ISSN 1980-5411 (On-line version)
Production, 28, e20180024, 2018
DOI: 10.1590/0103-6513.20180024

Research Article

Production line performance analysis within a MTS/MTO manufacturing framework: a queueing theory approach

Iman Ghalehkhondabi[a]*, Gursel Suer[b]

[a]Our Lady of the Lake University, Business Department, San Antonio, TX, United States of America

[b]1 Ohio University, Department of Industrial and Systems Engineering, Athens, OH, United States of America

*ighalehkhonda@ollusa.edu

Abstract

Paper aims: Mixing the Make-To-Stock (MTS) and Make-To-Order (MTO) strategies to benefit from the both manufacturing systems in an environment with impatient customers.

Originality: This is the first research article which uses the queuing theory to find the best place of the Order Penetration Point (OPP) in a production line with impatient customers.

Research method: Two scenarios are studied in this paper. 1- Semi-finished products are produced and stored in a buffer. No semi-finished product will be completed until a specific order comes to the system. This strategy leads to idle cost, but there is savings obtained in terms of eliminating investment in finished goods inventory and its holding cost. We can calculate the total cost of the system and find the optimal machine for the buffer of semi-finished products. 2- We use both MTS and MTO for completing the semi-finished products. When there is no customer in the system, semi-finished products are completed based on the MTS strategy and finished products are sent to a warehouse. But, when an order comes in for customization, a semi-finished product get assigned to that order and after finishing the current MTS job on each machine, this MTO job starts to complete the semi-finished product. To calculate system performance indexes, we use the Matrix Geometric Method (MGM) after modeling the systems with queuing theory concepts.

Main findings: Numerical examples show the convexity of total cost in terms of product completion percentage and number of customization lines after the OPP. Also, increasing the production rate leads to higher expected number of semi-finished products in the buffer.

Implications for theory and practice: Positioning OPP in manufacturing systems to compare different production strategies (MTS/MTO) has not been widely studied yet. This paper shows how manufacturing companies can apply the OPP to obtain benefits from both MTS and MTO strategies according to various cost parameters of the production lines. Using queuing theory concepts to model the problem under study helps to consider the external factors such as impatient customers and demand arrival uncertainty that can affect the performance measures of the system besides the internal factors such as production rate and inventory related costs. The idea of expanding the production line after a specific station and have more customization lines to improve customer satisfaction is studied in this paper as well.

Keywords

Performance analysis. Order Penetration Point (OPP). Make to Stock (MTS). Make to Order (MTO). Queuing theory.

How to cite this article: Ghalehkhondabi, I., & Suer, G. (2018). Production line performance analysis within a MTS/MTO manufacturing framework: a queueing theory approach. *Production*, *28*, e20180024. https://doi.org/10.1590/0103-6513.20180024.

Received: Apr. 16, 2018; Accepted: June 3, 2018.

1. Introduction

Balance of supply and demand has been one of the most important issues in manufacturing companies in recent decades. Manufacturing companies look for an inventory and supply management approach which can obtain the most benefit out of the investment in raw materials, work in process, and finished products

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

inventories (Donath, 2002). As discussed by Wu & Olson (2008) and Wang (2009), holding finished products, delayed orders, shortages, and disposing outmoded and expired products are the most common inventory costs in a manufacturing system.

Many of the above mentioned inventory costs are related to using MTS strategy, where the company produces finished products based on its market demand forecast. The most important feature in MTS strategy is to obtain higher order satisfaction rates in comparison to the other strategy, Make-to-Order (MTO). MTS strategy offers a lower possibility of customization, but less expensive products compared to MTO strategy. On the other hand, issues such as higher average order delay, higher average response time and due date setting are considered as concerns of MTO production systems. These systems offer more expensive products due to higher variety of customer customization (Soman et al., 2004).

The system under study tries to mix the Make-To-Stock (MTS) and Make-To-Order (MTO) strategies to benefit from the both manufacturing systems. This idea is developed by applying the Order Penetration Point (OPP) in production lines. The main idea of using such a point in manufacturing systems mostly addressed in research articles in the field of supply chain management. This point is defined as a point of a supply chain where each product is dedicated to a specific order of a known customer (Olhager, 2003). In order to emphasize that this point is directly related to the customer order, some researchers considered that as Customer Order Decoupling Point (CODP). The usage of OPP in production lines is depicted in Figure 1.

