acquisition is actually practically immediate. In some developing countries workers looking for a temporary job and companies in need of temporary labor gather in known venues early in the morning and the companies hire the workers that they are going to make use of that very day.

Changing the level of permanent capacity as a means of coping with demand fluctuations, such as hiring and/or firing permanent workers frequently, is not only very costly in general, but it may also have many negative impacts on a company. In the case of labor capacity, the social and motivational effects of frequent hiring and firing makes this tool even less attractive. Utilizing flexible capacity, such as hiring temporary workers from external labor supply agencies, is a means of overcoming these issues, and we consider this as one of the two main operational tools of coping with fluctuating demand, along with holding inventory. However, long-term changes in the state of the world can make permanent capacity changes unavoidable. Consequently, we consider the determination of the permanent capacity level to be a tactical decision that is made at the beginning of a finite planning horizon and not changed until the end of the horizon. The capacity-related decisions are the determination of the permanent workforce size to utilize through the planning horizon, and the number of temporary workers to hire in each period. The productivity of temporary workers is allowed to be different to that of permanent workers in our model. Our model also allows for the incorporation of fixed costs associated with: (i) initiating production in each period (including setup costs of production); and (ii) ordering contingent capacity. Finally, we note that hiring contingent workers may not be a feasible option for temporarily increasing production capacity in certain environments due to unavailability or special skill requirements. In these situations, overtime production using permanent capacity could be the means to create capacity flexibility. We also analyze this problem environment.

Capacity management problems have received significant attention in the relevant literature at all levels of the hierarchical decision-making process. Van Mieghem (2003) presents a survey of the literature and focuses on strategic decisions whereas Wu *et al.* (2005) focuses on tactical and operational level decisions. Capacity management problems include, among others, capacity planning in terms of the determination of the capacity levels of productive resources and the timing of the capacity adjustments. Since the problem of concern in this article is a tactical-level capacity planning problem coupled with production/inventory decisions, we mainly review articles in the capacity management literature that attempt to exploit the interactions between capacity planning and production/inventory decisions.

The papers by Bradley and Arntzen (1999), Atamturk and Hochbaum (2001) and Rajagopalan and Swaminathan (2001) are examples of research that deals with the joint capacity and inventory management problem at tactical and operational levels under a deterministic demand assumption. They provide formulations and solution approaches under different problem settings. However, these approaches do not apply to our problem since we consider stochastic demand in our model.

There are a number of studies in the literature that assume stochastic demand and are closely related to our work in terms of the problem environment. We discuss the major differences between these papers and our work after presenting a brief review of them. Bradley and Glynn (2002) deal with a continuous-review problem where the fixed capacity level of the productive resources are to be determined as well as the production quantities in a single-item, infinite-horizon environment. It is assumed that the item is replenished according to the base stock policy and that the capacity level is not subject to changes, permanent or temporary. Bradley (2004) extends this model to the case where the capacity level can be increased temporarily with the use of subcontracting, which is similar to the use of contingent capacity in our model. Their models are on a continuous time scale whereas ours is discrete, which might represent a contingent workforce environment more closely, where the decisions on hiring temporary workers are typically made on a period basis. Tan and Gershwin (2004) also deal with a similar make-to-stock environment where any demand that exceeds the current capacity is satisfied from one of the available subcontractors. The demand rate is assumed to have two states either high or low, and is dependent on the current backordering level. The decision variables are the production rates for in-house production and subcontracted production. The authors prove that there exist a series of threshold levels in the optimal policy for in-house production and for each subcontractor that indicate when there is sufficient surplus and when there is a need to use the subcontractors. Their model is a continuous-time model and does not accommodate setup costs for production and subcontracting. Another paper that is related to our work is by Kekre *et al.* (2004) (although what they refer to as strategic and tactical capacity planning stands for what we refer to as tactical and operational capacity planning, respectively), where the authors utilize stochastic programming with recourse to decide on the permanent capacity and to determine the production quantities in a similar problem environment. There are a number of major differences between their work and ours. Kekre *et al.* (2004) use linear production costs whereas we consider a neither concave nor convex production cost function. The modeling approaches are also different in the sense that we use stochastic dynamic programming which permits the use of optimal operating policies whereas their paper is not intended for policy characterization for inventory and contingent capacity decisions. Finally, our model is capable of handling a cheaper contingent capacity option, unlike their model.

Even though there are some similarities between the studies mentioned above and our work, we extend the literature in this stream of research in the following sense. We incorporate the fixed costs of production and fixed costs of acquiring contingent capacity in the problem environment under study for the first time in the literature, to the best of our knowledge. In addition, we utilize the optimal production/inventory decisions when we optimize the permanent capacity levels. Moreover, we provide an extension which can be used to solve the problem when flexible capacity is created through overtime production. We characterize the optimal policy in some cases of the problem and for the remaining cases we demonstrate through numerical studies that the optimal policy does not have a simple form. The existence of fixed costs in the system considerably changes the structure of the model and the analysis, causing the cost functions to be intractable in the most general case. Finally, we provide structurally different and useful managerial insights that stem from intricate tradeoffs between fixed costs and other problem parameters.

We also want to mention a number of studies from the capacity management literature that deal with related problems under different settings. Kouvelis and Milner (2002) and Pinker and Larson (2003) deal with problems where the starting capacity levels are optimized with inventory carry overs not allowed. Cheng *et al.* (2004) deal with a single-item problem without a contingent capacity option where the firm determines a fixed capacity level to be used in a medium-term planning horizon that cannot be changed through this horizon but with the option of expanding or contracting the capacity when starting the next planning horizon. The authors characterize the optimal capacity management policy. Zhang *et al.* (2004) exploit the tradeoff between capacity expansion and lost sales costs in an environment where multiple products and machine types exist, the demand is non-stationary and inventory holding and backordering are not allowed. Van Mieghem and Rudi (2002) deal with the problem of determining the optimal capacity and the base stock levels in a single-period multi-resource problem. The authors extend this problem to the multi-period case and show that the myopic policy is optimal when the unmet demand is lost and they also provide the conditions for which the myopic policy is still optimal for the backordering case. Angelus and Porteus (2002) deal with the joint capacity and inventory management problem of a short-life-cycle product and they characterize optimal policies under certain assumptions where demand is assumed to exhibit a stochastically increasing structure followed by a stochastically decreasing nature.

Finally, we note that Hu *et al.* (2004), Tan and Alp (2008) and Yang *et al.* (2005) deal with the characterization of optimal capacity planning and inventory decisions under problem settings in which the capacity level can be increased temporarily by the use of contingent resources, but they do not deal with the problem of optimization of the initial capacity level.

In this paper, we first develop a simple model to represent this relatively complicated problem and we characterize the solution when the fixed costs are negligible. For the case with positive fixed costs, we provide the solution to the single-period problem and elaborate on the general characteristics of the solution to the multi-period problem. Finally, we develop several useful managerial insights. In particular, we investigate the sensitivity of the optimal solution to changes in the parameters, we study the effects of operating under a suboptimal permanent capacity level and we explore the parameter settings where capacity flexibility is more valuable.

The rest of the paper is organized as follows. We present our dynamic programming model in Section 2. The optimal policy for the integrated problem is discussed in Section 3. In Section 4 we provide an extension of the model which assumes overtime production as the means of flexibility. Our computations that result in managerial insights are presented in Section 5. We conclude the paper in Section 6.

2. Model formulation

In this section, we present a finite-horizon dynamic programming model to formulate the problem under consideration. Unmet demand is assumed to be fully backlogged. The relevant costs in our environment are the inventory holding and backorder costs, the unit cost of permanent and contingent capacity, the fixed cost of production and the fixed cost of ordering contingent capacity, all of which are non-negative. We assume that there is an infinite supply of contingent workers, raw material is always available and the lead time of production and acquiring contingent capacity can be neglected. The notation is introduced as need arises, but we summarize our major notation in Table 1 for ease of reference.

We consider a production cost component which is a linear function of the permanent capacity in order to represent the costs that do not depend on the production quantity (even when there is no production), such as the salaries of permanent workers. That is, each unit of permanent capacity costs c_{p} per period, and the total cost of permanent capacity per period is Uc_{p}, for a permanent capacity of size U, independent of the production quantity. We do not consider material-related costs in our analysis. In order to synchronize the production quantity with the number of workers, we redefine the “unit production” as the number of actual units that an average permanent worker can produce; that is, the production capacity due to U permanent

Table 1. Summary of notation

Variable	*Definition*
T	Number of periods in the planning horizon
K_p	Fixed cost of production
K_c	Fixed cost of ordering contingent capacity
c_p	Unit cost of permanent capacity per period
c_c	Unit cost of contingent capacity per period
h	Inventory holding cost per unit per period
b	Penalty cost per unit of backorder per period
α	Discounting factor ($0 < \alpha \leq 1$)
Q_t	Number of items produced in period t
W_t	Random variable denoting the demand in period t
$G_t(w)$	Distribution function of W_t
U	Size of the permanent capacity
x_t	Inventory position at the beginning of period t before ordering
y_t	Inventory position in period t after ordering
$f_t(U, x_t)$	Minimum total expected cost of operating the system in periods $t, t+1, \ldots, T$, given the system state (U, x_t)

workers is U units per period. We also define the cost of production using temporary workers in the same unit basis, where the cost for flexible workers is related to their productivity. In particular, let c'_c be the hiring cost of a temporary worker per period, and let c''_c denote all other relevant variable costs associated with production by temporary workers per period. It is possible that the productivity rates of permanent and temporary workers are different. Let γ be the average productivity rate of temporary workers, relative to the productivity of permanent workers; that is, each temporary worker produces γ units per period. Assuming that this rate remains approximately the same over time, the unit production cost using temporary workers, c_c, can be written as $c_c = (c'_c + c''_c)/\gamma$. It is likely that $0 < \gamma < 1$, but the model holds for any $\gamma > 0$.

