

An integrated approach to inventory and flexible capacity management under non-stationary stochastic demand and set-up costs

Citation for published version (APA):

Tan, T., & Alp, O. (2005). *An integrated approach to inventory and flexible capacity management under non-stationary stochastic demand and set-up costs*. (BETA publicatie : working papers; Vol. 132). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2005

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

An Integrated Approach to Inventory and Flexible Capacity Management under Non-stationary Stochastic Demand and Set-up Costs

Tarkan Tan^{a,*}, Osman Alp^b

^a *Department of Technology Management, Eindhoven University of Technology,
P.O. Box 513, 5600MB Eindhoven, The Netherlands*

^b *Industrial Engineering Department, Bilkent University, 06800 Ankara, Turkey*

January 2005

Abstract: In a manufacturing system with flexible capacity, inventory management can be coupled with capacity management in order to handle fluctuations in demand more effectively. A typical example is the effective use of temporary workforce. In this paper, we discuss an integrated model for inventory and flexible capacity management under non-stationary stochastic demand with the possibility of positive set-up costs, both for initiating production and ordering contingent capacity. We analyze the characteristics of the optimal policies for the integrated problem. We also evaluate the value of utilizing flexible capacity under different settings, which enable us to develop managerial insights.

Keywords: Inventory; Production; Stochastic Processes; Capacity Management; Flexible Capacity

1. Introduction

A crucial problem that manufacturing companies face is how to cope with volatility in demand. For make-to-stock environments, holding safety stocks is the traditional remedy for handling the stochasticity in demand. If there is non-stationarity in demand, such as seasonality, then adjusting production capacity dynamically is another possible tool. While there is an extensive amount of literature on both of these measures, our aim is to contribute to the relatively limited research that considers both at the same time, which may be necessary especially if the demand is both non-stationary and stochastic. Consequently, in this paper we consider a periodic review make-to-stock production environment under non-stationary stochastic demand.

*Corresponding author

E-mail addresses: t.tan@tm.tue.nl (T. Tan), osmanalp@bilkent.edu.tr (O. Alp)

In most of the traditional production/inventory literature, either an infinite production capacity is assumed or a given finite capacity is considered as a constraint rather than a decision variable. We relax this assumption in the sense that the flexible capacity level in each period is to be optimized, as well as the amount of production. Capacity can be defined as the total productive capability of all productive resources utilized, such as 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 internal resources of the 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 capacity acquired temporarily as the contingent capacity. Flexible capacity management refers to adjusting the total production capacity in any period with the option of utilizing contingent resources on top of the permanent ones.

The capacity decisions can be in all hierarchies of decision making: strategic, tactical, and operational. Examples of such decisions would be, 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 operational level. For the ease of exposition in some parts we refer to the workforce capacity setting, not to mean that our analysis is confined to that environment. Consequently, the problem we consider can be viewed as one where the production is mostly determined by the workforce size. This workforce size is flexible, in the sense that temporary (contingent) workers can be hired in any period and they are paid only for the periods they work. The productivity of contingent workers are allowed to be different than that of permanent workers in our model. The size of the permanent workforce is assumed to be pre-determined in the tactical level and hence considered as fixed for a given planning horizon. Our model also allows for the incorporation of set-up costs associated with (i) initiating production in each period (production set-up cost) and (ii) ordering contingent capacity.

In this paper, we first show that the integrated flexible capacity and inventory management problem that we consider can be transformed into a typical production/inventory problem with neither concave nor convex production costs. Then we characterize the single period optimal ordering policy of such problems completely and elaborate on the general

characteristics of the multi-period problem. Finally, we develop several managerial insights on the value of utilizing flexible capacity, determining the range of problem parameters where it is of more value.

The rest of the paper is organized as follows. We present a review of relevant literature in Section 2 and present our dynamic programming model in Section 3. The optimal policy for the integrated problem is discussed in Section 4 and the value of utilizing flexible capacity is analyzed in Section 5. We summarize our conclusions and suggest some possible extensions in Section 6.

2. Related Literature

The problem that we deal with have interactions with a number of related fields. In specific, the following three fields have the closest connection: (i) integrated production/capacity management, (ii) workforce planning and flexibility, and (iii) capacitated production/inventory models. Instead of providing a detailed literature survey on each of these fields, we cite examples of related work from each of them and discuss some of the similarities and differences between those problems and the one that we consider.

An excellent survey of strategic capacity management problems mainly focusing on the capacity expansion decisions is presented by Van Mieghem (2003). The author explains prominent issues in formulating and solving various capacity management problems.

Atamtürk and Hochbaum (2001) deal with an integrated capacity and inventory management problem under a finite planning horizon and deterministic demand where trade-offs between capacity expansions, subcontracting, production, and inventory holding are exploited. Angelus and Porteus (2002) also deal with an integrated problem for a short-life-cycle product where the demand has a stochastically increasing and then decreasing structure. Authors show that the optimal capacity level follows a target interval policy. In another work that deals with integrated problems, Dellaert and de Kok (2004) show that integrated capacity and inventory management approaches outperform decoupled approaches.

Hu et al (2004) deal with an environment similar to ours: There is a fixed permanent production capacity, but it can temporarily be increased by using contingent capacity. Unlike ours, the problem is modelled on a continuous time framework under a demand rate that is Markov-modulated, and no setup costs are considered. Tan and Gershwin (2004) also deal with a similar problem with a similar approach. The differences are the existence of several

subcontracting opportunities with different cost and capacity structures and the demand being dependent on the lead time distribution during out of stock periods.

In the workforce planning and flexibility field, Holt et al. (1960) in their seminal work present models that exploit the trade-off between keeping large sized permanent workforce levels capable of satisfying peak season demands and frequent adjustment of the workforce level to cope with fluctuations. Indeed, this very idea of aggregate production planning problem constitutes the essence of our problem too, nevertheless we consider non-stationary stochastic demand, unlike the deterministic demand assumption of aggregate production planning models. Wild and Schneeweiss (1993) analyze and compare four “instruments” to cope with fluctuating demand when the capacity is defined in terms of work force level: variation of monthly working time, use of overtime, employment of temporary workers, and leasing a work force. A hierarchical model based on dynamic programming is presented for making rational decisions on the selective use of these instruments.

Milner and Pinker (2001) deal with the design of contracts between firms and external labor supply agencies for hiring long term and temporary workers under demand and supply uncertainty in a single period environment. In a related work, Pinker and Larson (2003) consider the problem of managing permanent and contingent workforce level under uncertain demand in a finite planning horizon where inventory holding is not allowed. The sizes of regular and temporary labor are decision variables that are fixed throughout the planning horizon, but the capacity level may be adjusted by setting the number of shifts for each class of workers.

The papers by Federgruen and Zipkin (1986) and Kapuscinski and Tayur (1998) are two examples of the research stream on capacitated production/inventory problems with stochastic demand, where no setup costs of production exist. In this case, it is shown that base stock type policies are optimal. Gallego and Scheller-Wolf (2000) consider the setup cost of production under a similar environment and characterize the optimal policy partially. Our model is an extension and a generalization of this stream of research, in which we provide the explicit solution of a single period problem.

3. 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 rele-

vant costs in our environment are inventory holding and backorder costs, unit cost of permanent and contingent capacity, set-up cost of production, and fixed cost of ordering contingent capacity, all of which are non-negative. We assume that there is an infinite supply of contingent workers, and that 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 the ease of reference.