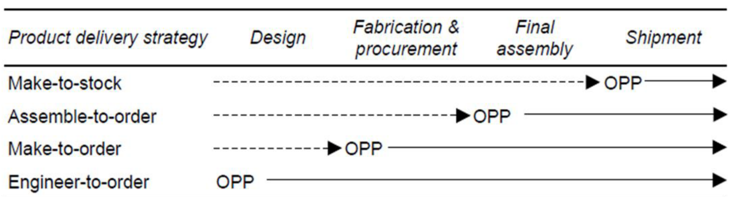


Figure 1. Different product delivery strategies relate to different order penetration points (OPP). The dotted lines depict the production activities that are forecast-driven, whereas the straight lines depict customer-order-driven activities (Olhager, 2003).

The positioning of OPP is widely studied as a challenging issue in supply chain management in recent decades. Nonetheless, considering this concept in manufacturing systems to compare different production strategies (MTS/MTO) has not been widely studied yet. In this regard, by doing a search in “scholar.google.com” to find an article with the words “Manufacturing” and “Order Penetration Point” in the title, we can’t find even one article. According to the best knowledge of authors (doing research in this area about 10 years), the most related study is a book chapter by Süer & Lobo (2011), where they considered the hybrid strategy in cellular systems. Their study compared the connected and disconnected cellular systems in a case study of a medical device manufacturing company and simulated different scenarios to compare two cellular manufacturing companies.

The motivation of this paper is to show how manufacturing companies can apply the OPP to obtain benefits from both MTS and MTO strategies according to various cost parameters of the production lines. Moreover, using queuing theory concepts to model the problem under study helps to consider the external factors such as impatient customers and demand arrival uncertainty that can affect the performance measures of the system besides the internal factors such as production rate and inventory related costs. In real production environments, it is possible that a company wants to expand the production line after a specific station and have more customization lines to improve customer satisfaction. This multiple customization line idea is studied in this study as well.

The rest of the paper is organized as follows. The related literature of different applications of OPP in both academia and industry is reviewed in Section 2. Section 3 is dedicated to describe the focus of this study and define the notation to model the problem. Section 4 formulates the problem and presents the solution method. A numerical example with a vast sensitivity analysis is presented in Section 5, and Section 6 concludes the study and discusses the future study paths.

2. Literature review

In the past two decades, developing new production strategies such as MTS/MTO manufacturing gained a high attention from the researchers to work out the lack of compliance between real market demand and demand forecasts. The differentiation point between MTO and MTS strategies is known as Order Penetration Point

(OPP) (Olhager, 2003) in the literature and determining the OPP has become an important strategic decision in supply chains and production systems according to its capabilities to reduce the inventory management costs. The terminology of Decoupling Point (DP) (Sharman, 1984) and Customer Order Decoupling Point (CODP) (Hoekstra & Romme, 1992) is noted in the literature for the same concept.

According to Herer et al. (2002), the OPP distinguishes the customer response part of a manufacturing system from the forward strategic planning part. From the strategic viewpoint, there are two encouraging factors to place the OPP in the final processes of a production line: decreasing the lead time and improving the production efficiency. Table 1 presents the advantages and disadvantages of shifting the OPP toward the end of the production line:

Table 1. Strategic issues, reasons and negative effects of shifting the OPP forwards (Olhager, 2003).

Competitive advantage addressed	Reasons for forward shifting	Negative effects
Delivery speed	Reduce the customer lead time	Rely more on forecasts (risk of obsolescence)
Delivery reliability	Process optimization (improved manufacturing efficiency)	Reduce product customization (to maintain work in process and inventories levels)
Price		Increase work-in-process (due to more items being forecast-driven)

On the other hand, the main reason to moving OPP toward the upper-hand processes of a production line is to gain more knowledge about the customer order specifications in manufacturing period, which enhance the possibility of customization and reducing the amount of work-in-process. Table 2 presents the advantages and disadvantages of shifting the OPP toward the first machines of a production line.