For the sake of generality, we allow for non-negative fixed costs, both for production and contingent capacity ordering. Let K_p denote the fixed cost of production and K_c denote the fixed cost of ordering contingent capacity. K_p is charged whenever production is initiated, even if the permanent workforce size is zero and all production is due to temporary workers. Therefore, together with the structure of the unit permanent capacity costs, this implies that it is never optimal to order contingent capacity unless the permanent capacity is fully utilized. On the other hand, K_c is charged only when temporary workers are ordered, independent of the amount. Fixed costs of contacting external labor supply agencies and training costs may be among the drivers of K_c. We ignore the costs that may be associated with acquiring permanent capacity that is incurred at the beginning of the planning horizon, nevertheless we discuss in Section 6 how such costs can be incorporated in the analysis.

Under these settings, it turns out that the production quantity of a period, Q_t, is sufficient to determine the number of temporary workers to be hired in that period, m_t, for any level of permanent capacity determined at the beginning of the planning horizon. In particular, $m_t = [(Q_t - U)^+/\gamma]$, ignoring integrality, where $(\cdot)^+$ denotes the value of the argument inside if it is positive and assumes a value of zero otherwise. Consequently, the problem translates into a production/inventory problem where the level of capacity is a decision variable and the production cost is piecewise linear, which is neither convex nor concave under positive fixed costs. See Fig. 1 for an illustration. Note that when K_p and K_c are both zero, this function is convex.

The order of events is as follows. At the beginning of the first period, the permanent capacity level U is determined. At the beginning of each period t, the initial inventory level x_t is observed, the production decision is made and the inventory level is raised to y_t by utilizing the necessary capacity means; that is, if $y_t \leq x_t + U$ then only permanent capacity is utilized, otherwise a contingent capacity of size $m_t = [(y_t - x_t - U)^+/\gamma]$ is hired on top of full permanent capacity usage. At the end of period t, the demand d_t is met/backlogged, resulting in $x_{t+1} = y_t - d_t$. We denote the random variable corresponding to the demand in period t as W_t and its distribution function as $G_t(w)$. The state of the system consists of the permanent capacity level and the initial inventory level, (U, x_t). Denoting the minimum cost of operating the system from the beginning of period t until the end of the planning horizon as $f_t(U, x_t)$, we use the following dynamic programming formulation to solve the integrated Capacity and Inventory Management Problem (CIMP):

$$
\begin{aligned}
f_t(U, x_t) = Uc_p + \min_{y_t : x_t \leq y_t} \{ & K_p\delta(y_t - x_t) \\
& + K_c\delta(y_t - x_t - U) + [y_t - x_t - U]^+ c_c \\
& + L_t(y_t) + \alpha E[f_{t+1}(U, y_t - W_t)]\} \\
& \text{for } t = 1, 2, \ldots, T,
\end{aligned}
$$

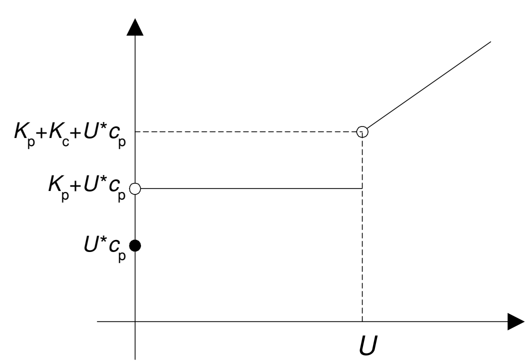

Fig. 1. Production cost function under positive fixed costs.

where $L_t(y_t) = h\int_0^{y_t}(y_t - w)\mathrm{d}G_t(w) + b\int_{y_t}^{\infty}(w - y_t)\mathrm{d}G_t(w)$ is the regular loss function, $\delta(\cdot)$ is the function that attains a value of one if its argument is positive, and zero otherwise, and the ending condition is defined as $f_{T+1}(U, x_{T+1}) = 0$.

3. Analysis

3.1. *Analysis with no fixed costs*

We first handle the case where the fixed costs are negligible. Tan and Alp (2005) show that the optimal operational policy for any given permanent capacity level is of state-dependent order-up-to type, where the optimal order-up-to level, $y_t^*(x_t)$, is

$$y_t^*(x_t) = \begin{cases} y_t^{\mathrm{c}} & \text{if} \quad x_t \leq y_t^{\mathrm{c}} - U, \\ x_t + U & \text{if} \quad y_t^{\mathrm{c}} - U \leq x_t \leq y_t^{\mathrm{u}} - U, \\ y_t^{\mathrm{u}} & \text{if} \quad y_t^{\mathrm{u}} - U \leq x_t \leq y_t^{\mathrm{u}}, \\ x_t & \text{if} \quad y_t^{u} \leq x_t, \end{cases} \quad (1)$$

and y_t^{u} and y_t^{c} are the minimizers of the functions $J_t^{\mathrm{u}}(y) = L_t(y) + \alpha E[f_{t+1}(y - W_t)]$ and $J_t^{\mathrm{c}}(y|x) = J_t^{\mathrm{u}}(y) + c_{\mathrm{c}}(y - x - U)$, respectively.

First we provide a convexity result which is useful in determining the optimal level of the permanent capacity.

Lemma 1. *Let $X = \mathcal{R} \times \mathcal{R}^+$. Note that X is a convex set. Let $Y(x, U)$ be a non-empty set for every $(x, U) \in X$, $C = \{(x, U, y) : (x, U) \in X, y \in Y(x, U)\}$ is a convex set, and the function $g(x, U, y)$ is a convex function on C. Then, $f(x, U) = \min_{y \in Y(x)}\{g(x, U, y)\}$ is also convex on X.*

Proof. See the Appendix. ■

Theorem 1. *Let $X = \mathcal{R} \times \mathcal{R}^+$. Then, $f_t(x_t, U)$ is convex on X.*

Proof. See the Appendix. ■

Consequently, one can search for the optimal permanent capacity level using this convexity result for any starting inventory level. We next consider the single-period problem and provide the solution explicitly, which is a newsboy-type solution. Although the optimal tactical-level capacity determination problem implies a multi-period setting in the problem environment we have discussed, there are some useful insights that can be gained from the analytical solution that the single-period problem brings. We suppress the time subscript in the analysis of the single-period problem.

Theorem 2. *The optimal permanent capacity level of the single-period problem is given by*

$$U^* = \begin{cases} 0 & \text{if } c_{\mathrm{p}} \geq c_{\mathrm{c}} \\ \left(G^{-1}\left(\dfrac{b - c_{\mathrm{p}}}{h + b}\right) - x\right)^+ & \text{if } c_{\mathrm{p}} < c_{\mathrm{c}} \end{cases}$$

Proof. See the Appendix. ■

Note that U^* is independent of c_{c} as long as $c_{\mathrm{p}} < c_{\mathrm{c}}$, because no contingent capacity would be used in the single-period problem in that case. Note also that U^* is decreasing in c_{p}, which means that expensive permanent resources result in a smaller permanent capacity. If $c_{\mathrm{p}} < c_{\mathrm{c}}$, then $y^{\mathrm{c}} - x \leq U^* \leq y^{\mathrm{u}} - x$, where $y^{\mathrm{u}} = G^{-1}(b/(h+b))$ and $y^c = G^{-1}((b - c_{\mathrm{c}})/(h+b))$. Consequently, $y^*(x) = x + U^*$, which implies that in a single-period problem the optimal policy is first to install a permanent capacity of U^* and then to produce in full terms without hiring any contingent capacity. If the unit cost of permanent capacity exceeds that of the contingent one, then it is optimal to hold no permanent capacity at all and to utilize only the contingent resources to produce up to y^{c}. In the multi-period problem, the optimal permanent capacity still takes the value of zero, as shown later in Theorem 4, when $c_{\mathrm{c}} \leq c_{\mathrm{p}}$.

Finally in this section, we analyze the behavior of the optimal permanent capacity level as a function of the number of periods in the planning horizon. We show in Table 2, by the use of a stationary problem instance, that there exists no monotonic relation between the two. In this particular example, the optimal capacity level first increases and then decreases converging to $U^* = 10$ as the length of the planning horizon increases. We also observe that there are other problem instances where U^* either monotonically increases or decreases where such a relation depends on the problem parameters. In any case, the solution converges after a number of periods.

3.2. *Analysis with fixed costs*

In this section, we analyze the problem when both fixed costs are positive. First we present our results on the optimal capacity level of a single-period problem. Upon solving the single-period problem we elaborate on the structural properties of the optimal solution in the multi-period problem.

Although it is more likely that the problem environment that we have discussed has a multi-period structure, there are also some problem environments where the single-period model is appropriate. When the demand for the product is mostly observed in a condensed time interval, such as the Christmas period, the single-period model is clearly relevant. Another application would be a product selection problem where the item(s) to produce in a finite

Table 2. Optimal capacity levels vs length of the horizon when $h = 1, b = 7, c_{\mathrm{p}} = 1.5, c_{\mathrm{c}} = 3, \alpha = 0.99$ and W_t is Poisson with $E[W_t] = 10$ for all t

	T										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	⋯	*50*
U^*	11	12	12	11	11	10	10	10	10	10	⋯	10

planning horizon is (one) to be selected. In that case, this problem could be handled by representing the total planning horizon as a single period. In particular, such a firm may be capable of producing a variety of items and would like to determine the capacity to dedicate for the production of each alternative, if any. In such environments, product selection decisions might be based on the expected total costs of producing items (along with the associated revenues) depending on their probabilistic demand behaviors, starting inventories, productivity of permanent and temporary resources and operating costs. The single-period problem could be solved for all possible alternatives (with K_p denoting the fixed cost of production changeover for that alternative) in order to make the decision as to how much capacity should be dedicated for that alternative, if any.

Let Q^p denote the production that is conducted solely by making use of permanent capacity, let Q^c denote the production that is conducted solely by making use of contingent capacity, and let $Q = Q^p + Q^c$ denote the total production. Then we can write the following Lemma.

Lemma 2. *In the optimal solution to the single-period problem, $U = Q^p$.*

Proof. See the Appendix. ■

Lemma 2 states that the permanent capacity to be installed, if any, must be equal to the production that is required to be conducted with that capacity. That is, there should be no underutilization in the optimal solution.

