SCHEDULING A MAKE-TO-STOCK QUEUE: INDEX POLICIES AND HEDGING POINTS

Michael H. Veatch
Department of Mathematics, Gordon College

Lawrence M. Wein
Sloan School of Management, M.I.T.

Abstract

A single machine produces several different classes of items in a make-to-stock mode. We consider the problem of scheduling the machine to regulate finished goods inventory, minimizing holding and backorder or holding and lost sales costs. Demands are Poisson, service times are exponentially distributed, and there are no delays or costs associated with switching products. A scheduling policy dictates whether the machine is idle or busy, and specifies the job class to serve in the latter case. Since the optimal solution can only be numerically computed for problems with several products, our goal is to develop effective policies that are computationally tractable for a large number of products. We develop index policies to decide which class to produce, including Whittle's "restless bandit" index, which possesses a certain asymptotic optimality. Several idleness policies are derived, and the best policy is obtained from a heavy traffic diffusion approximation. Nine sample problems are considered in a numerical study, and the average suboptimality of the best policy is less than 3%.

January 1994

In a make-to-stock production facility with multiple products, one of the goals of the scheduling policy is to regulate finished goods inventory. Too small of an inventory risks incurring backorder or lost sales costs, while too large of an inventory increases holding costs. The target inventory level, called the base or safety stock, is vitally linked to randomness in the system and capacity constraints that limit the ability to respond to unexpected demand. Accordingly, a realistic model of the make-to-stock system should include queueing effects. The queueing framework combines the dynamic stochastic nature of the scheduling problem, often studied in inventory systems, with the capacity constraint, usually dealt with through deterministic production scheduling.

Although make-to-order environments, where production occurs after customer orders are received, have been analyzed extensively as queueing control problems (see, for example, Klimov 1974), little work has been done on the problem of scheduling a multiclass make-to-stock queue. Wein (1992) develops a scheduling policy for the make-to-stock system based on a heavy traffic approximation that results in a Brownian motion control problem. Zheng and Zipkin (1990) and Zipkin (1992) and Zipkin (1992) propose and analyze a simple "longest queue" policy that is optimal for a system with identical product classes operating under independent base stock policies. Ha (1993) partially characterizes the optimal policy for the two-product case. While promising computational results have been reported for the Brownian and longest queue policies, no performance guarantees are available and both have obvious deficiencies in the structure of the control.

This paper develops and tests several new scheduling policies. The system considered is a multiclass $M/M/1$ make-to-stock queue. Preemptive resume scheduling is allowed and there are no set-up costs or times when switching classes. In the backorder version of the problem, which was addressed in the references cited above, the objective is to minimize holding and backorder costs. A lost sales problem is also considered, where demands that cannot be met from inventory are lost and a cost incurred. A long-run average cost criterion is used. Almost all of the policies reduce the system in some manner to a more tractable single-product subproblem.

Three *index policies* are considered, where an index is computed for each class as a function of the inventory in that class. The class with the smallest index is produced; if all indices are positive, the machine is idle. Index policies have the advantage of being computationally tractable even for a large number of classes. An index policy proposed by Paul Zipkin in a personal correspondence performed best among those considered. The most innovative, and one that also performs well, is a "restless bandit" index, defined for a general problem in Whittle (1988). This index has the property that it is asymptotically optimal as the number of classes goes to infinity (and the utilization of each class goes to zero). We also discuss the Gittens index for this problem, neglecting its "restlessness," and point out a connection between the Gittens and restless bandit indices. A third index is developed by computing the value function for the system under an allocation policy that allows each class to use a fraction of the service capability.

We will show that index policies perform well at determining which class to produce, but do poorly at deciding when to idle. The problem is that each index is computed without knowledge of the other classes, and hence, without knowing the system utilization. Several other approximations are proposed for when to idle. One method decomposes the system into single-class subproblems with the same utilization as the original system. Another aggregates the system into a single product class. The most elegant, and the most accurate, idling decision is derived using a heavy traffic diffusion approximation. We analyze the approximating Brownian control problem for the lost sales case, complementing the backorder case treated in Wein. A fourth idleness policy is derived by computing the inventory distribution assuming that the longest queue policy is used, and then decomposing into single-class subproblems.

Numerical results are presented that compare all of the proposed policies with optimal policies for two- and three-product problems. Combinations of index and idleness policies are found that perform well; the average suboptimality for our best policy is 2% for 6 lost sales problems and 4% for 3 backorder problems. The structure of optimal policies is also investigated.

The rest of the paper is organized as follows. Section 1 formulates the problem mathematically and discusses the structure of scheduling policies. After this some readers may wish to go to Section 5 on numerical results. Section 2 solves the single-product version of the problem. Index policies are derived in Section 3 and hedging points (idleness policies) in Section 4. Some concluding remarks are made in Section 6.

1 Dynamic Scheduling Problem

Consider a multiclass, make-to-stock M/M/1 queueing system: a machine can produce K different classes of items; each finished item is placed in its respective inventory, $X_k(t)$ for class k, $k = 1, \ldots, K$ at time t; this inventory services an exogenous demand. In the backorder version of the problem, demand that cannot be met from inventory is backordered and recorded as negative inventory. In the lost sales problem, class k demands that occur when $X_k(t) = 0$ are ignored and a cost incurred. We have several reasons for considering the lost sales problem. It has received less attention than the backorder problem in the literature, is at least as appropriate in many applications, provides an interesting example of Brownian motion analysis, and is the only version to which the restless bandit index of Section 2.3 applies. This index is unique in possessing some form of asymptotic optimality.

The demands for each class are independent Poisson processes with rates λ_k and the production times for class k items are independent and exponentially distributed with mean $m_k = 1/\mu_k$. In Section 3.3, we will briefly consider general inter-demand and production time distributions. For the backorder problem, stability of the system requires that $\rho < 1$, where $\rho = \sum \rho_k$ and $\rho_k = \lambda_k/\mu_k$ (all indices range over $1, \ldots, K$ unless otherwise noted).

The scheduling decision is whether to produce product $1, \ldots, K$ or to idle at each time t. An admissible scheduling policy π is a function $\zeta(X, t)$ that takes on the values $0, 1, \ldots, K$ (zero denoting idle) and is nonanticipating with respect to X. Let Π denote the class of admissible policies. Production of an item can be interrupted and resumed; no set-up costs or times are incurred when switching from one class to another. Because the system is memoryless, a Markov policy, depending only on the current state $X(t) = (X_1(t), \ldots, X_K(t))$, will be optimal. Under these assumptions, multiple machines or other models that allow partial production effort would not change the "all or nothing " form of optimal policies.

The objective is to minimize holding and backorder costs, incurred at the rate $c^{BO}(x) = \sum(h_k x_k^+ + b_k x_k^-)$ in state x, for the backorder problem, or holding and lost sales costs, $c^{LS}(x) = \sum(h_k x_k^+ + s_k 1_{\{x_k=0\}})$, for the lost sales problem. Note that s_k is the cost rate for a stockout, corresponding to an expected cost per lost sale of $l_k = s_k/\lambda_k$. We assume that $h_k > 0$ and $b_k(\text{or} s_k) > 0$ for all k. The infinite-horizon cost, discounted at the rate $\alpha > 0$, is $V^\pi(x) = E_x \int_0^\infty e^{-\alpha t} c(X(t))dt$. Here E_x denotes expectation given the initial state $X(0) = x$ and policy π. We will uniformize the process as in Lippman (1975). Let $\overline{\mu} = \max_k\{\mu_k\}$ and $\Lambda = \sum \lambda_k + \overline{\mu} + \alpha$. The optimal cost function, $V(x) = \min_{\pi \in \Pi} V^\pi(x)$, satisfies the dynamic programming optimality equations

$$V(x) = \mathbf{T}V(x), \tag{1}$$

$$\mathbf{T}V(x) = \frac{1}{\Lambda}\left[c(x) + \sum \lambda_k V(x - e_k) + \overline{\mu}V(x) + \min\{0, \min_k\{\mu_k \Delta_k V(x)\}\}\right], \tag{2}$$

where $\Delta_k V(x) = V(x + e_k) - V(x)$ and e_k is the unit vector with kth component equal to one. For the lost sales problem, replace $x - e_k$ with x in (2) when $x_k = 0$.

In this formulation the summation represents demands, with a class k demand causing the transition $x \to x - e_k$; idleness, represented by the $\overline{\mu}$ term and choosing zero in the min, keeps the system in state x; and production of class k, corresponding to the Δ_k term, causes the transition $x \to x + e_k$. Putting the equations in this form illustrates the relationship between the optimal policy and $V(x)$: the class with minimal $\mu_k \Delta_k V(x)$ is produced unless they are all positive, in which case the machine idles.

It is more convenient to deal with an undiscounted, long-run average cost criterion. In this case $\alpha = 0$; the optimal average cost rate (gain) g and relative value function $V(x)$ satisfy

$$V(x) + g/\Lambda = \mathbf{T}V(x), \tag{3}$$

where we arbitrarily set $V(0) = 0$.

Optimal policies for this problem can only be found numerically, and only when the number of classes is small. Hence, we are led to consider heuristic methods. All of these heuristics lead to *monotone* policies. For a given policy, let $\mathcal{B}_k$ be the set of states in which class k is produced and $\mathcal{I}$ the set in which the machine is idle. It will prove more convenient to deal with policies which are extended by specifying a class preference for states $x \in \mathcal{I}$. A natural extension is to choose the class with minimal $\mu_k \Delta_k V(x)$ in state x, where V is the value function for this policy (or a more easily computed proxy function). Let $\overline{\mathcal{B}}_k$ be the set of states in which class k is preferred. Note that $\mathcal{B}_k \subseteq \overline{\mathcal{B}}_k$ and $\{\overline{\mathcal{B}}_k\}$ partition $\mathbf{Z}^k$. The policy is *monotone* if it satisfies

1. Monotone switching: $\overline{\mathcal{B}}_k = \{x : x_k < s(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_K)\}$ for some increasing switching surface s, $-\infty \le s \le \infty$, and

2. Monotone idling: $\mathcal{I} = \{x : x_1 < s(x_2, \ldots, x_K)\}$ for some decreasing idling surface s, $0 \le s \le \infty$. (The idling surface could be written as a function of any $K - 1$ of the state variables.)

