

Chapter 8

Performance Analysis of a Two-Station MTO/MTS Production System

Kuo-Hwa Chang and Yang-Shu Lu

Abstract We consider a two-station hybrid MTO/MTS production system with random ordinary and specific demands, in which the first station is a MTS system providing the finished standard products for ordinary demands. These finished products also serve as the semi-finished products to specific demands. The second station performs some additional work on the standard products for specific demands. In our system, the MTS system is controlled under the base-stock policy. To evaluate the system, we consider the fill rate of the ordinary demands and the response time of the specific demands. Our objective is to study the relation between base-stock level and the fill rate of the ordinary demands and the response time of the specific demands. We analyze our system by modeling it as an inventory-queue system. Based on these analyses, we can determine the optimal base-stock level numerically under the constraints on the fill rate of the ordinary demands and the response time of the specific demands.

8.1 Introduction

Traditionally, a production system can be distinguished into make-to-order (MTO) or make-to-stock (MTS) systems. MTO products are usually made to customer specifications as nonstandard and custom products, however, MTS products are standard and delivered from inventory (stock). That is, a MTS production stocks the finished products in advance whereas a MTO system starts producing only when it receives orders from the demand. Assembly manufacturing plays a very important role in the

K.-H. Chang

Department of Industrial Engineering, Chung Yuan Christian University, Chung-Li 320, Taiwan
e-mail: kuohwa@cycu.edu.tw

Y.-S. Lu

Department of Industrial Engineering, Chung Yuan Christian University, chung-Li 320, Taiwan
e-mail: g9402404@cycu.edu.tw

global supply chain of consumer products, such as laptop computers. Assemblers, in addition to fulfilling the ordinary demands for the standard products by adopting MTS production, are often asked to take care of the specific demands for the custom products and to adopt MTO production. In usual cases, ordinary demands are the planned orders and should be satisfied immediately, however, there is a time window for the specific demand.

To the assembler, it is not profitable to maintain a solo MTO production line exclusively for the specific demand. In some case, custom products share almost all the parts of the standard products, therefore, the assembler usually considers embedding the MTO lines into the mainstream MTS lines, which become a hybrid production system. The corresponding design and the control issues for the hybrid lines are important to management.

In this chapter, we assume the custom products can be made by alternating the existing standard ones with little work. We consider a two-station hybrid MTO/MTS production system (see Fig. 8.1) with random ordinary and specific demands, in which the first station (station 1) is a MTS system providing the finished standard products for the ordinary demands. There is a base-stock level for the finished standard products.

These standard products also serve as semi-finished products to the specific demands collected at station 2 where the additional work on the finished standard product is performed to fulfill the corresponding specific demands.

When an ordinary demand arrives at station 1, if there are finished standard products, it will take one of them and leave and, at the same time, this satisfied ordinary order will send a production order to station 1 for a new standard product; if there are no finished products, this ordinary demand will be lost. When a specific order arrives, it will send a request (order) to station 1 for acquiring a finished standard

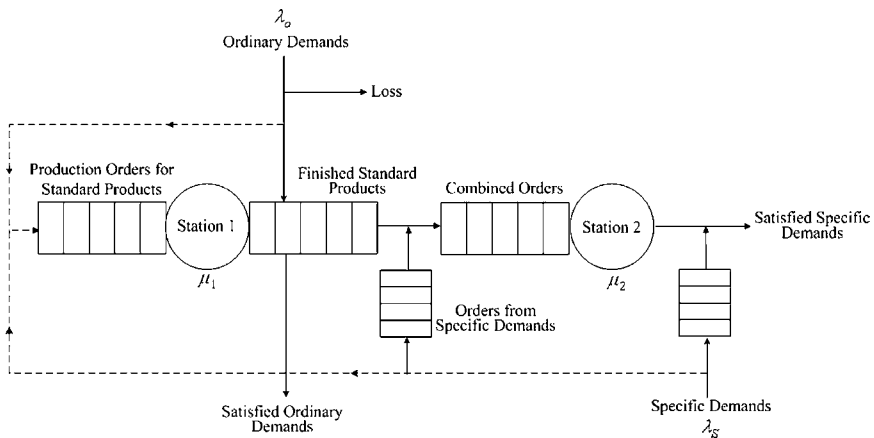


Fig. 8.1 Two-station MTO/MTS hybrid system.

product in stock and, at the same time, station 1 will also send a production order to itself for producing another new standard product; if it obtains a finished standard product, then this order along with the corresponding finished product will become a combined order and will enter station 2 on a First-Come First-Served (FCFS) basis for the additional work.

Among the research on hybrid systems, Soman, van Donk, and Gaalman [1] review the studies on the hybrid MTO/MTS production and mention that such systems can often be seen in the food industry. Krishnamurthy, Suri, and Vernon [2] use simulation to analyze a MTO/MTS hybrid system in which a base-stock controlled MTS production system supplies finished product to multiple MTO production systems. It also compares the performance of MRP and Kanban for a multistage, multiproduct manufacturing system. Adan and Ven der Wal [3] present two single-station systems.