The following complementarity result is useful in solving the single-period problem.

Lemma 3. *In the optimal solution to the single-period problem, $Q^p Q^c = 0$.*

Proof. See the Appendix. ■

Lemma 3 states that the production will be due to one type of capacity only, either permanent or contingent. That is, the optimal solution is either to set the permanent capacity to the level of desired production and not to utilize any contingent capacity, or to set the permanent capacity level to zero and conduct all production with contingent capacity, depending on the cost parameters.

Let $y^p = G^{-1}((b - c_p)/(h + b))$ and recall that $y^c = G^{-1}((b - c_c)/(h + b))$, and $y^u = G^{-1}(b/(h + b))$. Define two auxiliary functions as

$$s^c(x) = \min\{s : L(s) = K_p + K_c + L(y^c) + c_c(y^c - x)^+\},$$
$$s^p(x) = \min\{s : L(s) = K_p + c_p(y^p - x)^+ + L(y^p)\}.$$

The following theorem characterizes the optimal capacity level and the production quantity of a single-period problem.

Theorem 3. *Let $\tilde{U} = y^p - x$ for a starting inventory level of x. Then, the optimal capacity and order-up-to levels, (U^*, y^*), of a single-period problem can be characterized as follows.*

Case 1. $c_p \le c_c$:

$$(U^*, y^*) = \begin{cases} (\tilde{U}, y^p) & \text{if } x \le s^p(x), \\ (0, x) & \text{otherwise}. \end{cases}$$

Case 2. $c_p > c_c$:

$$(U^*, y^*) = \begin{cases} (0, y^c) & \text{if } x \le s^c(x) \text{ and } s^p(x) \le s^c(x), \\ (\tilde{U}, y^p) & \text{if } x \le s^p(x) \text{ and } s^c(x) \le s^p(x), \\ (0, x) & \text{otherwise}. \end{cases}$$

Proof. See the Appendix. ■

Theorem 3 suggests that

1. If the unit cost of permanent capacity is less than that of contingent capacity, then no contingent capacity should be utilized.
2. The optimal permanent capacity level is either $\tilde{U}$ or zero.
3. Contingent capacity will only be utilized if its unit cost is cheaper than that of the permanent capacity and the required production quantity is high enough to compensate for the additional fixed cost that will be incurred for utilizing contingent capacity, in which case no permanent capacity will be installed. In that case the inventory level after production will be higher than that with the alternative option of producing with the permanent capacity.

The following result is an implementation of Theorem 3 when the starting inventory level is zero.

Corollary 1. *When $c_p \le c_c$ and the starting inventory level is zero then:*

$$U^* = \begin{cases} y^p & \text{if } E[W] \ge \big(c_p y^p + K_p + L(y^p)\big)/b, \\ 0 & \text{otherwise}. \end{cases}$$

When the starting inventory level is zero, it is optimal to make production and install permanent capacity if the expected demand is greater than a prespecified value given by the problem parameters and distribution of demand. Otherwise production is not economic.

We note that the single-period model with $c_c < c_p$ fits also to a case where "contingent capacity" refers to the alternative of conducting the production in a developing country with cheaper production costs. In that case, K_p mimics the investment that is required independent of the country of investment (such as the specific machinery that needs to be procured), K_c mimics the additional costs that would be undertaken to begin production in that developing country (such as the costs of the additional research required to invest there, the additional risks taken, etc.), and the single-period production quantity mimics the total production that will be produced. In that case, the investment should only take place if the required production amount is large enough to recoup the additional investment costs. If that is the case, our solution suggests that all the production should be performed there, which brings the inventory to a level that is higher than that of producing with the alternative option of more expensive local production.

When we have more than one period in the planning horizon, it is not possible to obtain closed-form expressions for the optimal capacity levels. Nevertheless, we present the solution of a special case in the following theorem without proof.

Theorem 4. *If* $K_{\mathrm{p}} \geq 0$, $K_{\mathrm{c}} = 0$, *and* $c_{\mathrm{c}} \leq c_{\mathrm{p}}$ *then* $U^{*} = 0$.

This theorem states that when the contingent resources are cheaper than permanent resources, it is optimal to outsource all production or to produce in house with only contingent resources.

In some cases the myopic (single-period) solution provided in Corollary 1 gives the optimal or near-optimal solution to the general multi-period problem, especially when K_{p} is low. Moreover, the optimal solution is $U^{*} = 0$ in both problems when K_{p} is extremely high relative to the other cost parameters of the problem. Nevertheless, for the values of K_{p} in between, the system may prefer to avoid paying K_{p} every period in the multi-period problem and to produce in large quantities when production is initiated to cover the demand of a number of periods (as discussed later in this section) possibly by setting $U^{*} = 0$, while the single-period problem does not have such an option and may go for a high level of permanent capacity, creating a gap between those two solutions. Moreover, Lemma 3 provides a property of the myopic solution that does not necessarily hold in the multi-period case. Finally, y^{p} is independent of the cost parameters of the contingent capacity. Consequently, the myopic solution does not necessarily solve the multi-period problem and it may even be far from the optimal solution. It may be possible to devise a heuristic solution to the multi-period problem that makes use of the myopic solution, nevertheless we focus on some other aspects of the problem in this paper.

It is shown in Section 3.1 that the expected total costs of the system is convex in the permanent capacity levels when there are no fixed costs. This result enables us to easily determine the optimal permanent capacity level. One could expect a similar behavior of the expected total cost function under the existence of positive fixed costs, since extremely high and low levels of capacity would still be more costly than an intermediate level. However, this intuition turns out to be incorrect, as we demonstrate in Fig. 2. This figure denotes the expected total costs of the system for varying permanent capacity levels with problem parameters of $K_{\mathrm{p}} = 20$, $K_{\mathrm{c}} = 30$, T = 12, b = 10, h = 1, $c_{\mathrm{c}} = 3.5$, $\alpha = 0.99$ and a seasonal Normal demand pattern with a cycle of four periods with expected demand values of 15, 10, 5 and 10, respectively and a coefficient of variation, CV value of 0.1.

The reason why the convex structure does not hold anymore is as follows. If the system is working under a very low or zero permanent capacity, then the only way of avoiding both of the fixed costs in every period (other than cumulative backlogging of the demand) is to operate with elevated inventory levels by making use of contingent capacity in large amounts every time production is initiated, followed by a number of periods with no production. In that case, a marginal increase in permanent capacity may increase the system costs despite the decreased production costs in the periods where there is positive production, because that capacity will be paid in the periods with no production as well. However, as the permanent capacity level becomes sufficiently high, the total costs may decrease due to decreased production costs, since permanent capacity would be utilized most of the time. Finally, as the permanent capacity becomes excessively high, the system costs will increase due to low utilization. Moreover, the expected total cost function does not necessarily have only one local minimum in this region[1]. In particular, a certain relatively low permanent capacity level that makes the best use of contingent capacity may be a local minimum, whereas some higher permanent capacity level(s) may decrease or eliminate the need of contingent capacity (and hence its fixed cost) resulting in another local minimum. The reason why the cost may be partly increasing in between is that the additional permanent capacity in between may not be large enough to eliminate or significantly decrease the need of contingent capacity while resulting in increased permanent capacity costs.

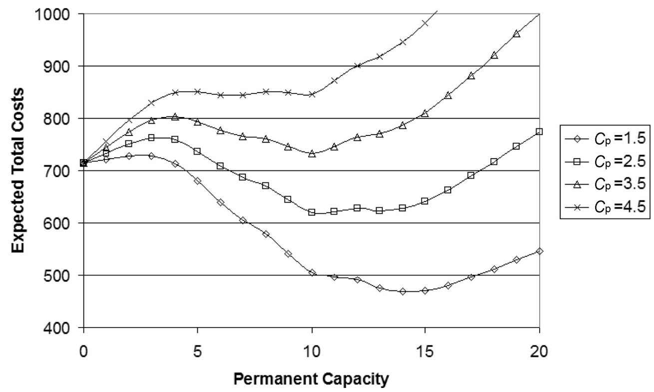

Fig. 2. Expected total costs plotted against permanent capacity.

We illustrate the reason for the cost behavior that we discussed above for low permanent capacity levels in Fig. 3. This figure depicts the expected production quantities under two different values of low permanent capacity, namely $U = 0$ and 2, where the other parameters are the same as those we reported for Fig. 2, with $c_{\mathrm{p}} = 1.5$. $E[Per]$ denotes the expected production by permanent resources, and $E[Con]$ denotes that by contingent resources. Expected production quantities are found by simulating the system by

[1]In all the numerical tests that we conducted, other than zero we faced no more than two local minima. However, we restrain ourselves from stating that this is necessarily so for other problem parameters as well.

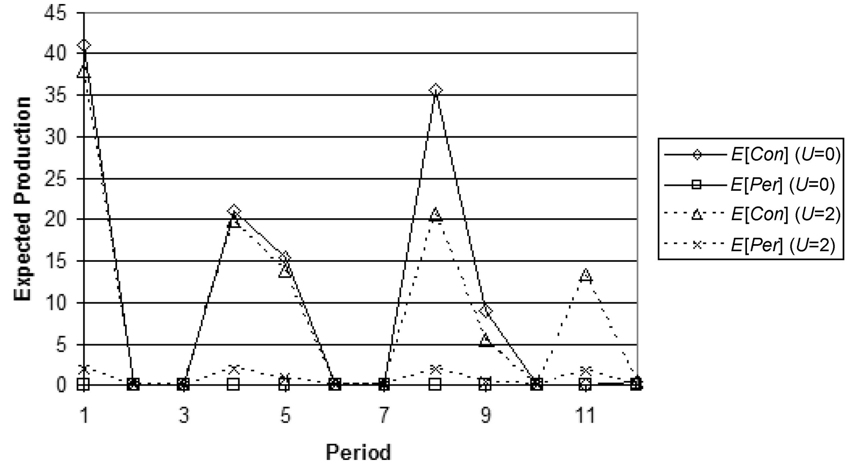

Fig. 3. Expected production under a low permanent capacity.

using the optimal policies. Observe that no production takes place in a number of the periods in both cases.