See Veatch and Wein (1992) for a discussion of monotonicity. Ha (1992) proves that the optimal policy is monotone for the case of two products with $\mu_1 = \mu_2$ and backordering. Based on numerical results, we suspect that optimal policies are monotone in all cases; however, standard techniques from these papers and some of the references therein have not yielded a general proof.

Figs. 1 and 2 illustrate monotonicity in two dimensions. When there are just two classes we see that the switching surfaces become a switching curve between producing class 1 and class 2. We have introduced $\overline{\mathcal{B}}_k$ so that this curve extends indefinitely in the positive direction rather than ending when it enters the idleness region. In a continuous space, the notion of a switching curve generalizes to K dimensions as the boundary that all of the $\overline{\mathcal{B}}_k$ have in common. In our discrete space, define the switching curve as $\{x : x - e_k \in \overline{\mathcal{B}}_k$ or, for the lost sales problem, $x_k = 0$ for all $k\}$. These points can be put into a sequence $\{x^n\}$ as follows. Choose any point on the curve and call it x^0. The forward iteration is $x^{n+1} = x^n + e_k$ for $x^n \in \overline{\mathcal{B}}_k$; the backward iteration is $x^{n-1} = x^n - e_i$, where i is the coordinate for which $x^n - e_i$ is on the switching curve. For the lost sales problem, the backward iteration terminates at the origin. This i must be unique, for if $i \neq j$, then either $x^n - e_i - e_j \in \overline{\mathcal{B}}_i$ so that $x^n - e_i$ is not on the switching curve or $x^n - e_i - e_j \in \overline{\mathcal{B}}_j$ so that $x^n - e_j$ is not on the switching curve. Monotonicity implies that $\{x^n\}$ includes all points on the switching curve.

A very useful implication of monotonicity is that the idleness policy can be described by a single point x^*, usually called the *hedging point* (see, for example, Kimemia and Gershwin 1983). The hedging quantity x^*_k is a base stock or "order up to" level that is never exceded. Assuming that $\mathcal{I}$ is nonempty, x^* is finite; in this case it is the only recurrent state in $\mathcal{I}$ and the recurrent class is $\{x : x \leq x^*\}$. On its recurrent states, a policy is fully defined by the switching surfaces and hedging point. The surfaces, or the sets $\overline{\mathcal{B}}_k$, define the preference among classes. The hedging point lies on the switching curve and defines the idleness policy. Characterizing the idleness policy by a single point that can be found by a one-dimensional search along the switching curve will allow us to construct heuristic policies.

Several classes of policies have been considered in other papers. Zheng and Zipkin view $x^* - X$ as a multiclass queue containing demands that have not yet been restocked. They consider a (non-Markov) FCFS policy and a longest queue first (LQ) policy for serving this demand queue. For two identical products they show that the LQ policy, which is symmetric, is optimal and marginally better that FCFS. For two non-identical products, we call a switching curve that lies along $x_1 = x_2$ for $x_k \le \min\{x_1^*, x_2^*\}$, then extends vertically or horizontally to the hedging point, an LQ switching curve. An *offset* LQ switching curve lies along $x_1 = x_2 + (x_1^* - x_2^*)$. However, the offset LQ switching curves tested in Section 5.2 are modified for consistancy with the following property, established by Ha in the two-product case. If the class k with maximal $b_k\mu_k$ is backordered,then it is optimal to produce this class. In other words, the switching curve cannot have $x_k < 0$ for this class.

Wein proposes a "$b\mu$ $h\mu$" rule for the backorder problem that is reminiscent of the $c\mu$ rule: if there are classes in danger of being backordered (i.e., $X_k(t) < \epsilon_k$, where ϵ_k is a parameter), produce the class within this set with the largest $b_k\mu_k$; otherwise, produce the class with the smallest $h_k\mu_k$. For $\epsilon_k = 0$, the switching curve follows one of the negative coordinate axes and one of the positive axes. The idling policy is obtained as the solution to a Brownian control problem.

The primary goal of this paper is to develop more sophisticated, yet easily computable, switching curves/surfaces and additional idleness policies that perform better than the policies described above.

2 Single-Product Problem

In this section, the dynamic scheduling problem is solved for the case of only one product. Several of our heuristic policies make use of these results. The index denoting product class will be suppressed when convenient. In later sections tildes will be added to denote the single-product subproblem $(\tilde{\mu}, \tilde{\lambda}, \tilde{\rho})$ when needed to distinguish them from the original problem. The optimality equation for the backorder problem becomes

$$V(x) + g/\Lambda = \frac{1}{\Lambda}\left[c(x) + \lambda V(x-1) + \mu V(x) + \min\{0, \mu\Delta V(x)\}\right], \tag{4}$$

where $\Lambda = \lambda + \mu$ and $\Delta V(x) = V(x+1) - V(x)$. The optimal policy will be of threshold form for some base stock level $B \geq 0$: the machine is busy in states $x < B$ and idle in states $x \geq B$. Under this policy, denoted $\pi(B)$, $B - X$ is an $M/M/1$ queue. Letting $f(\cdot)$ and $F(\cdot)$ be the steady-state p.d.f. and d.f. of $B - X$, the corresponding gain is

$$\begin{aligned} g^{\pi(B)} &= \sum_{x=-\infty}^{B} c(x)\Pr\{X = x\} && (4) \\ &= h\sum_{x=0}^{B-1}(B-x)f(x) + b\sum_{x=B+1}^{\infty}(x-B)f(x). && (5) \end{aligned}$$

Differencing,

$$\begin{aligned} g^{\pi(B+1)} - g^{\pi(B)} &= h\sum_{x=0}^{B} f(x) - b\sum_{x=B+1}^{\infty} f(x) && (5) \\ &= (h+b)F(B) - b. && (6) \end{aligned}$$

The d.f. F is nondecreasing, so $g^{\pi(B)}$ is convex in B. It attains its minimum at the smallest B for which (6) is positive, namely $B = \min\{x : F(x) > b/(h+b)\}$. Since

$F(x) = 1 - \rho^{x+1}$, it follows that

$$B = \lfloor \ln[h/(h+b)]/\ln\rho \rfloor. \quad (7)$$

As discussed in Section 1, the decision of which class to produce depends on the relative values of $\mu_k \Delta_k V(x)$. We will use $\mu \Delta V(x)$ from certain single-product subproblems as a proxy for $\mu_k \Delta_k V(x)$. To find ΔV, first compute the gain using $f(x) = (1-\rho)\rho^x$ in (5), giving

$$g = \frac{hB - h(B+1)\rho + (h+b)\rho^{B+1}}{1-\rho}. \quad (8)$$

Since the optimal policy is $\pi(B)$, the steady-state equations obtained from (4) are

$$V(x) + g/\Lambda = \frac{1}{\Lambda}[c(x) + \lambda V(x-1) + \mu V(x+1)], \quad x < B, \quad \text{and} \quad (9)$$

$$V(x) + g/\Lambda = \frac{1}{\Lambda}[c(x) + \lambda V(x-1) + \mu V(x)], \quad x \geq B, \quad (10)$$

or in terms of differences

$$\Delta V(x) = [(x+1)h - g]/\lambda, \quad x \geq B-1, \quad \text{and} \quad (11)$$

$$\Delta V(x) = \frac{c(x+1) - g}{\lambda} + \frac{\Delta V(x+1)}{\rho}, \quad x < B-1. \quad (12)$$

A simple recursive computation gives $\mu \Delta V(x)$. Optimality in (4) implies that $\Delta V(x) \leq 0$ for $x < B$ and $\Delta V(x) \geq 0$ for $x \geq B$.

For the lost sales problem, a search procedure is used to find the optimal base stock level B. For a given B, $B - X$ is a finite $M/M/1/B$ queue. Assuming that $\rho < 1$, its steady-state p.d.f. is $f_B(x) = \rho^{B-x} f_B(B)$ and $f_B(B) = (1-\rho)/(1-\rho^{B+1})$.

The gain is

$$
\begin{aligned}
g^{\pi(B)} &= s f_B(0) + h \sum_{x=1}^{B} x f_B(x) && (11) \\
&= s\rho^B f_B(B) + h f_B(B) \rho^B \sum_{x=1}^{B} x (1/\rho)^x && (12) \\
&= s\rho^B f_B(B) + h f_B(B) \left[\frac{B - (B+1)\rho + \rho^{B+1}}{(1-\rho)^2} \right]. && (13)
\end{aligned}
$$

It can be shown that $g^{\pi(B)}$ is convex in B. A search over the values $x = 0, 1, \ldots$ is used to find B. The recursion obtained for ΔV is

$$
\begin{aligned}
\Delta V(0) &= (g - s)/\mu \ \text{ and} && (14) \\
\Delta V(x) &= \frac{g - hx}{\mu} + \rho \Delta V(x-1), \ \ 0 < x < B, && (15)
\end{aligned}
$$

which has the solution

$$
\mu \Delta V(x) = -\frac{s\rho^x (1 - \rho^{B-x})}{1 - \rho^{B+1}} + \frac{h[B - x - (B+1)\rho^{x+1} + (x+1)\rho^{B+1}]}{(1 - \rho^{B+1})(1 - \rho)}. \qquad (16)
$$

If B is optimal we again have $\Delta V(x) \leq 0$ for $x < B$ and $\Delta V(x) \geq 0$ for $x \geq B$.

