Using Queuing Approach for Locating the Order Penetration Point in a Two-Echelon Supply Chain with Customer Loss

Abbas Hajfathaliha
Faculty of Industrial Engineering, Department of Engineering, Shahed University, Tehran, Iran
PO Box: 33191-18651, Opposite of Imam Khomeini tabernacle-Persian Gulf Avenue-Tehran-Iran
Tel: 98-21-3396-7166 E-mail: hajfathaliha@shahed.ac.ir

Ebrahim Teimoury
Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
PO Box: 16846-13114, University street-Hengam street-Resalat square-Tehran-Iran
Tel: 98-21-7391-3022 E-mail: teimoury@iust.ac.ir

Iman Ghaleh Khondabi
Faculty of Industrial Engineering, Department of Engineering, Shahed University, Tehran, Iran
PO Box: 33191-18651, Opposite of Imam Khomeini tabernacle-Persian Gulf Avenue-Tehran-Iran
Tel: 98-912-339-7104 E-mail: ig.khondabi@gmail.com

Mehdi Fathi (Corresponding author)
Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
PO Box: 16846-13114, University street-Hengam street-Resalat square-Tehran-Iran
Tel: 98-912-419-5892 E-mail: mfathi@iust.ac.ir

Abstract

An order penetration point (OPP) is the boundary between make-to-order (MTO) policy and make-to-stock (MTS) policy. This study is dedicated to OPP strategic decision making which is concerned with the upstream and downstream decisions of OPP in supply chains. In this paper, we consider a two-echelon production supply chain in which in first production stage, supplier manufactures the semi-finished products on a MTS policy for a manufacturer in second production stage that will customize the products on a MTO policy. The manufacturer desires to find the optimal fraction of processing time fulfilled by its supplier and its optimal semi-finished products buffer storage capacity in OPP. Also, impatient customers are considered in the study. In order to calculate system performance indexes, we employ matrix geometric method and queuing steady state equations. Afterward, optimal solutions are obtained applying an optimization mathematical model with enumeration and direct search techniques.

Keywords: Queuing system, Supply chain, Impatient customers, Order Penetration Point, Make To Stock/Make To Order queue, Matrix Geometric Method

1. Introduction

In terms of supply chain design, customer satisfaction lead-time and inventory costs are common problems which supply chain managers encounter. In a production supply chain system, there are various generators of uncertainties such as demands' arrival times, processing times, transportation lead times, reliability of the machines, and the capability of the operators. Customers consider the response time of the order completion as a service performance measure. According to the new business model of Internet/telephone ordering and quick response time requirement, MTO business model is growing quickly. MTS production system can meet customer orders fast, but confront inventory risks associated with short product life cycles and unpredictable demand. On the other hand, MTO producers can provide a variety of products and custom orders with lower inventory risks,

but usually have longer customer lead times. Moreover, in MTS production, products are stocked in advance, while in MTO production, a product only starts to be manufactured when an order of demand is received. In some cases, custom products share approximately all the parts of the standard products and can be produced by alternating the existing standard parts with some further works, thus, the assembler usually consider embedding the MTO processes into the main stream MTS lines, which forms a hybrid production system.

There are some articles of making decision on OPP which appeared in the literature with many names such as Decoupling Point (DP), Delay Product Differentiation (DPD) and product customization postponement. The term DP, in the logistics framework was first introduced by Sharman (1984) where he argued the DP's dependency on a balance between product cost, competitive pressure and complexity. The trade-off between aggregation of inventory (or inventory pooling) and the costs of redesigning the production process is studied by Aviv and Federgruen (2001) where they do not consider congestion impacts. Gupta and Benjaafar (2004) consider the impact of capacity restrictions and congestion. They propose a common framework to examine MTO, MTS and DPD systems in which production capability is considered. Furthermore, they analyze the optimal point of postponement in a multi-stage queuing system. The DPD problem in manufacturing systems is studied by Jewkes and Alfa (2009). They decide on where to locate the point of differentiation in a manufacturing system and what size of semi-finished products inventory storage should be considered. In addition, they present a model to realize how the degree of DPD affects the tradeoff between customer order completion postponement and inventory risks, when both stages of production have non-negligible time and the production capacity is limited. Also, the concept of order decoupling zone is introduced by Wikner and Rudberg (2005) instead of the DP concept.