The first system deals with MTS and MTO demands with base-stock control. Production is pre-empted by the MTO demand. The second system deals with the specific demands with base-stock control for the semi-finished products. Production is in two phases. The first phase is to produce semi-finished products and the second phase is to perform the further work on the semi-finished products in stock according to specific demands. Nguyen [4] considers a single-station hybrid production system for multiple MTS orders and multiple MTO orders. MTS orders are satisfied from the inventory controlled by base-stock policy and they are lost if there is no inventory. He models it as a mixed queueing network and approximates the performances under heavy traffic conditions by using the corresponding limiting theorem. Federgruen and Katalan [5] consider a single-station system producing some MTS products and one MTO product.

For the MTS products the base-stock policies with general periodic sequence are considered. By using an M/G/1 model with vacations, the impacts of various priority rules for the MTO products are studied. Carr and Duenyas [6] consider a single-station hybrid production system for the MTS order and MTO order. The MTS orders are satisfied from the finished-product inventory. There is no backorder for the MTS order and unsatisfied MTS orders are lost. They apply admission control on the MTO orders and sequencing on jobs at the workstation. They use the Markov decision process to find an optimal policy to maximize the average profit rate and obtain the corresponding switching curves. Arreola-Risa and DeCroix [7] consider a single-station system producing multiple products with base-stock inventory policies. They study the optimality conditions to decide which products are make-to-stock and which are make-to-order (with base-stock level zero) in order to have the minimum average cost per unit and minimum average cost rate per unit, respectively. Rajagopalan [8] also considers a single-station system for the MTS order and MTO order.

The inventory control policy for the MTS products is a (q, r) policy. Production orders for both MTS and MTO items are served on a FCFS basis. The objective is to partition the MTO/MTS items in order to minimize the inventory costs of MTS products while satisfying the constraint that the percentage of orders of MTO products fulfilled within lead time must be over a prespecified service level. The system

is modeled as an M/G/1 system. The corresponding optimization problem is modeled as a nonlinear integer program and is solved by a heuristic procedure.

To evaluate our system, we consider the fill rate (on the other side, the loss rate) of the ordinary demands and the in-time rates for the specific demands. In-time rates are defined as the probability that the waiting times of specific demands in the system, called the response times, are less than the predetermined lead time. Our system is analyzed by modeling it as an inventory-queue system. For studying the fill rate of the ordinary demands, we consider station 1 separately. We model it as an inventory queue with two classes of demands: ordinary demands and specific demands. By assuming the Markovian property, the limiting probabilities are obtained and the corresponding fill rate under base-stock control policy can also be obtained. For studying the response time for the specific demand, we study the recursive equations for approximating the response times. From these recursive equations, we can express the response times from their preceding demands and, furthermore, we can estimate the approximated distribution of the response time of specific demands. Combining the above analyses, we can further determine the optimal base-stock level under the constraints on the fill rate of the ordinary demands and the in-time rates for the specific demands according to some cost structure. We call the requirements on the fill rate for ordinary demands and the in-time rate for specific demands the corresponding required qualities of services.

The remainder of this chapter is organized as follows. In [Sect. 8.2](#), we present the inventory-queue model of our hybrid system. Our model is analyzed and the closed-form expressions for the fill rate and the distribution of the response times are obtained. In [Sect. 8.3](#), we verify our approximations obtained in [Sect. 8.2](#) and present some numerical examples. We conclude our study in [Sect. 8.4](#).

8.2 Model Description

We consider a two-station hybrid production system in which the ordinary demands arrive at station 1 according to a Poisson process with rate λ_o and specific demands arrive at station 2 according to a Poisson process with λ_s . We assume the exponential service times at each station with respective rates μ_1 and μ_2 . Station 1 (MTS system) is controlled under the base-stock policy with base-stock level S .

Let B_1 be the number of production orders for standard products in the queue or under processing at station 1; B_2 be the number of orders from specific demands at the end of station 1; B_3 be the number of specific demands waiting in the demand queue at the end of station 2; N_1 be the number of finished standard products in stock at station 1; and N_2 be the number of combined orders in the queue or under processing at station 2. Note that only one of N_1 and B_2 can be positive. We have the following relations.

Proposition 8.1.

$$B_3 = B_2 + N_2, \quad (8.1)$$

$$B_1 + N_1 - B_2 = S. \quad (8.2)$$

Proof. Relation (8.1) is well-known for a base-stock system. In fact, (8.2) is true for a base-stock system with only one kind of demand. Before we prove (8.2) by induction, we discuss the changes on B_1 , B_2 , and N_2 after any state transition. We first consider the case when $N_1 > 0$ ($B_2 = 0$). In this case, if a demand arrives, whether it is a specific demand or an ordinary demand, N_1 will be decreased by 1 and B_1 will be increased by 1, however, B_2 is still zero and (8.2) still holds; if a standard product is produced, then B_1 will be decreased by 1 but N_1 will be increased by 1.

Now we consider the case when $N_1 = 0$ ($B_2 \geq 0$). In this case, there will be no arriving ordinary demands that can be satisfied. If a specific demand arrives, both B_1 and B_2 will be increased by 1; if a standard product is produced and $B_2 = 0$ then B_1 will be decreased by 1 but N_1 will be increased by 1; if a standard product is produced and $B_2 > 0$ then B_1 will be decreased by 1 but B_2 will decrease by 1, and N_1 is still zero. All the changes mentioned above still make (8.2) hold.

We now prove (8.2) by induction on state transitions. Initially, $N_1 = S$, $B_1 = 0$, and $B_2 = 0$. After the first transition, (8.2) still holds from the assertion for the case $N_1 > 0$. Suppose that, after the k th transition (8.2) holds, then (8.2) will still hold after the $(k + 1)$ st transition based on the above assertions.