Table 1: Summary of Notation

T	: Number of periods in the planning horizon
U	: Size of available permanent capacity
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 ($0 < \alpha \leq 1$)
s_t	: Size of contingent capacity ordered in period t
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
x_t	: Inventory position at the beginning of period t before ordering
y_t	: Inventory position in period t after ordering
x_t^u	: Inventory position in period t after full permanent capacity and no contingent capacity are used for production ($x_t^u = x_t + U$)
$f_t(x_t)$: Minimum total expected cost of operating the system in periods $t, t + 1, \dots, T$, given the system state x_t
γ	: productivity of contingent resources relative to the permanent resources

Changing the level of permanent capacity as a means of coping with demand fluctuations, such as hiring and firing permanent workers, is not only very costly in general, but it may also have many negative impacts on the company, especially if the demand is highly variable. In 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 tools of coping with fluctuating demand, along with holding inventory. Yet, long-term changes in the state of the world can make permanent capacity changes unavoidable. Consequently, we consider the determination of the perma-

nent capacity level as a tactical decision that is made at the beginning of a finite horizon and not changed until the end of the horizon. This decision is kept out of the scope of this study since we focus only on the operational decisions. Therefore, a finite horizon dynamic programming model becomes an appropriate choice to formulate this problem for a given permanent capacity level.

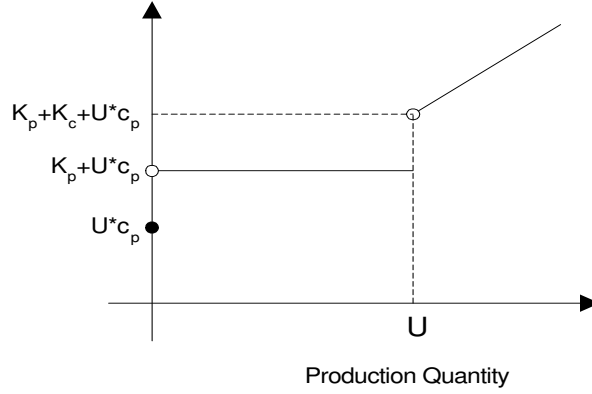
We consider a production cost component which is a linear function of permanent capacity in order to represent the costs that do not depend on the production quantity (even when there is no production), such as the salaries of permanent workers. That is, each unit of permanent capacity costs c_p per period, and the total cost of permanent capacity per period is $U * c_p$, 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 workers is U “unit”s per period. We also define the cost of production by temporary workers in the same unit basis in the following way. Let c'_c be the hiring cost of a temporary labor 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 in time, the unit production cost by 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 positive fixed costs, both for production and contingent capacity ordering. Let K_p denote the production set-up cost and K_c denote the fixed cost of ordering contingent capacity. K_p is charged whenever the production is initiated, even if the permanent workforce size is zero and all production is due to temporary workers. 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 note that it is more likely to have $K_p > K_c$, but the model holds for all values of K_p and K_c .

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 ordered in that period, s_t , for any given level of permanent capacity, via $s_t = [(Q_t - U)^+/\gamma]$, 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 the classical production/inventory problem with a piecewise linear production cost (made up of labor costs), which is neither convex nor concave under positive fixed costs. See Figure 1 for an illustration. Note that when K_p and K_c are both zero, this function is convex.

Figure 1: Production Cost Function Under Positive Set-up Costs



The order of events is as follows. 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 $y_t \leq x_t + U = x_t^u$ then only permanent capacity is utilized, otherwise a contingent capacity of size $s_t = [(y_t - x_t^u)^+ / \gamma]$ is hired on top of full permanent capacity usage. At the end of the period, the demand is met/backlogged. Consequently, denoting the minimum cost of operating the system from the beginning of period t until the end of the planning horizon as $f_t(x_t)$, we use the following dynamic programming formulation to solve the problem.

$$\begin{aligned}
 \text{(CIMP): } f_t(x_t) = & U c_p + \min_{x_t \leq y_t} \left\{ K_p \delta(y_t - x_t) + K_c \delta(y_t - x_t^u) + [y_t - x_t^u]^+ c_c \right. \\
 & \left. + L_t(y_t) + \alpha E [f_{t+1}(y_t - W_t)] \right\} \quad \text{for } t = 1, 2, \dots, T
 \end{aligned}$$

where $L_t(y_t) = h \int_0^{y_t} (y_t - w) dG_t(w) + b \int_{y_t}^{\infty} (w - y_t) dG_t(w)$ is the regular loss function, and $\delta(\cdot)$ is the function that attains the value 1 if its argument is positive, and zero otherwise. We assume the ending condition to be $f_{T+1}(x_{T+1}) = 0$. We note that the optimal contingent

capacity usage is independent of c_p , because the permanent capacity is fixed and cannot be changed, and hence the costs of holding that capacity is “sunk”.

Remark 1 *When $K_p = K_c = 0$ and $c_c \rightarrow \infty$, CIMP boils down to a capacitated production/inventory problem. Similarly, when $K_p > 0$ and either $K_c \rightarrow \infty$ or $c_c \rightarrow \infty$, CIMP boils down to a capacitated production/inventory problem with production setup costs.*

4. Integrated Flexible Capacity and Inventory Management

The characteristics of the problem and the optimal solution show significant differences depending on whether the set-up costs are strictly positive or not. Therefore, we analyse those two cases separately in the following two subsections.

4.1 Analysis with no Set-up Costs

As discussed in Section 3, the special case of no set-up costs translates into the classical production/inventory problem with piecewise linear, convex production cost and linear holding and backordering costs. The optimal policy of the multi-period production/inventory problem with a (strictly) convex production cost function is characterized in Karlin (1958) as a modified version of the order-up-to type policy, where the order-up-to level (which is not necessarily equal to the reorder level and can be less than that) is a function of the inventory level. It is also shown that as the initial inventory level increases, the quantity ordered decreases while the order-up-to level increases. For the single-period problem with a piecewise linear (hence non-strict) convex production cost, Porteus (1990) discusses that the optimal policy is of order-up-to type, where the order-up-to level is piecewise linear increasing in initial inventory level. He refers to this policy as a “finite generalized base stock policy”, since there are a finite number of distinct base stock levels. For the single-period problem with piecewise linear non-convex production cost, he suggests evaluating the optimal solutions within linear regions for a given initial inventory level (x) and picking the best among them, because of the difficulty of obtaining an explicit solution for all x due to the dependence of the optimal ordering policies on x .

Let $J_t^p(y) = L_t(y) + \alpha E[f_{t+1}(y - W_t)]$ and $J_t^c(y|x) = J_t^p(y) + c_c(y - x^u)$, where $x^u = x + U$.

Then the problem can be stated as

$$\text{(CIMP-NS): } f_t(x_t) = U c_p + \min_{x_t \leq y_t} \{J_t(y_t|x_t)\}$$

where

$$J_t(y|x) = \begin{cases} J_t^p(y) & \text{if } y \leq x^u \\ J_t^c(y|x) & \text{if } y \geq x^u \end{cases}$$

for $t = 1, 2, \dots, T$. Recalling that $f_{T+1}(x) = 0$, $J_T^p(y)$ is convex in $y \in \mathfrak{R}$ since the loss function $L_T(y)$ is a convex function. $J_T^c(y|x)$ is also a convex function in $y \in \mathfrak{R}$ for a given value of x . Hence, the first order condition is sufficient for the minimization, where

$$\frac{dJ_T^p(y)}{dy} = (h+b)G_T(y) - b \quad \text{and} \quad \frac{dJ_T^c(y|x)}{dy} = (h+b)G_T(y) - b + c_c.$$

Let y_t^p and y_t^c be the minimizers of the functions $J_t^p(y)$ and $J_t^c(y|x)$, respectively, for $t = 1, 2, \dots, T$. Then, assuming that G is an invertible function, we have

$$y_T^p = G_T^{-1} \left(\frac{b}{h+b} \right) \quad \text{and} \quad y_T^c = G_T^{-1} \left(\frac{b-c_c}{h+b} \right).$$

Theorem 1 *Optimal ordering policy of CIMP-NS at any period $t = 1, 2, \dots, T$ is of state-dependent order-up-to type, where the optimal order-up-to level, $y_t^*(x_t)$, is*

$$y_t^*(x_t) = \begin{cases} y_t^c & \text{if } x_t \leq y_t^c - U \\ x_t + U & \text{if } y_t^c - U \leq x_t \leq y_t^p - U \\ y_t^p & \text{if } y_t^p - U \leq x_t \leq y_t^p \\ x_t & \text{if } y_t^p \leq x_t \end{cases}.$$

Proof : See Appendix.