Ahmadi and Teimouri (2008) study where to locate OPP in an Auto Export supply chain by using dynamic programming. A notable literature review in positioning DPs and study the positioning multiple DPs in a supply network can be seen in Suna et al. (2008) but their positioning model does not take any decisions about optimal semi-finished buffer size and optimal fraction of processing time fulfilled by the upstream of DP. Also, many applications and methods for determining the OPP are surveyed in Olhager (2003), Yang and burns (2003), Yang et al. (2004), Rudberg and Wikner (2004) and Mikkola and Larsen (2004).

Our model tries to find equilibrium customer service levels with inventory costs, such as developed models in the literature. However, our model differs from the studied articles in several ways. First, we use a two stage MTS/MTO production model in a two-echelon supply chain. Second, our model gives the optimal semi-finished products warehouse capacity and optimal fraction of processing time fulfilled by the supplier in an integrated model and lastly it considers the impatient customers effect on OPP decision making.

The supply chain which is considered as a basic model in this paper is composed of two production stages and an infinite source of customers (who can balk and renege) with stochastic demand arrival times. In first production stage, supplier produces semi-finished products on a MTS policy for a manufacturer in second production stage that will customize the products on a MTO policy. More explicitly, the semi-finished products get completed due to the customers order.

We deal with the $M/M/1/N$ queue with finite capacity of impatient customers. The behavior of impatient customers which upon arrival may or may not go in the queue for service depending on the number of customers in the system and those which on going to the queue depart the queue without being served, are investigated in this study. The queuing systems with balking and reneging discussed in many articles. Examples of such studies can be seen in Wang et al. (2007), Yue and Sun (2008) and Al-Seedy et al. (2009).

Due to balance the costs of customer order fulfillment delay and inventory costs, manufacturer tries to find the optimal fraction of processing performed by the supplier and its optimal semi-finished products buffer storage. The primary objectives of the study are:

Developing the steady-state solutions for the $M/M/1/N$ queuing system with finite capacity, reneging and balking.

Developing a cost model to identify the optimal semi-finished products warehouse capacity in order to minimize the steady-state related inventory expected costs.

Obtaining the optimal fraction of processing time fulfilled in the first echelon of the chain which minimizes the steady-state expected costs.

The remainder of this paper is organized as follows. We review the problem description and formulation in Section 2. Also, the queuing aspect and performance evaluation indexes are studied in this Section. Section 3 is dedicated to system performance measurement. Our mathematical model is presented in Section 4. Section 5

represents a numerical example and finally, we conclude our study in Section 6.

2. Problem description and formulation

We consider a two-echelon production supply chain where the demands arrive according to a Poisson process with rate λ. The production time of workstation is assumed to be exponentially distributed with rate μ. We suppose that the supplier has an infinite source of raw materials and never faces shortage. The manufacturer has to determine the optimal storage capacity of semi-finished products (S).

Product supplier produces a semi-finished product (P completed $(0 < P < 1)$) to being delivered to the manufacturer. The manufacturer then completes the remaining $1-P$ fraction according to a particular customer order. We assume that the order satisfying time from the manufacturer to each demand is exponentially distributed with mean $1/(\mu/1-P)\,(>0)$ which is independent from the congestion in manufacturer. Also replenishment lead time for supplier is exponentially distributed with mean $1/(\mu/P)\,(>0)$. It is notable that the supplier is not necessarily in a different organization from the producer; the ‘‘supplier” and ‘‘manufacturer” may be two successive stages in a same organization. We chose to model P as a continuous variable so that we can gain greater insights into the overall relationship between P and system performance. The assumption also facilitates our computational analysis. We, therefore, present results as if the producer can implement any values of P. If this is not the case, our model enables to quickly identify the best choice of P among a finite number of feasible alternatives. According to market characteristics studied by Jewkes and Alfa (2009), there is a value per unit of semi-finished products $V(P)$ which is monotonically increasing with P, rationally. See Figure 1. for a diagram which is depicting our model.

A customer on arrival finds n customers (including the one being served) in the manufacturer, either decides to enter the system with probability θ_n or do not enter the system (balk) with probability $1-\theta_n$. If a customer enters the system but find the server busy, it enters the finite waiting queue but may get impatient and run off the system without getting service. The certain time that a customer wait for service to begin before getting impatient, is random variable which is distributed as a negative exponential distribution with parameter β. The manufacturer can serve just one customer at a time and the service process assumed to be independent of customer arrivals. The problem is to find the best OPP for our supply chain and its optimal semi-finished products storage capacity according to defined system parameters.