For the fill rate of the ordinary demands, we consider the subsystem corresponding to station 1. Let the state be (m, n) where m denotes the number of finished standard products at station 1 and n denotes the number of orders from specific demands in stock at the end of station 1. That is, $m = N_1$ and $n = B_2$. The possible states are actually $(m, 0)$ where $0 \leq m \leq S$ and $(0, n)$ for all $n \geq 0$. The corresponding transition rate diagram is shown in Fig. 8.2. Note that if l is the number of production orders for the standard products in the queue or under processing, then, from (8.2), we have $l = n - m + S$. Our objective here is to find the fill rate, denoted by P_f , for the ordinary demands and the corresponding effective arrival rate, denoted by λ_e , where $\lambda_e = P_f \lambda_o$.

Define $P(m, n)$ to be the limiting probability of state (m, n) ; then the balance equations are as follows:

$$\begin{aligned} (\lambda_s + \lambda_o)P(S, 0) &= \mu_1 P(S - 1, 0), \\ (\lambda_s + \lambda_o + \mu_1)P(m, 0) &= \mu_1 P(m - 1, 0) + (\lambda_s + \lambda_o)P(m + 1, 0), \quad 1 \leq m \leq S - 1, \\ (\lambda_s + \mu_1)P(0, 0) &= (\lambda_s + \lambda_o)P(1, 0) + \mu_1 P(0, 1), \\ (\lambda_s + \mu_1)P(0, n) &= \lambda_s P(0, n - 1) + \mu_1 P(0, n + 1), \quad 1 \leq n \leq \infty. \end{aligned}$$

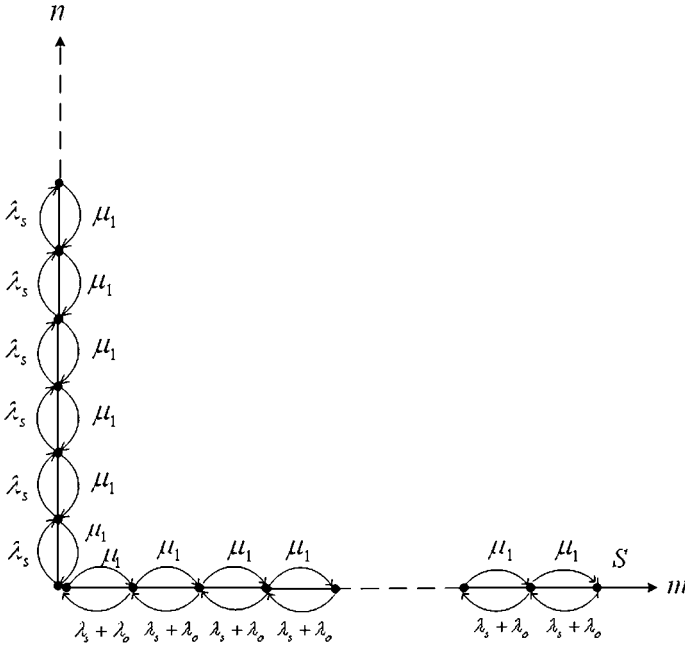


Fig. 8.2 State transition diagram for a two-station MTO/MTS hybrid system.

We have the further expressions for any $P(m, n)$.

$$P(m, 0) = \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S-m} P(S, 0), \quad 0 \leq m \leq S-1,$$

$$P(0, n) = \left(\frac{\lambda_s}{\mu_1} \right)^n \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S P(S, 0), \quad 1 \leq n \leq \infty.$$

By the law of total probabilities,

$$\sum_{m=0}^S P(m, 0) + \sum_{n=1}^{\infty} P(0, n) = 1. \tag{8.3}$$

If $\lambda_s/\mu_1 < 1$, then limiting probabilities exist and

$$P(S, 0) = \left\{ \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S+1}}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) + \left(\frac{\lambda_s}{\mu_1} \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S \right) \right\}^{-1}.$$

Let L_{N_1} denote the expected number of finished standard products in stock and L_{B_2} be the expected number of orders from specific orders at the end of station 1; then P_f , L_{N_1} , and L_{B_2} can be obtained as follows:

$$P_f = \sum_{m=1}^S P(m, 0) = \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) \left\{ \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S+1}}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) + \left(\frac{\frac{\lambda_s}{\mu_1} \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \frac{\lambda_s}{\mu_1}} \right) \right\}^{-1}, \quad (8.4)$$

$$L_{N_1} = \sum_{m=0}^S mP(m, 0) = P(S, 0) \left(1 \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S-1} + 2 \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S-2} + \dots + (S-1) \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right) + S \right) \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} - \frac{S}{\left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) = \frac{\left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} - \frac{S}{\left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right)}{\left(1 - \frac{1}{\left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) \left\{ \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S+1}}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) + \left(\frac{\frac{\lambda_s}{\mu_1} \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \frac{\lambda_s}{\mu_1}} \right) \right\}}, \quad (8.5)$$

$$L_{B_2} = \sum_{n=0}^{\infty} nP(0, n) = \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S \left\{ \frac{\left(\frac{\lambda_s}{\mu_1} \right)}{\left(1 - \left(\frac{\lambda_s}{\mu_1} \right) \right)^2} \right\} = \frac{\left\{ \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S+1}}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) + \left(\frac{\frac{\lambda_s}{\mu_1} \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \frac{\lambda_s}{\mu_1}} \right) \right\}}{\left\{ \left(\frac{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^{S+1}}{1 - \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)} \right) + \left(\frac{\frac{\lambda_s}{\mu_1} \left(\frac{\lambda_s + \lambda_o}{\mu_1} \right)^S}{1 - \frac{\lambda_s}{\mu_1}} \right) \right\}}. \quad (8.6)$$