Table 2. Strategic issues, reasons and negative effects of shifting the OPP backwards (Olhager, 2003).

Competitive advantage addressed	Reasons for backward shifting	Negative effects
Product range	Increasing the degree of product customization	Longer delivery lead times and reduced delivery reliability (if production lead times are not reduced)
Product mix flexibility	Reduce the reliance on forecasts	Reduced manufacturing efficiency (due to reduced possibilities to process optimization)
Quality	Reduce or eliminate WIP buffers	
	Reduce the risk of obsolescence of inventories	

The concept of using hybrid manufacturing beside pure MTO in a scheduling problem is developed by Wu et al. (2008) to maintain high utilization of machines. Their research proposed a scheduling method for a hybrid manufacturing system with the characteristics of machine-dedication. The goal of the presented model is to achieve a high throughput for MTS products and high on-time delivery rate for MTO products.

A recent study by Olhager & Prajogo (2012) compared Make-To-Order and Make-To-Stock firms by examining the data from 216 Australian manufacturing firms. It is shown that MTO systems exhibits a specified impact for integration of suppliers on improving the performance of business, but this is not the case for supplier rationalization and lean practices.

Queuing theory concepts are applied to find the OPP in a study by Teimoury et al. (2012). They considered a multiproduct MTS/MTO manufacturing system which has probabilistic distribution for both customer arrival and manufacturing semi-finished products. Zhang et al. (2013) used queuing theory for dynamic pooling of the MTS and MTO operations. In their model, a subset of machines is dynamically switched between two strategies. They quantified the performance measures of the system by developing analytical formula. This approach minimizes the total costs of the system with customer satisfaction constraints.

The performance analysis issue in MTS/MTO manufacturing systems is studied by Almehdawe & Jewkes (2013). This study is more dedicated to inventory costs including replenishment and batch ordering policies.

Table 3 shows a brief introduction of recent studies of OPP in different aspects of production and service systems:

The viewpoint of current study is similar to Teimoury et al. (2012), where the authors used queuing Matrix Geometric Method (MGM) to find the OPP in a two-stage supply chain. The mathematical models in the literature

Table 3. Some of the OPP published articles from 2008 to 2016.

Authors	Problem	Objective Function	Case Study	Environment	Demand-Product Structure	OPP	Modelling/ Solution Approach
Sun et al. (2008)	Positioning multiple decoupling points in a supply network	Min. setup cost, inventory holding cost, stock-out cost and asset specificity cost	No	MTS/MTO	Stochastic demand, multiple parts of a product	Decision variable	Mathematical modeling/Optimization
Rafiei & Rabbani (2011)	Order partitioning and locating OPP	Comparing various performance measures	Yes	MTS, MTO, MTS/MTO	Exact demand, Product families	Decision variable	Fuzzy analytic network process
Kalantari et al. (2011)	Order acceptance or rejection	Min. production cost, outsourcing cost and lateness/ earliness penalty	No	MTS/MTO	Fuzzy demand, Multiproduct	Located at the middle of production line	Fuzzy TOPSIS, Mixed integer mathematical programming
Sharda & Akiya (2012)	Selecting the inventory management policy	Max. on time fulfillment, Min. production and inventory cost	Yes	Combined MTS/ Postponement	Stochastic demand, Multiproduct	Decision variable	Discrete event simulation
Zhang et al. (2013)	Finding the optimal plan of switching servers between MTS and MTO	Min. total cost of system, satisfying service constraints	No	MTS, MTO	Stochastic demand, Single product	Non applicable	Queueing theory/Direct search procedure
Almehdawe & Jewkes (2013)	Applying batch ordering policy to optimize a MTS/MTO manufacturing system	Min. holding cost, batch ordering cost and penalty cost of fulfillment delay	No	MTS/MTO	Stochastic demand, Single product	Located at the middle of production line	Matrix-analytic method, optimization model/Enumeration
Shidpour et al. (2014)	Comparing single and multiple customer order decoupling point	Max. sold product income and Min. holding cost, production cost and back order cost	Yes	MTS/MTO	Stochastic demand, Multiproduct	Located at the middle of production line	Multi-objective mathematical programming
Liu et al. (2014)	Time scheduling of logistics service	Min. operation cost, tardiness and Max. service providers satisfaction	No	MTS/MTO	Deterministic demand, Stochastic multiprocess time	Decision variable	Multi-objective Programming/Genetic Algorithm
Ghalehkhondabi et al. (2016)	Multiple OPPs in a supply chain with uncertain demands in two consecutive echelons	Min. manufacturing cost, holding cost, and waiting cost.	No	MTS/MTO	Stochastic demand, Multiproduct	Decision variable	Mathematical modeling/Optimization