3 Index Policies

The class of index policies is attractive because its computational complexity only grows linearly in the number of classes. An index policy is defined by assigning an index $\nu_k(x_k)$ to class k and producing the class with smallest index. If the indices are nondecreasing, the policy has the monotonicity property of Section 1 and has a switching curve. An idleness policy must also be defined. We will call a policy *pure index* if the idleness region is $\mathcal{I} = \{x : \nu_k(x_k) \geq 0 \text{ for all } k\}$. This definition is motivated by the fact that our indices try to approximate $\mu_k \Delta_k V(x)$, which is the

Table 1: Proposed Indices and Idleness Policies.

	Backorders	Lost Sales
Index		
Value function approx. ($\mu\Delta V$)	$\surd$	$\surd$
Service time look-ahead (STLA)	$\surd$	$\surd$
Restless bandit	—	$\surd$
Idleness Policy		
Server allocation	hedging point	hedging point
Aggregate product	workload	workload
Longest queue (LQ) hedging point	hedging point	—
Brownian motion	workload	workload

rate at which expected cost changes when producing class k. If cost increases when each class is produced, then idleness is preferred. For policies that are not pure index, the idleness policy must be specified in addition to the index.

Several indices, listed in Table 1, are developed and tested in this paper. The most obvious approach is to approximate the optimal value function $V(x)$. In Section 2.1, $\Delta_k V(x)$ is approximated using the value function for separate single-product problems where service capacity is allocated across product classes. The service time look-ahead (STLA) index of Section 2.2, proposed by Zipkin (1990), replaces $V(x)$ with the expected cost rate after one service time. It is related to the fully myopic policy of using the cost rate after one transition, which produces the $b\mu$ $h\mu$ rule. Section 3.3 presents an index based on a Lagrangian approach to the multi-armed "restless" bandit problem in Whittle.

3.1 Value Function Approximation

An index called $\mu\Delta V$ is obtained using a value function approximation. Consider a single-product subproblem for class k with production rate $\tilde{\mu} = (\rho_k/\rho)\mu_k$ and demand rate $\tilde{\lambda} = \lambda_k$. The index is $\nu(x) = \tilde{\mu}\Delta\tilde{V}(x)$ (derived in Section 2). In this subproblem, the machine's availability in any time interval has been allocated across classes according to the fraction ρ_k/ρ of the system utilization ρ. Modifying μ_k changes the subproblem utilization from ρ_k, which tends to zero as the number of classes increases, to the system utilization ρ. Another motivation is that the allocation and the optimal subproblem policies can be viewed as a crude policy for the original system. The $\mu\Delta V$ index policy is equivalent to performing one value iteration starting with the value function of the crude policy, giving an improved policy.

3.2 Service Time Look-Ahead

Static priority policies such as the $h\mu$ rule are fully myopic in the sense that they minimize the cost rate $c(\cdot)$ of the next state. A service time look-ahead (STLA) policy, proposed by Paul Zipkin, considers the expected cost rate after one service time. It can be viewed as a myopic policy if we represent the server by a Poisson process that is always on, with a decision made at the end of each service time as to whether to load an item or run empty. This situation, which does not allow preemption, forces one to look ahead somewhat and is preferable to only considering the cost rate of the next state. This policy is reminiscent of the transportation time look-ahead policy of Miller (1974) for the decision of which base to send an item repaired at a central depot, where transportation time takes on the role of service time.

For a given class, again suppress the index k and let $g(x) = E[c(x - D(S))]$, where S is the service time and $D(t)$ is the number of demands in the interval $(0, t]$.

Then $g(x)$ is the expected cost rate after one service time if the server is running empty. If the server is actually producing, the cost rate is $g(x+1)$; hence, the rate at which serving this class increases the expected cost rate, assuming no preemption, is $\mu\Delta g(x)$. This quantity is the STLA index.

Zipkin evaluates the index for the backorder problem using the fact that $D(S)+1$ has a geometric distribution with parameter $p = \mu/(\lambda+\mu)$. Letting $q = 1-p$, $\Pr\{D(S) = j\} = q^j p$, $j \geq 0$ and for $x \geq 0$

$$\begin{aligned} g(x) &= h\sum_{j=1}^{x} jq^{x-j}p + b\sum_{j=1}^{\infty} jq^{x+j}p && (16)\\ &= h(xp - q + q^{x+1})/p + bq^{x+1}/p. && (17)\end{aligned}$$

Letting $\Delta g(x) = g(x+1) - g(x)$, the index is

$$\mu\Delta g(x) = -b\mu q^{x+1} + h\mu(1-q^{x+1}), \quad x \geq 0. \qquad (18)$$

Note that, as $x \to \infty$, (18) approaches $h\mu$, and for large inventories the STLA policy gives the $h\mu$ rule. For $x < 0$, $g(x) = E[b(-x + D(S))] = b(-x + q/p)$, and the index is

$$\mu\Delta g(x) = -b\mu, \quad x < 0, \qquad (19)$$

which yields the $b\mu$ rule. Turning to the lost sales problem,

$$\begin{aligned} g(x) &= s\sum_{j=x}^{\infty} q^j p + h\sum_{j=1}^{x} jq^{x-j}p && (19)\\ &= sq^x + h\left(\frac{xp - q + q^{x+1}}{p}\right). && (20)\end{aligned}$$

The index is

$$\mu\Delta g(x) = -s\mu pq^x + h\mu(1-q^{x+1}). \qquad (21)$$

3.3 Restless Bandit Analysis

In this section we use Whittle's work on "restless bandits" to obtain an index for the lost sales problem and analyze its properties. Whittle defines a *restless bandit* problem as a resource allocation problem similar to a multi-armed bandit except that the arms not being played, called passive, continue to change state according to a Markov law that is different than the law governing their transitions when active. Passive arms can also incur costs. In the scheduling problem there are $K+1$ arms, one for each class plus an idleness arm. There must be exactly one active arm at each time, where active means that the machine is assigned to that class. To define an index, introduce a passive tax, or cost of not producing, ν in the single-product subproblem with the same parameters as class k. Then (4) becomes

$$V(x) + g(\nu)/\Lambda = \frac{1}{\Lambda}\left[c(x) + \lambda V(x-1) + \mu V(x) + \min\{\nu, \mu\Delta V(x)\}\right]. \qquad (22)$$

The restless bandit index $\nu(x)$ is defined as the value of ν that achieves indifference in the min in (22) for a given value of x, say $x = B$. Indifference in (22) implies that the threshold policies $\pi(B)$ and $\pi(B+1)$ are both optimal. Although the index satisfies $\Delta V(B) = \nu(B)/\mu$, it cannot be found by simply computing $V(B)$ for these policies because $V(\cdot)$ depends on ν. Instead, the index $\nu(B)$ is found by solving for ν in the equation

$$g^{\pi(B)}(\nu) = g^{\pi(B+1)}(\nu). \qquad (23)$$

For the lost sales case, the contribution of ν to the gain is

$$g^{\pi(B)}(\nu) = g^{\pi(B)}(0) + f_B(B)\nu. \qquad (24)$$

Combining (23) and (24), then using (27) and simplifying gives

$$\nu(B) \quad = \quad \frac{g^{\pi(B+1)}(0) - g^{\pi(B)}(0)}{f_B(B) - f_{B+1}(B+1)} \qquad (24)$$

$$= -\frac{s}{\rho} + \frac{h}{(1-\rho)^2}\left[\rho^{-B-1} - 1 - (1-\rho)(B+1)\right]. \qquad (25)$$

Notice that the lost sales term in (25) is constant, meaning that the penalty paid for lost sales in this index scheme is the same whether lost sales are being incurred ($x = 0$) or not ($x > 0$). The reason for this result is that the cost s is only incurred in state 0 and can be thought of as a terminal cost for the process that stops upon entering state 0. This process stops with probability one from any initial state, so the expected cost is constant. Discounting would change this result. The lost sales term tends to dominate the index for small inventory positions (less than the hedging point), and the class with minimal $s_k\mu_k/\lambda_k$ is produced when x is small. This min $s\mu/\lambda$ rule will be seen again in Section 4.3. For large inventory positions (beyond the hedging point), the switching curve is approximately a straight line with slope $\mathrm{d}x_i/\mathrm{d}x_j = \ln\rho_j/\ln\rho_i$, which is a weighted version of the LQ policy. Idleness can be determined using a pure index policy if we set $\nu = 0$ for the idleness arm.

Unlike the Gittens index for the standard multi-armed bandit problem, the restless bandit index does not give an optimal policy. Under certain conditions, however, an asymptotic optimality holds. Let $\mathbf{X}$ be the state space for an arm and $D(\nu) \subseteq \mathbf{X}$ be the set of states in which the optimal policy for tax ν is active.

Definition 1. *An arm is indexable if $D(\nu)$ increases monotonically from $\emptyset$ to $\mathbf{X}$ as ν increases from $-\infty$ to ∞.*

Consider a problem with the constraint that exactly m of n arms must be active at any one time, and a relaxed-constraint problem where a time average of m arms must be active. If all arms are indexable, then as $m, n \to \infty$ with m/n fixed, the optimal per-project gain is asymptotically the same as that for the relaxed-constraint problem. Furthermore, Whittle conjectures and Weber and Weiss (1990) prove under an additional technical condition that the index policy defined above is asymptotically optimal for the relaxed-constraint problem. Hence, the index policy is asymptotically optimal for the exactly-m problem.