Because we want to find the base-stock levels where the predetermined qualities of services can be satisfied, we should look at the limiting behaviors of the fill rate, L_{N_1} and L_{B_2} as S goes to infinity. We need to study them in two cases: $\lambda_s + \lambda_o < \mu_1$ and $\lambda_s + \lambda_o \geq \mu_1$. After some algebra, the limits are found and presented in the following theorem.

Proposition 8.2. As $S \rightarrow \infty$,

(a) If $\lambda_s + \lambda_o < \mu_1$, then

$$P_f \rightarrow 1, \quad (8.7)$$

$$L_{N_1} \rightarrow \infty, \quad (8.8)$$

$$L_{B_2} \rightarrow 0. \quad (8.9)$$

(b) If $\lambda_s + \lambda_o \geq \mu_1$, then

$$P_f \rightarrow \frac{\mu_1 - \lambda_s}{\lambda_o}, \quad (8.10)$$

$$L_{N_1} \rightarrow \frac{\frac{\lambda_s + \lambda_o}{\mu_1}}{\left(\frac{\lambda_s + \lambda_o}{\mu_1} - 1\right) \left(\left(\frac{\lambda_s + \lambda_o}{\mu_1}\right) + \left(\frac{\lambda_s}{\mu_1 - \lambda_s}\right) \left(\frac{\lambda_s + \lambda_o}{\mu_1} - 1\right)\right)}, \quad (8.11)$$

$$L_{B_2} \rightarrow \frac{\frac{\lambda_s}{\mu_1} \left(\frac{\lambda_s + \lambda_o}{\mu_1} - 1\right)}{\left(\left(\frac{\lambda_s + \lambda_o}{\mu_1}\right) - \left(\frac{\lambda_s}{\mu_1}\right)\right) \left(1 - \left(\frac{\lambda_s}{\mu_1}\right)\right)}. \quad (8.12)$$

Proof. The proofs for the results in part (a) are straightforward and they are omitted. For part (b), we only prove the convergence of (8.10). The proofs on the convergences of the other two can be conducted in the similar way. The term on the right side of (8.4) can be rewritten in terms of $\rho = (\lambda_o + \lambda_s)/\mu_1$ and $\rho_s = \lambda_s/\mu_1$ as follows:

$$\frac{(1 - \rho^S)/(1 - \rho)}{(1 - \rho^{S+1})/(1 - \rho) + \rho_s \rho^S/(1 - \rho_s)}. \quad (8.13)$$

After applying l'Hôpital's rule, it can be shown that (8.13) converges to

$$\frac{-1/(1 - \rho)}{-\rho/(1 - \rho) + \rho_s/(1 - \rho_s)} \quad (8.14)$$

which can be simplified to $(\mu_1 - \lambda_s)/\lambda_o$.

According to Proposition 8.2, when the capacity of the workstation is large enough to handle all the traffic, the fill rate will converge to 1 and the expected number of the order from specific demands at the end of station 1 will converge to zero as the base-stock level increases to infinity. This implies that, in the case of $\lambda_s + \lambda_o < \mu_1$, we are able to find a base-stock level to satisfy the predetermined service qualities. When the capacity of the workstation is not enough to handle all the traffic, all of these three converge to constants as we increase the base-stock level. Equation (8.10) implies λ_e converges to $\mu_1 - \lambda_s$. Note that the specific demands will eventually be served. This means that the maximal capacity that the system can offer to the ordinary demands is the residual capacity, $\mu_1 - \lambda_s$. In this case, $(\mu_1 - \lambda_s)/\lambda_o$

can be considered as the upper bound of the fill rate and it can be used to check the feasibility of the system. Also note that, in this case, although both L_{N_1} and L_{B_2} converge to constants, from (8.2) the expected number of production orders for standard products in front of station 1 will go to infinity as the base-stock increases to infinity.

Let L_{B_1} be the expected number of production orders for standard products in front of station 1. From Proposition 8.1 and 8.2 we have the following limiting results for L_{B_1} for the case $\lambda_s + \lambda_o < \mu_1$. Note that L_{B_1} will diverge when $\lambda_s + \lambda_o \geq \mu_1$.

Proposition 8.3. *If $\lambda_s + \lambda_o < \mu_1$,*

$$L_{B_1} \rightarrow \frac{(\lambda_s + \lambda_o)/\mu_1}{1 - (\lambda_s + \lambda_o)/\mu_1}, \quad \text{as } S \rightarrow \infty. \quad (8.15)$$

It is intuitive that the right term in (8.15) is the expected number of customers in the system in an M/M/1 queue because all the ordinary demands will be satisfied as the stock level becomes large and the arrival rates of the production orders from the ordinary demands will be λ_o . For studying the response time of the specific demand, we consider the case when $\lambda_s + \lambda_o < \mu_1$.