Establishing the exact form of the expected total cost function requires full characterization of the optimal ordering policies, which is difficult even for special cases. Therefore, we apply explicit enumeration in our computations.

Similar to the case of no fixed costs, there are problem instances where a non-monotonic relation between the optimal capacity levels and the length of the planning horizon exists in the case of positive fixed costs. Table 3 lists examples under stationary demand. When $c_p = 1.5$, the optimal capacity level of a single-period problem is 12, which is slightly higher than the expected demand, whereas it jumps to 20 for a two-period problem. In this case, keeping a larger permanent capacity level but initiating production only once is preferred to keeping a lower permanent capacity level and initiating production twice in the optimal solution as a result of the tradeoffs between the cost components of this specific problem instance.

For planning horizons with more than three periods, we observe a structure in which the optimal capacity level alternates between zero and 16 until a certain length of the planning horizon. In these situations, the expected total costs of two local minima (see Fig. 2) $U = 0$ and $U = 16$ are very close to each other and they exhibit structurally different operating characteristics. In Table 4, the expected production using permanent or contingent resources in each period are presented for the same problem instance with $T = 5$. When $U = 0$, all production is due to contingent resources and in order to alleviate the effect of large fixed costs in the system, there is only one major production setup scheduled in the first period followed by occasional production instances towards the end of the horizon. As a matter of fact, the optimal operating policy is to hold a permanent capacity of size 16 and make more frequent production runs using permanent resources and almost neglect the use of contingent capacity. As T increases the solution converges to $U^* = 0$. In Fig. 4, we present the normalized expected total costs for some values of T in this problem instance (we normalize the cost of each T to a unit cost at $U = 0$) where the convergence can easily be observed.

When $c_p = 1$ in Table 3, we observe a fluctuating structure, which converges to a high permanent level due to a relatively cheaper permanent capacity cost, whereas the optimal permanent capacity level becomes zero starting from the two-period problem when $c_p = 2$. Consequently, we conclude that the optimal permanent capacity size may

Table 3. Optimal capacity levels vs length of the horizon when $K_p = 50$, $K_c = 10$, $h = 1$, $b = 10$, $c_c = 3$, $\alpha = 0.99$ and W_t is Poisson with $E[W_t] = 10$ for all t

		T											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	$\cdots$	*50*
$c_p = 1$	U^*	13	21	16	21	18	20	18	20	19	19	$\cdots$	19
$c_p = 1.5$	U^*	12	20	15	0	16	16	0	0	16	0	$\cdots$	0
$c_p = 2$	U^*	12	0	0	0	0	0	0	0	0	0	$\cdots$	0

Table 4. Expected production by permanent and contingent resources for a five-period problem

	U = 0		*U = 16*	
t	*E[Perm prod]*[a]	*E[Cont prod]*[a]	*E[Perm prod]*	*E[Cont prod]*
1	0	45	16	0
2	0	0	13.91	0
3	0	0.01	6.49	0.01
4	0	1.72	11.18	0.09
5	0	1.26	3.56	0

[a]Expected production with permanent resources.
[b]Expected production with contingent resources.

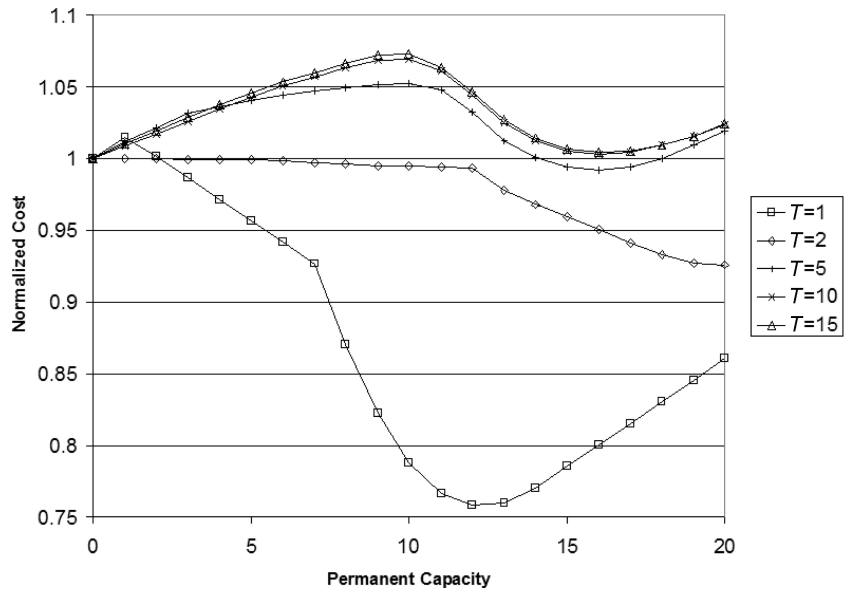

Fig. 4. Normalized costs plotted against permanent capacity levels for different T.

show significant differences as the length of the planning horizon changes when a short horizon problem is considered, whereas this does not hold for a long planning horizon.

As this example points out, it is very difficult to characterize the optimal structure of the capacity levels in problems with fixed costs as there are many alternative potential ways of coping with the demand uncertainty with the help of the flexibility inherent in the system.

4. An extension: overtime production

In industries where the majority of the production must be executed by specially skilled or trained workers, the use of contingent labor may not be a viable option to temporarily increase the permanent capacity. In such an environment, the use of overtime production could be the means to create flexibility in capacity. In this section, we provide a modification of the basic model presented in Section 2 that can be used to solve the integrated CIMP under the option of overtime use. We also present a brief analysis of this model.

Overtime production is defined as production that is performed using permanent resources in addition to production during regular working hours. Owing to capacitated permanent resources and time limitations, the total amount of overtime production is also limited. We reflect this limit by defining an overtime coefficient, η, which provides the upper bound on the total production quantity in any given period. We define η as the ratio of the maximum total production (including overtime) to production with permanent capacity, which implies that $\eta \geq 1$. The integrated CIMP model can be modified for this purpose as follows resulting in the following model, which we refer to as CIMP-OT:

$$\begin{aligned} f_t(U, x_t) = Uc_\mathrm{p} + \min_{y_t : x_t \leq y_t \leq x_t + \eta U} \{ & K_\mathrm{p}\delta(y_t - x_t) \\ & + K_\mathrm{c}\delta(y_t - x_t - U) + [y_t - x_t - U]^+ c_\mathrm{c} \\ & + L_t(y_t) + \alpha E[f_{t+1}(U, y_t - W_t)]\} \\ & \text{for } t = 1, 2, \ldots, T. \end{aligned}$$

Similar to CIMP, we solve CIMP-OT for the single-period setting, the solution of which is presented in the Appendix. The following theorem establishes the convexity of the minimum cost function in the multi-period setting, when the fixed costs in the system are zero.

Theorem 5. *Let $X = \mathcal{R} \times \mathcal{R}^+$. Then $f_t(x_t, U)$ is convex on X when both K_p and K_c are zero.*

Proof. See the Appendix. ■

There is an important structural difference between CIMP and CIMP-OT in the sense that the model may prefer to set the permanent capacity level to zero in CIMP (for example, as stated in Theorem 4), which refers to production with only contingent resources. However, such an action is not sensible in overtime situations since setting the permanent capacity level to zero merely refers to backordering of the whole demand, which could only be optimal at relatively very low values of the unit backordering cost with respect to other problem parameters.

The multi-period minimum expected total cost function ($f_t(U, x_t)$) of CIMP-OT is neither convex nor quasiconvex under positive fixed costs. The function has a decreasing structure as U starts increasing from zero and has an increasing structure for very large values of U. In between, we observe that there may exist more than one local

Table 5. The set of parameters used in the computations

K_p	0, 10, 20, 30, 40, 50, 60
K_c	0, 10, 20, 30, 40
h	1
b	3, 4, 5, 10, 50
c_p	1.5, 2.5, 3.5, 4.5
c_c	1.5, 2.5, 3.5, 4.5
α	0.99
T	12

minima, similar to the case of CIMP as discussed in Section 3.

Finally, we note here that the non-monotonic behavior of the optimal permanent capacity level with respect to an increase in the length of the planning horizon still persists when overtime production is considered.

5. Computations

In this section we present the results of our computational study that was conducted to gain insights on the characteristics of the problem. In our computations, we used the parameter set presented in Table 5. A seasonal demand stream having a cycle of four with expected demand values of 15, 10, 5 and 10 was used. The demand distribution was assumed to be Poisson, Normal and Gamma. We used three different values of CV for the Normal distribution: 0.1, 0.2 and 0.3. Similarly, we used three different values of CV for the Gamma distribution: 0.5, 1 and 1.5. While investigating the effect of a change in one parameter, we kept the other parameters unchanged. In order to avoid trivialities, we assumed that the starting inventory level was zero. We provide intuitive explanations to all of our results below and our findings are verified through several numerical studies. However, one should be careful in generalizing them, as for any experimental result, especially for extreme values of problem parameters. In the results that we present, we use the term "increasing" ("decreasing") in the weak sense to mean "non-decreasing" ("non-increasing").

We first present our computational analysis for CIMP in Section 5.1. Then we briefly state our computational analysis for CIMP-OT, mainly stressing the results that are not similar, in Section 5.2.

5.1. *Contingent labor usage*

5.1.1. *Optimal capacity level*

In this section we investigate the sensitivity of the optimal permanent capacity level as the problem parameters change. Our first observation is on the effects of the fixed cost of production, as we illustrate for a certain parameter setting ($K_c = 10, b = 10, c_p = 1.5, c_c = 3.5$, Normal distributed demand with $CV = 0.3$) in Table 6. Namely, we observe that as the fixed cost of production increases, the optimal permanent capacity level also increases up to a certain threshold point. Until this threshold level, an increased fixed cost of production calls for a higher permanent capacity level, in case it is economical to hold a positive permanent capacity in the first place, so that production does not need to be initiated every period. After this threshold, production is initiated only a few times due to a high fixed cost and all production becomes due to contingent resources. In this region, the optimal permanent capacity level becomes zero because paying for the permanent resources in the non-productive periods becomes too costly.