2.1 The Markov chain

The system can be described by a quasi birth-and-death Markov process with states (n,k). Consider the continuous Markov chain $\{(n,k),\ 0 \le n \le N,\ 0 \le k \le S\}$ where n is the number of customers in the system including the one being served and k is the number of semi-finished products storage which are available in the supplier buffer. The state space with transition rates is depicted in Figure. 2.

2.2 Steady state results

We described the state of the system by the pairs $\{(n,k),\ 0 \le n \le N,\ 0 \le k \le S\}$. Now, let θ_n denote the probability that a customer enters the queue when there are n customers in the system. θ_n is defined as follows (Bhat 2008):

$$\theta_n = \begin{cases} 1 & n = 0, \\ e^{-n/(\mu/1-P)} & 1 \le n \le N-1, \\ 0 & n = N \end{cases}$$

Since the waiting time before getting impatient is a random variable which follows an exponential distribution with mean $1/\beta$ and customers decisions are independent of each other, the average renege rate is $n\beta$. In order to develop the steady-state probabilities $\pi_{(n,k)}$, $0 \le n \le N$, $0 \le k \le S$ we get help from the matrix geometric method which was first introduced by Neuts (1981). The generator matrix of the under study Markov chain is given as:

$$Q=\begin{bmatrix} B_0 & A_0 & & & & \\ C_1 & B_1 & A_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & C_{N-1} & B_{N-1} & A_{N-1} & \\ & & & C_N & B_N & \end{bmatrix}$$

Where A_n, B_n and C_n are block square matrixes of order $S+1$ which are displayed in Appendix. According to the finite capacity of the queue, there is no need to check the stability condition for the system under consideration. It is notable that A_n, $B_n\ (n \neq 0)$ and C_n giving the rate at which the number of the customer orders in the system increase by one, stay at the same level, or decrease by one. B_0 is the matrix rate at which the customer orders in the system moves from zero to one.

We define the steady-state probability vector $\pi = [\pi_0, \pi_1, ..., \pi_N]$ for the Markov chain $\{(n, k),\ 0 \le n \le N, 0 \le k \le S\}$. Each π_n can be calculated using $\pi Q = 0$ and $\sum_{n=0}^{N} \pi_n = 1$, where $\pi_n = [\pi_{(n,0)}, \pi_{(n,1)}, ..., \pi_{(n,S)}]$ is a $1 \times (S+1)$ row vector. $\pi_{(n,k)}$ denotes the steady-state probability associated with the condition that there are n customers and k semi-finished products in the system.

Referring to the state transition diagram for the finite $M/M/1/N$ queuing system with balking and reneging which is shown in Figure 2, the following balance equations are derived:

$$(\theta_n\lambda + \frac{\mu}{P})\pi_{(n,k)} = \beta\pi_{(n+1,k)} \qquad n=0, k=0$$

$$(\theta_n\lambda + \frac{\mu}{P})\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{\mu}{1-P}\pi_{(n+1,k+1)} + \frac{\mu}{P}\pi_{(n,k-1)} \qquad n=0, 1\le k\le S-1$$

$$\theta_n\lambda\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{\mu}{P}\pi_{(n,k-1)} \qquad n=0, k=S$$

$$(\theta_n\lambda + n\beta + \frac{\mu}{P})\pi_{(n,k)} = (n+1)\beta\pi_{(n+1,k)} + \frac{\mu}{1-P}\pi_{(n+1,k+1)} + \theta_{n-1}\lambda\pi_{(n-1,k)} \qquad 1\le n\le N-1, k=0$$

$$(\theta_n\lambda + n\beta + \frac{\mu}{1-P} + \frac{\mu}{P})\pi_{(n,k)} = (n+1)\beta\pi_{(n+1,k)} + \frac{\mu}{1-P}\pi_{(n+1,k+1)} + \theta_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{P}\pi_{(n,k-1)} \qquad 1\le n\le N-1, 1\le k\le S-1$$

$$(\theta_n\lambda + n\beta + \frac{\mu}{1-P})\pi_{(n,k)} = (n+1)\beta\pi_{(n+1,k)} + \theta_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{P}\pi_{(n,k-1)} \qquad 1\le n\le N-1, k=S$$

$$(n\beta + \frac{\mu}{P})\pi_{(n,k)} = \theta_{n-1}\lambda\pi_{(n-1,k)} \qquad n=N, k=0$$

$$(n\beta + \frac{\mu}{1-P} + \frac{\mu}{P})\pi_{(n,k)} = \theta_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{P}\pi_{(n,k-1)} \qquad n=N, 1\le k\le S-1$$

$$(n\beta + \frac{\mu}{1-P})\pi_{(n,k)} = \theta_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{P}\pi_{(n,k-1)} \qquad n=N, k=S$$