of MTS/MTO manufacturing commonly try to find an optimal balance between the customer satisfaction levels and inventory costs, but to the best knowledge of authors, system idleness and multiple customization lines have not been studied in the literature yet.

3. Problem description

The following notations are used for problem description and mathematical modeling:

Decision Variables:

θ Percentage of product completion in the first g stations

T Number of production lines after OPP

g Number of stations before OPP

Parameters:

λ Arrival rate of the customers to the system

m Number of stations in the production line

μ Production rate of the line with one machine in each station
S Buffer storage capacity of semi-finished products
$V(\theta)$ Per unit value of a θ% completed semi-finished product
DD The mean due date for specific customer order completion
α Setup rate per machine when MTO production is applied
τ Constant fraction of the overall customer order completion delay
C_K The holding cost of semi-finished product
C_H The holding cost of completed products (Produced on pure MTS)
C_{LO} The cost of losing one customer (Due to balking or reneging)
C_B The backorder cost when there is no semi-finished product in the buffer to satisfy the orders
C_D The cost of fulfilling an order after its due date
C_I The cost of idle machines in the MTO line
C_T The cost of constructing a new customization line after the semi-finished products buffer

We consider a production line with m stations, where customer orders arrive to the line following a Poisson distribution with rate λ. The production time of the line with one machine in each station (m machines) is assumed to be exponentially distributed with parameter μ (notations are chosen same as many articles in the literature by purpose) (Jewkes & Alfa, 2009). There is not a setup time for the machines in MTS strategy as they are producing routinely and no change is need. If a machine is assigned to customize a semi-finished product, there would be a same setup time for each product with the rate of α on each machine (all of the processing and setup times follow exponential distribution). It is assumed that there is no restriction on raw material. The manufacturer prefers to produce semi-finished products which are produced on the first g stations (g<m) by MTS and complete the remaining processes (m-g stations) according to specific order of coming customers to the line (MTO). More precisely, in the first g machines θ% of the product is completed and (1-θ)% will be completed on the remaining (m-g) machines. The semi-finished products are stocked in a buffer at OPP place with adequate capacity, but there is not a buffer in stations before or after OPP. More explicitly, the line is balanced and there is no need to have a buffer in each station.

According to above explanations, the time needed to complete a semi-finished product based on a specific demand follows an exponential distribution with mean $1/(\mu/(1-\theta))+1/(\alpha/(m-g))$ which is not dependent on the present customers in the system, and the time needed to complete a semi-finished product based on MTS follows an exponential distribution with mean $1/(\mu/(1-\theta))$. Following the same logic, producing semi-finished products in first g machines follows an exponential distribution with mean $1/(\mu/\theta)$. It is notable θ refers to the cumulative product completion percentage by first g machines. We also use a value per unit function $V(\theta)$, which is introduced by Jewkes & Alfa (2009). This function is an increasing function of θ, which is used to obtain the inventory holding costs.

It is notable that the customization of semi-finished products can be done by T number of production lines after OPP. More explicitly, in order to increase the rate of service in completing semi-finished products and decreasing the waiting time, the manufacturing plant can resume the production after OPP with more than one production line, which is a decision variable in the current study. The cost of constructing a new line is a decreasing function of θ, due to the fact when θ increases, a higher percentage of product is completed before OPP and the number of remaining processing stations is reduced. It is logical that for constructing less number of stations, lower investment is needed.