We first express the respective response times at both stations by the recursive equations. Note that when a specific order arrives, it will wait for its custom product by sending an order (request) to the inventory of station 1 for a finished product, and, at the same time, station 1 will also send a production order to itself for a standard product.

Let $\{A_n, n = 1, 2, \dots\}$ be the arrival process of the specific demands, where A_n denotes the arrival time of the n th specific demand. Let U_n be the interarrival time between the n th and $(n-1)$ st arrivals; then, by our assumptions, U_n s are i.i.d. exponential random variables with rate λ_s . Note that $\{A_n, n = 1, 2, \dots\}$ is also the arrival process of orders at the end of station 1. Let $\{A'_n, n = 1, 2, \dots\}$ be the arrival process of production orders for the standard products in front of station 1. Note that these orders can be initiated by either specific demands or satisfied ordinary demands. Let U'_n be the interarrival time between the n th and $(n-1)$ st arrivals and we approximate U'_n s as i.i.d. exponential random variables with rate $\lambda_s + \lambda_e$. Suppose that there are already d satisfied ordinary demands that left the system when the n th specific order arrives; then $A_n = A'_{n+d}$.

If the response time at station 1 of the n th specific order is positive, then it means that when the n th specific order arrives, there are no finished standard products available and it will wait for the product made by the $n+d-S$ production orders for the standard products. And, before it obtains this standard product, there will be no other ordinary demands that can be satisfied. Therefore, in this case, the response time of the n th specific demand at station 1, denoted by R_n^1 , is

$$\begin{aligned} R_n^1 &= A'_{n+d-S} + W'_{n+d-S} - A'_{n+d} \\ &= W'_{n+d-S} - \sum_{k=n+d-S+1}^{n+d} U'_k, \end{aligned} \quad (8.16)$$

where W'_n is the waiting time in the system of the n th production order for a standard product at station 1. We approximate the underlying queueing system of W'_n 's by an M/M/1 queue with arrival rate $\lambda_s + \lambda_e$ and service rate μ_1 . Also note that $\sum_{k=n+d-S+1}^{n+d} U'_n$ is distributed as a gamma distribution with parameters S and $\lambda_s + \lambda_e$. Therefore, for $t > 0$, the density function for the response time, denoted by $f_{R^1}(t)$, can be obtained as

$$\begin{aligned}
 f_{R^1}(t) &= \int_{v=0}^{\infty} \frac{(\lambda_s + \lambda_e)^S v^{S-1}}{(S-1)!} e^{-(\lambda_s + \lambda_e)v} (\mu_1 - (\lambda_s + \lambda_e)) e^{-(\mu_1 - (\lambda_s + \lambda_e))(v+t)} dv \\
 &= \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S (\mu_1 - (\lambda_s + \lambda_e)) e^{-(\mu_1 - (\lambda_s + \lambda_e))t}, \quad t > 0.
 \end{aligned}$$

Furthermore, the probability that the response time is zero, denoted by $P(R^1 = 0)$, is equal to

$$1 - \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S. \tag{8.17}$$

Let R_n^2 denote the response time of the n th specific demand at station 2 and W_n^2 denote the waiting time in system of the n th combined order at station 2; then

$$R_n^2 = R_n^1 + W_n^2.$$

The arrival process of the combined orders to station 2 is not a Poisson process, thus we consider the queueing system corresponding to the combined orders at station 2 as a GI/M/1 queue. The waiting time in system of a combined order, denoted by W^2 , has the density

$$f_{W^2}(t) = \mu_2(1 - \alpha)e^{-\mu_2(1-\alpha)t}, \quad t \geq 0,$$

where α is a solution of $\alpha = F^*(\mu_2(1 - \alpha))$ and F^* is the Laplace transform of the interarrival time of a combined order to station 2 (see Kulkarni [9]). Because departures from station 1 may be triggered by ordinary demands or specific demands and only the departing specific demands will enter station 2, we approximate the arrival process to station 2 as the departure process of an M/M/1 base-stock inventory-queue with arrival rate λ_s and service rate μ_1 . Form Buzacott, Price, and Shanthikumar [10], we have

$$\begin{aligned}
 F^*(\tau) &= (1 - (\lambda_s/\mu_1)^{S+1}) \frac{\lambda_s}{\lambda_s + \tau} + (\lambda_s/\mu_1)^{S-1} \frac{\mu_1}{\mu_1 + \tau} - (\lambda_s/\mu_1)^{S-1} \\
 &\quad \cdot (1 - (\lambda_s/\mu_1)^2) \frac{\lambda_s + \mu_1}{\lambda_s + \mu_1 + \tau}.
 \end{aligned}$$

The density of the response time of a specific demand is then

$$f_{R^2}(t) = \int_{v=0}^t f_{R^1}(v) f_{W^2}(t-v) dv + P(R^1 = 0) f_{W^2}(t).$$