In order to solve $\pi Q = 0$, it is not possible to define a constant rate matrix R such that $\pi_n = \pi_{n-1}R = \pi_1 R^{n-1}$ as discussed in Neuts (1981), because of the asymmetric structure of Qs' sub-matrixes.

So, due to the finite number of sub-matrixes in generator matrix Q, balance equations are solved directly by MATLAB 7.1 (the language of technical computing), in order to calculate the steady-state probabilities.

3. System performance measures

In this section we derive a number of performance measures of the system under consideration in the steady-state.

3.1 Mean inventory level

Let $E(I)$ represent the expected number of semi-finished products in the system. Then we have

$$E(I)=\sum_{n=0}^{N}\sum_{k=1}^{S}k\pi_{(n,k)}$$

3.2 Mean backorder level

Let $E(B)$ denote the mean number of backorders in the steady-state. Then we have

$$E(B)=\sum_{n=1}^{N}n\pi_{(n,0)}$$

3.3 Mean customer order fulfillment delay

Let $E(W)$ represent the average customer order fulfillment delay in the steady-state. In order to calculating $E(W)$ we have to obtain the mean customer orders in the system $E(L)$ that is achievable as follows

$$E(L)=\sum_{k=0}^{S}\sum_{n=1}^{N}n\pi_{(n,k)}$$

Then via Little's law we have

$$E(W)=E(L)/\lambda(1-\sum_{k=0}^{S}\pi_{(N,k)})$$

3.4 Mean rate of customer loss

Let $E(BA)$, $E(RE)$ and $E(LO)$ denote the average balking rate, the average reneging rate and the average rate of customer loss. Using the concept of Ancker and Gafarian (1963) these average rates are obtained as follows

$$E(BA)=\sum_{k=0}^{S}\sum_{n=1}^{N}(1-\theta_n)\lambda\pi_{(n,k)}$$

$$E(RE)=\sum_{k=0}^{S}\sum_{n=1}^{N}n\beta\pi_{(n,k)}$$

$$E(LO)=E(BA)+E(RE)$$

4. Mathematical model

We use the following notations for the mathematical formulation of our model in order to choose the optimal fraction of process fulfilled by the supplier of each product type and their optimal semi-finished products buffer storage capacity in OPP.

Variables:

P Percentage of product completion in first production stage

S Optimal storage capacity of semi-finished products

Parameters:

$V(P)$ The value per unit of semi-finished products (dollars/unit)

τ Constant fraction of the overall customer order completion delay

μ Production rate for each product unit

C_H The holding cost for semi-finished product (dollars/dollar/unit time)

C_W The cost of customer order fulfillment delay (dollars/unit/unit time)

C_B The fixed backorder cost per order per unit time

C_{LO} The loss cost of one customer (renege or balk) per unit time

C_C The cost of establishing semi-finished products storage capacity (dollars/unit/unit time)

Objective function:

The objective function includes the following costs:

Holding semi-finished products in buffer storage ($C_H V(P) E(I)$).

Product customization delay ($C_W E(W)$)

Backordering customer orders fulfillment ($C_B E(B)$).

Customer loss ($C_{LO} E(LO)$).

Providing storage capacity for the semi-finished products ($C_C S$).

The mathematical formulation of the model is as follows:

$$\underset{P,S}{Min}\ TC(P,S) = C_H V(P) E(I) + C_W E(W) + C_B E(B) + C_{LO} E(LO) + C_C S \quad (1)$$

subject to :

$$\frac{(1-P)}{\mu} \geq \tau E(W) \quad (2)$$

$$0 < P < 1.0 \quad (3)$$

$$S = 1,2,... \quad (4)$$

The model (8) minimizes the total expected cost including the expected semi-finished products holding cost, expected cost of delay in customer order completion, backordering cost, customer loss related cost and the cost of establishing storage capacity for semi-finished products. Constraint (9) represents that the mean customization time (service level constraint) for a product $\frac{(1-P)}{\mu}$ must be at least more than a constant value (τ) of the total expected customer order completion delay. Constraints (10) and (11) represent the ranges of the model variables. The outputs of represented model are the optimal fraction of process fulfilled by the supplier and the optimal semi-finished products buffer storage capacity.

