Zero-Inventory Conditions For a Two-Part-Type Make-to-Stock Production System

Michael H. Veatch * Francis de Véricourt †

October 9, 2002

Abstract

We consider the dynamic scheduling of a two-part-type make-to-stock production system using the model of Wein [12]. Exogenous demand for each part type is met from finished goods inventory; unmet demand is backordered. The control policy determines which part type, if any, to produce at each moment; complete flexibility is assumed. The objective is to minimize average holding and backorder costs. For exponentially distributed interarrival and production times, necessary and sufficient conditions are found for a zero-inventory policy to be optimal. This result indicates the economic and production conditions under which a simple make-to-order control is optimal. Weaker results are given for the case of general production times.

Key Words: Make-to-Stock Queue; Hedging Points; Just-in-Time

1 Introduction

Safety stock serves a vital role in buffering against supply and demand uncertainty in production systems. Even in a lean manufacturing environment

*Department of Mathematics, Gordon College, Wenham, MA 01984, veatch@gordon.edu

†Fuqua School of Business, Duke University, Durham, NC 27708, fdv1@duke.edu

with short cycle times, a make-to-stock environment may be required to rapidly and reliably fill orders. Assuming that building stock in advance is feasible, operational and economic parameters determine whether it should be built in advance or whether a make-to-order policy is preferable.

For a simple model with one production stage, a single part type, and backordering costs the tradeoff is known: Zero inventory is optimal if the probability of backordering a demand is less than a certain cost ratio. In multi-part-type systems, the decision is made more complex by the variety of sequencing controls that are possible. We give necessary and sufficient conditions on the parameters of a two part-type system for a zero-inventory policy to be optimal. The model we consider is that of Wein [12]. A single production facility produces stock in order to satisfy arriving demands.

For tractability, the production system is modeled by an exponential server and the demands arrive according to independent Poisson processes. However, partial results for general production times are presented in the last section. This model is the make-to-stock version of a multi-class queue. We consider any make-to-stock policy; i.e., production capacity can be dynamically allocated between part types. The objective is to minimize average holding and backorder costs.

The structure of the optimal policy is only partially known for this problem. Wein [12] provides insights into the structure of the optimal policy using a Brownian approximation for this multi-class make-to-stock queuing control problem. Ha [3] shows that the optimal policy is a hedging point policy in the case of infinite horizon discounted cost. A hedging point policy is characterized by switching and idling curves which split the state space in three regions: one where the machine is idle and the two others that determine the part type to produce. The intersection of these two curves is called the hedging point. The hedging point and a switching curve are sufficient to describe the stationary behavior of the policy. de Véricourt et al. [10] show that the optimal switching curve is a straight line in a particular subset of the state space. Numerical studies indicate that the remainder of the curve has no simple structure. Heuristics have been proposed by Wein [12]. Veatch and Wein [9] and Peña-Perez and Zipkin [4] propose effective heuristics to approximate this curve along with the optimal hedging point.

However, if zero inventory (produce only when demands are waiting in the system) is optimal, Ha [3] implies that the production capacity is allocated among the part types according to a "$c\mu$" rule. This completely specifies the control of the system. In this paper, we derive necessary and sufficient conditions under which no on-hand inventory is held. These conditions provide a full characterization of the optimal policy for some specific values of the parameters. They are of practical interest, for they indicate when the system should be controlled in a simple make-to-order mode.

Similar conditions for zero inventory have been derived for the continuous flow control problem (Bieleki and Kumar [1], Perkins and Srikant [5], Presman et al. [6], Veatch and Caramanis [8], Yee and Veatch [13]). In this model, the flow of discrete parts is approximated by a continuous "fluid". The randomness in the system is only due to machine failures, which are captured by a Markov process. In this paper we partly follow the approach of Veatch and Caramanis [8], evaluating different coupled trajectories generated by policy perturbations.

The rest of the paper is organized as follows. The dynamic scheduling problem is presented in Section 2 with the optimality equations. In Section 3, we describe the general structure of the optimal policy and we define the zero-inventory policy. Necessary conditions for zero inventory are derived in Section 4 and shown to be sufficient in Section 5. An extension to general production time is discussed in Section 6. Finally, the conditions of Sections 4 and 5 are studied numerically in Section 7.

2 The Dynamic Scheduling Problem

Consider a production system with a single flexible machine that produces two part types in a make-to-stock mode. We assume that raw materials are always available in front of the machine. Each finished item is placed in its respective inventory. When a demand arrives to the system, it is satisfied with the on-hand inventory of the required part type, if it is not empty. The demand is backordered otherwise. Type i demand arrives according to an independent Poisson process with rate λ_i, $i = 1, 2$. The production times of the products are exponentially distributed with rates μ_i for $i \in \{1, 2\}$.

At any time, a control policy specifies whether to produce part type 1 or 2, or to idle the machine. The production of a part can be interrupted and resumed, so that a preemptive discipline is permitted . Since the system is memoryless, for the control of the system we can consider only Markov policies, which only depend on the current state.

We denote by $\mathbf{x}(t) = (x_1(t), x_2(t))$ the state of the system where $x_i(t)$ is the surplus, (or negative of the backlog if demands are backordered) of type

i. We also denote by $\mathbf{X}(t) = (X_1(t), X_2(t))$ the associated random variable. Let C^π be the control associated with a Markov policy π.

$$C^\pi(\mathbf{x}) = \begin{cases} 0 & \text{when the action is to idle} \\ 1 & \text{when the action is to produce type 1} \\ 2 & \text{when the action is to produce type 2} \end{cases}$$

We consider a unit holding cost h_i and a non-zero unit backorder cost b_i per unit of time for type i. In the rest of the paper we also assume without loss of generality that $\mu_1 b_1 > \mu_2 b_2$. In the state $\mathbf{x}$, the system incurs a cost rate of $c(\mathbf{x}) = \sum_{i=1}^{2} c_i(x_i)$ with $c_i(x_i) = h_i x_i^+ + b_i x_i^-$ (where $x_i^+ = \max(x_i, 0)$ and $x_i^+ = -\min(x_i, 0)$). The objective is then to find the policy which minimizes the long run average cost

$$\limsup_{t\to\infty} \frac{1}{t} E^\pi_{\mathbf{x}_0}[\int_0^t c(\mathbf{X}(t))dt]. \tag{1}$$

where $E^\pi_{\mathbf{x}_0}$ denotes the expectation given the control policy π and initial state $\mathbf{x}_0$.