Table 6. Optimal capacity levels for different values of fixed cost of production

	K_p						
	0	*10*	*20*	*30*	*40*	*50*	*60*
U^*	11	12	13	15	0	0	0

For the remaining capacity cost parameters, we have the following observations. When the unit cost of permanent capacity increases, the optimal capacity level decreases since the utilization of contingent resources becomes more critical, as discussed in Section 5.1.3. On the other hand, as the cost of contingent resources (fixed and/or variable) decrease, the proportion of the production that is conducted by the contingent resources increases and hence the optimal permanent capacity levels decrease. We also note at this point that there is a correspondence between the unit cost of temporary workers, c_c, and the productivity rate of them, γ, as explained in Section 2. As γ, increases, we have lower c_c values. Hence, the effect of γ on the optimal permanent capacity (and on all other measures discussed throughout this section) can be deduced from the effect of a change in c_c on that corresponding measure.

One might expect that the optimal permanent capacity level increases as the variability of the demand increases. This was investigated using the parameters listed in Table 7. We observe that when the cost parameters related to contingent capacity are so high that the system tries to avoid using contingent capacity, the optimal permanent capacity level increases as the variability of the demand increases, as illustrated in problem 1 (*Prob.* 1) of Table 8, for $CV \leq 0.5$.

Table 7. The problem parameters

	Prob. 1	*Prob. 2*	*Prob. 3*	*Prob. 4*	*Prob. 5*	*Prob. 6*	*Prob. 7*	*Prob. 8*
K_p	0	0	20	0	30	60	0	0
K_c	20	10	0	60	∞	20	40	40
b	10	10	10	10	10	10	10	5
c_p	1.5	1.5	1.5	1.5	1.5	1.5	2.5	2.5
c_c	3.5	2.5	2.5	4.5	∞	3.5	1.5	1.5

Table 8. Optimal capacity levels for different CV values

CV	*Prob. 1*	*Prob. 2*	*Prob. 3*	*Prob. 4*	*Prob. 5*	*Prob. 6*	*Prob. 7*	*Prob. 8*
0.1	11	10	12	13	15	17	10	10
0.2	13	10	11	13	15	17	10	11
0.3	14	10	10	14	16	18	9	11
0.5	14	10	0	15	17	19	1	6
1	13	8	0	15	23	0	0	0
1.5	12	6	0	15	28	0	0	0

Nevertheless, an interesting observation that we make about our computations is that most of the time the interactions between the problem parameters are so intricate that the relation between the coefficient of variation of the demand and the optimal capacity levels exhibit completely different behaviors for different parameter settings. As shown in Table 8, the direction of the change in the optimal capacity level (if any) may vary as the variability of the demand increases. Even for the above-mentioned example of expensive contingent capacity, the optimal permanent capacity levels show a decrease when the coefficient of variation of demand is further increased (see *Prob.* 1 of Table 8 for $CV \geq 0.5$), because the expensive option of flexible capacity turns out to be reasonably attractive under a high demand variability, compared to holding very high levels of permanent capacity. For increasingly expensive contingent capacity, we observe a monotonic behavior as illustrated in *Prob. 4*. For the extreme case in which the contingent capacity is restrictively expensive, the problem reduces to one without contingent capacity, in which case the increase in the optimal permanent capacity level in demand variability is structural, as illustrated in *Prob. 5*, but this special case is of little interest in this paper.

The decrease in problems 2 and 3 can be explained in a similar fashion. For instance, in problem 2 the contingent capacity is not very expensive, yet requires a fixed cost. As the demand variability increases, the contingent capacity needs to be employed more often, which makes the option of keeping a smaller permanent capacity and producing more with the contingent capacity more economic (the total expected cost for $CV = 1.5$ decreases from 814.94 to 804.87 when U decreases from ten to six). We also note that a zero permanent capacity is always an option that can be explored if performing the production with contingent capacity becomes more economic than holding any permanent capacity at all and utilizing contingent capacity frequently (such as in *Prob.* 3), which may also be due to high fixed costs of production (such as in *Prob.* 6), as discussed earlier in this section. Finally, we note that when $c_c < c_p$, the optimal permanent capacity will always be zero, unless $K_c > 0$. When $K_c > 0$, the optimal permanent capacity may increase or decrease as the demand variability increases, possibly eventually becoming zero, due to similar arguments as discussed above (see problems 7 and 8).

In conclusion, while it is difficult to draw "hard conclusions" on the behavior of the optimal permanent capacity level as a function of demand variability, we summarize our observations as follows. Since the increase in demand variability requires a higher capacity flexibility, the optimal permanent capacity level tends to increase in demand variability as long as contingent capacity is relatively more expensive in comparison to holding a higher level of permanent capacity that will be kept partially idle, and it tends to decrease otherwise, zero permanent capacity always being an option.

5.1.2. *Effect of operating at suboptimal permanent capacity levels*

We also evaluated the effect of operating at suboptimal permanent capacity levels under different problem parameters. We measure this effect by the percentage penalty of installing a suboptimal capacity which is defined as $\%PSC = (f_1(U,0) - f_1(U^*,0))/f_1(U^*,0)$ where U^* is the optimal capacity level. By definition, the structure of the $\%PSC$ function is expected to be very similar to the structure of the expected total cost function (see Fig. 2). Possible and typical behaviors are presented in Fig. 5 for different values of K_p. We observe similar characteristics for other problem parameters.

When deciding on the values of operating parameters in a typical production/inventory environment, an intuitive solution would be to base operating decisions on expectations of the demand together with its variability. However, such a solution would incur significantly higher costs when the value of the capacity flexibility is underestimated. For example, for situations where the value of flexibility is high and the optimal permanent capacity is zero, such an approach results in high percentage penalty values, as can be observed in Fig. 5. When $K_p = 60$ for example, the percentage penalty of operating with any permanent capacity level between ten and 15 (recall that the expected demand is ten) results in percentage penalty values around 15%. We note that for some other problem instances this penalty may be even more severe in the same range and is observed up to 40% in our computational study.

As for the sensitivity of operating at suboptimal permanent capacity levels, we make the following observations as the demand variability changes[2]. If the system has a high level of suboptimal permanent capacity, $\%PSC$ decreases as the demand variability increases. This is because the expected system cost function under a high permanent capacity level is relatively more robust to changes in demand variability than that under the optimal permanent capacity, since the contingent capacity will mostly not be utilized

[2]In some cases the optimal permanent capacity itself changes as the demand variability changes, but then the costs in those switching points for a given CV are close to each other, and our discussions still hold in general.

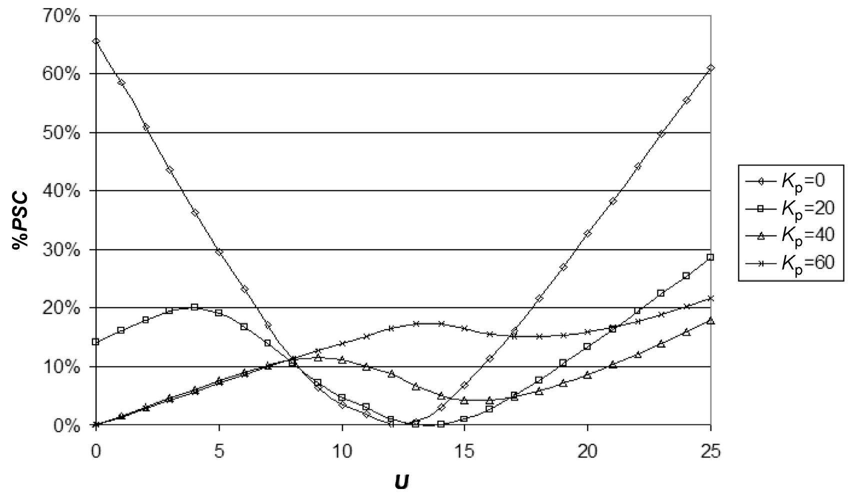

Fig. 5. Operating at suboptimal capacity levels under different fixed costs of production when $K_c = 20$, $h = 1$, $b = 5$, $c_p = 1.5$, $c_c = 2.5$ and a Normal demand with $CV = 0.1$.

anyway. For example, when production with the contingent capacity is more economic than that with the permanent capacity, the optimal permanent capacity size is zero and this discussion will hold for any permanent capacity level higher than a sufficiently high threshold value. Similarly, when the system is operating with a low suboptimal permanent capacity when the optimal one is high, then expensive contingent capacity is going to be often utilized in rather a consistent manner, which results in the expected system cost function being more robust to changes in demand variability than that under the optimal permanent capacity. Consequently, $\%PSC$ decreases as the demand variability increases in this case, as well. Nevertheless, in some cases where the penalty of operating with a suboptimal permanent capacity is not high, the variability degrades the performance of this close-to-optimal system more than it does the optimal system, and hence $\%PSC$ increases as the demand variability increases. A typical example for this case is a low permanent capacity level when the fixed costs are negligible and $c_p = c_c$, in which case $U^* = 0$. For instance, when $U = 6$ with $c_p = c_c = 3.5$ and $b = 10$, $\%PSC = 0.75\%$ for $CV = 0.1$, and it increases in CV. However, it is difficult to draw concrete conclusions as to when $\%PSC$ increases in the demand variability, due to the intricate relations between the problem parameters.

5.1.3. *The value of utilizing flexible capacity*

In this section, our aim is to investigate the general behavior of the value of flexible capacity (VFC) under different problem parameters. We define VFC as $ETC_{IC} - ETC_{FC}$ where ETC_{IC} and ETC_{FC} are the expected total costs of operating in an inflexible environment (where no contingent resources are available) and in a flexible environment, under the respective optimal permanent capacity levels. Similarly, the percentage value of utilizing flexible capacity is defined as $\%VFC = VFC/ETC_{IC}$.