Replacing the values of mean performance measures, we get the following expected total cost function:

$$TC(P,S) = C_H V(P) \sum_{n=0}^{N}\sum_{k=1}^{S} k\pi_{(n,k)} + C_W \sum_{k=0}^{S}\sum_{n=1}^{N} n\pi_{(n,k)} \Big/ \lambda(1 - \sum_{k=0}^{S} \pi_{(N,k)} + C_B \sum_{n=1}^{N} n\pi_{(n,0)}$$
$$+ (C_{LO} \times (\sum_{k=0}^{S}\sum_{n=1}^{N} (1-\theta_n)\lambda\pi_{(n,k)} + n\beta\pi_{(n,k)})) + C_C S$$

According to recursive computation of the π's, it is quite difficult to show the convexity of the expected total cost function. However we present a numerical example to prove the computability of the results derived in this study.

5. Numerical example

In this section a numerical example is used to show the relation between $TC(P,S)$ and variables P and S, also system analysis is done for a variety of parameters. In Matlab implementation, P is varied between 0.01 and 0.99 which increments by 0.01. We assume our semi-finished storage capacity cannot exceed 10 units and our queue length must not exceed 8 due to physical constraints.

The following parameters are assumed for this example: $\lambda = 0.7$, $\mu = 1$, $\tau = 0.03$, $V(P) = 0.5P$,

$\beta = 0.05$, $C_H = 0.1$, $C_W = 0.5$, $C_B = 0.2$ and $C_{LO} = 1.8$, $C_C = 0.2$. We assume that every semi-finished product will be compatible with customer demand and no product will be discarded. Table1 shows the optimal OPP for various product storages. According to the fact that the supplier is not able to provide enough semi-finished products in $S = 1$ to carry on the demand, this state is not stable. For $S = 2$ it is optimal to produce $P = 0.01$ of each product according to system stability and least cost.

As S increase to 3, we can see total cost increases according to the cost of establishing one more storage capacity but when S increases to 4, there is an unusual trend in both total cost and optimal OPP. As we can see, it is optimal to place OPP in $P = 0.96$ which is completely far away from OPP with $S = 3$. This difference is reasonable due to customer loss and order completion delay costs. According to concave structure of total cost function, the total cost increases with increasing P , but after a specific value it starts to decrease. This is the reason which makes us to move from P initial values in $S = 3$ to its end values in $S = 4$ according to Figure 3.

It is notable that in Figure 3 the zero total costs are related to infeasible points. This example tell us that we can obtain the least cost of manufacturing with the storage capacity of 5 and with this buffer, the OPP must be placed after $P = 0.98$ which means %98 of product must be performed by the first echelon of the chain and the remained %2 must be completed due to customer order in second echelon of the chain. In the next subsections, we analyze the system manner under different parameter variations.

5.1 Effect of queue capacity

In our model, we suppose that our queuing system has a finite capacity (N) due to physical constraints. N can affect our model because all generator matrixes are related to queue capacity. We can consider N as a model characteristic where higher values of N can increase the values of customer queue length and higher values of delay for customers and lower values may decrease the serviced customers and increase the customer loss. In our example we can't use upper values than 9 for queue capacity so we have conducted some lower values than 8 to understand how the value of N would affect our total cost.

As shown in Table2, we can see it is not possible to find a specific trend for total cost for each constant value of S by increasing N. But in a comparison to N=8 we can see lower values of N have lower total costs.

For fixed values of N, total cost increases by increasing P except N=6. In this column we see total cost is decreasing from P=2 to P=4 according to model constraint which doesn't permit $P = 0.99$ to be feasible. In examined values of N, none of initial values of P are optimal, unlike N=8 where P values like 0.01 could be optimal for some S. This could be one of the increased N effects which decrease customer loss and make it possible to do more manufacturing process in second echelon of supply chain in lower values of S.

5.2 Effect of demand variation

If a supply chain manager has made his decisions about semi-finished product storage and establishing the OPP, changes in product demand may force him to change the decisions. As shown in Table3, when the demand becomes larger, S=4 is better than S=2 because of preventing semi-finished product shortages. But, this is not the case about S=6 due to increased cost of establishing buffer storage.