We assume the independence of the response time at station 1 and the waiting time in the system at station 2. After some algebra we have, for $t > 0$,

$$\begin{aligned}
 f_{R^2}(t) &= \left(1 - \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S \left(\frac{\mu_2(1-\alpha)}{\mu_2(1-\alpha) - \mu_1 + (\lambda_s + \lambda_e)} \right) \right) \\
 &\quad \cdot (\mu_2(1-\alpha))e^{-\mu_2(1-\alpha)t} + \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S (\mu_1 - (\lambda_s + \lambda_e)) \\
 &\quad \cdot \frac{\mu_2(1-\alpha)}{\mu_2(1-\alpha) - \mu_1 + (\lambda_s + \lambda_e)} e^{-(\mu_1 - (\lambda_s + \lambda_e))t}
 \end{aligned} \tag{8.18}$$

and the expected response times of specific demands

$$\begin{aligned}
 E[R^2] &= \int_0^\infty t f_{R^2}(t) dt \\
 &= \frac{1 - \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S \left(\frac{\mu_2(1-\alpha)}{\mu_2(1-\alpha) - \mu_1 + (\lambda_s + \lambda_e)} \right)}{(\mu_2(1-\alpha))} \\
 &\quad + \frac{\left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S \left(\frac{\mu_2(1-\alpha)}{\mu_2(1-\alpha) - \mu_1 + (\lambda_s + \lambda_e)} \right)}{\mu_1 - (\lambda_s + \lambda_e)}.
 \end{aligned}$$

Finally, we have a closed form for the c.d.f of R^2 as follows:

$$\begin{aligned}
 F_{R^2}(u) &= \int_0^u f_{R^2}(t) dt \\
 &= \left(1 - \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S \left(\frac{\mu_2(1-\alpha)}{\mu_2(1-\alpha) - \mu_1 + (\lambda_s + \lambda_e)} \right) \right) \left(1 - e^{-\mu_2(1-\alpha)u} \right) \\
 &\quad + \left(\frac{\lambda_s + \lambda_e}{\mu_1} \right)^S \left(\frac{\mu_2(1-\alpha)}{\mu_2(1-\alpha) - \mu_1 + (\lambda_s + \lambda_e)} \right) \\
 &\quad \cdot \left(1 - e^{-(\mu_1 - (\lambda_s + \lambda_e))u} \right).
 \end{aligned} \tag{8.19}$$

Note that if T is our maximal lead time for a specific demand, then $F_{R^2}(T)$ will be the corresponding in-time rate.

8.3 Numerical Results

In this section, we conduct numerical studies to verify our results and we are then interested in finding an optimal base-stock level to minimize the total cost subject to the requirements on the fill rate and in-time rate. Based on our results of the

closed-form expressions for the fill rate (8.4) and in-time rate (8.19), we first verify our results by comparing our results and the results from simulations through examples (Examples 8.1 and 8.2). Note (8.4) can be expressed in terms of $\rho_s = \lambda_s/\mu_1$ and $\rho_o = \lambda_o/\mu_1$. In the following example, we test our fill rate results with the results from simulations based on various ρ_s and ρ_o .

Example 8.1. We consider four cases with $\rho_s = 0.5$ and 0.3 and $\rho_o = 0.2$ and 0.4 under three base-stock levels, 1, 2, and 3, to verify our results on the fill rate, P_f . The comparison results are shown in Table 8.1. Our results (indicated by ‘‘Approx.’’) and those obtained from simulations (indicated by ‘‘Sim.’’) are very close to each other.

In the next example, we verify our approximations (indicated by ‘‘Approx.’’) on in-time rates with the results from simulations (indicated by ‘‘Sim.’’).

Example 8.2. Let $\lambda_o = 0.05$, $\lambda_s = 0.02$, $\mu_1 = 0.1$, and $\mu_2 = 0.075$. We assume that the requested lead time of the specific demand is 70. The comparisons between the results obtained from our approximations for the in-time rate and the expected response time $E[R^2]$ on $S = 1, 2, 3, 4, 5$, and 6 are shown in Table 8.2.

In the following two examples, we verify our limiting results on P_f , L_{N_1} , and L_{B_2} obtained from Proposition 8.2. According to Proposition 8.2, we discuss this matter in two cases: $\lambda_s + \lambda_o < \mu_1$ in Example 8.3 and $\lambda_s + \lambda_o \geq \mu_1$ in Example 8.4.

Example 8.3 ($\lambda_s + \lambda_o < \mu_1$). Let $\lambda_o = 5$, $\lambda_s = 3$, $\mu_1 = 10$, and $\mu_2 = 11$. The results on various base-stock levels S are shown in Table 8.3. As we can see, P_f converges to 1; L_{N_1} is getting large and L_{B_2} converges to zero.

Table 8.1 Comparison results on fill rates P_f of ordinary demands on various ρ_s and ρ_o .

ρ_o	S		ρ_s		ρ_o	S		ρ_s	
			0.5	0.3				0.5	0.3
0.2	1	Sim.	0.4153	0.5850	0.4	1	Sim.	0.3573	0.4999
		Approx.	0.4167	0.5833			Approx.	0.3571	0.5000
	2	Sim.	0.6342	0.8082		2	Sim.	0.5400	0.7078
		Approx.	0.6343	0.8077			Approx.	0.5398	0.7083
	3	Sim.	0.7670	0.9073		3	Sim.	0.6505	0.8176
		Approx.	0.7615	0.9074			Approx.	0.6502	0.8172

Table 8.2 Comparison results on in-time rates and mean response times.

S	In-Time Rate		S	$E[R^2]$	
	Sim.	Approx.		Sim.	Approx.
1	0.941	0.940	1	26.539	26.565
2	0.943	0.938	2	25.371	25.533
3	0.944	0.942	3	24.338	24.340
4	0.951	0.948	4	22.954	23.138
5	0.953	0.953	5	22.046	22.051
6	0.962	0.960	6	21.040	21.132

Table 8.3 P_f , L_{N_1} , and L_{B_2} on various base-stock levels ($\lambda_o = 5$, $\lambda_s = 3$, $\mu_1 = 10$, $\mu_2 = 11$).