To establish indexability for the lost sales problem, one needs to show that $\nu(x)$ is nondecreasing and well-defined for all x. These properties follow from (25). However, the above result is for problems with $m \to \infty$ active arms, i.e., m machines. It says that, for an m-machine K-class problem (with $K + m$ arms so that all the machines can idle), the index policy is approximately optimal for a large number of machines and classes. The one-machine problem differs from an m-machine problem with production rates divided by m only in the higher production rate that can be applied to a class. If $K\rho_k$ is bounded for all k as $K \to \infty$, i.e., each class has a small utilization, then the production rate limit should be irrelevant and it is reasonable to expect that asymptotic optimality holds for the one-machine problem with a large number of classes. We have not tried to prove this rigorously.

In contrast, the backorder problem is not indexable; $\nu(x)$ does not exist (i.e., equals $-\infty$) for all x. The difficulty is that ν is a Lagrange multiplier for the constraint on the time-average number of active arms. For the backorder problem, any stable policy must serve a time-average of ρ classes, so relaxing this constraint does not change the optimal value and the Lagrange multiplier does not exist. In fact, no scheduling problem with a fixed utilization will be indexable.

A second characteristic of the restless bandit index is that the hedging point x^* defined by the pure index policy lies on the asymptotes of the optimal idleness region, $x_k^* = \min\{x_k : x \in \mathcal{I}\}$, so that x^* is a lower bound on the optimal hedging point. (By monotonicity, $\mathcal{I}$ consists of all points above a decreasing function. It can be shown that points in $\mathcal{I}$ are nonnegative, therefore $\mathcal{I}$ has vertical and horizontal asymptotes.) This property follows from the fact that for large x_j, $j \neq k$, the optimal control is the same as for the class k subproblem in isolation.

4 Hedging Points

Several idleness policies, listed in Table 1, are developed in this section. An idleness policy can be specified by a hedging point or a workload threshold. Workload is defined as the total expected production time represented by the stock in the system. Given a switching curve, the workload threshold defines a hedging point. The first two methods reduce to solving a single-product subproblem (Section 4.1). A third method (Section 4.2) uses the analysis of the LQ policy for the backorder problem in Zipkin (1992). A workload threshold can also be found through a Brownian motion approximation. The backorder problem is analyzed in Wein; we treat the lost sales problem in Section 4.3.

4.1 Single-Product Methods for Hedging Points

In this section, we present two hedging point approximations that reduce to single-product subproblems, solved in Section 2. The first method, called allocated server, creates a subproblem for each class by allocating the server as in Section 3.1. The subproblem parameters for class k are $\tilde{\mu} = (\rho_k/\rho)\mu_k$ and λ_k, h_k, b_k, and s_k unchanged. The second method, called aggregate product, aggregates all products into a single class. The subproblem parameters are $\tilde{\lambda} = \sum \lambda_k$, $\tilde{\mu} = \sum \lambda_k/\rho$, $\tilde{h} = \sum(\rho_k/\rho)h_k$, $\tilde{b} = \sum(\rho_k/\rho)b_k$, and $\tilde{s} = \sum(\rho_k/\rho)s_k$. To interpret stock levels for this aggregate product, we use the workload concept of Section 4.3. A stock level of B represents an expected production time, or workload, of $w = B/\tilde{\mu}$. This workload can then be combined with a switching curve to set stock levels for the original products.

All of these subproblems have the same utilization as the original system ($\tilde{\rho} = \rho$). Decomposing the system into single-product problems with the same utilization is analogous to decomposing a multi-server queue into parallel single-server queues.

Performance degrades, total queue length increases, and the optimal stock level increases to compensate. Thus, the allocated server hedging point should be larger than optimal. Aggregating the system into a single product neglects the variation in X_k given $\sum X_k$, i.e., the inability to maintain the desired allocation of inventory among products. Because variability is neglected, this hedging point should be smaller than optimal.

4.2 Hedging Points Based on Longest Queue Policies

Zipkin (1992) analyzes the longest queue (LQ) policy for identical products in the backorder problem. In this section, we use his result to approximate the steady-state distribution of the demand queue for non-identical products using an offset LQ switching curve. This distribution is then used to find a hedging point which we call the LQ hedging point. Five approximations are involved:

1. Assume an offset LQ switching curve is optimal,
2. Use Zipkin's approximation of the demand queue variance,
3. Adjust for non-identical products,
4. Fit a distribution to the mean and variance, and
5. Decompose the idleness decision into single-product subproblems.

Step 2 is exact for two products and step 1 is exact for identical products.

Zipkin's variance estimate is as follows. Let $N_k(t) = x_k^* - X_k(t)$ be the class k demand queue and $N = \sum N_k$. The steady-state variance of N_k can be decomposed as

$$\sigma_k^2 = \frac{\text{Var}(N) + E(D_k^2)}{K^2}, \tag{26}$$

where $E(D_k^2)$ measures the variability of N_k given N. Since N is the number of customers in an $M/M/1$ queue, $\mathrm{Var}(N) = \rho/(1-\rho)^2$. Use the approximation

$$E(D_k^2) \approx (K-1)\rho[1+\rho+(1-2\alpha_k)\rho^2+(1-2\alpha_k)^2\rho^3], \tag{27}$$

where $\alpha_k = \rho_k/\rho$.

For products with nonidentical λ or μ we define the effective number of products for class k as $K_k^{\mathrm{eff}} = 1/\alpha_k$, i.e., the number of identical products needed to achieve the system utilization, and use K_k^{eff} in place of K in (26). Also, estimate the mean of N_k as

$$E(N_k) = \frac{E(N)}{K_k^{\mathrm{eff}}} = \frac{\alpha_k \rho}{1-\rho} = \frac{\rho_k}{1-\rho}, \tag{28}$$

which is exact for identical products. Given the first two moments of N_k, we will fit a distribution. Since N is geometric, it is reasonable to use a distribution that has a geometric tail. For simplicity, we use a geometric distribution shifted a units to the right, $f(x) = pq^{(x-a)-1}$, $x = a+1, a+2, \ldots$. Fitting to (26) and (28) gives

$$q = 1 - \frac{\sqrt{4\sigma_k^2+1}-1}{2\sigma_k^2} \quad \text{and} \tag{29}$$

$$a = E(N_k) - \frac{1}{1-q}. \tag{30}$$

Finally, we use the distribution of N_k as if it were for some single-product subproblem with threshold control. The optimal stock level has already been found for a geometric distribution; (7) gives

$$B = \lfloor \ln[h/(b+h)]/\ln q \rfloor + a. \tag{31}$$

4.3 Brownian Motion Analysis

Wein develops a policy for the backorder problem using a diffusion approximation. The machine is idle when the workload $W(t) = \sum m_k X_k(t)$ exceeds the threshold

$$c = \frac{\sum \lambda_k m_k^2 (v_{sk}^2 + v_{dk}^2)}{2(1-\rho)} \ln(1 + b/h), \tag{32}$$

where $b = \min\{b_k \mu_k\}$, $h = \min\{h_k \mu_k\}$, and v_{sk} and v_{dk} are the service and inter-demand time coefficients of variation for class k. In the case of exponential distributions, $v_{sk} = v_{dk} = 1$. The switching curve (or surfaces) is determined by the $b\mu$ $h\mu$ rule, so that all of the workload is held in the class with minimal $h_k \mu_k$. This switching curve is not as realistic as the curves based on index policies found in Section 3. Only the idleness policy is used in our numerical tests. However, Wein shows that modifying this curve by holding a small amount of inventory in every other class can give fairly good results.

We derive an analogous policy for the lost sales problem. The essential difference is that the heavy traffic condition, ρ slightly less than one, is replaced by the conditions $\rho \approx 1$ and the ratio of holding to lost sales cost is small, $h/s \ll 1$. The approach taken is to (1) formulate the scheduling problem in terms of cumulative processes, (2) define an approximate Brownian motion control problem, (3) reformulate to give a more tractable control problem called the workload formulation, (4) solve the workload formulation for an initial throughput, and (5) calculate a new throughput from this solution and iterate until the throughput converges. For consistency with much of the heavy traffic scheduling literature, inventory will be denoted $Z_k(t)$, not $X_k(t)$, in this section.

4.3.1 The Scheduling Problem

Define the renewal counting processes

$$\begin{aligned} S_k(t) &= \text{number of class } k \text{ service completions after serving class } k \text{ for time } t, \\ D_k(t) &= \text{number of class } k \text{ demands that occur during the first } t \text{ units of nonstockout time, i.e., times } s \text{ when } Z_k(s) > 0, \end{aligned}$$

with rates and coefficients of variation μ_k, λ_k, v_{sk}, and v_{dk}, respectively. It is implicit in this definition of D_k that the demand process is turned off during stockouts, i.e., interdemand time is measured with respect to nonstockout time. For exponential distributions, the Poisson demand process of Section 1, which is not turned off, is equivalent. The scheduling problem of Section 1 can be posed as finding processes T_k that are nonanticipating with respect to Z to

$$\min \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T \sum h_k Z_k(t)\mathrm{d}t + \sum s_k \overline{A}_k(T)\right] \tag{33}$$

subject to

$$Z_k(t) = S_k(T_k(t)) - D_k(A_k(t)), \tag{34}$$

$$A_k(t) = \int_0^t 1_{\{Z_k(s)>0\}}\mathrm{d}s, \tag{35}$$

$$\overline{A}_k(t) = t - A_k(t), \tag{36}$$

$$I(t) = t - \sum T_k(t), \quad \text{and} \tag{37}$$

$$T_k \ \text{ and } \ I \ \text{ nondecreasing with } \ T_k(0) = 0. \tag{38}$$

Here $T_k(t)$ is the cumulative time that class k is produced in $(0, t]$, $I(t)$ is the idle time in $(0, t]$, and $A_k(t)$ is the time in $(0, t]$ during which class k demands can arrive.