First we analyze the effect of unit costs of permanent and contingent capacity on VFC. In all problem instances that we solved, we observe that VFC has an increasing structure as the unit cost of keeping permanent capacity (c_p) increases (Fig. 6). The average $\%VFC$ is observed as 7.02%, 18.87% and 30.33% in all problem instances with a Normal demand when c_p is 1.5, 2.5 and 3.5, respectively. We can also observe from Fig. 6 that the value of flexibility increases as the unit cost of contingent resources decreases. This indicates that one should search for more possibilities to use contingent resources since their relative costs decrease.

As to the effect of the fixed costs of production on the value of flexibility (see Fig. 7), we first note that VFC (as well as $\%VFC$) exhibits a unimodal behavior after some particular value of K_p. Because, for moderately large values of K_p, the inflexible system starts to prefer to completely backorder all demand, that is perform no production at all, whereas the flexible system may still be better off by using contingent resources. Similarly, for very large values of K_p both systems are better off completely backordering with a zero permanent capacity, and hence they converge to each other so that VFC becomes zero. For the rest, we do not necessarily observe a steady behavior. In almost all of the cases, we observe either a monotonic increase or a monotonic decrease followed by a monotonic increase, prior to the unimodal behavior explained above, such as the example provided in Fig. 7. The decrease in that specific example is caused by ETC_{FC} increasing with a higher rate than ETC_{IC} for an increase in K_p while K_p is low, since the optimal permanent capacity levels do not change (at least significantly). Similarly, the increase is caused by the inflexible system's inability to react to increased fixed cost of production, which results in either underutilization of a large permanent capacity, or incurring the fixed cost of production frequently, before complete backordering. Nevertheless, we encountered some problem instances where VFC (as well

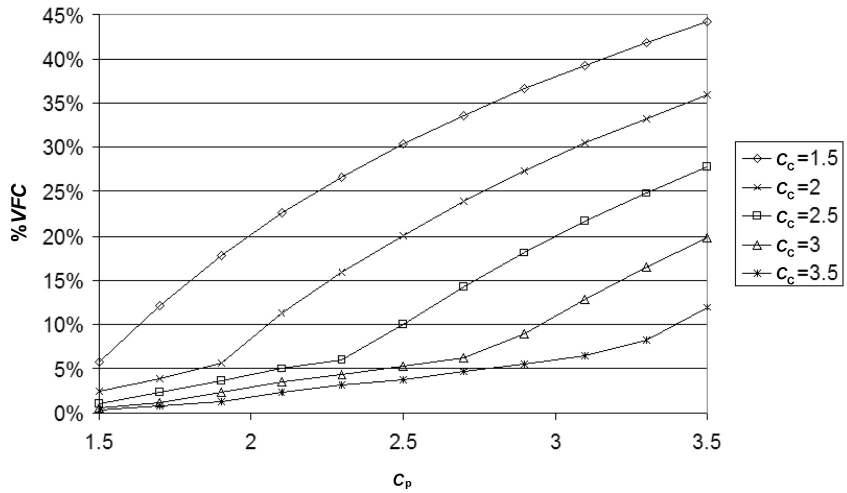

Fig. 6. Effect of unit permanent capacity cost on the value of flexible capacity under a normal demand when $K_p = 10$, $K_c = 10$, $h = 1$ and $b = 5$.

as $\%VFC$) fluctuates as the value of K_p increases. Finally, we note that VFC increases as the fixed cost of contingent resources decreases as shown in Fig. 7.

For a given level of permanent capacity, Tan and Alp (2005) demonstrate some monotonicity results for the value of flexibility as a function of CV or b. Nevertheless, similar to the case in Section 5.1.2, we observe that there are no monotonicity results for VFC (and $\%VFC$) as CV or b increases. For example, there are some problem instances where the value of flexibility decreases (and there are some others where it increases) as the variability in the system increases, even when both of the fixed costs are zero. The reason for the non-monotonicity is the system's ability to adapt itself to changes in CV or b by optimizing the permanent capacity level accordingly.

5.2. *Overtime production*

We conducted a computational study for the overtime model, using the data set of Table 5, in order to reveal the characteristics of the problem, especially the ones that are different. We set $\eta = 1.4$ in all of our computations.

As K_p increases, we observe a non-decreasing behavior in the optimal permanent capacity level similar to CIMP, but it never takes a zero value under higher K_p values contrary to what we observed for CIMP, since producing only with contingent resources is not an option in this case. Moreover, there is no monotonic relation between c_c and U^* and between K_c and U^* in CIMP-OT, as opposed to CIMP due to the interactions between the fixed costs and the capacitated nature of the problem. On the other hand, our observations

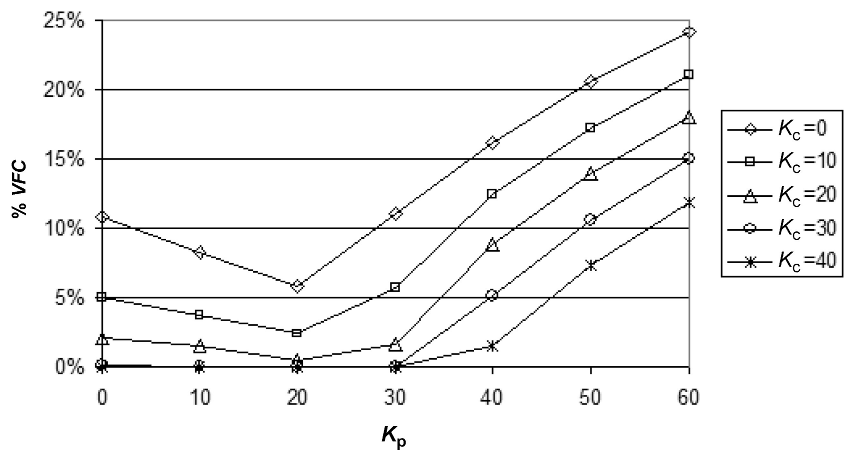

Fig. 7. Effect of fixed costs on the value of flexible capacity when $h = 1$, $b = 5$, $c_p = 2.5$, $c_c = 3.5$ and a normal demand with $CV = 0.1$.

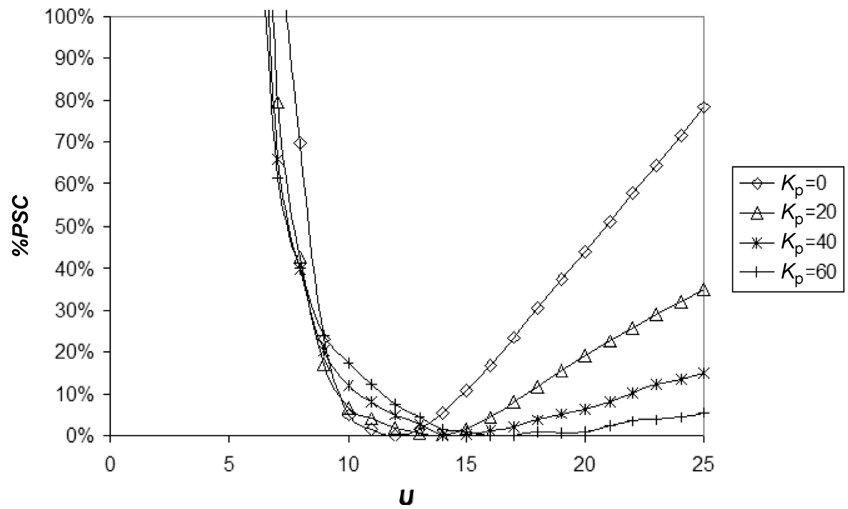

Fig. 8. Operating at suboptimal capacity levels under different fixed costs of production in the overtime model when $K_c = 20$, $h = 1$, $b = 5$, $c_p = 1.5$, $c_c = 2.5$ and a Normal demand with $CV = 0.1$.

remain similar for the relations between c_p and U^*, and between CV and U^* in both of the models.

The penalty of operating at suboptimal capacity levels exhibits a different structure in many problem instances since setting U^* to the zero level is not a viable option. Figure 8 depicts the relation for the same problem instances as Fig. 5. Indeed, in all problem instances we observe very high $\%PSC$ values under low permanent capacity levels due to limited production opportunity in CIMP-OT. However, starting from moderate permanent capacity levels (a level of ten in the problem instances of Fig. 8 for example), we observe lower $\%PSC$ values in CIMP-OT when compared to CIMP, because higher costs are incurred in CIMP-OT due to capacitated production and hence the penalty of not acquiring the optimal capacity is relatively lower.

For the overtime production case, there is still a considerable value of flexibility, however, the magnitude is less than that of the external contingent capacity case, due to the limit on the flexibility. We make similar observations as in Section 5.1.3 as to the relation between the problem parameters and the value of flexibility, except for a few cases that violate monotonicity due to the interactions of parameters brought by the capacitated nature of the problem.

6. Conclusions

In this paper the problem of determining the permanent capacity level in a make-to-stock environment under nonstationary stochastic demand with the option of a temporary increase of capacity via contingent resources such as temporary labor or overtime production was considered. A dynamic programming model was built to represent this problem, where the possibility of incurring distinct fixed costs to initiate production and to order contingent capacity is also incorporated. We ignored the fixed costs that may be associated with installing permanent capacity, which are incurred only at the beginning of the planning horizon. However, under the existence of such costs (that are independent of the permanent capacity size), one can first solve the problem by the proposed model and find the corresponding optimal permanent size. Then, the total minimum expected cost of this solution plus the fixed cost of installing this permanent capacity level could be compared with the minimum total expected costs obtained by solving CIMP when U is set to zero. The alternative with a lower expected total cost would give the optimal solution.

For the multi-period problem when the fixed costs are negligible, we showed that the expected total cost of the system is convex in the permanent capacity level and the starting inventory, using which the optimal permanent capacity level for any starting inventory level can be searched. The convexity result is intuitive, since too low levels of permanent capacity would result in elevated production and/or backorder costs and too high levels of permanent capacity would result in a low utilization of capacity. Nevertheless, this is not necessarily true for the case with positive fixed costs. If the system is working under too low a permanent capacity, then a marginal increase in permanent capacity may increase the system costs, because that capacity will be paid in the periods with no production as well, which may occur in order to avoid incurring fixed costs in every period.