The optimal average cost rate g^* and the relative value function $v(\mathbf{x})$ satisfy the following dynamic programming optimality equations (see Veatch and Wein [9]):

$$\begin{aligned} v(\mathbf{x}) + \frac{g^*}{\Lambda} &= \frac{1}{\Lambda}\Big[c(\mathbf{x}) + \lambda_1 v(\mathbf{x} - \mathbf{e}_1) + \lambda_2 v(\mathbf{x} - \mathbf{e}_2) + \mu v(\mathbf{x}) \\ &\quad + \min(0, \mu_1 \Delta_1 v(\mathbf{x}), \mu_2 \Delta_2 v(\mathbf{x}))\Big], \end{aligned} \tag{2}$$

where $\mathbf{e}_1$ is the unit vector along x_1, $\mathbf{e}_2$ is the unit vector along x_2, $\Delta_i v(\mathbf{x}) = v(\mathbf{x} + \mathbf{e}_i) - v(\mathbf{x})$, $\mu = max(\mu_1, \mu_2)$ and $\Lambda = \lambda_1 + \lambda_2 + \mu$.

We denote by $\lambda = \lambda_1 + \lambda_2$, the total arrival rate. $\rho_i = \lambda_i/\mu_i$ is the utilization rate for type i, and $\rho = \rho_1 + \rho_2$ is the total utilization rate. In the rest of the paper, we assume that ρ is less than one.

An optimal policy satisfying (2) exists if a stable policy incurring a finite cost exists (see for instance Weber and Stidham [11]). Note then that a priority policy which states to produce if and only if a demand is waiting is equivalent to a multi-class priority queue. Since $\rho < 1$ this system is stable and an optimal policy exists.

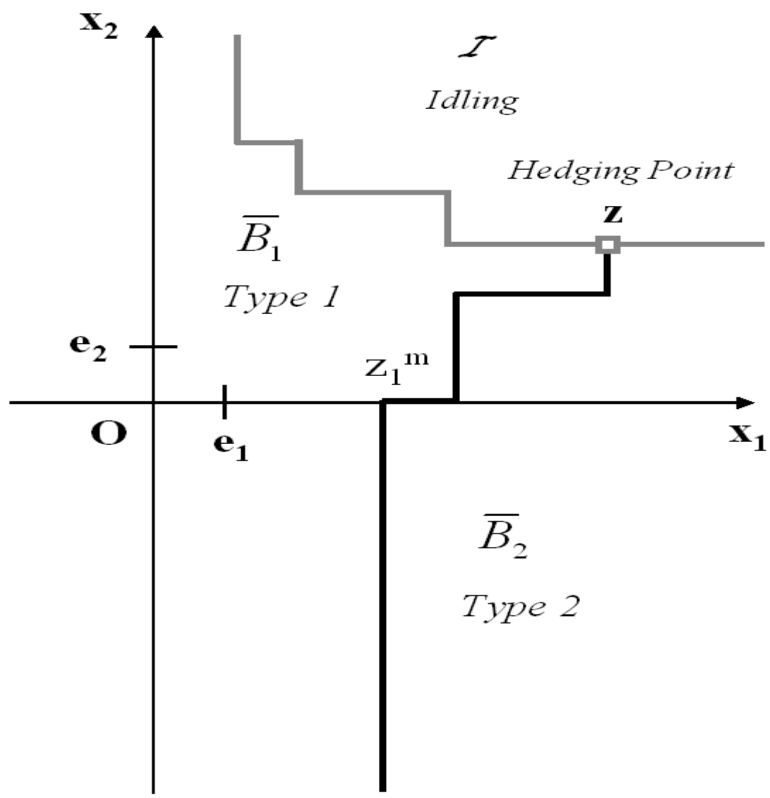

Figure 1: General Structure of an Optimal Policy

3 Structure of the Optimal Policy

In this section we present some properties for the optimal policies that will be needed to establish zero-inventory conditions. They are based on the conjecture that the optimal policy has a hedging point form. This form has been shown by Ha [3] when $\mu_1 = \mu_2$, and has been extensively numerically verified when $\mu_1 \neq \mu_2$.

A hedging point policy is characterized by a hedging point and a switching curve (Figure 1). The hedging point $\mathbf{z} = (z_1, z_2) \geq 0$ determines the on-hand inventory levels from which the machine does not produce. The switching curve splits the state space in two regions, according to which type is prefered. More precisely, following Veatch and Wein [9], we denote by $\mathcal{I}$ the set of states in which the control policy idles the machine. Let $\bar{\mathcal{B}}_i$ be the set of states in which the preferred part type for production is part i regardless of the idling decision: $\bar{\mathcal{B}}_i = \{\mathbf{x} : \min_{j=1,2} \mu_j \Delta_j v(\mathbf{x}) = \mu_i \Delta_i v(\mathbf{x})\}$. Then a hedging point policy is such that $\bar{\mathcal{B}}_1 = \{\mathbf{x} : x_1 < s_B(x_2)\}$ for an increasing curve $s_B(x_2)$ where $-\infty \leq s_B(x_2) \leq \infty$ and $\mathcal{I} = \{\mathbf{x} : x_1 < s_I(x_2)\}$ for some decreasing curve, $s_I(x_2)$, where $0 \leq s_I(x_2) \leq \infty$. The hedging point $\mathbf{z}$ is the unique state such that $s_B(z_2) = s_I(z_2) = z_1$. Thus, $\bar{\mathcal{B}}_2$ includes the states on the switching curve and region $\mathcal{I}$ includes the idling curve and the hedging point. Because states with $x_1 > z_1$ or $x_1 > z_2$ are transient, the stationary behavior of a hedging point policy is entirely characterized by its hedging point and the portion of its switching curve $s_B(x_2)$ with $x_2 \leq z_2$.