4.3.2 The Brownian Control Problem

Let $\alpha_k = \rho_k/\rho$. Following Wein and Dai and Harrison (1991), define $Y_k(t) = \alpha_k t - T_k(t)$ and

$$X_k(t) = S_k(T_k(t)) - \mu_k T_k(t) - [D_k(A_k(t)) - \lambda_k A_k(t)] + (\mu_k \alpha_k - \lambda_k)t. \tag{39}$$

By (34), (36), and (39),

$$Z_k(t) = X_k(t) - \mu_k Y_k(t) + \lambda_k \overline{A}_k(t). \tag{40}$$

Note that $\overline{A}_k$ increases only when $Z_k = 0$. Letting $L_k(t) = \lambda_k \overline{A}_k(t)$, we see that L_k is the lower regulator for Z_k (see Harrison 1985), and the term $s_k \overline{A}_k(T)$ in (33) equals $l_k L_k(T)$, where $l_k = s_k/\lambda_k$ is the cost per lost sale.

A heavy traffic limit argument can be used to approximate X_k; see Veatch (1992) for details. We assume the heavy traffic condition that there exists a large integer n such that $\sqrt{n}|1-\rho|$ is small. Notice that, in contrast to traditional open queueing systems, our lost sales case allows for the possibility that $\rho > 1$. However, with lost sales, this heavy traffic condition is not sufficient to obtain a limiting Brownian control problem. As in Krichagina, Lou and Taksar (1992), we must also have that the relative cost of lost sales, l_k/h_k, tends to infinity; in particular, $l_k/h_k = O(n)$. We will see that this is precisely the increase in lost sales cost needed to make the optimal base stock level $O(\sqrt{n})$ as needed. It is the *scaled processes* $X_k(nt)/\sqrt{n}$, where n is large, for which a limit argument is constructed. Accordingly, we fix n and introduce the scaled processes and cost parameters

$$\hat{Z}_k(t) = \frac{Z_k(nt)}{\sqrt{n}}, \quad \hat{X}_k(t) = \frac{X_k(nt)}{\sqrt{n}}, \quad \hat{Y}_k(t) = \frac{Y_k(nt)}{\sqrt{n}}, \quad \hat{L}_k(t) = \frac{L_k(nt)}{\sqrt{n}}, \tag{41}$$

$$\hat{I}(t) = \frac{I(nt)}{\sqrt{n}}, \quad \hat{h}_k = \sqrt{n}h_k \quad \text{and} \quad \hat{l}_k = l_k/\sqrt{n}. \tag{42}$$

For a given policy, define $\gamma_k = \lim_{t\to\infty} T_k(t)/t$ and $P_k = \lim_{t\to\infty} \overline{A}_k(t)/t$. Then γ_k is the *actual* utilization for class k, reduced from ρ_k by lost sales, and $\mu_k\gamma_k$ is the class k throughput. The lost sales rate for class k is $\lambda_k P_k$, P_k is the probability of a stockout, and

$$\gamma_k = (1 - P_k)\rho_k \leq \rho_k. \tag{43}$$

Upon replacing $T_k(t)$ and $A_k(t)$ in (39) by their mean values $\gamma_k t$ and $(1 - P_k)t$, as proposed in Harrison (1988), it can be shown by weak convergence arguments that $X_k(nt)/\sqrt{n}$ is well approximated by a Brownian motion process. For simplicity of notation, we also use $\hat{X}_k$ to denote the Brownian motion. By (39), the drift of the Brownian motion $\hat{X}_k$ is $\sqrt{n}(\mu_k\alpha_k - \lambda_k)$ and the variance of $\hat{X}_k$ depends on the policy and equals $\mu_k\gamma_k v_{sk}^2 + \lambda_k(1-P_k)v_{dk}^2 = \mu_k\gamma_k(v_{sk}^2 + v_{dk}^2)$. Replace T by nT in (33) to get

$$\min \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T \sum \hat{h}_k \hat{Z}_k(t)\mathrm{d}t + \sum \hat{l}_k \hat{L}_k(T)\right]. \tag{44}$$

The requirement that T_k be nondecreasing can be dropped because the scaled control $T_k(nt)/\sqrt{n}$ increases with respect to t at a mean rate, or drift, $\sqrt{n}\gamma_k$. If $\gamma_k > 0$, then the scaled control has infinite drift as $n \to \infty$; if $\gamma_k = 0$, then $T_k(t) = 0$ is optimal. In either case, the scaled control is nondecreasing for large n. Hence, for small $|1-\rho|$ and large l_k/h_k, we are led to approximate the scheduling problem (33)-(39) by the following Brownian control problem (BCP): find processes $\hat{Y}_k$ that are nonanticipating with respect to $\hat{X}$ to achieve the objective (44) subject to

$$\hat{Z}_k(t) = \hat{X}_k(t) - \mu_k\hat{Y}_k(t) + \hat{L}_k(t), \tag{45}$$

$$\hat{I}(t) = \sum \hat{Y}_k(t), \tag{46}$$

$$\hat{L}_k(t) = -\inf_{0\leq s\leq t}\{\hat{X}_k(s) - \mu_k\hat{Y}_k(s)\}, \tag{47}$$

$$\hat{I} \text{ nondecreasing and } \hat{Y}_k(0) = 0. \tag{48}$$

The symbols $\hat{Z}_k, \hat{Y}_k, \hat{L}_k$ and $\hat{I}$ now refer to the Brownian approximation. As with other heavy traffic approximations, the BCP may be accurate even when these conditions are not met.

4.3.3 The Workload Formulation

To achieve a state space collapse, we reformulate as in Wein. Let $\hat{B}(t) = \sum m_k \hat{X}_k(t)$, a Brownian motion with drift $\delta = \sqrt{n}(1-\rho)$ and variance $\sigma^2 = \sum m_k \gamma_k (v_{sk}^2 + v_{dk}^2)$. The workload formulation (WF) is to find processes $\hat{Z}_k$, $\hat{I}$, and $\hat{L}_k$ that are non-anticipating with respect to $\hat{B}$ to minimize (44) subject to

$$\sum m_k \hat{Z}_k(t) = \hat{B}(t) - \hat{I}(t) + \hat{\mathcal{L}}(t), \tag{49}$$

$$\hat{\mathcal{L}}(t) = \sum m_k \hat{L}_k(t), \tag{50}$$

$$\hat{Z}_k(t) \geq 0, \quad \text{and} \tag{51}$$

$$\hat{I} \text{ and } \hat{L}_k \text{ nondecreasing.} \tag{52}$$

As the following theorem asserts, WF is a relaxation of BCP with the same optimal objective function value, and we can solve WF instead of BCP.

Theorem 1. *(i) Every feasible policy $\hat{Y}$ for BCP corresponds to a feasible policy $(\hat{Z}, \hat{I}, \hat{L})$ for WF of equal cost. (ii) Every optimal policy $(\hat{Z}, \hat{I}, \hat{L})$ for WF corresponds to a feasible policy $\hat{Y}$ for BCP of equal cost.*

A proof is given in Section 3.3.4 after the optimal policy is derived.

4.3.4 Solving the Workload Formulation

The workload formulation will be solved in two steps. First, an optimal $\hat{Z}$ and $\hat{L}$ is found in terms of $\hat{I}$, then an optimal $\hat{I}$ is found. Define $\hat{W}(t) = \sum m_k \hat{Z}_k(t)$ and classes i and j satisfying $\hat{h} \equiv \hat{h}_i \mu_i = \min\{\hat{h}_k \mu_k\}$ and $\hat{l} \equiv \hat{l}_j \mu_j = \min\{\hat{l}_k \mu_k\}$. It is optimal to set $\hat{\mathcal{L}}(t) = -\inf_{0 \le s \le t}\{\hat{B}(s) - \hat{I}(s)\}$, since this is the minimal $\hat{\mathcal{L}}$ that satisfies $\hat{W}(t) \ge 0$, implied by (51), and cost is increasing in $\hat{\mathcal{L}}$. Then the optimal $\hat{Z}$ at each t is a solution to the linear program

$$\min \quad \sum \hat{h}_k \hat{Z}_k(t) \tag{53}$$

$$\text{subject to} \quad \sum m_k \hat{Z}_k(t) = \hat{B}(t) - \hat{I}(t) + \hat{\mathcal{L}}(t) \;\text{ and} \tag{54}$$

$$\hat{Z}_k(t) \ge 0, \tag{55}$$

namely, the $h\mu$ rule

$$\hat{Z}_k^*(t) = \begin{cases} \mu_k \hat{W}(t), & k = i \\ 0, & k \ne i. \end{cases} \tag{56}$$

The optimal cost is $\hat{h}\hat{W}(t)$.

Similarly, the optimal $\hat{L}$ at each t is a solution to

$$\min \quad \sum \hat{l}_k \hat{L}_k(t) \tag{57}$$

$$\text{subject to} \quad \sum m_k \hat{L}_k(t) = \hat{\mathcal{L}}(t) \;\text{ and} \tag{58}$$

$$\hat{L}_k(t) \ge 0, \tag{59}$$

namely, the "$l\mu$" rule

$$\hat{L}_k^*(t) = \begin{cases} \mu_k \hat{\mathcal{L}}(t), & k = j \\ 0, & k \ne j. \end{cases} \tag{60}$$

Note that $\hat{L}^*$ is nondecreasing as required by (52). The optimal cost is $\hat{l}\hat{\mathcal{L}}(t)$.