As it is seen in Table3, the optimal total cost is increasing in λ . The reason may be the growth of system busy time which cause more order completion time and more impatient customers loss. Also, increasing system busy time may cause a higher backorder cost which lastly makes a greater total cost.

6. Conclusions

OPP is the boundary between MTO and MTS policies. In this article an optimization model is presented to determine OPP in a two-echelon supply chain. The affects of product customization postponement on customer order completion delay and inventory risks are discussed. The service procedure is modeled as a finite queueing system with customer loss (renege and balk). In order to evaluate performance measures, a simple queueing model and explicitly matrix geometric method is applied. Proposed model aims to obtain the near optimal order penetration point in a supply chain and optimal level of buffer storage capacity applying an integrated total cost function. Computability of the proposed model is proved with a numerical example. Following issues can be as future research possibilities:

1. Analyzing the problem assuming the queueing system has no finite capacity

2. Relaxing the assumptions of exponentially distributed arrivals and service times

3. Considering more than one manufacturer and customers jockeying between these manufacturers

Acknowledgement

The authors would like to thank Professor Attahiru S. Alfa and Professor Hartanto Wong for their valuable comments, which have improved the content and format of the paper.

References

Al-Seedy, R.O., El-Sherbiny, A.A., El-Shehawy, S.A., & Ammar, S.I. (2009). Transient solution of the M/M/c queue with balking and reneging. *Computers & Mathematics with Applications,* 57: 1280-1285.

Ancker, C.J., & Gafarian, A.V. (1963). Some Queuing Problems with Balking and Reneging. *Operations Research*, 11: 88-100.

Ahmadi, M., & Teimouri, E. (2008). Determining the Order Penetration Point in Auto Export Supply Chain by the Use of Dynamic Programming. *Journal of Applied Sciences,* 8 (18): 3214-3220.

Aviv, Y., & Federgruen, A. (2001). Design for postponement: A comprehensive characterization of its benefits under unknown demand distributions. *Operations Research,* 49 (4), 578–598.

Bhat, U.N. (2008). *An Introduction to Queueing Theory:Modeling and Analysis in Applications.* Birkhauser: Boston.

Gupta, D., & Benjaafar, S. (2004). Make-to-order, make-to-stock, or delay product differentiation? A common framework for modeling and analysis. *IIE Transactions,* 36, 529–546.

Jewkes, E.M., & Alfa, A.S. (2009). A queuing model of delayed product differentiation. *European Journal of Operational Research,* 199, 734-743.

Mikkola, J.H., & Larsen, T.S. (2004). Supply chain integration: implications for mass customization, modularization and postponement strategies. *Production Planning & Control,* 15(4), 352-361.

Neuts, M.F. (1981). *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach.* Johns Hopkins University Press: Baltimore.

Olhager, J. (2003). Strategic positioning of the order penetration point. *International Journal of Production Economics*, 85, 319–329.

Rudberg, M., & Wikner, J. (2004). Mass customization in terms of the customer order decoupling point. *Production Planning & Control,* 15(4), 445-458.

Sharman, G. (1984). The rediscovery of logistics. *Harvard Business Review,* 62 (5), 71–79.

Suna, X.Y., Jib, P., Suna, L.Y., & Wang, Y.L. (2008). Positioning multiple decoupling points in a supply network. *International Journal of Production Economics,* 113, 943–956.

Wang, K.H., Ke, J.B., & Ke, J.C. (2007). Profit analysis of the M/M/R machine repair problem with balking, reneging, and standby switching failures. *Computers and Operations Research,* 34: 835-847.

Wikner, J., & Rudberg, M. (2005). Introducing a customer order decoupling zone in logistics decision making. *International Journal of Logistics: Research and Application,* 8(3), 211-224.

Yang, B., & Burns, N.D. (2003). Implications of postponement for the supply chain. *International Journal of Production Research*, 41(9), 2075-2090.

Yang, B., Burns, N.D., & Backhouse, C. J. (2004). Postponement: a review and an integrated framework. *International Journal of Operations & Production Management,* 24(5), 468-487.

Yue, D.Q., & Sun, Y.P. (2008). Waiting time of M/M/c/N queuing system with balking, reneging, and multiple synchronous vacations of partial servers. *Systems Engineering - Theory & Practice,* 28: 89-97.