We will use the following result to compare the expectation of $W = X_1/\mu_1 + X_2/\mu_2$, the aggregated workload of the system, for different policies. Property 1 follows from the underlying nature of hedging point policies; see Lemma 2 of de Véricourt et al. [10].

Property 1 *Consider two hedging point policies* π_1 *and* π_2 *and their respective hedging points* $\mathbf{z}^{\pi_1}, \mathbf{z}^{\pi_2}$. *Then,*

$$E^{\pi_1}[W] - E^{\pi_2}[W] = \frac{z_1^{\pi_1}}{\mu_1} + \frac{z_2^{\pi_1}}{\mu_2} - \frac{z_1^{\pi_2}}{\mu_1} - \frac{z_2^{\pi_2}}{\mu_2}$$

where E^{π_i} *is the the steady-state expectation given the hedging point policy* π_i, $i \in \{1, 2\}$.

We will also use the fact that, given $\mu_1 b_1 > \mu_2 b_2$, the optimal switching curve in the region $x_2 < 0$ is a vertical line whose position is expressed by a simple equation (see Figure 1). Applying Theorem 1 of de Véricourt and al. [10], the optimal hedging point policy satisfies

$$C^{\pi^*}(\mathbf{x}) = \begin{cases} 1 & \text{if } x_1 < z_1^m, x_2 < 0 \\ 2 & \text{if } x_1 \geq z_1^m, x_2 < 0, \end{cases}$$

$$z_1^m = \left\lfloor \frac{\ln\left(\frac{h_1 + b_2 \frac{\mu_2}{\mu_1}}{h_1 + b_1}\right)}{\ln \rho_1} \right\rfloor \tag{3}$$

and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

4 Zero-inventory Conditions

When $x_2 \geq 0$, no general equation for the switching curve is known. Nevertheless, if the optimal hedging point is at $\mathbf{0}$, the optimal switching curve is fully characterized by the straight line $x_1 = 0$. The system is then in a make-to-order mode. We denote by policy ζ, this zero-inventory policy whose hedging point is at $\mathbf{0}$. The associated control is defined by

$$C^{\zeta}(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad \mathbf{x} = \mathbf{0} \\ 1 & \text{if} \quad x_1 < 0 \\ 2 & \text{if} \quad x_1 = 0, x_2 < 0 \end{cases} \tag{4}$$

We present four conditions that must hold for ζ to be optimal.

The first condition comes from the partial characterization of the optimal switching curve. If policy ζ is optimal then z_1^m must be equal to zero, and we obtain from (3),

Condition 1

$$h_1\mu_1 + b_2\mu_2 > (h_1 + b_1)\lambda_1$$

To derive the other conditions, we evaluate the effect of three policy perturbations depicted in Figure 2. If policy ζ is optimal, these perturbations increase the average cost.

4.1 Right Shift

We consider the policy π defined by a right shift of policy ζ. The hedging point of policy π is $(1, 0)$ and the switching curve is the line $x_1 = 1$. It is clear that $E^{\pi}[b_2X_2^-] = E^{\zeta}[b_2X_2^-]$; the only difference in average cost is due to X_1. The control of the first part type corresponds to a hedging point policy in a single part-type system, where the hedging point is one for policy π and zero for policy ζ, so $P^{\pi}(X_1 = x_1) = P^{\zeta}(X_1 = x_1 + 1)$ for $x_1 \leq 0$. The incremental average cost is

$$\begin{aligned} g^{\pi} - g^{\zeta} &= h_1P^{\pi}(X_1 = 1) - b_1P^{\pi}(X_1 \leq 0) \\ &= h_1(1 - \rho_1) - b_1\rho_1. \end{aligned} \tag{5}$$

If policy ζ is optimal, then (5) must be nonnegative, which is the second condition:

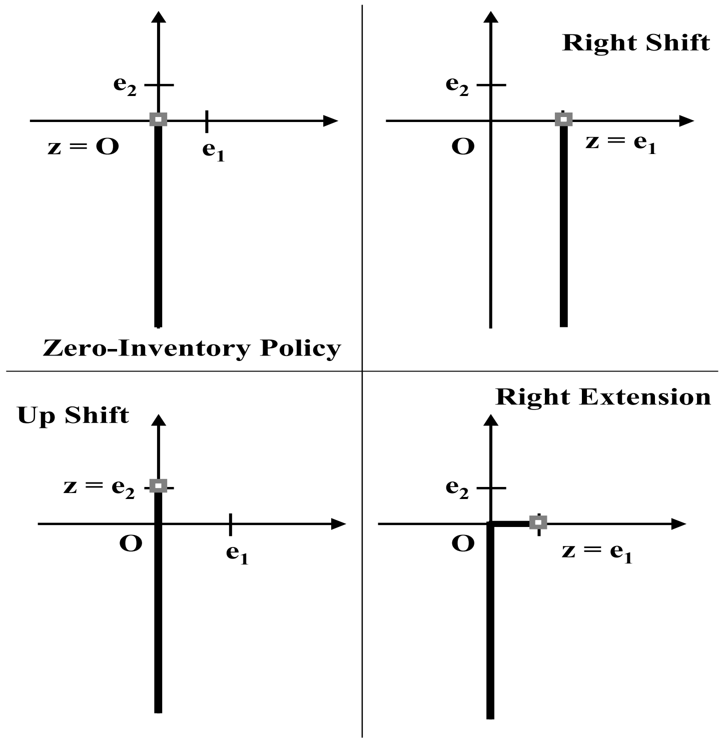

Figure 2: The Hedging Point Perturbations

Condition 2

$$\rho_1 \leq \frac{h_1}{h_1 + b_1}$$

4.2 Up Shift

Here policy π is defined by a shift up of policy ζ. The hedging point of policy π is $(0, 1)$ and the switching curve is the line $x_1 = 0$. For this perturbation the only difference in average cost is due to X_2. Let $\gamma_2 = P^\pi(X_2 = 1)$. Then

$$g^\pi - g^\zeta = h_2\gamma_2 - b_2(1 - \gamma_2) \tag{6}$$

Now, $z_2^\pi - X_2$ has the same probability law as the number of low priority customers in a priority queue. Hence, γ_2 is the stationary probability of no low priority customer in an M/M/1 priority queue (with our parameters). Expression (36) for γ_2 is obtained in the Appendix. Again, if policy ζ is optimal, then (6) is nonnegative:

Condition 3

$$1 - \gamma_2 \leq \frac{h_2}{h_2 + b_2}$$

4.3 Right Extension

Finally, we assume that policy π is defined by a right extension of policy ζ. The hedging point of policy π is state $\mathbf{e}_1$ and the switching curve is the straight line $x_1 = 0$ extended by the segment $[\mathbf{0}, \mathbf{e}_1]$. The associated control is then given by

$$C^{\pi}(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad \mathbf{x} = \mathbf{e}_1 \\ 1 & \text{if} \quad x_1 < 0 \qquad \text{or} \qquad \mathbf{x} = \mathbf{0} \\ 2 & \text{if} \quad 0 \leq x_1 \leq 1 \quad \text{and} \quad x_2 < 0 \end{cases} \tag{7}$$

Let

$$\tilde{c}(x_1) = (h_1 + \frac{\mu_2}{\mu_1} b_2) x_1^+ + (b_1 - \frac{\mu_2}{\mu_1} b_2) x_1^- \tag{8}$$

and recall that $w = x_1/\mu_1 + x_2/\mu_2$. Then, for $x_2 \leq 0$,

$$\begin{aligned} c(x) &= h_1 x_1^+ + b_1 x_1^- - b_2 x_2 \\ &= (h_1 + \frac{\mu_2}{\mu_1} b_2) x_1^+ + (b_1 - \frac{\mu_2}{\mu_1} b_2) x_1^- - \frac{\mu_2}{\mu_1} b_2 x_1 - b_2 x_2 \\ &= \tilde{c}(x_1) - \mu_2 b_2 w \end{aligned} \tag{9}$$

and

$$g^{\pi} - g^{\zeta} = E^{\pi}[\tilde{c}(X_1)] - E^{\zeta}[\tilde{c}(X_1)] - \mu_2 b_2 (E^{\pi}[W] - E^{\zeta}[W]). \tag{10}$$

From Property 1,

$$E^{\pi}[W] - E^{\zeta}[W] = \frac{1}{\mu_1}. \tag{11}$$

Now, we only need the marginal distribution of X_1 under policies π and ζ. Observe that, under π, the X_1 transition intensities are independent of X_2 except those between $X_1 = 0$ and $X_1 = 1$. Thus, the marginal distribution of X_1 obeys the same balance equations under π and ζ in $\{X_1 \leq 0\}$ and the probabilities are the same up to normalization:

$$P^{\pi}(X_1 = x_1 | X_1 \leq 0) = P^{\zeta}(X_1 = x_1), x_1 \leq 0. \tag{12}$$

Using (8) and (12), and the exponential assumptions,

$$\begin{aligned} E^{\pi}[\tilde{c}(X_1)] - E^{\zeta}[\tilde{c}(X_1)] &= (h_1 + \frac{\mu_2}{\mu_1} b_2) P^{\pi}(X_1 = 1) \\ &\quad + (b_1 - \frac{\mu_2}{\mu_1} b_2) E^{\zeta}[X_1^-](P^{\pi}(X_1 < 1) - 1) \\ &= \left(h_1 - \frac{\rho_1 b_1 - \frac{\mu_2}{\mu_1} b_2}{1 - \rho_1} \right) P^{\pi}(X_1 = 1). \end{aligned} \tag{13}$$

Let $\gamma_2^{'} = P^{\pi}(X_1 = 1)$. Next we show that $\gamma_2^{'}$ is equal to the probability of having no class 1 customer in a priority queue where the high priority is given to class 2. Consider $\mathbf{X}^{\pi}$, a trajectory generated by policy π.

Observe that $\mathbf{e}_1 - \mathbf{X}^{\pi}$ is the trajectory of a queue with class 2 having high priority except that class 1 is served first in states where $x_1 \leq 0$ and $x_2 \leq -1$. A queue that gives priority to class 2 can be constructed by changing the order of service so that class 2 is served first in these states. This resequencing does not change the times at which the $\mathbf{e}_1 - \mathbf{X}^{\pi}$ trajectory reaches $\mathbf{0}$; hence, it does not change the times at which the $\mathbf{X}^{\pi}$ trajectory reaches and leaves states with $x_1 = 1$. Once again, if policy ζ is optimal then (10) is nonnegative. Using (11), (13) and (37) in the Appendix for $\gamma_2^{'}$, the last condition is

Condition 4

$$\left(h_1 - \frac{\rho_1 b_1 - \frac{\mu_2}{\mu_1} b_2}{1 - \rho_1} \right) \gamma_2^{'} - \frac{\mu_2}{\mu_1} b_2 \geq 0$$

Yee and Veatch [13] suggest that condition 3 will be the same for n part types but there will be additional complex conditions in the place of Condition 4.

5 Sufficient Conditions

In the previous section, we derived four necessary zero-inventory conditions. Here, we outline an argument that conditions 3 and 4 are also sufficient. We first show that conditions 1 and 2 are redundant:

Property 2 *Condition 4 ⇒ Condition 2 ⇒ Condition 1*

Proof: The stationary probability of no low priority customer in a priority queue is less than the probability of the same event when the high priority class has been withdrawn. In particular, $\gamma_2^{'} \leq 1 - \rho_1$. Combining this bound with Condition 4

$$h_1 - \frac{\rho_1 b_1}{1-\rho_1} \geq \frac{\mu_2}{\mu_1} b_2 \left(\frac{1}{\gamma_2^{'}} - \frac{1}{1-\rho_1} \right) \geq 0$$

and Condition 2 is verified. Furthermore, Condition 2 can be written $h_1\mu_1 \geq (h_1 + b_1)\lambda_1$, which implies Condition 1. □

Suppose conditions 3 and 4 are satisfied. From Property 2, Condition 1 is also satisfied and the optimal switching curve is given by $x_1 = 0$, in the region $x_2 < 0$. Furthermore, Condition 3 and 4 imply that an up or right extension of the switching curve increases the average cost. Following Veatch and Caramanis [8], we consider the class of policies π using the optimal switching curve but various hedging points along this curve. It can be shown that g^π decreases monotonically as the hedging point moves toward the optimal hedging point. Hence, Condition 3 and 4 imply that the zero-inventory policy is optimal in this class of policies, and in fact optimal among all policies.