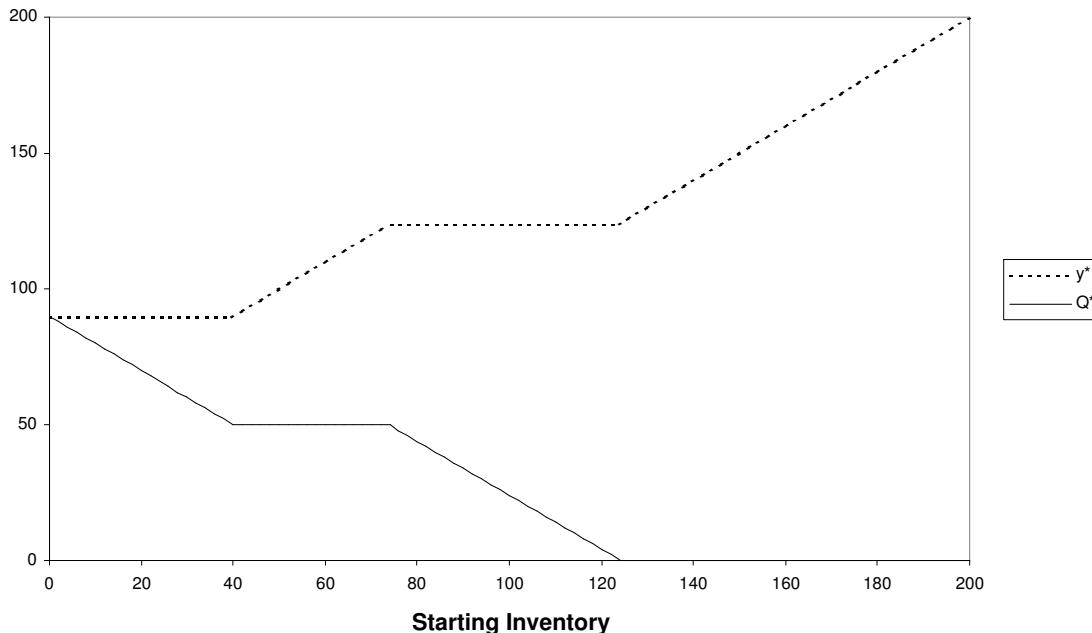
We note that although y_t^c and y_t^p are independent of x_t , $y_t^*(x_t)$ is a function of x_t .

Property 1 *The optimal production quantity in any given period is a non-increasing function of the starting inventory level. Similarly, the optimal order-up-to level in any given period is a non-decreasing function of the starting inventory level.*

This monotonic relation is also discussed by Porteus (1990) and is illustrated in Figure 2 for a single period problem instance.

Property 2 *The two critical order-up-to points, y_t^c and y_t^p , of CIMP-NS have a non-increasing structure for a stationary demand stream as the number of periods remaining in the planning horizon decreases.*

Figure 2: Optimal Production Quantities and Order-up-to Levels vs Starting Inventory



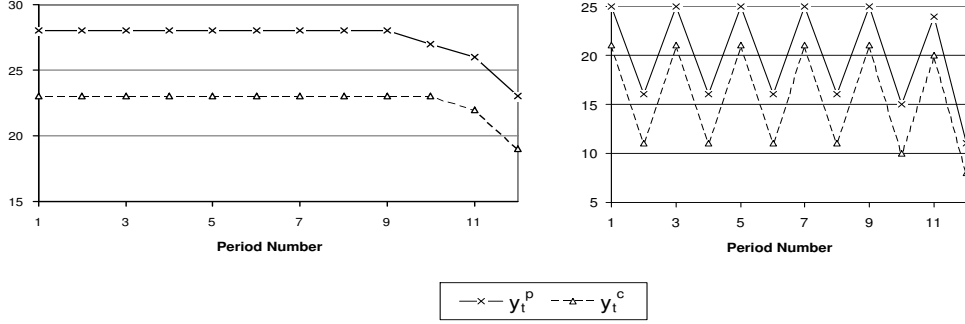
Karlin (1958) and Porteus (1990) present general discussion of similar results for various production/inventory problems. Figure 3 shows this behavior for two problem instances: one with a stationary demand and the other with a periodic demand of two seasons.

4.2 Analysis with Set-up Costs

When the setup costs in the problem, K_p and K_c , are positive, CIMP translates into a finite horizon production/inventory problem with neither concave nor convex production costs. In what follows, we characterize the optimal ordering policies for the last period of the planning horizon (single period problem) and discuss the complications of deriving general optimal policies for multiple periods.

We first note that there exist unique numbers $s^c(x)$, $s^u(x)$, and $s^p(x)$ for a given value of x , due to the convexity of the functions $J_T(y|x)$ and $J_T^p(y)$, such that

Figure 3: Optimal Control Parameters vs Time for Stationary and Non-stationary Demand



$$\begin{aligned}
 s^c(x) &= \min\{s : J_T(s|x) = J_T^c(y_T^c|x) + K_p + K_c\}, \\
 s^u(x) &= \min\{s : J_T^p(s) = J_T^p(x + U) + K_p\}, \\
 s^p(x) &= \min\{s : J_T^p(s) = J_T^p(y_T^p) + K_p\}.
 \end{aligned}$$

In order to simplify the exposition of the optimal policy in the last period, we introduce two auxiliary functions, $s(x)$ and $S(x)$, as follows.

$$s(x) = \begin{cases} \max\{s^u(x), s^c(x)\} & \text{if } x^u \leq y_T^c \\ s^u(x) & \text{if } y_T^c \leq x^u \leq y_T^p \\ s^p(x) & \text{if } y_T^p \leq x^u \end{cases}$$

$$S(x) = \begin{cases} y_T^c & \text{if } s(x) = s^c(x) \\ x^u & \text{if } s(x) = s^u(x) \\ y_T^p & \text{if } s(x) = s^p(x) \end{cases}$$

Theorem 2 *The optimal ordering policy of the last period is*

$$y_T^*(x) = \begin{cases} S(x) & \text{if } x \leq s(x) \\ x & \text{otherwise} \end{cases} .$$

Proof : See Appendix.

In this optimal policy, there exists a state dependent reorder level which is a function of the starting inventory level, x . This function, $s(x)$, takes the value of one of the critical levels $s^c(x)$, $s^u(x)$, and $s^p(x)$, where production does not pay off above that level due to the existence of setup costs. Critical levels $s^c(x)$, $s^u(x)$, and $s^p(x)$ can be considered as the

“reorder” levels for production with contingent capacity, production with full capacity, and production with idle capacity, respectively. Therefore, if the starting inventory level is below the reorder level, the optimal order-up-to level is given by y_T^c, x^u , or y_T^p depending on the value that $s(x)$ takes, either $s^c(x)$, or $s^u(x)$, or $s^p(x)$, respectively. Otherwise, no orders are placed.

The following result characterizes the structure of the reorder level function with respect to the starting inventory level, x .

Theorem 3 $s(x)$ is non-decreasing in x .