The customers of this production line have the characteristics of impatient customers including balking and reneging. A customer on arrival to the line may find n customers already in the line (including the one currently being served). Arriving customer may decide to enter the line with probability P_n or do not enter the line with $(1-P_n)$ (which is known as balking in Queuing Theory). According to Bhat (2008), P_n can be obtained as follows:

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

The customers who are in the queue may get impatient and run out of the line without being served (which is known as reneging in Queuing Theory). This waiting time follows an exponential distribution with mean $1/\beta$ (Teimoury et al., 2012). Due to the independence of customers' decision and principals of exponential distribution, it is concluded that the average renege rate in the current system is $n\beta$ (where n is the number of customers in the system).

The object is to achieve the following goals under two scenarios: 1- Obtaining the best place for semi-finished products buffer (OPP). 2- Comparing the cost of MTS/MTO manufacturing and MTS/MTO manufacturing plus MTS when there is no specific order in the system. 3- Finding the optimal number of customization lines after the place of OPP.

3.1. Scenario 1

In the first scenario, semi-finished products are produced by the first g machines (g<m) and stored in a buffer after machine g. The percentage of completion can be considered by θ and the production rate of the system is μ(For whole m machines). So, the production rate to the end of machine g is $\frac{\mu}{\theta}$. No semi-finished product will be completed until a specific order comes to the system. Customers enter the system with a rate of λ. This strategy leads to idle cost, but there is savings obtained in terms of eliminating investment in finished goods inventory and its holding cost. We can calculate the total cost of the system and find the optimal value of g. This value shows us where to put the buffer of semi-finished products. The first scenario manufacturing system can be shown in Figure 2.

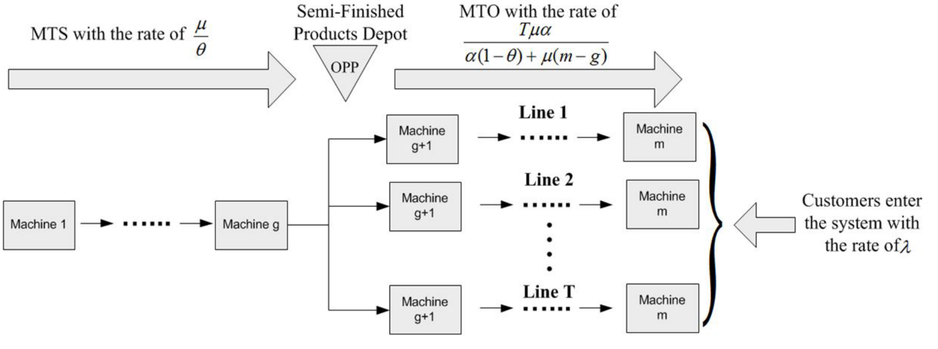


Figure 2. The pure Make to Stock/Make to Order production system (MTS = Make to Stock; MTO = Make to Order.

3.2. Scenario 2

In the second scenario, we use both MTS and MTO for completing the semi-finished products. When there is no customer in the system, semi-finished products are completed based on the MTS strategy and finished products are sent to a warehouse. But, when an order comes in for customization, a semi-finished product get assigned to that order and after finishing the current MTS job on each machine, this MTO job starts to complete the semi-finished product. The second scenario manufacturing system can be shown as Figure 3.

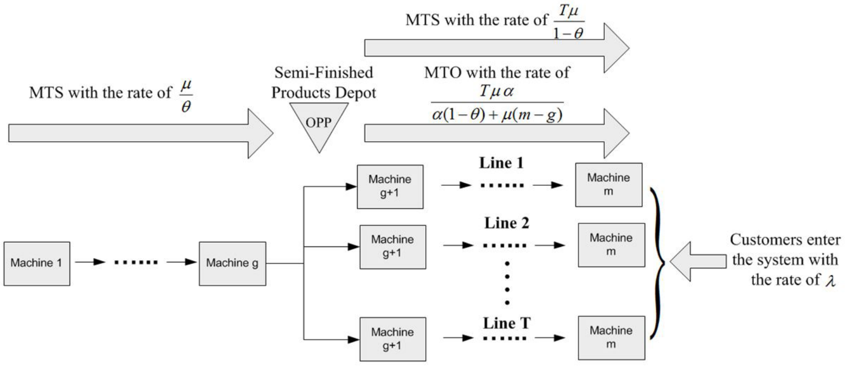


Figure 3. The Make to Stock/Make to Order production system with both Make to Stock and Make to Order completion strategies after Order Penetration Point (MTS = Make to Stock; MTO = Make to Order).