Our computational analyses pointed out some useful managerial insights. In particular, our computations revealed that the optimal permanent capacity: (i) decreases as the costs of the contingent resources decrease; (ii) increases as the fixed cost of production increases until a threshold level, after which it is economically better to conduct all of the production with contingent resources; and (iii) decreases as the unit cost of permanent capacity increases.

Consequently, the optimal permanent capacity level may be equal to, greater than, or less than the expected average demand, with the possibility of zero as well. We have shown (in some cases analytically and in some cases numerically) that there are many problem instances where the optimal permanent capacity level is zero. This solution requires special attention from a managerial perspective since the parameter settings that result with $U^* = 0$ indicate the situations where the optimal course of action is to outsource all production or to produce in house with only contingent resources. Also, note that U^* may turn out to be zero even though the outsourcing option or the contingent resources are more expensive or less productive. For this case with more expensive contingent resources and $U^* = 0$, the solution in the multi-period problem refers to having many periods where no production takes place and the demand is mainly satisfied from stock that results from bulk production in some periods. Moreover, we have numerically shown that if this optimal course of action is not taken and a positive permanent capacity level is installed then the consequences of taking such a suboptimal action may be very costly.

One might expect that the optimal permanent capacity level increases as the variability of demand increases. However, we show in our computations that the optimal permanent capacity level does not necessarily exhibit a monotonic behavior as the variability of the demand increases. In particular, since the increase in demand variability requires a higher capacity flexibility, the optimal permanent capacity level tends to increase in demand variability as long as contingent capacity is relatively more expensive in comparison to holding a higher level of permanent capacity that will be kept partially idle, and it tends to decrease otherwise, zero permanent capacity always being an option.

As for the sensitivity of operating at suboptimal permanent capacity levels as the demand variability changes, we conclude that if the system has a high level of suboptimal permanent capacity, $\%PSC$ decreases as the demand variability increases. Similarly, when the system is operating with a low suboptimal permanent capacity when the optimal one is high, $\%PSC$ again decreases as the demand variability increases. Nevertheless, in some cases where the penalty of operating with a suboptimal permanent capacity is not high, $\%PSC$ increases as demand variability increases.

There exist relative values of problem parameters where introducing flexibility reduces the costs of the system significantly, even when the corresponding inflexible system is operated with an optimal capacity level: (i) lower costs of contingent capacity; and (ii) higher unit cost of keeping permanent capacity. Finally, no monotonicity results can be deduced for the value of flexibility as backorder costs and demand variability change, due to the system's ability to adapt itself by optimizing the permanent capacity level accordingly.

Acknowledgements

The authors would like to thank the Capacity Management Group (Will Bertrand, Nico Dellaert and Simme Douwe Flapper) of the Technology Management Department at the Technische Universiteit Eindhoven for introducing us the problem and for several useful suggestions and comments. The authors would also like to thank Nesim Erkip and Refik Güllü for helpful comments on an earlier version of this paper.

References

Angelus, A. and Porteus, E. (2002) Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand. *Management Science*, **48**(3), 399–413.

Atamturk, A. and Hochbaum, D. (2001) Capacity acquisition, subcontracting, and lot sizing. *Management Science*, **47**(8), 1081–1100.

Bradley, J. (2004) A Brownian approximation of a production-inventory system with a manufacturer that subcontracts. *Operations Research*, **52**(5), 765–784.

Bradley, J. and Arntzen, B. (1999) The simultaneous planning of production, capacity, and inventory in seasonal demand environments. *Operations Research*, **47**(6), 795–806.

Bradley, J. and Glynn, P. (2002) Managing capacity and inventory jointly in manufacturing systems. *Management Science*, **48**(2), 273–288.

Cheng, L., Subrahmanian, E. and Westerberg, A. (2004) Multi-objective decisions on capacity planning and production-inventory control under uncertainty. *Industrial & Engineering Chemistry Research*, **43**, 2192–2208.

Hu, J., Vakili, P. and Huang, L. (2004) Capacity and production management in a single product manufacturing system. *Annals of Operations Research*, **125**, 191–204.

Kekre, S., Rao, U., Sarthi, S. and Yi, J. (2004) Risk-based integration of strategic and tactical capacity planning. *Annals of Operations Research*, **132**, 85–108.

Kouvelis, P. and Milner, J. (2002) Supply chain capacity and outsourcing decisions: The dynamic interplay of demand and supply uncertainty. *IIE Transactions*, **34**(8), 717–728.

Pinker, E. and Larson, R. (2003) Optimizing the use of contingent labor when demand is uncertain. *European Journal of Operational Research*, **144**, 39–55.

Porteus, E. (2002) *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA.

Rajagopalan, S. and Swaminathan, J. (2001) A coordinated production planning model with capacity expansion and inventory management. *Management Science*, **47**(11), 1562–1580.

Tan, B. and Gershwin, S. (2004) Production and subcontracting strategies for manufacturers with limited capacity and volatile demand. *Annals of Operations Research*, **125**, 205–232.

Tan, T. and Alp, O. (2008) An integrated approach to inventory and flexible capacity management under non-stationary stochastic demand and setup costs. OR Spectrum, to appear.

Van Mieghem, J. (2003) Capacity management, investment, and hedging: review and recent developments. *Manufacturing & Service Operations Management*, **5**(4), 269–302.

Van Mieghem, J. and Rudi, N. (2002) Newsvendor networks: inventory management and capacity investment with discretionary activities. *Manufacturing & Service Operations Management*, **4**(4), 313–335.

Wu, S., Erkoc, M. and Karabuk, S. (2005) Managing capacity in the high tech industry: a review of literature. *The Engineering Economist*, **50**, 125–158.

Yang, J., Qi, X. and Xia, Y. (2005) A production-inventory system with Markovian capacity and outsourcing option. *Operations Research*, **53**(2), 328–349.

Zhang, J., Roundy, R., Cakanyildirim, M. and Huh, W. (2004) Optimal capacity expansion for multi-product, multi-machine manufacturing systems with stochastic demand. *IIE Transactions*, **36**(1), 23–36.

Appendix

Proof of Lemma 1. (Due to Porteus (2002). Let (x, U) and $(\bar{x}, \bar{U})$ be arbitrary elements of X. Let $0 \le \theta \le 1$ and $\bar{\theta} = 1 - \theta$. Let $\delta \ge 0$ be an arbitrary number. There must exist $y \in Y(x, U)$ and $\bar{y} \in Y(\bar{x}, \bar{U})$ such that $g(x, U, y) \le f(x, U) + \delta$ and $g(\bar{x}, \bar{U}, \bar{y}) \le f(\bar{x}, \bar{U}) + \delta$. Then:

$$\begin{aligned}\theta f(x, U) + \bar{\theta} f(\bar{x}, \bar{U}) &\ge \theta g(x, U, y) + \bar{\theta} g(\bar{x}, \bar{U}, \bar{y}) - \delta \\ &\ge g(\theta x + \bar{\theta}\bar{x}, \theta U + \bar{\theta}\bar{U}, \theta y + \bar{\theta}\bar{y}) - \delta \\ &\ge f(\theta x + \bar{\theta}\bar{x}, \theta U + \bar{\theta}\bar{U}) - \delta.\end{aligned}$$

Last inequality follows since δ is an arbitrary number and must hold for $\delta = 0$. ■

Proof of Theorem 1. Let $Y(x, U) = [x, \infty)$. Y is nonempty for every $(x, U) \in X$. Set C as defined in Lemma 1 as a convex set. Note that $f_{T+1}(\cdot) = 0$. Proof is by induction. For period T, let $J_T(x, U, y) = [y_t - x_t - U]^+ c_c + L_T(y_T)$. We first note that, for any pair of $(x^1, U^1, y^1) \in C$ and $(x^2, U^2, y^2) \in C$, and every scalar $\lambda \in [0, 1]$, defining $(x, U, y) \equiv \lambda(x^1, U^1, y^1) + (1 - \lambda)(x^2, U^2, y^2)$, we have $[y - x - U]^+ = [\lambda[y^1 - x^1 - U^1] + (1 - \lambda)[y^2 - x^2 - U^2]]^+ \le \lambda[y^1 - x^1 - U^1]^+ + (1 - \lambda)[y^2 - x^2 - U^2]^+$, which shows that $[y - x - U]^+$ is convex on C. Hence, J_T is convex on C as a summation of two convex functions is also convex. By Lemma 1, $f_T(x_t, U)$ is convex on X. By noting that $\alpha E f_t(x_t, U)$ is convex on X if f_t is convex as expectation of the function f is basically a convex combination and by using regular inductive arguments, we see that $f_t(x_t, U)$ is convex on X for every period t. ■

Proof of Theorem 2. The optimal replenishment policy stated in Equation (1) implies that

$$\begin{aligned} &f(U, x) \\ &= U c_p + \min_{y: x \le y} \{(y - x - U)^+ c_c + L(y)\} \\ &= U c_p + \begin{cases} c_c(y^c - x - U) + L(y^c) & \text{if } U \le y^c - x \\ L(x + U) & \text{if } y^c - x \le U \le y^u - x \\ L(y^u) & \text{if } 0 \le y^u - x \le U \\ L(x) & \text{if } y^u - x \le 0 \le U \end{cases} \\ &= U c_p + \begin{cases} c_c(y^c - x - U) + L(y^c) & \text{if } U \le y^c - x \\ L(x + U) & \text{if } y^c - x \le U \le y^u - x \\ L(\max\{y^u, x\}) & \text{if } y^u - x \le U. \end{cases} \end{aligned}$$

Then,

$$\frac{\partial f(U, x)}{\partial U} = c_p + \begin{cases} -c_c & \text{if } U \le y^c - x, \\ (h + b)G(x + U) - b & \text{if } y^c - x \le U \le y^u - x, \\ 0 & \text{if } y^u - x \le U. \end{cases}$$

When $c_p \ge c_c$, $\partial f(U, x)/\partial U$ is positive in the first and third regions, and because of convexity it must also be positive in the second region. Therefore, if $c_p \ge c_c$ then $U^* = 0$.