Appendix

$$B_n = \begin{bmatrix} B_{n_{0,0}} & B_{n_{0,1}} & & & \\ & B_{n_{1,1}} & B_{n_{1,2}} & & \\ & & \ddots & \ddots & \\ & & & B_{n_{S-1,S-1}} & B_{n_{S-1,S}} \\ & & & & B_{n_{S,S}} \end{bmatrix} \tag{A.1}$$

$$\left.\begin{aligned} B_{n_{k,k}} &= \begin{cases} -(\frac{\mu}{P}+\theta_n\lambda) & 0 \le k \le S-1 \\ -\theta_n\lambda & k = S \end{cases} \\ B_{n_{k,k+1}} &= \frac{\mu}{P} \qquad 0 \le k \le S-1 \end{aligned}\right\} \quad n = 0 \tag{A.2}$$

$$\left.\begin{aligned} B_{n_{k,k}} &= \begin{cases} -(\frac{\mu}{P}+\theta_n\lambda+n\beta) & k = 0 \\ -(\frac{\mu}{P}+\theta_n\lambda+n\beta+\frac{\mu}{1-P}) & 1 \le k \le S-1 \\ -(\theta_n\lambda+n\beta+\frac{\mu}{1-P}) & k = S \end{cases} \\ B_{n_{k,k+1}} &= \frac{\mu}{P} \qquad 0 \le k \le S-1 \end{aligned}\right\} \quad 1 \le n \le N-1 \tag{A.3}$$

$$\left.\begin{aligned} B_{n_{k,k}} &= \begin{cases} -(\frac{\mu}{P}+n\beta) & k = 0 \\ -(\frac{\mu}{P}+n\beta+\frac{\mu}{1-P}) & 1 \le k \le S-1 \\ -(n\beta+\frac{\mu}{1-P}) & k = S \end{cases} \\ B_{n_{k,k+1}} &= \frac{\mu}{P} \qquad 0 \le k \le S-1 \end{aligned}\right\} \quad n = N \tag{A.4}$$

$$A_n = \theta_n\lambda\mathbf{I}_{k\times k}, \quad 0 \le n \le N-1 \tag{A.5}$$

$$C_n = \begin{bmatrix} n\beta & & & \\ \mu & n\beta & & \\ & \ddots & \ddots & \\ & & \mu & n\beta \\ & & & \mu \quad n\beta \end{bmatrix}, \qquad 1 \le n \le N \tag{A.6}$$

Table 1. $TC(S, P^*(S))$ and $P^*(S)$ as a function of S

S	$P^*(S)$	$TC(S, P^*(S))$
1	Non stable	Non stable
2	0.01	1.95
3	0.01	2.11
4	0.96	1.89
5	0.98	1.61
6	0.99	1.64
7	0.99	1.87
8	0.99	2.11
9	0.99	2.39
10	0.99	2.59

ISSN 1833-3850 E-ISSN 1833-8119

Table 2. Optimal solution for various S and N

S	$N=2$		$N=4$		$N=6$	
	$P^*(S)$	$TC(S,P^*(S))$	$P^*(S)$	$TC(S,P^*(S))$	$P^*(S)$	$TC(S,P^*(S))$
1	Non stable	Non stable	Non stable	Non stable	Non stable	Non stable
2	0.99	0.8	0.99	0.87	0.94	1.86
3	0.99	0.93	0.99	0.9	0.97	1.31
4	0.99	1.12	0.99	1.1	0.99	1.18
5	0.99	1.39	0.99	1.36	0.99	1.37
6	0.99	1.6	0.99	1.57	0.99	1.58
7	0.99	1.87	0.99	1.86	0.99	1.86
8	0.99	2.12	0.99	2.11	0.99	2.1
9	0.99	2.4	0.99	2.4	0.99	2.39
10	0.99	2.6	0.99	2.6	0.99	2.6

Table 3. Optimal solution for various λ

λ	$S=2$		$S=4$		$S=6$	
	$P^*(S)$	$TC(S,P^*(S))$	$P^*(S)$	$TC(S,P^*(S))$	$P^*(S)$	$TC(S,P^*(S))$
0.3	0.99	1.05	0.99	1.12	0.99	1.61
0.4	0.01	1.49	0.99	1.17	0.99	1.61
0.5	0.01	1.64	0.98	1.33	0.99	1.61
0.6	0.01	1.79	0.97	1.58	0.99	1.62