6 General Symmetric Production Times

Until now, we have considered an exponential production time. However, some of the previous conditions can be extended when an arbitrary distribution is assumed. We consider here the same system as defined in Section 2 except that the production time does not depend on the part type and we only specify its mean $1/\mu$. We also assume that the type of a part is chosen at the end of its production. In this case, the optimal control is not Markovian any more, and can depend on the time. We restrict our study to the class of hedging point policies (which may not be optimal).

Property 1 still holds for this system. To see this, note that, for $i \in \{1, 2\}$, $z_1^{\pi_i} + z_2^{\pi_i} - X_1 - X_2$ corresponds to the number of customers in an M/G/1 queue.

If we compute γ_2 for a priority M/G/1 queue, conditions 2 and 3 still hold. However, Equation (12) of Section 4.3 is not true in general, and the derivation of Condition 4 cannot be extended directly. In the following we propose to derive this last condition using coupled trajectories.

We consider two trajectories $\mathbf{X}^\pi$ and $\mathbf{X}^\zeta$ generated by policy π of Section 4.3 and policy ζ. The starting points of the trajectories are their respective hedging points. We consider then sequences of realizations of the arrival time of type 1 and 2 demands, $t_1^1, t_2^1, \ldots, t_k^1, \ldots$, and $t_1^2, t_2^2, \ldots, t_k^2, \ldots$ respectively, and of realizations of the production time $t_1^p, t_2^p, \ldots, t_k^p, \ldots$. We couple $\mathbf{X}^\pi$ and $\mathbf{X}^\zeta$ by assuming that they are generated by these common realizations of the underlying random variables.

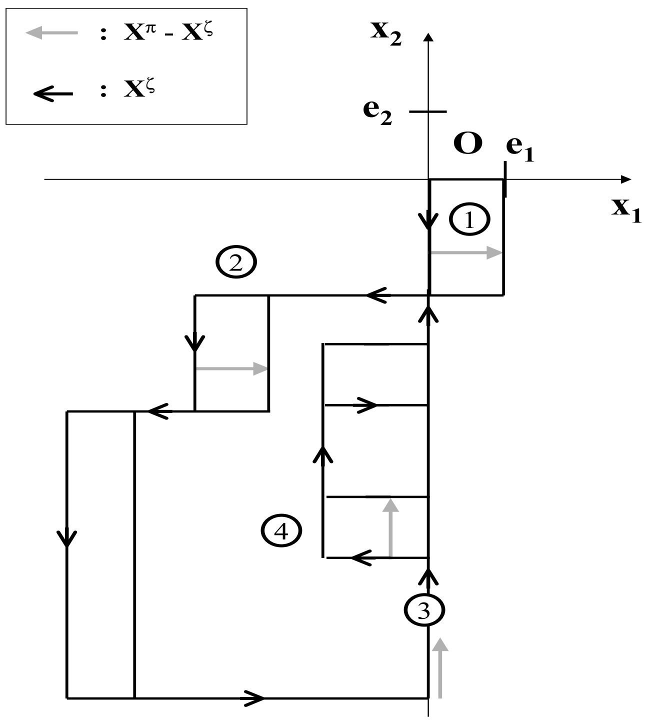

Figure 3: Different Segments of the Coupled Trajectories for the Right Extension

The coupled trajectories can be divided into 4 segments(see Figure 3):

- Segment 1:

  In this segment, $X_1^\zeta = 0$, $\mathbf{X}^\pi - \mathbf{X}^\zeta = \mathbf{e}_1$ (so that $X_1^\pi = 1$). Both policies produce type 2 when $X_2^\pi < 0$, and idle when $X_2^\pi = 0$. As a result, the coupled trajectories remain in this segment when type 1 demands arrive or production times are completed. However, when a type 2 demand arrives to the system, the coupled trajectories move into Segment 2.

- Segment 2:

  In this segment, $X_1^\zeta < 0$, and $\mathbf{X}^\pi - \mathbf{X}^\zeta = \mathbf{e}_1$. If $X_1^\pi < 0$, then both policies produce type 1 and the trajectories stay in Segment 2. If $X_1^\pi = 0$ and $X_2^\pi = 0$ both policies still produce type 1, but when a production time is complete, the trajectories move into Segment 1. If $X_1^\pi = 0$ and $X_2^\pi < 0$, then $C_1^\zeta = 1$ and $C_1^\pi = 2$, and when a production time is completed, the trajectories move into Segment 3.

- Segment 3: In this segment, $X_1^\zeta = 0$, $\mathbf{X}^\pi - \mathbf{X}^\zeta = \mathbf{e}_2$ (so that we also have $X_1^\pi = 0$). Both policies produce type 2 if $X_2^\pi < 0$. If $X_2^\pi = 0$, $C_1^\zeta = 2$ and $C_1^\pi = 1$, and when a production time is completed, the trajectories move into Segment 1, at their hedging points. When a type 1 demand arrives to the system, the coupled trajectories move into Segment 4. Otherwise, they stay in Segment 3.

- Segment 4:

  In this segment, $X_1^\zeta < 0$, $\mathbf{X}^\pi - \mathbf{X}^\zeta = \mathbf{e}_2$. Both policies state to produce type 1, so that both trajectories stay in Segment 4 until a service completion makes them enter Segment 3.