Next we solve for $\hat{I}$. Substituting $\hat{Z}^*$ and $\hat{L}^*$ into WF gives

$$\min \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T \hat{h}\hat{W}(t)\mathrm{d}t + \hat{l}\hat{\mathcal{L}}(T)\right] \tag{61}$$

$$\text{subject to} \quad \hat{W}(t) = \hat{B}(t) - \hat{I}(t) + \hat{\mathcal{L}}(t) \ \text{ and} \tag{62}$$

$$\hat{\mathcal{L}}(t) = -\inf_{0\le s\le t}\{\hat{B}(s) - \hat{I}(s)\}. \tag{63}$$

A natural choice for $\hat{I}$ is to keep $\hat{W}$ in the interval $[0, \hat{c}]$; standard arguments (see, for example, Menaldi and Robin 1984 and Taksar 1985) can be used to show that such a policy is optimal. Let $\hat{I}$ be the unique function satisfying $\hat{I}(t) = \sup_{0\le s\le t}[\hat{B}(s) + \hat{\mathcal{L}}(s) - \hat{c}]^+$, (63), and $\hat{I}$ increasing only when $\hat{W}(t) = \hat{c}$. Then $\hat{W}$ is a regulated Brownian motion (RBM) on $[0, \hat{c}]$ with the same parameters as $\hat{B}$. Let us begin with the case $\delta > 0$, corresponding to $\rho < 1$. From Harrison (1985) p.90, the steady-state p.d.f. of $\hat{W}$ is $p(x) = \nu e^{\nu x}/(e^{\nu\hat{c}} - 1)$, $0 \le x \le \hat{c}$, where $\nu = 2|\delta|/\sigma^2$, and the lower control rate is

$$\beta \equiv \lim_{t\to\infty} \frac{\hat{\mathcal{L}}(t)}{t} = \frac{\delta}{e^{\nu\hat{c}} - 1}. \tag{64}$$

If we restrict ourselves to RBM policies, then (61) - (63) reduces to finding $\hat{c}$ to minimize

$$\begin{aligned} \phi(\hat{c}) &= \int_0^{\hat{c}} \hat{h}xp(x)\mathrm{d}x + \hat{l}\beta && (63) \\ &= \int_0^{\hat{c}} \frac{\hat{h}\nu x e^{\nu x}}{e^{\nu\hat{c}} - 1}\mathrm{d}x + \frac{\hat{l}\delta}{e^{\nu\hat{c}} - 1} && (64) \\ &= \frac{\hat{l}\delta + \hat{h}\hat{c}e^{\nu\hat{c}}}{e^{\nu\hat{c}} - 1} - \frac{\hat{h}}{\nu}. && (65) \end{aligned}$$

Setting $\phi'(\hat{c}) = 0$ yields

$$e^{\nu\hat{c}} - \nu\hat{c} - 1 - \nu\delta\hat{l}/\hat{h} = 0. \tag{66}$$

We now reverse the scalings of the parameters in this equation so that it is expressed solely in terms of the original problem parameters. The solution $\hat{c}$ to (66) is an upper control limit for the scaled workload. If we denote the workload in the original system by $W(t)$, where $\hat{W}(t) = W(nt)/\sqrt{n}$, then our proposed idleness policy for the original system is to idle when $W(t) = c = \sqrt{n}\hat{c}$. In terms of c, (66) can be written, using (42) and $\delta = \sqrt{n}(1-\rho)$, as

$$e^{\frac{2(1-\rho)c}{\sigma^2}} - \frac{2(1-\rho)c}{\sigma^2} - 1 - \frac{2(1-\rho)^2 l}{\sigma^2 h} = 0, \tag{67}$$

where $h \equiv \min\{h_k\mu_k\}$ and $l \equiv \min\{l_k\mu_k\}$. This equation can be solved numerically for the proposed workload threshold c in terms of the original problem parameters.

Now consider the case $\delta < 0$, corresponding to $\rho > 1$. The p.d.f. of $\hat{W}$ is $p(x) = \nu e^{-\nu x}/(1 - e^{-\nu\hat{c}})$, $0 \le x \le \hat{c}$, and the lower control rate is

$$\beta = \frac{|\delta|}{1 - e^{-\nu\hat{c}}}. \tag{68}$$

The cost rate is

$$\begin{aligned} \phi(\hat{c}) &= \int_0^{\hat{c}} \frac{\hat{h}\nu x e^{-\nu x}}{1 - e^{-\nu\hat{c}}}\mathrm{d}x + \frac{\hat{l}|\delta|}{1 - e^{-\nu\hat{c}}} && (68) \\ &= \frac{\hat{l}|\delta| - \hat{h}\hat{c}e^{-\nu\hat{c}}}{1 - e^{-\nu\hat{c}}} + \frac{\hat{h}}{\nu}, && (69) \end{aligned}$$

and is minimized at

$$e^{-\nu\hat{c}} + \nu\hat{c} - 1 - \nu|\delta|\hat{l}/\hat{h} = 0, \quad \text{or} \tag{70}$$

$$e^{\frac{2(1-\rho)c}{\sigma^2}} + \frac{2(1-\rho)c}{\sigma^2} - 1 - \frac{2(1-\rho)^2 l}{\sigma^2 h} = 0. \tag{71}$$

Finally, for the case $\mu = 0$, corresponding to $\rho = 1$, the p.d.f. of $\hat{W}$ is $p(x) = 1/\hat{c}$, $0 \le x \le \hat{c}$, and the lower control rate is $\beta = \sigma^2/(2\hat{c})$. The cost rate is $\phi(\hat{c}) = \hat{h}\hat{c}/2 + \hat{l}\sigma^2/(2\hat{c})$, and is minimized at $\hat{c} = \sqrt{\sigma^2\hat{l}/\hat{h}}$, or

$$c = \sqrt{\sigma^2 l/h}. \tag{72}$$

These three results are consistent. As $\rho \to 1$, the small-exponent approximation $e^x \approx 1 + x + x^2/2$ can be used in (67) and (71), giving (72).

We end this section by proving the theorem.

Proof of theorem.

(i) Given $\hat{Y}$ feasible for BCP, let $\hat{Z}_k$ satisfy (45), $\hat{I}$ satisfy (46), $\hat{L}_k$ satisfy (47), and $\hat{\mathcal{L}}$ satisfy (50). Then

$$\begin{aligned} \sum m_k \hat{Z}_k(t) &= \sum m_k \hat{X}(t) - \sum \hat{Y}_k(t) + \sum m_k \hat{L}_k(t) &(72)\\ &= \hat{B}(t) - \hat{I}(t) + \hat{\mathcal{L}}(t), &(73) \end{aligned}$$

i.e., (49) holds. Also, (45), (47), and (48) imply (51) - (52), and $(\hat{Z}, \hat{I}, \hat{L})$ is feasible for WF.

(ii) Given $(\hat{Z}, \hat{I}, \hat{L})$ optimal for WF, let $\hat{Y}$ satisfy (45), namely $\hat{Y}_k(t) = m_k[\hat{X}_k(t) - \hat{Z}_k(t) + \hat{L}_k(t)]$. Then

$$\sum \hat{Y}_k(t) = \hat{B}(t) - \sum m_k \hat{Z}_k(t) + \hat{\mathcal{L}}(t) = \hat{I}(t), \tag{74}$$

i.e., (46) holds. Substituting $\hat{Y}_k$ into the r.h.s. of (47) gives

$$-\inf_{0 \le s \le t} \{\hat{Z}_k(s) - \hat{L}_k(s)\}. \tag{75}$$

For $k \neq j$, $\hat{L}_k(s) = 0$, (75) reduces to $\hat{Z}_k(0) = 0$, and (47) holds. Now consider $k = j$. If $j \neq i$, then $\hat{Z}_j(s) = 0$ and, since $\hat{L}_j$ is nondecreasing, (75) is just $\hat{L}_j(t)$, i.e., (47) holds. If $j = i$, then by WF optimality, $\hat{L}_j$ increases only at times s when $\hat{\mathcal{L}}(s)$ is increasing. But $\hat{\mathcal{L}}$ is the lower regulator for $\hat{W}$, so at these times $\hat{W}(s) = \hat{Z}_j(s) = 0$. Since $\hat{Z}_j(s) \geq 0$ and $\hat{L}_j$ is nondecreasing, it follows that (75) is the largest value of $\hat{L}_j$, namely, $\hat{L}_j(t)$, and (47) holds. Optimality also ensures that $\hat{I}(0) = 0$ and (48) holds, so $\hat{Y}$ is feasible for BCP. □

Table 2: Throughput Iteration for Lost Sales Case 2.

Iteration	Initial γ_2	c	Final γ_2
1	.45	10.8	.4069
2	.4069	10.5	.4084
3	.4084	10.5	.4083

4.3.5 Updating the Throughput

The Brownian motion variance σ^2 appearing in (66) and (70) depends on the unknown throughputs γ_k. As in Dai and Harrison, we overcome this difficulty by iteratively computing c and γ. A reasonable initial value is $\gamma_k = \rho_k$. Given γ_k, compute σ^2, use (67), (71) or (72) to compute c, and (64) or (68) to compute β. To update γ, recall that all lost sales are attributed to class j by the $l\mu$ rule, so that $\hat{\mathcal{L}}(t) = m_j \hat{L}_j(t)$ and the lost sales rate for class j is $\lambda_j P_j = \beta/m_j$. From (43), we obtain

$$\gamma_j = \rho_j - \beta. \tag{76}$$

It is possible for (76) to give $\gamma_j < 0$, meaning that there are more lost sales than class j arrivals. A reasonable allocation of these lost sales is to set $\gamma_j = 0$, $\beta = \beta - \rho_j$, and repeatedly apply (76) to the class with next smallest $l_k \mu_k$. Using the new γ, the calculations can be repeated. Convergence is reached rapidly, as demonstrated in Table 2.

5 Numerical Results

Dynamic programming value iteration was used to compute optimal policies for undiscounted problems with two and three products. The recurrent states are those below the hedging point, $x \leq x^*$. For the lost sales problem, the recurrent class is finite,

$0 \leq x \leq x^*$. For the backorder problem, the state space was truncated. Larger and larger state spaces were tested until the results were insensitive to increasing the state space. State spaces up to about 30 by 30 and up to 2000 value iterations were required to achieve three digit accuracy. The lost sales problem generally ran much faster.