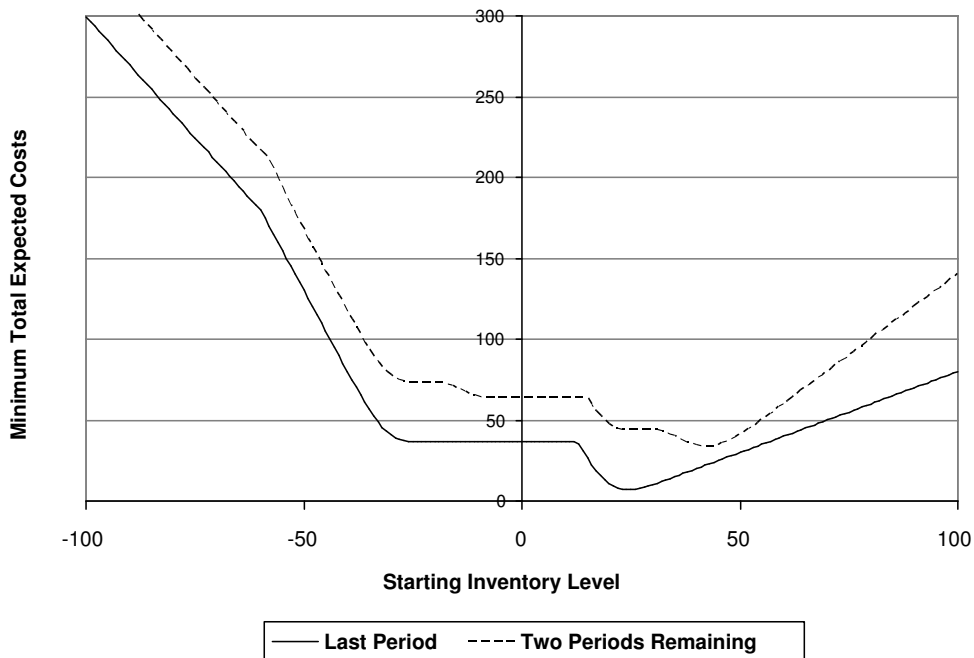
Proof : See Appendix.

The optimal policy of the last period cannot simply be generalized to multiple periods. Such a generalization would require the convexity -or quasiconvexity- of the function $J_t(y|x)$ over y for every period t . However, while $f_T(x)$ is a quasiconvex function, its expectation ($E[f_T(y - W_T)]$) is not necessarily quasiconvex. Similarly, even if $E[f_T(y - W_T)]$ is quasiconvex, $J_{T-1}(y|x)$ is not necessarily quasiconvex, since the summation of convex and quasiconvex functions, as well as the convex combinations of quasiconvex functions may fail to be quasiconvex. Therefore, regular inductive arguments do not hold. Figure 4 illustrates this point for a particular problem instance. The solid line in this figure is the graph of $f_T(x)$, which is quasiconvex, and the dashed line is the graph of $f_{T-1}(x)$, which fails to be quasiconvex, because the function increases in the region $-30 \leq x \leq -20$.

Gallego and Scheller-Wolf (2000) consider a capacitated production/inventory problem under setup costs of production, which is a special case of CIMP as mentioned in Section 3. They state that the full characterization of the optimal policies for this special case is extremely difficult but the authors could generate four regions of the starting inventory level where optimal production decisions are characterized to a certain extent. The authors also show that there may exist further sub-intervals in two of these four regions where the optimal production quantities may fluctuate with the increase of the starting inventory level. We conjecture the optimal ordering policies of CIMP to be even more complicated than that of this special case.

We observe also in CIMP that there is no monotonic relation between the starting inventory level (x) and the optimal production quantity. This result is depicted in Figure 5, which is the result of a 3-period problem instance. We note that, for this problem instance

Figure 4: Starting Inventory vs Minimum Expected Total Costs



the optimal production quantity has a non-increasing structure in x up to $x = -11$, where the optimal production quantity is 35. However, when $x = -10$ the optimal decision is to produce 50 units, and this violates the monotonic structure.

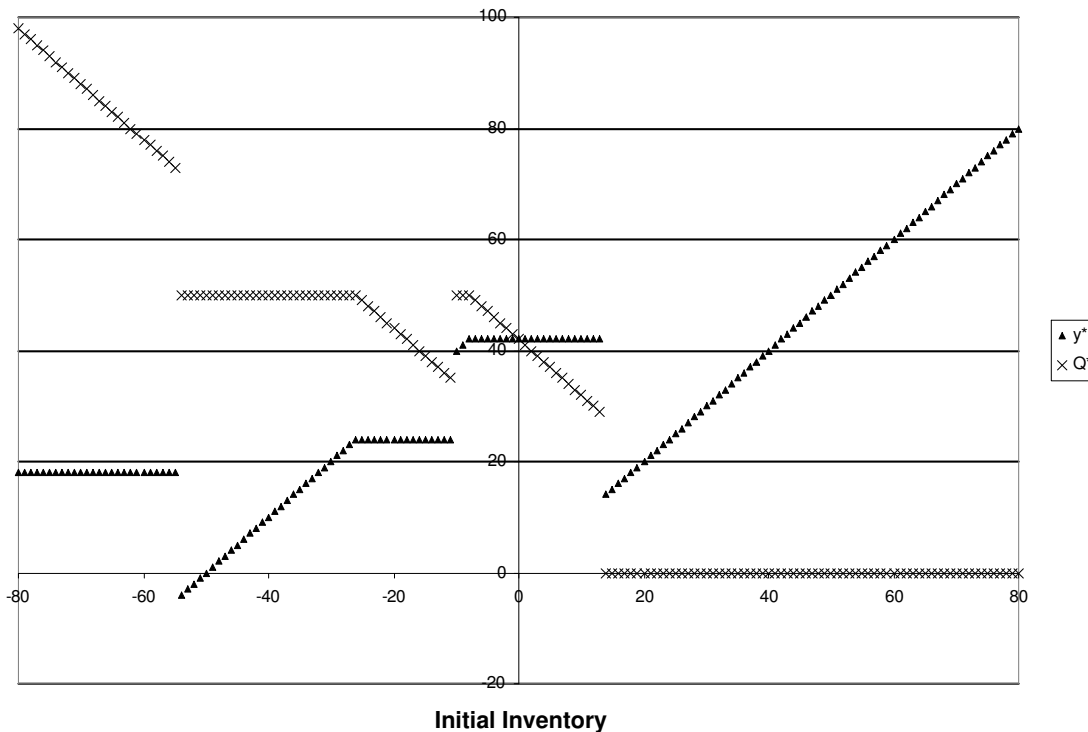
Figure 5 also shows the optimal order-up-to levels with respect to x . We observe that there are 3 distinct values of optimal order-up-to levels on this figure other than x , and full permanent capacity production, $x + U$. Note that there exist only two distinct critical order-up-to points (other than x and $x + U$) in the optimal policy stated for the last period.

5. Value of Flexible Capacity

Our purpose in this section is to investigate the value of utilizing flexible capacity and the sensitivity of it as system parameters change. We compare a flexible capacity (FC) system with an inflexible one (IC), where the contingent capacity can be utilized in the former but cannot in the latter.