From the previous definition of the four segments, we have

$$\mathbf{X}^\pi - \mathbf{X}^\zeta = \begin{cases} \mathbf{e}_1 & \text{in Segment 1 and 2} \\ \mathbf{e}_2 & \text{in Segment 3 and 4} \end{cases} \tag{14}$$

$$c(\mathbf{X}^\pi) - c(\mathbf{X}^\zeta) = \begin{cases} h_1 & \text{in Segment 1} \\ -b_1 & \text{in Segment 2} \\ -b_2 & \text{in Segment 3 and 4} \end{cases} \tag{15}$$

If we denote by p_1, p_2, p_3 and p_4 the stationary probabilities of being in Segment 1, 2, 3 and 4 respectively, we obtain from (15)

$$g^\pi - g^\zeta = h_1 p_1 - b_1 p_2 - b_2(p_3 + p_4). \tag{16}$$

It remains then to derive expressions for p_1, p_2, p_3 and p_4.

First, it can be shown that p_1 is the probability of having no class 1 customer in a priority queue where the high priority is given to class 2 (see Section 4.3). It follows that

$$p_1 = \gamma_2^{'}, \tag{17}$$

where $\gamma_2^{'}$ is given by (37) of the Appendix.

Furthermore, Segments 1 and 3 are the segments of trajectory $\mathbf{X}^{\zeta}$ where $X_1^{\zeta} = 0$. Hence we have $p_1 + p_3 = P(X_1^{\zeta} = 0) = 1 - \rho_1$ and we obtain

$$p_3 = 1 - \rho_1 - \gamma_2^{'}. \tag{18}$$

We then evaluate p_2 and p_4 using the following two equations

$$\begin{aligned} p_1 + p_2 + p_3 + p_4 &= 1 && (19) \\ \frac{p_3}{p_1} &= \frac{p_4}{p_2}. && (20) \end{aligned}$$

The first equation is straightforward. To show the second one, we follow Veatch and Caramanis [8]. Let $N_i(t)$ be the number of type 1 demand arrivals in $(0, t]$ occuring while trajectories are in Segment 1,2, 3 and 4 respectively. Let $T_i(t)$ be the time in $(0, t]$ spent in Segment i. Using a weak law of large numbers, one can show that the proportion $N_3(t)/N_1(t)$ approaches $T_3(t)/T_1(t)$ when $t \to \infty$ with probability 1. Furthermore, each arrival of type 1 demand makes the trajectories move from Segment 1 into Segment 2, or from Segment 3 into Segment 4. It follows that $T_4(t)/T_2(t)$ and $N_3(t)/N_1(t)$ approach the same limit when $t \to \infty$. Since $T_3(t)/T_1(t)$ and $T_4(t)/T_2(t)$ also approach p_3/p_1 and p_4/p_2 respectively, we obtain Equation (20).
It follows from (17), (18), (19) and (20) that

$$\begin{aligned} p_2 &= \frac{\rho_1}{1 - \rho_1} \gamma_2^{'} && (21) \\ p_4 &= \rho_1 - \frac{\rho_1}{1 - \rho_1} \gamma_2^{'} && (22) \end{aligned}$$

Combining (17), (21), (18) and (22) with (16) we can derive the difference of average costs

$$g^{\pi} - g^{\zeta} = h_1 \gamma_2^{'} - b_1 \frac{\rho_1}{1 - \rho_1} \gamma_2^{'} - b_2 (1 - \frac{\gamma_2^{'}}{1 - \rho_1}). \tag{23}$$

This leads to a direct extension of Condition 4 to the symmetrical non-exponential case. Hence, using (33) and (34) of the Appendix to compute γ_2 and $\gamma_2^{'}$, Conditions 2-4 for the zero-inventory policy to be optimal, become:

$$\rho_1 \leq \frac{h_1}{h_1 + b_1} \tag{24}$$

$$\frac{1}{\lambda_2}[\rho\lambda - \lambda_1 + (1-\rho)\lambda_1\sigma_{BP1}(\lambda_2)] \leq \frac{h_2}{h_2 + b_2} \tag{25}$$

$$\frac{1-\rho}{\lambda_1}(h_1 - \frac{\rho_1 b_1 - b_2}{1-\rho_1})[\lambda - \lambda_2\sigma_{BP2}(\lambda_1)] \geq b_2 \tag{26}$$

with σ_{BPi}, the Laplace transform of the high priority busy period in a M/G/1 queue, where class i has the high priority. Following the proof of Property 2, one can also show that (26) implies (24) .

As mentioned earlier, the optimal policy is not Markovian, and since (25) and (26) consider hedging point policy perturbations, these conditions are not sufficient. It may be possible to prove the sufficiency if we restrict the study to the class of hedging point policies.

7 Numerical results and insights

In this section we explore the zero-inventory conditions numerically and draw some insights. Our results suggest additional conditions under which make-to-order (MTO) is attractive. For a single part-type, it is well known that when utilization is low or holding costs are large relative to backorder costs, no inventory should be held (Condition 2). Condition 3 is similar, but uses costs for the low priority part type and replaces the combined utilization (which the low priority part sees) with the smaller $1 - \gamma_2$. Condition 4 captures the interaction between part types. To examine when Condition 4 applies, we first note that it holds asymptotically as $h_1/b_2 \to \infty$. Even for h_1 of comparable size to h_2, it may hold if $1 - \gamma_2^{'}$ is small (suggesting that the utilization is low), and either the discrepancy between part type priorities ($b_1\mu_1 - b_2\mu_2$) is small or the high priority part type has low utilization ($\rho_1 \ll 1$).