There is a major difference between the two scenarios, where semi-finished products are present in the system but there is no customer to get service. Here, the conflict between the idle cost and inventory cost should be balanced by choosing one of the above mentioned scenarios. In Section 4, main parts of problem formulation and solution approach are presented only for first scenario to avoid the extra writing. Some of important issues of second scenario are explained as well.

4. Problem formulation and system performance measures

In order to obtain the cost of performing under each scenario, there is a need to find the steady state measures of the system. As the arrival orders and producing semi-finished products have a probabilistic nature, bi-dimensional queue is used to model the system.

4.1. State transition diagrams and balance equations

The system consists of probabilistic arrival of customers and semi-finished products, so each of these arrivals can create a queue in the production line. Such a system can be modeled by a quasi birth and death Markov process with the states of (n,k). If n is the number of customers in the system with the one currently being served and k is the number of semi-finished products which are available in the system, we can consider the Markov chain $\{(n,k), 0 \le n \le N, 0 \le k \le S\}$ to present the possible states of present products and customers in the system.

In order to model the long term costs of current production line, it is necessary to obtain the probabilities of steady states $\pi_{(n,k)}$ for both scenarios. According to two-dimensional nature of the queue, we use the Matrix Geometric Method (MGM) and queuing balance equations to find the steady probabilities of product-customer states in the production line (Neuts, 1981). Due to the basis of MGM we need to define a generator matrix for the presented Markov chain as follows:

$$Q = \begin{bmatrix} B_0 & A_0 & & & \\ C_1 & B_1 & A_1 & & \\ & \ddots & \ddots & \ddots & \\ & & C_{N-1} & B_{N-1} & A_{N-1} \\ & & & C_N & B_N \end{bmatrix} \tag{2}$$

In this matrix A_n, B_n and C_n are block square matrices with the order of $S+1$. These matrices are presented in Appendix A for more explanations. A_n displays the rate of increasing one customer in the production line, $B_n\,(n \ne 0)$ displays the rate at which the number of customers stay at the same level (no order enters or exits) and C_n denotes the rate of decreasing one customer in the production line. Finally, B_0 denotes the rate of increasing customers from zero to one.

Each steady-state probability vector π for a constant value of k can be defined as $\pi = [\pi_0, \pi_1, \dots, \pi_N]$ where π_n is calculated by $\sum_{n=0}^{N} \pi_n = 1$ and $\pi Q = 0$. $\pi_n = [\pi_{(n,0)}, \pi_{(n,1)}, \dots, \pi_{(n,S)}]$ is a $1 \times (S+1)$ row vector where $\pi_{(n,k)}$ is the steady probability of a state in which n customers and k semi-finished products exist in the production line. In each scenario, balance equations are derived and by solving the equation $\pi Q = 0$, we can find the steady-state probabilities and following system performance measures. Using all of the above discussions, the first scenario steady state diagram can be depicted as in Figure 4.

Where a is the rate at which semi-finished products are converted to finished products and equals to $\dfrac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}$. The associated balance equations of Figure 3 are given in Equations 3 to 11:

$$(P_n\lambda + \frac{\mu}{\theta})\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}\pi_{(n+1,k+1)} \qquad n=0, k=0 \tag{3}$$

$$(P_n\lambda + \frac{\mu}{\theta})\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}\pi_{(n+1,k+1)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad n=0, 1 \le k \le S-1 \tag{4}$$

$$P_n\lambda\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad n=0, k=S \tag{5}$$

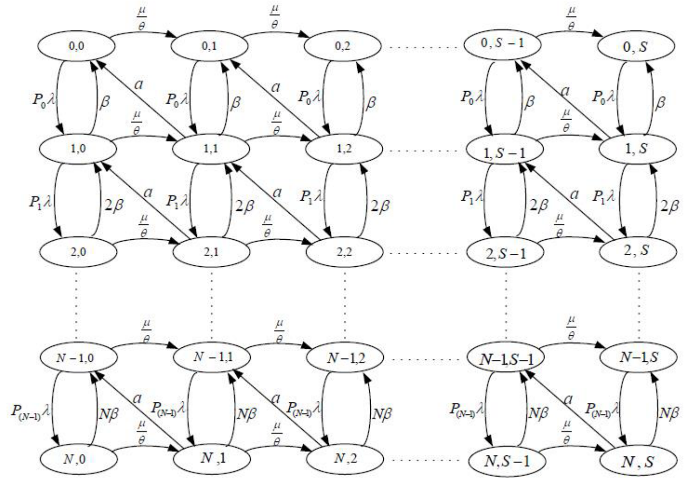


Figure 4. First scenario state transition rates diagram.