We define the (absolute) value of flexible capacity, VFC , as the difference between the optimal expected total cost of operating the IC system, ETC_{IC} , and that of the FC system,

Figure 5: Starting Inventory vs Optimal Production Quantity for 3 Periods to go

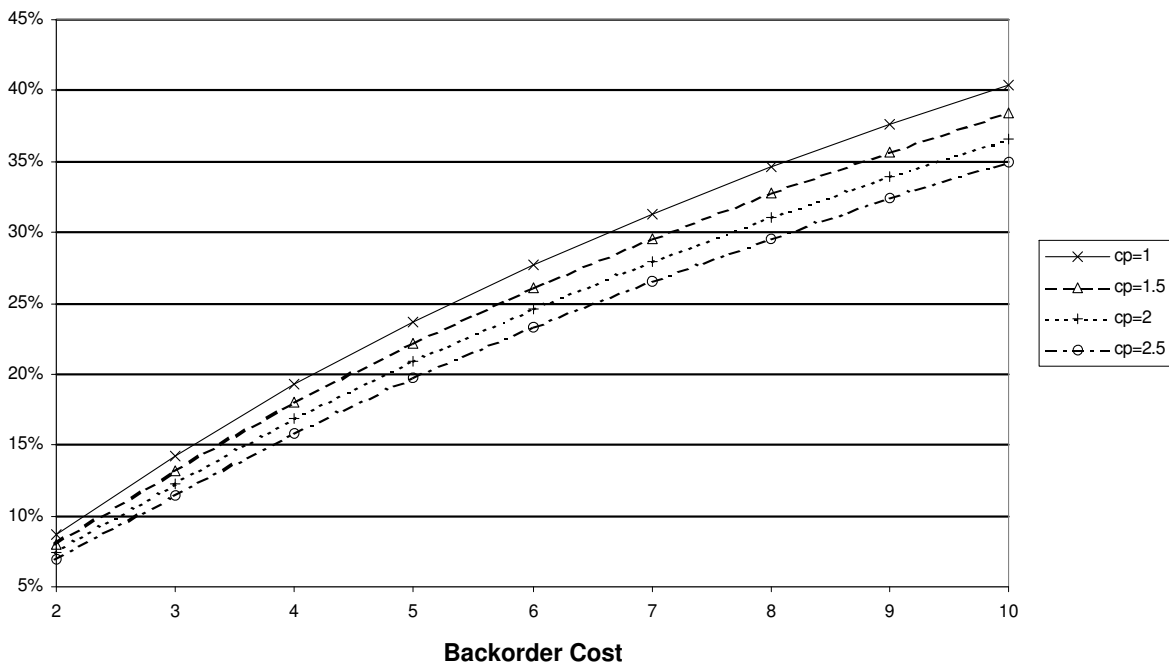


ETC_{FC} . That is, $VFC = ETC_{IC} - ETC_{FC}$. In order to reflect the relativity, we also define the relative value of flexible capacity as $\%VFC = VFC/ETC_{IC}$. We note that both VFC and $\%VFC$ are always non-negative.

The first conclusion that can be drawn is that the value of flexible capacity increases as the contingent capacity becomes less costly to utilize. It is easy to see why this relation holds: As c_c or K_c decreases while the other parameters are kept constant, ETC_{IC} remains the same, since flexible capacity is not utilized in this case, but ETC_{FC} decreases due to decreased costs. Consequently, both VFC and $\%VFC$ increases.

We conduct some numerical experiments to reveal the sensitivity of $\%VFC$ with respect to the change in the rest of the system parameters. In the IC case of these numerical examples, we simply take the contingent capacity cost high enough to assure that no contingent capacity is used (see Remark 1). We solve CIMP for the following set of input parameters, unless otherwise noted: $T = 12$, $U = 10$, $b = 5$, $h = 1$, $c_c = 2.5$, $c_p = 1.5$, $K_p = 40$, $K_c = 20$, $\alpha = 0.99$, and $x_1 = 0$. We consider demand that follows a seasonal pattern with

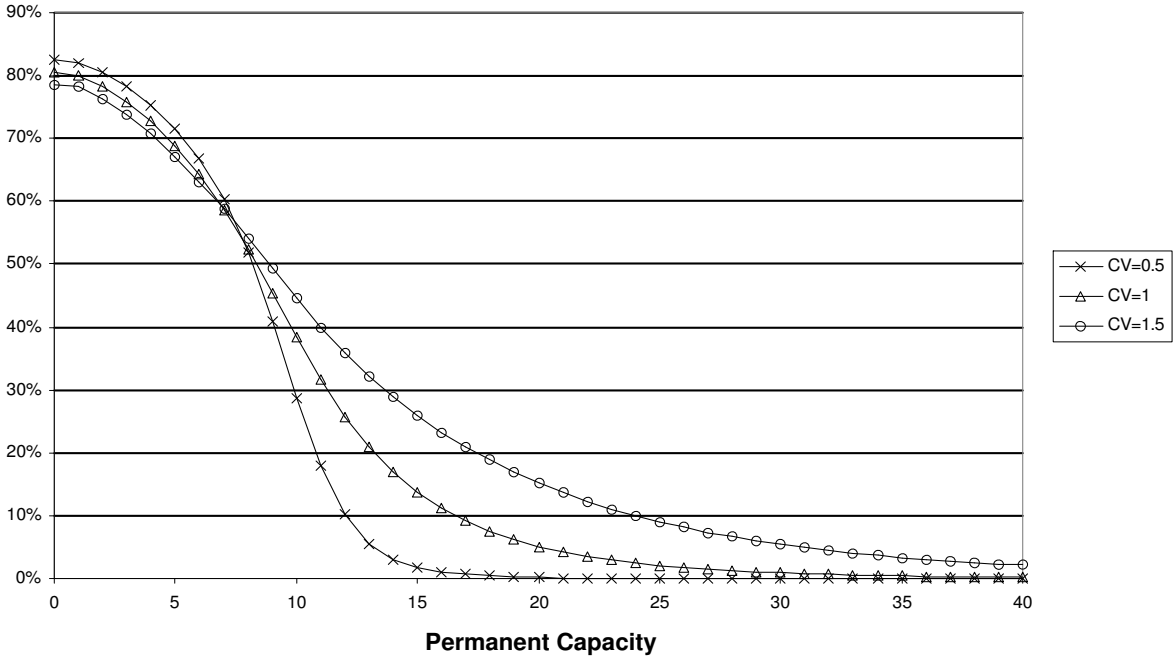
Figure 6: % Value of Flexible Capacity vs Backorder Cost under Different Permanent Capacity Costs



a cycle of 4 periods, where the expected demand is 10, 15, 10, and 5, respectively. Note that the capacity is the same as the average demand for this base set. We first consider Poisson distribution, but later we also consider Gamma distribution, in order to investigate the impact of variation. 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.

We first test the value of flexibility with respect to the backorder cost by varying the value of b between 2 and 10. Figure 6 verifies intuition in the sense that $\%VFC$ is higher when backorders are more costly, or equivalently when higher service levels are targeted. Another result that the same graph depicts is that $\%VFC$ is higher when the permanent capacity cost is lower. This result holds only for the relative value, not for the absolute value. Indeed, absolute VFC remains unchanged for all values of c_p for any given b , because the optimal contingent capacity usage is independent of c_p , as discussed in Section 3. Yet, the relative value, $\%VFC$, decreases for higher c_p , because both of the expected costs, hence

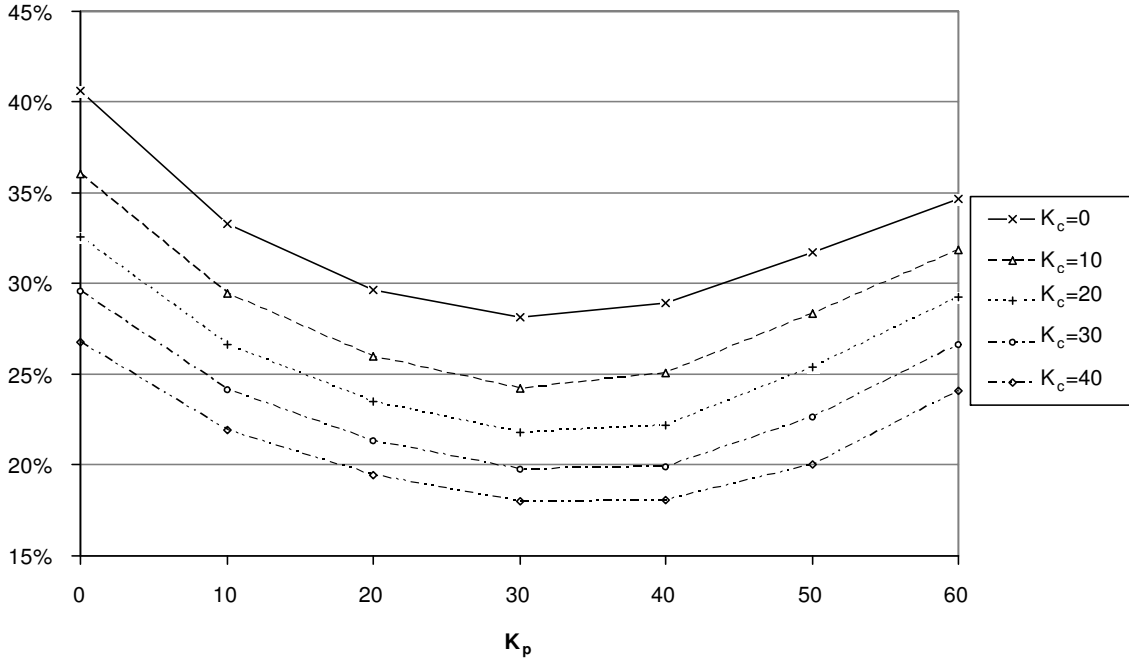
Figure 7: % Value of Flexible Capacity vs Permanent Capacity Size under Different CVs