Further insight is gained by considering the case where part types differ only in cost. Setting $\mu_1 = \mu_2$, Condition 4 can be written

$$\gamma_2^{'} \geq \frac{b_2}{h_1 + b_1 - \frac{b_1 - b_2}{1-\rho_1}} \tag{27}$$

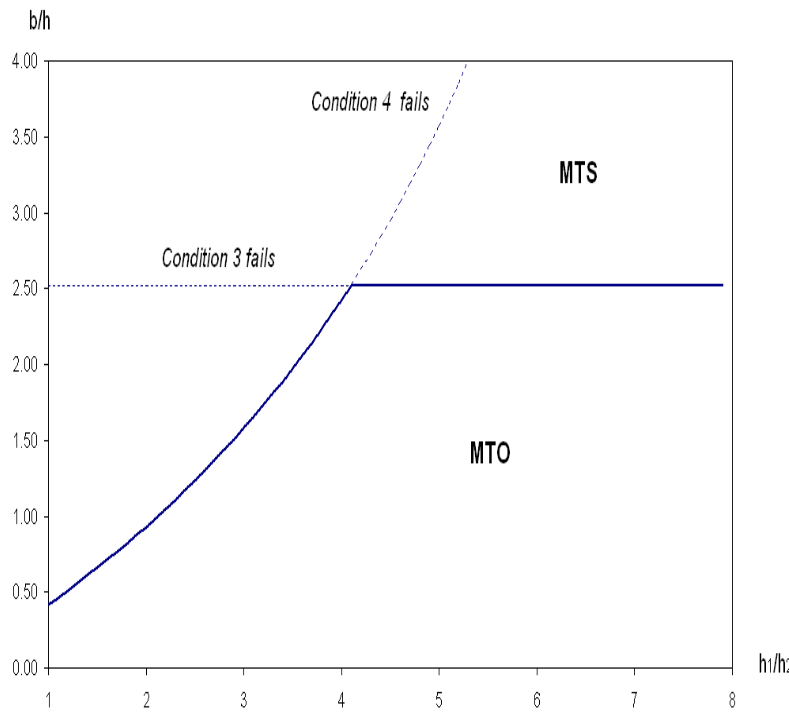

Figure 4: Optimality of the Zero-Inventory Policy

Typically $\rho_1 \ll 1$ when these conditions are met, so we will weaken (27) slightly by replacing ρ_1 by 0, giving

$$1 - \gamma_2^{'} \leq \frac{h_1}{h_1 + b_2} \tag{28}$$

Also setting $\lambda_1 = \lambda_2$, we have $\gamma_2 = \gamma_2^{'}$. Thus, Condition 4 is similar to Condition 3 but compares part types: the high priority holding cost must be large relative to the low priority backorder cost.

Specific parameter values where MTO is optimal are shown in Figure 4. In this example, the total utilization rate of the system is equal to 0.4. Part types are symmetrical in demand and production rates. We set $h_2 = 1$ and $b_1/h_1 = b_2/h_2 = b/h$ and vary b/h and h_1/h_2. Increasing b/h makes the system switches from MTO to make-to-stock (MTS). When part types are similar (h_1/h_2 close to 1), the system switch to MTS only when Condition

4 fails. This results in producing some stock in advance for part type one when there are no backlogs in the system, while maintaining a zero-inventory policy for part type 2. On the other hand, when h_1/h_2 is larger than 4.1 the switch to MTS occurs when Condition 3 fails, and results in building a safety stock for part type two.

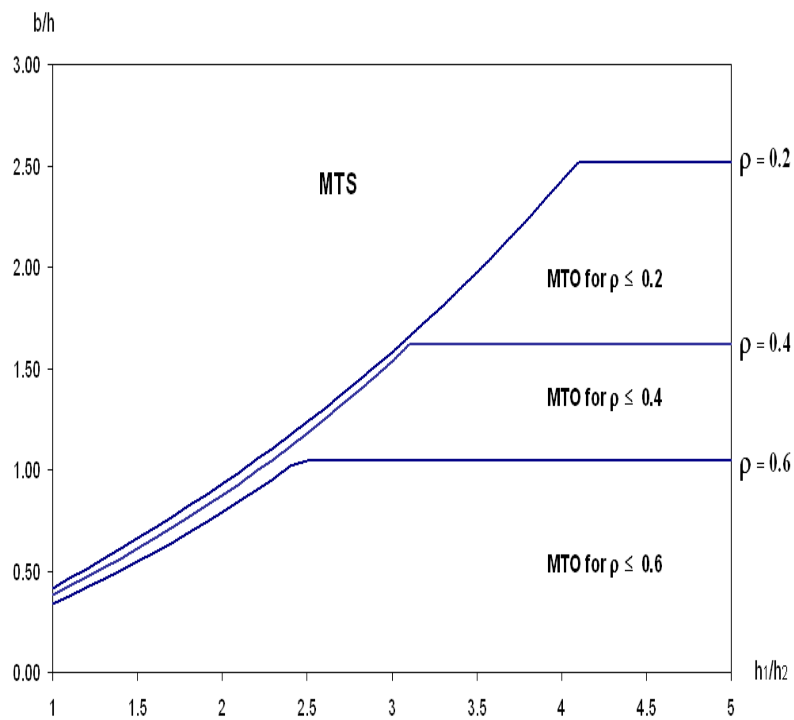

Figure 5: Optimality of the Zero-Inventory Policy for Different Utilization Rates

Figure 5 presents similar results for different values of the utilization rate. As ρ increases, the region where a zero-inventory policy is optimal gets smaller. The maximum value of h_1/h_2 for which a switch in production mode results in holding parts of type one (that is, when Condition 4 fails) also decreases. In other words, for make-to-order systems with moderate to high utilization increasing b/h precipitates holding safety stock for the less important part type in terms of backorder cost.

In practice, backorder costs are very difficult to measure and demand may be difficult to predict. It may be helpful to check Conditions 3 and 4 for a range of values of these parameters to determine if a product is in or near the MTO regime.

8 Conclusion

We have presented necessary and sufficient conditions for a simple make-to-order policy to be optimal in a two part type production system. Under these conditions, the optimal control policy is fully characterized.

Further research need to be carried out to address the case of more than two part types. Another relevant problem for practical applications is to know when one of the part types is made to order, while some stock of the other type is built in advance. Our numerical study provides some initial insights into this question. Conditions similar to ours for this problem would help understand how to choose part types requiring safety stocks when the production capacity is shared.

Appendix

A Probability of No Low Priority Customer

In this section, we derive the stationary probability of having no low priority customers in a priority M/G/1 queue with two classes of customers and preemptive resume. Customers of class 1 and 2 arrive according to Poisson processes with parameters λ_1 and λ_2, respectively. Class 1 has the highest priority. The service times have arbitrary distributions with mean $1/\mu_1$ and $1/\mu_2$. Let $\gamma_2 = P(X_2 = 0)$, $\lambda = \lambda_1 + \lambda_2$ and $\rho = \lambda_1/\mu_1 + \lambda_2/\mu_2$.