All compatible combinations of switching curves and idleness policies listed in Table 1 were tested. These candidate policies were evaluated using a value-iteration scheme to avoid directly solving a large linear system. The LQ and offset LQ switching curves described in Section 1 were also tested, as were the hedging points generated by pure STLA and restless bandit index policies. Finally, for STLA, restless bandit, and LQ switching curves, a one-dimensional search along the switching curve was conducted to find the best hedging point for that switching curve. This data point is used to determine how much of the suboptimality of a policy is due to the switching curve and how much is due to the idleness policy. These three switching curves are combined with hedging points by converting the hedging point to a workload threshold (see Section 3.3), then finding the point on the switching curve that matches or exceeds this workload. The $\mu\Delta V$ and offset LQ switching curves require a K-dimensional hedging point, not just a a one-dimensional workload, to be specified. Best hedging points were not found for these switching curves.

5.1 Lost Sales

Most of the testing was devoted to the two-product lost sales problem. Five test cases are defined in Table 3. We begin by comparing idleness policies. Table 4 shows the idleness policies for the test problems. Since $\mu_k = 1$ in these problems, the workload is just the sum of the hedging point coordinates, $w = \sum x_k^*$. The suboptimality, measured in terms of the cost rate g, is also shown. For convenience, all of the idleness policies are combined with the STLA switching curve, except for the

Table 3: Lost Sales Cases ($\mu = 1$)

Case	λ_1	λ_2	λ_3	h_1	h_2	h_3	s_1	s_2	s_3
1	.4	.5		1	1		60	80	
2	.45	.45		4	1		45	45	
3	.45	.45		1	1		90	360	
4	.3	.6		1	1		120	300	
5	.3	.4		2	1		60	40	
6	.2	.3	.4	1	1.25	1.5	40	75	120

Table 4: Lost Sales Idleness Policies

		Pure Index									
	Opt.	Restless		STLA		Allocated		Aggregate		Brownian	
Case	x^*	x^*	sub	x^*	sub	x^*	sub	w	sub	w	sub
1	(6,7)	(4,5)	15%	(4,4)	21%	(8,10)	9%	9	12%	12.4	0%
2	(3,6)	(2,4)	7%	(2,3)	14%	(4,7)	2%	5	14%	10.5	2%
3	(7,10)	(5,6)	41%	(4,5)	54%	(10,18)	23%	15	5%	13.9	8%
4	(7,13)	(4,8)	48%	(4,6)	54%	(11,17)	15%	15	9%	17.9	1%
5	(3,5)	(3,4)	2%	(3,3)	5%	(5,5)	8%	6	5%	6.5	0%
6	(5,5,6)	(3,3,4)	29%	(2,3,4)	28%	(7,8,10)	27%	9	28%	13.9	2%

pure restless bandit index hedging point, which is combined with its own switching curve. However, the ranking of the suboptimalities was the same for most of the cases when other switching curves were used. Note that the allocated server hedging points shown must be shifted onto the STLA switching curves, maintaining the same workload, before evaluating the policy.

The results show the Brownian idleness policy to be the clear winner, both in terms of accuracy of the workload threshold and the resulting gain. If the workload must be broken down into stock levels, the allocated server hedging point is the best candidate. As expected, the allocated server hedging point is too large and the aggregate product too small. Since the pure index policies disregard congestion

Table 5: Lost Sales Switching Curves

	LQ		Offset LQ		Restless		STLA		$\mu\Delta V$	
Case	w	sub	x^*	sub	x^*	sub	x^*	sub	x^*	sub
1	13	1%	(8,10)	11%	(6,7)	1%	(6,7)	0%	(8,10)	10%
	13	1%			(6,7)	1%	(6,7)	0%		
2	11	20%	(4,7)	13%	(5,6)	13%	(3,8)	2%	(4,7)	6%
	8	9%			(3,5)	1%	(3,6)	0%		
3	14	38%	(10,18)	35%	(7,7)	15%	(6,8)	8%	(10,18)	34%
	20	17%			(9,9)	6%	(8,10)	2%		
4	18	9%	(11,17)	20%	(6,12)	8%	(7,11)	1%	(11,17)	19%
	21	6%			(7,14)	6%	(8,12)	1%		
5	7	5%	(5,5)	16%	(3,4)	2%	(3,4)	0%	(5,5)	15%
	8	4%			(3,5)	1%	(3,5)	0%		
6	14	7%	(7,8,10)	26%	(4,4,6)	6%	(5,4,5)	2%	(7,8,10)	27%
	17	2%			(4,5,6)	4%	(6,4,5)	2%		

from other classes, their hedging points are too small, as predicted in Section 3.3. However, they perform much better in case 5, where $\rho = 0.7$, than the other cases, where $\rho = 0.9$.

Next, using the best available idleness policy (Brownian if only the workload is needed, allocated server if the hedging point is needed), the switching curves are compared in Table 5. The second row of figures within a case gives the best hedging point for the switching curve, where available. The best hedging point illustrates how much of the suboptimality is due to the switching curve. The results suggest that the best policy is the STLA index combined with the Brownian idleness policy. Its gain is within 8% of optimal for all five test cases. The restless bandit index combined with the Brownian idleness policy also does well. Much of the suboptimality for these policies is due to the idleness policy, even though we have used the best available policy. It appears that finding a good idleness policy is more difficult than finding a good switching curve. The potential savings from using index policies is best measured by comparing STLA to the LQ policy. Average suboptimality is reduced from 13%

to 2%.

The offset LQ and $\mu\Delta V$ switching curves perform poorly, possibly because they must use the less accurate allocated server hedging point. The simple LQ switching curve performs slightly better, but has trouble with asymmetric products, such as case 3. Its suboptimality is due to both the switching curve and the Brownian idleness policy that is used with it.

The shapes of various switching curves for case 4 are shown in Fig. 1. The STLA curve is slightly closer to the line of symmetry $x_1 = x_2$, i.e., the LQ switching curve, than optimal. The restless bandit curve starts much farther from symmetry and initially is nearly horizontal, reflecting the dominance of the lost sales term in (25). The $\mu\Delta V$ curve is parallel to the symmetry line but shifted too far away. The optimal idleness region is also shown. The curvature of its boundary determines how far the optimal hedging point, (7,13), is from the the pure restless bandit index hedging point, (4,8), the latter lying on the asymptotes of the idleness region.

The optimal switching curve was found to be close to the symmetry line for fairly different products. The curve is particularly insensitive to s, with differences of a factor of two scarcely affecting the curve. Significantly different product parameters were found to have the following effects on the switching curve:

1. $s_1 \ll s_2$. The curve is offset toward class 2 ($x_1 < x_2$).

2. $h_1 \gg h_2$. The curve starts at the origin and moves toward $x_1 < x_2$ (slope steeper than one).

3. $\lambda_1 < \lambda_2$. The curve starts at $x_1 > x_2$ (below the symmetry line) and moves to $x_1 < x_2$ (slope steeper than one). If lost sales costs are equal, $s_1/\lambda_1 = s_2/\lambda_2$, then the curve starts above the symmetry line.

4. Expensive product: $s_1 \ll s_2$ and $h_1/s_1 = h_2/s_2$. The curve starts at $x_1 < x_2$

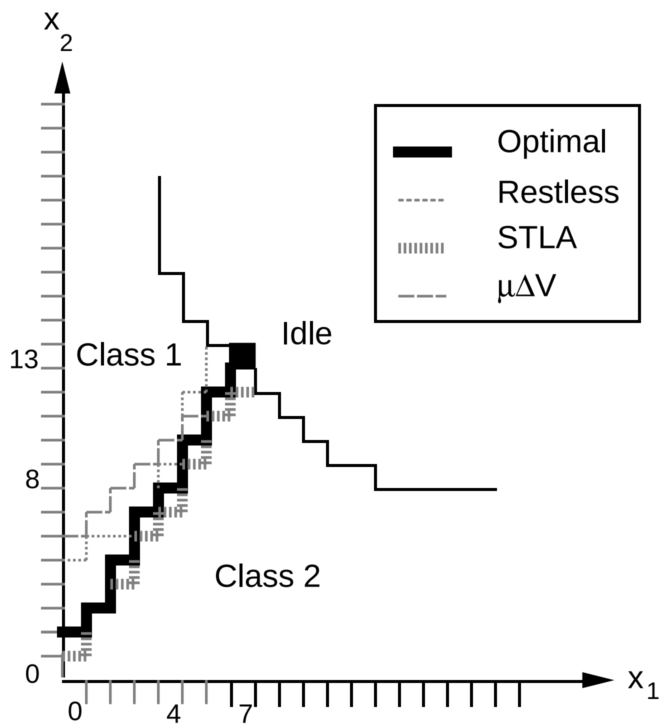

Figure 1: Shape of the Switching Curves— Lost Sales Case 4

(above the symmetry line) and is nearly horizontal.

We also made a few runs with three products. Case 6 in Tables 4 and 5 is consistent with the results for two products. Recall that the restless bandit index is predicted to become more accurate for a large number of products. The three-product cases we have tested neither support nor refute this prediction.

5.2 Backordering

Three backorder cases are described in Table 6. The idleness policies and their sub-optimality are shown in Table 7. Again, STLA switching curves are combined with the idleness policies. The results suggest that the Brownian and LQ policies are the most accurate, with Brownian a little better on case 1, which, surprisingly, is the case with the lowest traffic intensity. The aggregate product policy also performs fairly

Table 6: Two-Product Backorder Cases ($\alpha = 0, \mu = 1$)

Case	λ_1	λ_2	h_1	h_2	b_1	b_2
1	.3	.4	2	1	10	5
2	.45	.45	1	1	2	4
3	.45	.45	1.25	1	4	2

Table 7: Backorder Idleness Policies

	Opt.	Pure STLA		Allocated		LQ		Aggregate		Brownian	
Case	x^*	x^*	sub	x^*	sub	x^*	sub	w	sub	w	sub
1	(1,3)	(1,1)	23%	(5,5)	46%	(1,2)	6%	5	0%	4.2	0%
2	(4,4)	(0,1)	55%	(10,15)	99%	(4,6)	5%	13	15%	9.9	5%
3	(3,5)	(1,0)	48%	(13,10)	78%	(6,4)	7%	12	13%	9.9	7%

well. As expected, the allocated server hedging point is too large and the aggregate product too small.