When $c_p < c_c$, $\partial f(U, x)/\partial U$ is negative in the first region and positive in the third region. Therefore, the sign must switch from negative to positive at a particular point in the second region due to convexity. Consequently, noting also that U must be non-negative:

$$U^* = \left(G^{-1}\left(\frac{b - c_p}{h + b}\right) - x\right)^+.$$

■

Proof of Lemma 2. First note that $Q^p \le U$ by the definition of Q^p. The total cost of the capacity installed, $U c_p$, does not depend on the production quantity and the production cost increases by c_p per each unit of capacity that is installed but not used. Hence, $U = Q^p$. ■

Proof of Lemma 3. By contradiction. Let $\hat{Q} = \hat{Q}^p + \hat{Q}^c$ be an optimal solution, where $\hat{Q}^p > 0$ and $\hat{Q}^c > 0$. Then, the production cost associated with $\hat{Q}$ is $PC(\hat{Q}) = K_p + K_c + \hat{Q}^p c_p + \hat{Q}^c c_c$, since $U = Q^p$ from Lemma 2. Now consider two alternative ways of producing the same quantity: $\dot{Q} = \dot{Q}^p + \dot{Q}^c = \hat{Q}$ with $\dot{Q}^c = 0$ and $\ddot{Q} = \ddot{Q}^p + \ddot{Q}^c = \hat{Q}$ with $\ddot{Q}^p = 0$. Then, the production cost associated with $\dot{Q}$ is $PC(\dot{Q}) = K_p + \dot{Q}^p c_p$, and that with $\ddot{Q}$ is $PC(\ddot{Q}) = K_p + K_c + \ddot{Q}^c c_c$. If $c_p \le c_c$, then $PC(\dot{Q}) < PC(\hat{Q})$, since $K_c > 0$. If $c_p > c_c$ then $PC(\ddot{Q}) < PC(\hat{Q})$. Consequently, $\hat{Q}^p > 0$ and $\hat{Q}^c > 0$ cannot be an optimal solution. Note also that if $c_p > c_c$ and $K_c > (c_p - c_c)\hat{Q}^c$, then $PC(\dot{Q}) < PC(\hat{Q})$. Hence, $Q^p Q^c = 0$. ■

Proof of Theorem 3. Due to Lemmas 2 and 3, we have either one of the following in the optimal solution:

(i) production with only permanent capacity: $U > 0$, $Q^p = U$, and $Q^c = 0$;
(ii) production with only contingent capacity: $U = 0$, $Q^p = 0$ and $Q^c > 0$;
(iii) no production: $U = Q^p = Q^c = 0$.

By comparing the costs incurring in each of the above situations, we prove the optimality for each case. If the optimal solution has the form of (i) then the cost function is $f^{(i)}(U, x) = K_p + U c_p + L(x + U)$ since $y^* = x + Q^p = x + U$. The optimal U value that minimizes this function is given by $\tilde{U} = G^{-1}((b - c_p)/(h + b)) - x =$

$y^{\mathrm{p}} - x$. By noting that $U > 0$, the optimal cost function can be rewritten as $f^{(i)}(\tilde{U}, x) = K_{\mathrm{p}} + \tilde{U} c_{\mathrm{p}} + L(x + \tilde{U}) = K_{\mathrm{p}} + c_{\mathrm{p}}(y^{\mathrm{p}} - x)^{+} + L(y^{\mathrm{p}})$. In this case, we have $(U^{*}, y^{*}) = (\tilde{U}, x + \tilde{U}) = (\tilde{U}, y^{\mathrm{p}})$. If the optimal solution has the form of (ii) then the cost function is $f^{(ii)}(0, x) = \min_{y:y \geq x}\{K_{\mathrm{p}} + K_{\mathrm{c}} + c_{\mathrm{c}}(y - x) + L(y)\}$. It can be shown that $y^{\mathrm{c}} = G^{-1}((b - c_{\mathrm{c}})/(h + b))$ is the unconstrained minimizer of the function inside the minimization. Hence, provided that $Q^{\mathrm{c}} > 0$, we have $f^{(ii)}(0, x) = K_{\mathrm{p}} + K_{\mathrm{c}} + c_{\mathrm{c}}(y^{\mathrm{c}} - x) + L(y^{\mathrm{c}})$ and $(U^{*}, y^{*}) = (0, y^{\mathrm{c}})$. If the optimal solution has the form of (iii) then $f^{(iii)}(0, x) = L(x)$ and $(U^{*}, y^{*}) = (0, x)$.

Case 1. $c_{\mathrm{p}} \leq c_{\mathrm{c}}$: In such a case only (i) and (iii) are viable options. Since $L(y)$ is convex, $y^{\mathrm{p}} \leq y^{\mathrm{u}}$, and y^{u} is the minimizer of $L(y)$, if $x \leq s^{\mathrm{p}}(x) \leq y^{\mathrm{p}}$ then $f^{(iii)}(0, x) = L(x) \geq L(s^{\mathrm{p}}(x)) = K_{\mathrm{p}} + c_{\mathrm{p}}(y^{\mathrm{p}} - x)^{+} + L(y^{\mathrm{p}}) = f^{(i)}(\tilde{U}, x)$. Hence, $(U^{*}, y^{*}) = (\tilde{U}, y^{\mathrm{p}})$. The result of the other condition $(x > s^{\mathrm{p}}(x))$ can be shown in the same manner.

Case 2. $c_{\mathrm{p}} > c_{\mathrm{c}}$: In such a case all three options (i), (ii) and (iii) are viable. In this case we have $y^{\mathrm{p}} \leq y^{\mathrm{c}} \leq y^{\mathrm{u}}$. Note also that $s^{\mathrm{c}}(x) \leq y^{\mathrm{u}}$ by definition. If $x \leq s^{\mathrm{c}}(x)$ and $s^{\mathrm{p}}(x) \leq s^{\mathrm{c}}(x)$ then $L(s^{\mathrm{p}}(x)) = K_{\mathrm{p}} + c_{\mathrm{p}}(y^{\mathrm{p}} - x)^{+} + L(y^{\mathrm{p}}) = f^{(i)}(\tilde{U}, x) \geq L(s^{\mathrm{c}}(x)) = K_{\mathrm{p}} + K_{\mathrm{c}} + L(y^{\mathrm{c}}) + c_{\mathrm{c}}(y^{\mathrm{c}} - x)^{+} = f^{(ii)}(0, x)$. Similarly, $L(x) = f^{(iii)}(0, x) \geq L(s^{\mathrm{c}}(x)) = f^{(ii)}(0, x)$. Hence, $(U^{*}, y^{*}) = (0, y^{\mathrm{c}})$. The results of the other conditions can be shown in a similar way. ■

Proof of Theorem 5. Setting $Y(x, U) = [x, x + \eta U]$, we see that Y is a non-empty set and the set C as defined in Lemma 1 is convex. Therefore, Lemma 1 still applies to this case. The proof of the convexity in this case goes along with the lines of the proof of Theorem 1 and hence is omitted here. ■

Solution of CIMP-OT for the single period

Theorem A1. *The optimal permanent capacity level of the single-period problem when both* K_{p} *and* K_{c} *are zero is given by*

$$U^{*} = \begin{cases} \frac{1}{\eta}(y^{\eta} - x)^{+} & \text{if } c_{\mathrm{p}} \geq c_{\mathrm{c}} \\ (y^{\mathrm{p}} - x)^{+} & \text{if } c_{\mathrm{p}} < c_{\mathrm{c}} \end{cases}$$

where $y^{\eta} = G^{-1}((b - c_{\mathrm{p}} - c_{\mathrm{c}}(\eta - 1))/(h + b))$ and $y^{\mathrm{p}} = G^{-1}((b - c_{\mathrm{p}})/(h + b))$.

Proof. The proof goes along with the lines of the proof of Theorem 2 and hence is omitted here. ■

Theorem A2. *If* $c_{\mathrm{p}} y^{\mathrm{p}} - c_{\mathrm{c}} y^{\eta} + (c_{\mathrm{p}} - c_{\mathrm{c}})\frac{x(\eta - 1) - y^{\eta}}{\eta} \leq K_{\mathrm{c}} + L(y^{\eta}) - L(y^{\mathrm{p}})$*, then:*

$$(U^{*}, y^{*}) = \begin{cases} ((y^{\mathrm{p}} - x)^{+}, y^{\mathrm{p}}) & \text{if } x \leq s^{\mathrm{p}}(x), \\ (0, 0) & \text{otherwise.} \end{cases}$$

Otherwise

$$(U^{*}, y^{*}) = \begin{cases} (\frac{1}{\eta}(y^{\eta} - x)^{+}, y^{\eta}) & \text{if } x \leq s^{\eta}(x), \\ (0, 0) & \text{otherwise,} \end{cases}$$

where

$$s^{\eta}(x) = \min\left\{s : L(s) = \frac{1}{\eta}(y^{\eta} - x)^{+} c_{\mathrm{p}} + K_{\mathrm{p}} + K_{\mathrm{c}} + c_{\mathrm{c}}\left(y^{\eta} - x - \frac{1}{\eta}(y^{\eta} - x)^{+}\right) + L(y^{\eta})\right\}.$$

Proof. Omitted. ■

We note that, similar to the previous results, the optimal capacity level in the single-period problem is independent of c_{c} and η when $c_{\mathrm{p}} < c_{\mathrm{c}}$, since no overtime is utilized in this case.

Biographies

Osman Alp is an Assistant Professor in the Industrial Engineering Department of Bilkent University, Ankara, Turkey. He received his Ph.D in Industrial Engineering from the Middle East Technical University, Ankara, Turkey, in 2000. He then worked as a Postdoctoral Research Fellow and Sessional Instructor in the School of Business, University of Alberta, Canada. He joined the Industrial Engineering Department of Bilkent University in 2003. His current research interests include supply chain management, stochastic and deterministic inventory models and transportation of hazardous materials.

Tarkan Tan is an Assistant Professor in the Technology Management Department at Technische Universiteit Eindhoven, The Netherlands. He received his Ph.D in Industrial Engineering from the Middle East Technical University, Ankara, Turkey, in 2002. His research interests include capacity management, inventory theory and supply chain management.