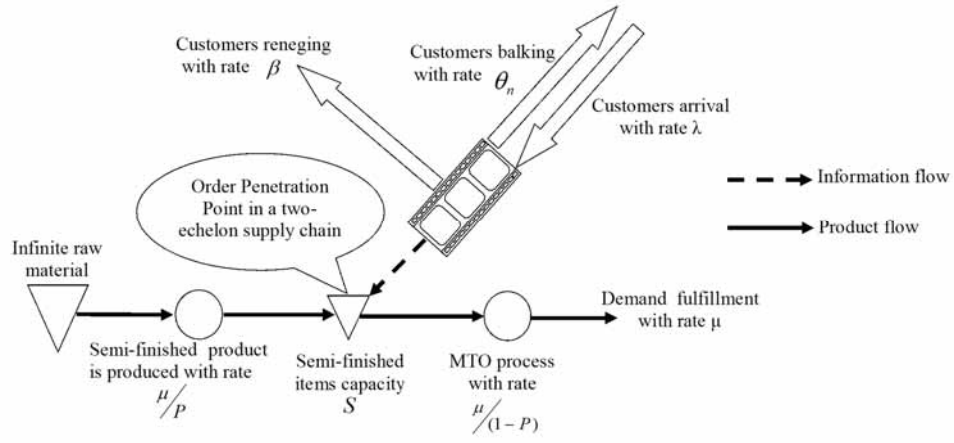

Figure 1. The hybrid MTO/MTS production supply chain system

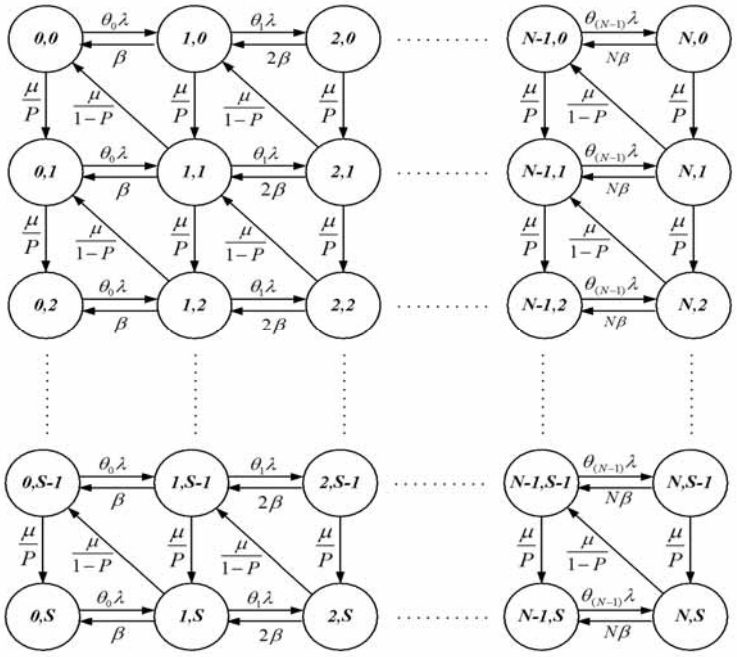

Figure 2. State transition rates diagram

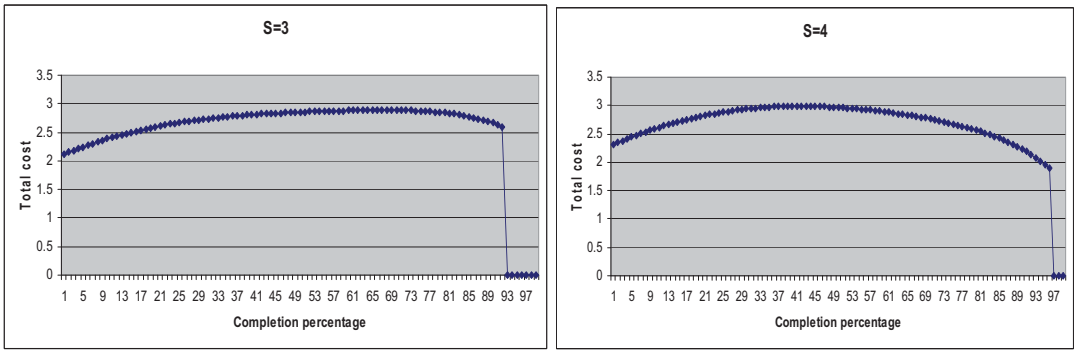

Figure 3. $TC(S,P)$ versus P