ETC_{IC} , increases.

Another result we obtain is that $\%VFC$ increases as the permanent capacity decreases, as can be expected, since the inadequacy of permanent capacity increases the requirement for contingent capacity. As shown in Figure 7, the value of flexibility becomes extremely high for low levels of permanent capacity, because of elevated backorder costs in case of IC. Next, we test the impact of change in demand variance. But since this is not possible with Poisson distribution, we assume Gamma distribution for demand. We alter the coefficient of variation (CV) values between 0.5 and 1.5. We present the resulting relation also in Figure 7. The results are rather surprising, in the sense that the value of flexibility does not always increase as variance increases. The reason for this is as follows. While both of the ETC_{IC} and ETC_{FC} terms do increase as demand variance increases, the increase in ETC_{IC} term is less than that in ETC_{FC} under low capacity. Because, most of the demand is backlogged anyway for the IC case, whereas the expected total cost increases more significantly for the flexible system. However, when the permanent capacity is sufficient to meet the demand on the average, the value of flexibility increases as demand variance increases. This value persists to be very significant under high demand variance, even when the permanent capacity is

Figure 8: % Value of Flexible Capacity vs Set-up Costs



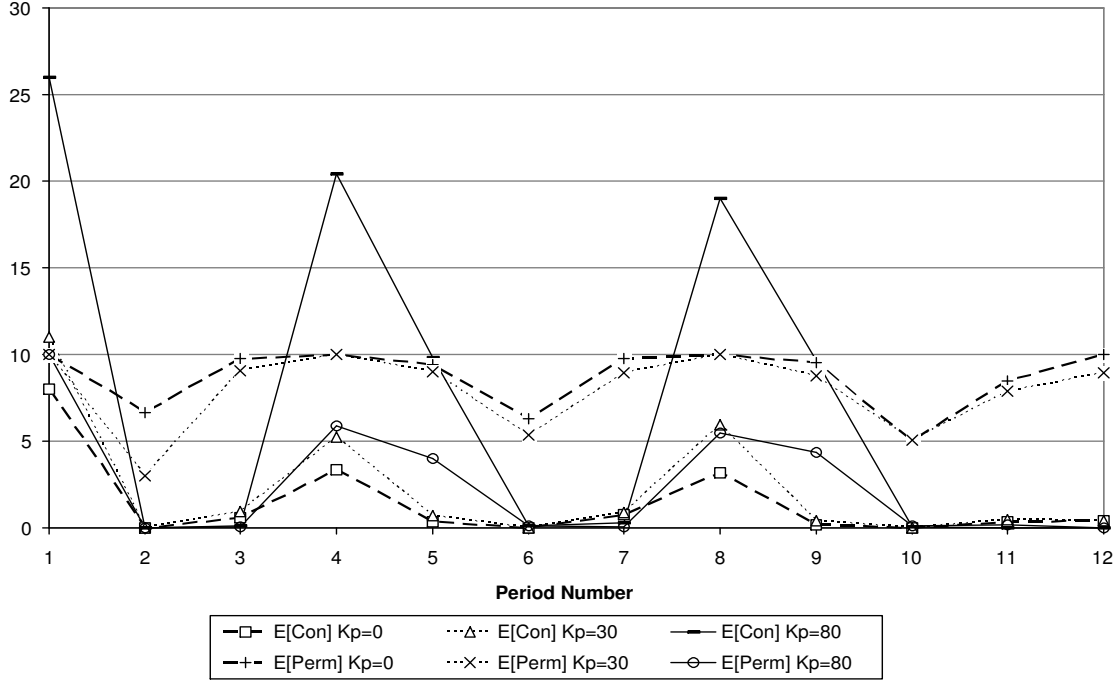
much higher than average demand. For example, when $CV = 1.5$, $\%VFC$ equals 15.3% for $U = 20$, which is twice the size of average demand. These results hold for the VFC as well as the $\%VFC$.

Finally, we investigate the impact of the set-up costs. Again, we assume the original experimental setting and parameter values with Poisson demand. We vary the values of set up costs for the permanent capacity (K_p) between 0 and 60, and that for the contingent capacity (K_c) between 0 and 40.

The resulting $\%VFC$ values are presented in Figure 8. As discussed before, $\%VFC$ decreases as K_c increases for any given level of K_p . However, the impact of a change in K_p is less straightforward to express, since the change in $\%VFC$ as K_p increases, for any given level of K_c , is not monotonic: it first decreases and then increases under the setting that we consider. Nevertheless, the absolute value of flexible capacity, VFC , does increase as K_p increases, because flexibility becomes more crucial as initiating production becomes more costly.

The reason why the relation is not monotonic in the $\%VFC$ case is as follows. For relatively small values of K_p , the rate of increase in ETC_{IC} as K_p increases is less than the

Figure 9: Expected Production Under Varying Production Set-up Costs



rate of increase in ETC_{FC} . But for relatively large values of K_p , the rate of increase in ETC_{IC} becomes the dominant one, because there is no way of avoiding high set-up costs in the inflexible system, except for backordering.

We illustrate these points by the expected production quantities under three different values of K_p , namely 0, 30, and 80, where $K_c = 20$. Expected production quantities are presented in Figure 9. Note that the structure of the solution shows a similarity among the cases where K_p is 0 and 30: There is a positive expectation of production by permanent workers in every period, which is higher when $K_p = 0$, and the usage of contingent workers is relatively limited, which is higher when $K_p = 30$. Nevertheless, the structure of the solution changes dramatically when $K_p = 80$, in the sense that lot sizes become very large to make benefit of economies of scale, once it is decided to initiate production, and hence the majority of production is conducted by temporary workers. Therefore, each period with a positive production is followed by a number of periods with no production and the system holds much higher inventories when $K_p = 80$. Hence, for larger values of K_p , frequent use of contingent resources make it possible to control the increase in expected costs (ETC_{IC}), whereas in an inflexible system such an opportunity does not exist. Consequently, the spread of the total

production shifts from permanent to contingent resources as K_p increases. In this problem instance, the percentage of total expected production performed by contingent resources are 14.05%, 21.37%, and 73.94% for K_p being 0, 30, and 80, respectively. This also indicates that it is beneficial for the companies to outsource a larger portion of their production activities as the setup cost of production increases.

6. Conclusions and Future Research

In this paper the integrated problem of inventory and flexible capacity management under non-stationary stochastic demand with the possibility of positive set-up costs, both for initiating production and ordering contingent capacity is considered. While a workforce capacity jargon is adopted in some parts of the paper, the model can be applied to any other bottleneck resource that defines the capacity as well, given that there exists the option of increasing this capacity by making use of contingent resources.