From Corollary 1 of Keilson and Servi [7] the pgf of the number of class 2 customers in the system is:

$$\pi_{S_2}(u) = \frac{(1-\tilde{\rho}_2)\alpha_{\tilde{T}_2}(\lambda_2 - \lambda_2 u)}{1 - \tilde{\rho}_2 \alpha^*_{\tilde{T}_2}(\lambda_2 - \lambda_2 u)} \pi_{B_2}(u), \tag{29}$$

where

π_{B_2}		is the pgf of the number of class 2 customers in the system given that no class 2 customer is in service
α_{T_2}		is the Laplace transform of T_2, the service time of class 2
σ_{BP1}		is the Laplace transform of the class 1 busy period
$\alpha_{\tilde{T}_2}$		is the Laplace transform of $\tilde{T}_2$, the time from the beginning of service of class 2 until the customer leaves the system, $\alpha_{\tilde{T}_2} = \alpha_{T_2}(s + \lambda_1 - \lambda_1 \sigma_{BP1}(s))$
$\tilde{\rho}_2$	$=$	$\lambda_2 E[\tilde{T}_2]$
$\alpha^*_{\tilde{T}_2}(s)$	$=$	$(1 - \alpha_{\tilde{T}_2}(\lambda_2))/(sE[\tilde{T}_2])$.

We will use the fact that

$$\gamma_2 = \pi_{S_2}(0) = (1-\tilde{\rho}_2)\frac{\alpha_{\tilde{T}_2}(\lambda_2)}{1 - \tilde{\rho}_2 \alpha^*_{\tilde{T}_2}(\lambda_2)} \pi_{B_2}(0). \tag{30}$$

But from Keilson and Servi [7], we also have,

$$\pi_{B_2}(0) = \frac{1-\rho_1}{\lambda_2}[\lambda - \lambda_1 \sigma_{BP1}(\lambda_2)] \tag{31}$$

$$\tilde{\rho}_2 = \frac{\rho_2}{1-\rho_1}. \tag{32}$$

Assembling (30), (31) and (32), we finally obtain

$$\gamma_2 = \frac{1-\rho}{\lambda_2}[\lambda - \lambda_1 \sigma_{BP1}(\lambda_2)]. \tag{33}$$

For a queue where class 2 has the high priority, the probability $\gamma_2^{'}$ of having no class 1 customer in the system can then directly be obtained from (33):

$$\gamma_2^{'} = \frac{1-\rho}{\lambda_1}[\lambda - \lambda_2 \sigma_{BP2}(\lambda_1)] \tag{34}$$

where σ_{BP2} is the Laplace transform of the class 2 busy period, when class 2 has the high priority.

B Case of the M/M/1

If the service time is exponentially distributed, we have from Gross and Harris [2],

$$\sigma_{BP1}(s) = \frac{2\mu_1}{\lambda_1 + \mu_1 + s + \sqrt{(\lambda_1 + \mu_1 + s)^2 - 4\lambda_1\mu_1}}. \tag{35}$$

From (33), (34) and (35), we obtain the following expressions for γ_2 and $\gamma_2^{'}$

$$\gamma_2 = \frac{1-\rho}{\lambda_2}\left[\lambda - \frac{2\lambda_1\mu_1}{\lambda + \mu_1 + \sqrt{(\lambda + \mu_1)^2 - 4\lambda_1\mu_1}}\right] \tag{36}$$

$$\gamma_2^{'} = \frac{1-\rho}{\lambda_1}\left[\lambda - \frac{2\lambda_2\mu_2}{\lambda + \mu_2 + \sqrt{(\lambda + \mu_2)^2 - 4\lambda_2\mu_2}}\right]. \tag{37}$$

References

[1] T. Bielecki, and P.R Kumar, Optimality of Zero-inventory Policies for Unreliable Manufacturing Systems, Operations Research 36 (1988) 532-546.

[2] D. Gross and C. M. Harris, Fundamentals of Queueing Theory, John Wiley and Sons (1985).

[3] A. Ha, Optimal Dynamic Scheduling Policy for a Make-to-Stock Production System, Operations Research 45 (1997) 42-53.

[4] A. Peña-Perez and P. Zipkin, Dynamic Scheduling Rules For a Multi-product Make-to-Stock Queue, Operations Research 45 (1997) 919-930.

[5] J. R. Perkins and R. Srikant, Hedging Policies for Failure-Prone Manufacturing Systems: Optimality of JIT Policies and Bounds on Buffer Levels, IEEE Transactions on Automatic Control (1998) 953-958.

[6] E. Presman, S. P. Sethi, H. Zhang and Q. Zhang, Optimality of Zero-Inventory Policies for an Unreliable Manufacturing System Producing Two Part Types, Dynamics of Continuous, Discrete and Impulsive Systems 4 (1998) 485-496.

[7] J. Keilson and L. D. Servi, the Distributional Form of Little's Law and the Furhmann-Cooper Decomposition, Operations Research Letters 9 (1990) 239-247.

[8] M. H. Veatch and M. C. Caramanis, Optimal Manufacturing Flow Controllers: Zero-Inventory Policies and Control Switching Sets", IEEE Transactions on Automatic Control 44 (1999) 914-921.

[9] M. H. Veatch and L. M. Wein, Scheduling a Make-to-Stock Queue: Index Policies and Hedging Points, Operations Research 44 (1996) 634-647.

[10] F. de Véricourt, F. Karaesmen, and Y. Dallery, Dynamic Scheduling in a Make-to-Stock System: A Partial Characterization of Optimal Policies, Operations Research 48 (2000) 811-819.

[11] R.R. Weber and S. Stidham, Optimal Control of Service Rates in Networks of Queues, Advances in Applied Probability 19 (1987) 202-218.

[12] L. M. Wein, Dynamic Scheduling of a Multi-class Make-to-Stock Queue, Operations Research 40 (1992) 724-735.

[13] M. J. Yee and M. H. Veatch, Just-in-Time Policies for Manufacturing Flow Controllers, IEEE Transactions on Automatic Control 45 (2000) 336-339.