To compare switching curves, the Brownian policy is used when only a workload is needed; the LQ hedging point is used when a hedging point is needed. Table 8 shows the results, where the second row within each case again gives the best hedging point for the switching curve. The STLA index combined with the Brownian idleness policy performs the best, with suboptimality of 7% or less. The LQ switching curve (which is symmetrical) combined with the Brownian idleness policy also performs well, but may do poorly on more asymmetric products. Case 3 suggests that the LQ hedging point of (6,4) gives a fairly accurate workload (10 versus an optimal workload of 8), used with the STLA switching curve in Table 7, but poor stock levels, used with the offset LQ switching curve in Table 8. The differences found between switching curves are not very dramatic; however, the results clearly show that choosing a good hedging point is very important.

Switching curves for case 3 are drawn in Fig. 2. The $\mu\Delta V$ curve is inaccurate

Table 8: Backorder Switching Curves

	LQ		Offset LQ		STLA		$\mu\Delta V$	
Case	w	sub	x^*	sub	x^*	sub	x^*	sub
1	5	7%	(1,2)	8%	(1,4)	0%	(2,1)	11%
	4	3%			(1,3)	0%		
2	10	6%	(4,6)	7%	(5,5)	5%	(3,7)	12%
	9	5%			(4,5)	4%		
3	10	7%	(6,4)	10%	(2,8)	7%	(7,3)	15%
	8	5%			(2,6)	5%		

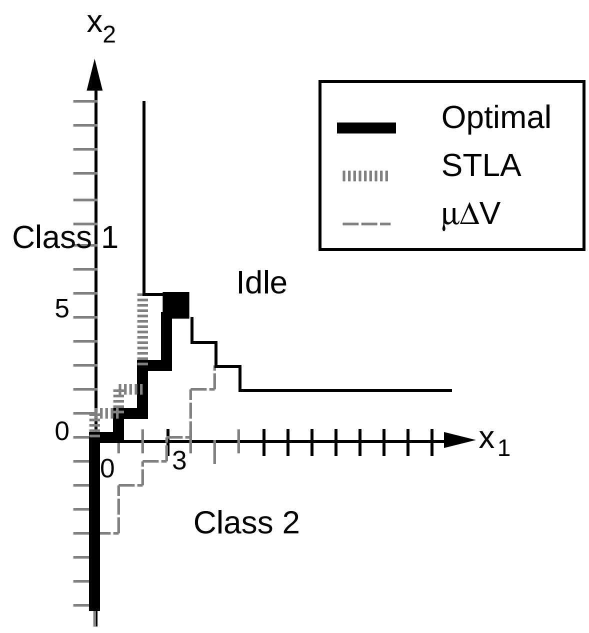

Figure 2: Shape of the Switching Curves— Backorder Case 3

primarily because of the LQ hedging point, (6,4), on which it is based. The STLA curve is slightly more asymmetrical than the optimal curve. In general, backorder curves are much more sensitive to asymmetric products, specifically to differences in λ and h. The 25% difference in holding costs in case 3 is responsible for moving the hedging point from symmetry to (3,5). Larger differences in h, such as in case 1, result in a hedging point near the axis. This explains the good performance of the "modified $h\mu$" rule reported in Wein, where hedging points one unit away from the axis, e.g., $(x_1, 1)$, were used.

6 Concluding Remarks

We have considered a scheduling problem for a multiclass make-to-stock queue with either backordering or lost sales. Since the optimal solution can only be numerically computed when there are a small number of products, our goal was to derive a scheduling policy, which consists of an index policy that specifies which class to produce and an idleness policy, that performs well and can be easily computed when there are a large number of products.

Our numerical results suggest that the STLA index policy derived by Zipkin (1990) coupled with the idleness policy derived by a Brownian analysis performs very well, at least for two- and three-product problems. Of all the index and idleness policies considered in this paper, these two are the only ones for which the exponential assumption can be easily relaxed. Our numerical results also suggest that both a good idleness policy and a good switching curve are required to attain performance that is close to optimal. Moreover, good idleness policies appear to be harder to derive than good switching curves.

We did not attempt any numerical computations for problems with a large number of products because the suboptimality of our proposed policies could not be calculated. It is possible that the accuracy of the Brownian policy would degrade with more products, since it is based on the $h\mu$ rule, which holds all inventory in one class. However, a Brownian workload threshold can be computed using any desired inventory mix, not just the mix dictated by the $h\mu$ rule. For example, Section 9 of Wein obtains an idleness policy for the backorder problem by performing a Brownian analysis under the LQ policy. This method may be more accurate for large problems that are relatively symmetric.

We noticed a relationship between the Gittens index for the multi-arm bandit problem and the restless bandit index that may be useful for other research. The Gittens index measures the value of playing an arm (producing a class) given that there are other arms of equal value, so that when its value drops we can "retire" to another arm of equal value. For restless problems, the arms change state while passive and there is no constant retirement value. As a result, the Gittens index may not be meaningful. The Gittens index for the discounted, backorder problem, computed in Veatch (1992), is not monotone and does not produce a coherent policy. Whittle's restless bandit index assumes that the cost of the machine is constant over time, and that each product can use the machine whenever it is cost-effective. For our problem, these assumptions are only accurate when there are many classes and many machines (or one machine and many low utilization classes). Another approach is to compute a Gittens index using a variable retirement cost, $M(x_k, \nu)$, that depends on the inventory x_k, as well as the "base" retirement cost ν. The traditional Gittens index uses $M(x_k, \nu) = \nu$. Using $M(x_k, \nu) = V(x_k, \nu)$, the optimal value function for the single-product subproblem with machine cost ν (see Section 3.3), gives the restless bandit index. In other words, the restless bandit index can be defined as a Gittens index with variable retirement cost. The question is how to specify $M(x_k, \nu)$ so that the Gittens index will be nondecreasing and produce accurate policies.

The restless bandit and generalized Gittens index may also be useful for attacking the related problem with set-up costs. When set-up costs are added, the state of the dynamic program must be augmented with the class currently being produced. The form of the optimal policy becomes much more complex, involving lot-sizing and scheduling, and good approximations have not been found. An index policy could be constructed by computing two indices for each class, measuring the value of starting and stopping production of that class.

Acknowledgements

We acknowledge helpful discussions with Avi Mandelbaum, Viên Nguyen, and Paul Zipkin. This research is supported by National Science Foundation grant DDM-9057297.

References

Dai, J.G. and J.M. Harrison 1991. Steady-State Analysis of RBM in a Rectangle: Numerical Methods and a Queueing Application. *Annals Appl. Prob.* **1**, 16-35.

Ha, A.Y. 1993. Optimal Dynamic Scheduling Policy for a Make-to-Stock Production System. Working paper series B #124, School of Organization and Management, Yale University, New London.

Harrison, J.M. 1985. *Brownian Motion and Stochastic Flow Systems.* John Wiley and Sons, New York.

Harrison, J.M. 1988. Brownian Models of Queueing Networks with Heterogeneous Customer Populations, in W. Fleming and P.L. Lions (eds.), *Stochastic Differential Systems, Stochastic Control Theory and Applications*, IMA Vol. **10**, Springer-Verlag, New York, 147-186.

Lippman, S. 1975. Applying a New Device in the Optimization of Exponential Queueing Systems. *Oper. Res.* **23**,687-710.

Kimemia, J.G. and S.B. Gershwin 1983. An Algorithm for the Computer Control of Production in Flexible Manufacturing Systems. *IIE Trans.* **15**, 353-362.

Klimov, G.P. 1974. Time Sharing Service Systems I. *Th. Appl. Prob.* **19**, 532-551.

Krichagina, E.V., S.X.C. Lou and M.I. Taksar 1992. Double Band Policy for Stochastic Manufacturing Systems in Heavy Traffic. Working paper, State Univ. of New York, Stony Brook.

Menaldi, J.L. and M. Robin 1984. Some Singular Control Problems with Long Term Average Criterion. *Lecture Notes in Control and Information Sciences*, Springer-Verlag, New York, 424-432.

Taksar, M.I. 1985. Average Optimal Singular Control and a Related Stopping Problem. *Math. Oper. Res.* **10**, 63-81.

Veatch, M.H. 1992. Queueing Control Problems for Production/Inventory Systems. Ph.D. dissertation, Sloan School of Management, MIT, Cambridge.

Veatch, M.H. and L.M. Wein 1992. Monotone Control of Queueing Networks. *Queueing Systems* **12**, 391-408.

Weber, R.R. and G. Weiss 1990. On an Index Policy for Restless Bandits. *J. Appl. Prob.* **27**, 637-648.

Wein, L.M. 1992. Dynamic Scheduling of a Multiclass Make-to-Stock Queue. *Oper. Res.* **40**, 724-735.

Whittle, P. 1988. Restless Bandits: Activity Allocation in a Changing World. In *A Celebration of Applied Probability*, ed. J. Gani, *J. Appl. Prob.* **25A**, 287-298.

Zheng, Y. and P. Zipkin 1990. A Queueing Model to Analyze the Value of Centralized Inventory Information. *Oper. Res.* **38**, 296-307.

Zipkin, P. 1992. Performance of the Smallest-Inventory (or Longest-Queue) Policy. Working paper, Graduate School of Business, Columbia Univ., New York.