For the environment that we consider, the equivalence of this problem with the classical production/inventory problem under piecewise linear production cost is shown. When the set-up costs are negligible, the optimal policy depends on the level of starting inventory and it is a variant of base-stock policy, in the sense that there are two order-up-to levels in each period: One of them can be attained by utilizing contingent capacity, and the other can be attained by utilizing only permanent capacity. There is also a region of starting inventory level where full permanent capacity and no contingent capacity should be utilized.

When the set-up costs are positive, the optimal policy for the single-period problem is shown to be a variant of (s, S) type policies where the policy parameters are functions of the starting inventory level. With the help of a problem instance, it is shown that this policy does not hold for the multi-period problem. This does not mean that the solution of the multi-period problem cannot be characterized, so it is yet an open question. Nevertheless, even a famous special case of this problem, namely the capacitated multi-period inventory/production problem with positive set-up cost, is yet an open question, which makes it likely that the optimal policy does not have a very simple structure.

Our computational analyses point out important directions for the use of flexibility. First of all, there exists relative values of problem parameters where flexibility is very important: (i) lower costs of contingent capacity, (ii) higher set-up costs of production, (iii) lower levels of permanent capacity, and (iv) higher costs of backorders. Moreover, for businesses with

demand volatility, the value of flexibility is extremely high even under abundant permanent capacity levels. Under such circumstances, businesses should pursue establishing long-term contractual relations with third-party contingent capacity providers (such as external labor supply agencies). Such long-term agreements would bring significant operational cost savings. On the other hand, the opposite range of parameters yield relatively low value for flexibility, which is also very important from a managerial point of view. Under such circumstances, there does not exist enough motivation to invest in capacity flexibility, since the existing resources are sufficient for reasonable management of operations.

This research can be extended in several ways. Relaxing our assumptions on the capacity usage, such as introducing an upper limit on contingent capacity, or introducing uncertainty on contingent and/or permanent capacity would enrich the model, as well as relaxing our assumption of zero lead times. Interactions with material availability; determining the optimal permanent capacity level at the beginning of the problem horizon; incorporating intentional changes in permanent capacity in strategic, tactical, or even operational level; exploring the structure of the optimal solution under seasonal demand; developing an efficient heuristic for the multi-period problem with set-up costs are among some other extension possibilities.

Appendix

Proof of Theorem 1. First, we need the following preliminary results for the proof:

Lemma 1 $y_T^c \leq y_T^p$.

Proof : $y_T^c \leq y_T^p$ since $\frac{b-c_c}{h+b} \leq \frac{b}{h+b}$ and $G_T(y)$ is a non-decreasing function. \square

Lemma 2 $J_T(y|x)$ is convex in y for any value of x . Moreover, y_T^* is the minimizer of $J_T(y|x)$ where

$$y_T^* = \begin{cases} y_T^c & \text{if } x \leq y_T^c - U \\ x + U & \text{if } y_T^c - U \leq x \leq y_T^p - U \\ y_T^p & \text{if } y_T^p - U \leq x \end{cases} .$$

Proof : First note that $J_T^p(y) = J_T^c(y)$ at $y = x + U$. Moreover, $\frac{dJ_T^p(y)}{dy} = L_T'(y) \leq L_T'(y) + c_c = \frac{dJ_T^c(y)}{dy}$ for any y . For $y \leq x + U$, $J_T(y|x)$ is convex because $J_T^p(y)$ is convex. At $y = x + U$, $J_T(y|x)$ takes the form of another convex function $J_T^c(y)$ and since the derivative of the new function is greater than that of the previous, the convexity is not violated after the change of shape.

By using the definition and the convexity of $J_T(y|x)$ and Lemma 1, it can be observed that the function takes different forms in intervals $x \leq y_T^c - U$, $y_T^c - U \leq x \leq y_T^p - U$, and $y_T^p - U \leq x$ and the minimizer of the function in each of these intervals are y_T^c , $x + U$, and y_T^p respectively. \square

Recall that $f_T(x) = Uc_p + \min_{x \leq y} \{J_T(y|x)\}$. Uc_p is a constant term, hence $f_T(x)$ is determined by only minimizing $J_T(y|x)$ over $x \leq y$. Therefore, the optimal policy for the last period follows from the evaluation of this minimum over intervals $x \leq y_T^c - U$, $y_T^c - U \leq x \leq y_T^p - U$, $y_T^p - U \leq x \leq y_T^p$, and $y_T^p \leq x$ one by one and by the use of Lemma 2.

In order this policy to hold for all other periods in the planning horizon, similar results as in Lemma 1 and 2 must hold for each of these periods: (i) $J_t(y|x)$ is convex over y for a given x , and (ii) $y_t^c \leq y_t^p$ for every period $t \leq T$.

By using the optimal policy of period T , it can be shown that

$$f_T(x) = \begin{cases} J_T^c(y_T^c) = c_c(y_T^c - x - U) + L_T(y_T^c) & x \leq y_T^c - U \\ J_T^p(x + U) = L_T(x + U) & y_T^c - U \leq x \leq y_T^p - U \\ J_T^p(y_T^p) = L_T(y_T^p) & y_T^p - U \leq x \leq y_T^p \\ J_T^p(x) = L_T(x) & y_T^p \leq x \end{cases}$$

and

$$\frac{df_T(x)}{dx} = \begin{cases} -c_c & x \leq y_T^c - U \\ L_T'(x + U) & y_T^c - U \leq x \leq y_T^p - U \\ 0 & y_T^p - U \leq x \leq y_T^p \\ L_T'(x) & y_T^p \leq x \end{cases}.$$

At $x = y_T^c - U$, we have $L_T'(x + U) = L_T'(y_T^c) = (h + b)G_T(y_T^c) - b = -c_c$, because $y_T^c = G_T^{-1}\left(\frac{b - c_c}{h + b}\right)$. But since L is convex and is minimized at y_T^p , we conclude that $df_T(x)/dx$ is continuous and non-decreasing, and consequently the functions $f_T(x)$ and $E[f_T(x)]$ are convex in x . Starting with period T and assuming that the optimal policy and the convexity of $E[f_k(x)]$ hold for all periods $k = T, T - 1, \dots, t + 1$, it follows that $J_t(y|x)$ is convex over y for a given x . Let $j(y) = dJ_t^p(y)/dy$ then $y_t^p = j^{-1}(0)$. Moreover, $y_t^c = j^{-1}(-c_c)$ since $dJ_t^c(y)/dy = j(y) + c_c$. Hence, due to the convexity of function J_t^p , $y_t^c \leq y_t^p$ and an optimal policy similar to the one stated in Lemma 2 directly follows. To conclude the proof, note that

$$f_t(x) = \begin{cases} J_t^c(y_t^c) & x_t \leq y_t^c - U \\ J_t^p(x_t + U) & y_t^c - U \leq x_t \leq y_t^p - U \\ J_t^p(y_t^p) & y_t^p - U \leq x_t \leq y_t^p \\ J_t^p(x_t) & y_t^p \leq x_t \end{cases}, \quad \frac{df_t(x)}{dx} = \begin{cases} -c_c & x_t \leq y_t^c - U \\ \frac{dJ_t^p(x_t + U)}{dx} & y_t^c - U \leq x_t \leq y_t^p - U \\ 0 & y_t^p - U \leq x_t \leq y_t^p \\ \frac{dJ_t^p(x_t)}{dx_t} & y_t^p \leq x_t \end{cases}$$

At $x_t = y_t^c - U$, $j(x_t + U) = j(y_t^c) = j(j^{-1}(-c_c)) = -c_c$. Hence, $f_t(x_t)$ is a convex function over x_t . This completes the proof. \square

Proof of Theorem 2:

U_{c_p} is a constant term. Note that $s^u(x) \leq x^u$ by definition of $s^u(x)$. We need to examine the following cases on the value of x .