$$(P_n\lambda + n\beta + \frac{\mu}{\theta})\pi_{(n,k)} = (n+1)\beta\pi_{(n+1,k)} + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}\pi_{(n+1,k+1)} + P_{n-1}\lambda\pi_{(n-1,k)} \qquad 1 \le n \le N-1, k=0 \tag{6}$$

$$(P_n\lambda + n\beta + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)} + \frac{\mu}{\theta})\pi_{(n,k)} = (n+1)\beta\pi_{(n+1,k)} + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}\pi_{(n+1,k+1)} + P_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad 1 \le n \le N-1, 1 \le k \le S-1 \tag{7}$$

$$(P_n\lambda + n\beta + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)})\pi_{(n,k)} = (n+1)\beta\pi_{(n+1,k)} + P_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad 1 \le n \le N-1, k=S \tag{8}$$

$$(n\beta + \frac{\mu}{\theta})\pi_{(n,k)} = P_{n-1}\lambda\pi_{(n-1,k)} \qquad n=N, k=0 \tag{9}$$

$$(n\beta + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)} + \frac{\mu}{\theta})\pi_{(n,k)} = P_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad n=N, 1 \le k \le S-1 \tag{10}$$

$$(n\beta + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)})\pi_{(n,k)} = P_{n-1}\lambda\pi_{(n-1,k)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad n=N, k=S \tag{11}$$

By all of the previous information, the second scenario steady state diagram is depicted in Figure 5.

Where a is the rate at which semi-finished products are converted to finished products and equals to $\frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}$. We can use the same balance equations as in scenario1 for scenario2, but Equations 3 to 5 have some changes due to additional MTS production rates. We can use Equations 12 to 14 for calculating the steady states probabilities when $n=0$.

$$(P_n\lambda + \frac{\mu}{\theta})\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}\pi_{(n+1,k+1)} + \frac{T\mu}{1-\theta}\pi_{(n,k+1)} \qquad n=0, k=0 \tag{12}$$

$$(P_n\lambda + \frac{\mu}{\theta} + \frac{T\mu}{1-\theta})\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}\pi_{(n+1,k+1)} + \frac{\mu}{\theta}\pi_{(n,k-1)} + \frac{T\mu}{1-\theta}\pi_{(n,k+1)} \qquad n=0, 1 \le k \le S-1 \tag{13}$$

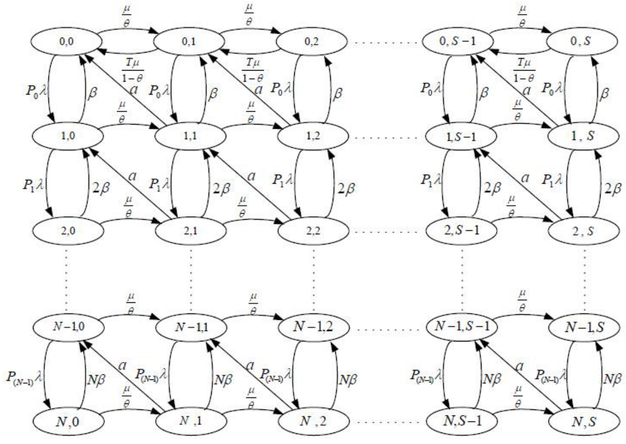


Figure 5. Second scenario state transition rates diagram.

$$(P_n\lambda + \frac{T\mu}{1-\theta})\pi_{(n,k)} = \beta\pi_{(n+1,k)} + \frac{\mu}{\theta}\pi_{(n,k-1)} \qquad n=0, k=S \tag{14}$$

By solving the queuing balance equations, we can find the probability of system idleness and also the probability of states that customers are waiting in the system but no semi-finished product exists in the buffer. By these probabilities, we can define a cost-based objective function which includes cost of idle machines, holding inventory and order fulfillment delay.