S	P_f	L_{N_1}	L_{B_2}	L_{B_1}	S	P_f	L_{N_1}	L_{B_2}	L_{B_1}
1	0.466	0.466	0.228	0.762	11	0.973	7.823	0.011	3.188
2	0.663	1.031	0.144	1.113	12	0.979	8.702	0.008	3.306
3	0.769	1.652	0.098	1.446	13	0.983	9.596	0.007	3.411
4	0.834	2.316	0.070	1.754	14	0.987	10.506	0.005	3.499
5	0.877	3.016	0.052	2.036	15	0.989	11.427	0.004	3.577
6	0.907	3.750	0.039	2.289	16	0.991	12.361	0.003	3.642
7	0.929	4.515	0.030	2.515	17	0.993	13.304	0.002	3.698
8	0.945	5.307	0.023	2.716	18	0.994	14.255	0.002	3.747
9	0.957	6.124	0.018	2.894	40	0.996	15.214	0.002	3.996
10	0.966	6.963	0.014	3.051	60	0.996	16.179	0.001	4.000

Table 8.4 P_f , L_{N_1} , and L_{B_2} on various base-stock levels ($\lambda_o = 20$, $\lambda_s = 7$, $\mu_1 = 10$, $\mu_2 = 11$).

S	P_f	L_{N_1}	L_{B_2}	L_{B_1}	S	P_f	L_{N_1}	L_{B_2}	L_{B_1}
1	0.100	0.100	2.100	3.000	11	0.150	0.238	1.983	12.745
2	0.132	0.167	2.025	3.858	12	0.150	0.238	1.983	13.745
3	0.143	0.204	1.998	4.794	13	0.150	0.238	1.983	14.745
4	0.147	0.223	1.988	5.765	14	0.150	0.238	1.983	15.745
5	0.149	0.231	1.985	6.754	15	0.150	0.238	1.983	16.745
6	0.149	0.235	1.984	7.749	16	0.150	0.238	1.983	17.745
7	0.149	0.237	1.983	8.746	17	0.150	0.238	1.983	18.745
8	0.150	0.237	1.983	9.746	18	0.150	0.238	1.983	19.745
9	0.150	0.238	1.983	10.745	19	0.150	0.238	1.983	20.745
10	0.150	0.238	1.983	11.745	20	0.150	0.238	1.983	21.745

Example 8.4 ($\lambda_s + \lambda_o \geq \mu_1$). Let $\lambda_o = 20$, $\lambda_s = 7$, $\mu_1 = 10$, and $\mu_2 = 11$. The situations with various base-stock levels S are in [Table 8.4](#). P_f , L_{N_1} , and L_{B_2} all converge to the same constants as estimated in [Proposition 8.2](#).

After verifying our estimations on the fill rate and in-time rate, in the next example we implement our results in finding the feasible base-stock levels where both requirements on the fill rate and in-time rate can be satisfied. In this example, we consider the case when $\lambda_s + \lambda_o < \mu_1$.

Example 8.5. Consider a system with $\lambda_o = 9$, $\lambda_s = 4$, $\mu_1 = 16$, and $\mu_2 = 15$. Suppose that the fill rate is required to be at least 0.9 and the in-time rate (with the required lead time 0.5) at least 0.95. We first try to find the base-stock levels where these qualities of services can be satisfied. The results on various base-stock levels are shown in [Table 8.5](#). We can see that these qualities of services are satisfied if S is greater than or equal to 6.

Now, we apply some cost structure by defining the following costs. Let C_1 denote the penalty cost for each unsatisfied ordinary demand; Let C_2 denote the penalty cost for each unsatisfied specific demand and u be the maximal allowable lead time. Let

Table 8.5 Fill rates and in-time rates on various S ($\lambda_o = 9, \lambda_s = 4, \mu_1 = 16,$ and $\mu_2 = 15$).

S	P_f	In-Time Rate	S	P_f	In-Time Rate
1	0.4800	0.9683	11	0.9724	0.9733
2	0.6731	0.9560	12	0.9779	0.9767
3	0.7757	0.9505	13	0.9822	0.9796
4	0.8381	0.9487	14	0.9857	0.9822
5	0.8795	0.9501	15	0.9885	0.9845
6	0.9083	0.9530	16	0.9907	0.9864
7	0.9291	0.9571	17	0.9925	0.9880
8	0.9446	0.9613	18	0.9939	0.9894
9	0.9564	0.9655	19	0.9950	0.9905
10	0.9654	0.9696	20	0.9960	0.9915

Table 8.6 TCs on various base-stock levels ($\lambda_o = 9, \lambda_s = 4, \mu_1 = 16, \mu_2 = 15, C_1 = \$5, C_2 = \$15,$ and $C_3 = \$1$).