Case I: $x^u \leq y_T^c$ and $s^c(x) \leq s^u(x) \leq x^u$

$J_T(y|x)$ is minimized at y_T^c by Lemma 2 and therefore is non-increasing over $x \leq y \leq y_T^c$. Since $s^c(x) \leq s^u(x)$, $J_T^c(y_T^c|x) + K_p + K_c = J_T(s^c(x)|x) \geq J_T(s^u(x)|x) = J_T^p(x^u) + K_p$. Hence, if production is viable then it must be limited to x^u , ordering contingent capacity does not pay off. If $x \leq s^u(x) = s(x)$, $J_T(x|x) \geq J_T(s^u(x)|x) = J_T^p(x^u|x) + K_p$ by the definition of $s^u(x)$. Therefore, $y_T^*(x) = x^u = S(x)$ since $s(x) = s^u(x)$. If $x \geq s^u(x) = s(x)$, $J_T(s^u(x)|x) \geq J_T(x|x)$ and therefore $y_T^*(x) = x$.

Case II: $x^u \leq y_T^c$ and $s^u(x) \leq s^c(x)$

$J_T(y|x)$ is minimized at y_T^c by Lemma 2 and therefore is non-increasing over $x \leq y \leq y_T^c$. As $s^u(x) \leq s^c(x)$, $J_T^p(x^u) + K_p = J_T(s^u(x)|x) \geq J_T(s^c(x)|x) = J_T^c(y_T^c|x) + K_p + K_c$. Hence, if production is viable then it must be up to y_T^c . If $x \leq s^c(x) = s(x)$, $J_T(x|x) \geq J_T(s^c(x)|x) = J_T^c(y_T^c|x) + K_p$ by the definition of $s^c(x)$. Therefore, $y_T^*(x) = y_T^c = S(x)$ since $s(x) = s^c(x)$. If $x \geq s^c(x) = s(x)$, $J_T(s^c(x)|x) \geq J_T(x|x)$ and therefore $y_T^*(x) = x$.

Case III: $y_T^c \leq x^u \leq y_T^p$

$J_T(y|x)$ is minimized at x^u by Lemma 2 and therefore is non-increasing over $x \leq y \leq x^u$. If $x \leq s^u(x) = s(x) \leq x^u$, $J_T(x|x) \geq J_T(s^u(x)|x) = J_T^p(x^u) + K_p$ by the definition of $s^u(x)$. Therefore, $y_T^*(x) = x^u = S(x)$ since $s(x) = s^u(x)$. If $x \geq s^u(x) = s(x)$ then $y_T^*(x) = x$.

Case IV: $y_T^p \leq x^u$

$J_T(y|x)$ is minimized at y_T^p by Lemma 2 and therefore is non-increasing over $x \leq y \leq y_T^p$. If $x \leq s^p(x) = s(x) \leq y_T^p$, then $J_T(x|x) \geq J_T(s^p(x)|x) = J_T^p(y_T^p) + K_p$ by the definition of $s^p(x)$. Therefore, $y_T^*(x) = y_T^p = S(x)$ since $s(x) = s^p(x)$. If $x \geq s^p(x) = s(x)$ then $y_T^*(x) = x$. \square

Proof of Theorem 3:

Case I: $s^u(x) \leq s^c(x)$, $x^u \leq y_T^c$

$s(x) = s^c(x)$. For $x \leq x + \Delta \leq x^u$:

$$c_c(y_T^c - x - U) + L(y_T^c) \geq c_c(y_T^c - x - \Delta - U) + L(y_T^c)$$

$$\begin{aligned}
J_T(y_T^c|x) + K_p + K_c &\geq J_T(y_T^c|x + \Delta) + K_p + K_c \\
J_T(s^c(x)|x) &\geq J_T(s^c(x + \Delta)|x + \Delta) \\
s^c(x) &\leq s^c(x + \Delta)
\end{aligned}$$

Case II: $s^c(x) \leq s^u(x)$, $x^u \leq y_T^c$

$s(x) = s^u(x)$. Since J_T^p is convex and minimized at $y_T^p \geq x + \Delta + U$, we can write for $x \leq x + \Delta \leq x^u$:

$$\begin{aligned}
J_T^p(x^u) + K_p &\geq J_T^p(x^u + \Delta) + K_p \\
J_T(s^u(x)|x) &\geq J_T(s^u(x + \Delta)|x + \Delta) \\
s^u(x) &\leq s^u(x + \Delta)
\end{aligned}$$

Case III: $y_T^c \leq x^u \leq y_T^p$

$s(x) = s^u(x)$. This case can be proved similar to Case II.

Case IV: $y_T^p \leq x^u$

$s(x) = s^p(x)$. For $x \leq x + \Delta \leq x^u$, $J_T(s^p(x)|x) = J_T(s^p(x + \Delta)|x + \Delta) = J_T^p(y_T^p) + K_p$. Therefore $s^p(x) = s^p(x + \Delta)$. \square

References

- Angelus, A., and E.L. Porteus. 2002. Simultaneous Capacity and Production Management of Short-Life-Cycle, Produce-to-Stock Goods Under Stochastic Demand. *Management Science*. **48**(3) 399-413.
- Atamtürk, A., and D.S. Hochbaum. 2001. Capacity Acquisition, Subcontracting, and Lot Sizing. *Management Science*. **47**(8) 1081-1100.
- Dellaert, N., and T. De Kok. 2004. Integrating Resource and Production Decisions in a Simple Multi-Stage Assembly System. *International Journal of Production Economics*. **90** 281-294.
- Federgruen, A., and P. Zipkin. 1986. An Inventory Model with Limited Production Capacity and Uncertain Demands I. The Average Cost Criterion. *Mathematics of Operations Research*. **11**(2) 193-207.
- Gallego, G., and A. Scheller-Wolf. 2000. Capacitated inventory problems with fixed order costs: Some optimal policy structure. *European Journal of Operational Research*. **126**(3) 603-613.

- Holt, C.C., F. Modigliani, J. F. Muth, and H. A. Simon. 1960. Planning Production, Inventories and Work Force. Prentice-Hall, Englewood Cliffs, NJ.
- Hu, J.Q., P. Vakili, and L. Huang. 2004. Capacity and Production Management in a Single Product Manufacturing System. *Annals of Operations Research*. **125** 191-204.
- Kapuscinski, R., and S. Tayur. 1998. A Capacitated Production-Inventory Model with Periodic Demand. *Operations Research*. **46** 899-911.
- Karlin, S. 1958. Optimal Inventory Policies for the Arrow-Harris-Marschak Dynamic Model. Arrow, K., S. Karlin, and H. Scarf, eds. *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, California, Chapter 9, 135–154.
- Milner, J.M., and E.J. Pinker. 2001. Contingent Labor Contracting Under Demand and Supply Uncertainty. *Management Science*. **47**(8) 1046-1062.
- Pinker, E.J., and R.C. Larson. 2003. Optimizing the Use of Contingent Labor When Demand is Uncertain. *European Journal of Operational Research*. **144** 39-55.
- Porteus, E.L. 1990. Stochastic Inventory Theory. Heyman, D.P., Sobel, M.J., eds. *Handbooks in OR & MS, Vol.2*. Elsevier, North-Holland, Chapter 12, 605–652.
- Tan, B., and S.B. Gershwin. 2004. Production and Subcontracting Strategies for Manufacturers with Limited Capacity and Volatile Demand. *Annals of Operations Research*. **125** 205-232.
- van Mieghem, J.A. 2003. Capacity Management, Investment, and Hedging: Review and Recent Developments. *Manufacturing & Service Operations Management*. **5**(4) 269-302.
- Wild, B., and C. Schneeweiss. 1993. Manpower capacity planning - A hierarchical approach. *International Journal of Production Economics*. **30-31** 95-106.