S	TC	S	TC
1	25.782	11	10.539
2	18.394	12	10.947
3	14.723	13	11.455
4	12.672	14	12.036
5	11.415	15	12.680
6	10.662	16	13.385
7	10.225	17	14.138
8*	10.049	18	14.929
9	10.067	19	15.759
10	10.236	20	16.611

C_3 denote the inventory cost rate per each finished standard product in stock. Then, the total cost rate, TC , is expressed as

$$TC = (1 - P_f) \lambda_o C_1 + (1 - F_{R^2}(u)) \lambda_s C_2 + LC_3. \tag{8.20}$$

Following Example 8.5, we are interested in finding an optimal base-stock level minimizing the total cost subject to the requirements on the fill rate and in-time rate.

Example 8.6. We consider the same case of $\lambda_o = 9, \lambda_s = 4, \mu_1 = 16,$ and $\mu_2 = 15$ with $C_1 = \$5, C_2 = \$15,$ and $C_3 = \$1$. Suppose the qualities of service are that the fill rate must be at least 0.9 and the in-time rate (with the required lead time 0.5) must be at least 0.95. The TCs on various base-stock levels S are shown in Table 8.6 and the corresponding figure is Fig. 8.3. From Example 8.5, we know that the feasible base-stock levels are those greater than or equal to 6. Among these feasible levels, we then obtain the optimal base-stock level at $S = 8$ with minimum total cost rate 10.049.

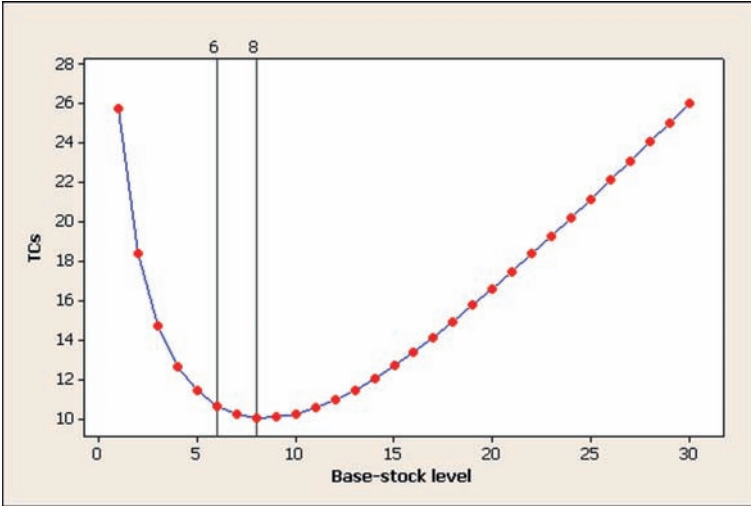


Fig. 8.3 TCs on various base-stock levels ($\lambda_o = 9, \lambda_s = 4, \mu_1 = 16, \mu_2 = 15, C_1 = \$5, C_2 = \$15,$ and $C_3 = \$1$).

8.4 Conclusions

In this chapter, we consider a two-station MTO/MTS hybrid production system dealing with ordinary and specific demands. We are interested in determining the fill rate of ordinary demands and response times of specific demands. By assuming the Markovian model, for station 1, we give the closed-form for the fill rate and some limiting results as the base-stock level increases, however, because of the intractability in analyzing station 2, we approximate station 2 as a GI/M/1 queue. The corresponding closed-form for the approximated in-time rate is obtained. These results of the fill rate and in-time rate can assist management in determining the optimal base-stock level efficiently. In future study, we may consider a multistation system for both the process producing the standard products and the process performing the additional work for the custom work.

Acknowledgments This work was supported by the National Science Council of Taiwan, ROC under the grant Contract NSC-95-2221-E-033-032.

References

1. C. A. Soman, D. P. van Donk, and G. Gaalman, Combined make-to-order and make-to-stock in a food production system, *International Journal of Production Economics*, vol. 90, pp. 223–235, 2004.

2. A. Krishnamurthy, R. Suri, and M. Vernon, Re-examining the performance of MRP and kanban material control strategies for multi-product flexible manufacturing systems, *The International Journal of Flexible Manufacturing Systems*, vol. 16, pp. 123–150, 2004.
3. I. J. B. F. Adan and J. van der Wal, Combining make to order and make to stock, *OR Spektrum*, vol. 20, pp. 73–81, 1998.
4. V. Nguyen, A multiclass hybrid production center in heavy traffic, *Operations Research*, vol. 46, Supp. 3, pp. S13–S25, 1998.
5. A. Federguruen and Z. Katalan, Impact of adding a make-to-order item to a make-to-stock production system, *Management Science*, vol. 45, no. 7, pp. 980–994, 1999.
6. S. Carr and I. Duenyas, Optimal admission control and sequencing in a Make-to-Stock/Make-to-Order production system, *Operations Research*, vol. 48, no. 5, pp. 709–720, 2000.
7. A. Arreola-Risa and G. A. DeCroix, Make-to-order versus make-to-stock in a production-inventory system with general production times, *IIE Transactions*, vol. 30, no. 8, pp. 705–713, 1998.
8. S. Rajagopalan, Make to order or make to stock: model and application, *Management Science*, vol. 48, no. 2, pp. 241–256, 2002.
9. V. G. Kulkarni, *Modeling, Analysis, Design, and Control of Stochastic Systems*. New York: Springer, 1999.
10. J. A. Buzacott, S. M. Price, and J. G. Shanthikumar, Service level in multistage MRP and base stock controlled production systems, in: G. Fandel, T. Gullledge, A. Jones (Ed.), *New Directions for Operations for Operations Research in Manufacturing System*, pp. 445–464. New York: Springer, 1992.