4.2. Performance evaluation measures

In order to construct optimization models for two scenarios under study, we need to derive some performance measures of the production line considering the steady state probabilities. Table 4 shows the applied performance measures in the optimization model.

Table 4. Performance measures of the production line.

Performance measure	Formula	Identifier
The expected number of semi-finished products in the buffer	$E(K) = \sum_{n=0}^{N}\sum_{k=1}^{S} k\pi_{(n,k)}$	(15)
The expected time at which the machines after semi-finished products buffer are idle	$E(I) = \sum_{k=0}^{S}\pi_{(0,k)}$	(16)
The mean number of products which are produced based on pure MTS when there is no specific order in the system	$E(H) = \sum_{k=1}^{S}\pi_{(0,k)}$	(17)
The mean number of customers in the system when there is no semi-finished product available to satisfy their order	$E(B) = \sum_{n=1}^{N} n\pi_{(n,0)}$	(18)
The expected number of orders in the line	$E(L) = \sum_{k=0}^{S}\sum_{n=1}^{N} n\pi_{(n,k)}$	(19)
The mean value of waiting time	$E(W) = E(L)/\lambda(1 - \sum_{k=0}^{S}\pi_{(N,k)})$	(20)
The mean value of balking customers	$E(BA) = \sum_{k=0}^{S}\sum_{n=1}^{N}(1-P_n)\lambda\pi_{(n,k)}$	(21)
The mean value of reneging customers	$E(RE) = \sum_{k=0}^{S}\sum_{n=1}^{N} n\beta\pi_{(n,k)}$	(22)
The expected number of lost customers	$E(LO) = E(BA) + E(RE)$	(23)

It is notable, when there is no customer in the system in the second scenario, one of the semi-finished products goes to the line to be completed based on MTS and after producing this product, if a customer with a specific order had come into the system, the line starts to produce on MTO. So, in calculating E(H), steady state probabilities are multiplied by k=1.

4.3. Model of the first Scenario

In order to construct the mathematical models in this study, the approach of Jewkes & Alfa (2009) and Teimoury et al. (2012) is applied. The objective function is based on the defined costs of the system. The total cost of the system should be minimized considering the operational constraints of the system. The mathematical formulation of Scenario 1 is as follows:

$$Min(\theta,T) = C_K V(\theta)E(K) + C_{LO}E(LO) + C_D(E(W) - DD) + C_I T(m-g)E(I) + C_B E(B) + C_T T \tag{24}$$

subject to:

$$\frac{\alpha(1-\theta)+\mu(m-g)}{T\mu\alpha} \geq \tau E(W) \tag{25}$$

$$0 < \theta < 1 \tag{26}$$

$$T = 1,2,3,... \tag{27}$$

Objective function 24 minimizes the total cost of holding semi-finished products in the buffer, cost of lost customers, cost of delay in fulfilling customer orders, cost of idleness for T lines, backorder cost and the cost of establishing T customization production lines. Constraint 25 denotes the service level of the product, where the expected customization time $\frac{\alpha(1-\theta)+\mu(m-g)}{T\mu\alpha}$ must be greater than or equal with a portion τ of the expected order fulfillment time. Constraints 26 and 27 denote the range of model variables. Applying enumeration method to find all possible values of the presented model lets us know the cost trend of the production line and find the best place to put the buffer of semi-finished products. Also, it is possible to find the optimal number of customization production lines to be built in the system. Objective function values in response of different decision variable values are used to compare the performance of the first Scenario and the second Scenario.

4.4. Model of the second Scenario

The mathematical formulation of Scenario 2 is as follows:

$$Min(\theta,T) = C_K V(\theta)E(K) + C_{LO}E(LO) + C_D(E(W) - DD) + C_H TE(H) + C_B E(B) + C_T T \tag{28}$$

subject to:

$$\frac{\alpha(1-\theta)+\mu(m-g)}{T\mu\alpha} \geq \tau E(W) \tag{29}$$

$$0 < \theta < 1 \tag{30}$$

$$T = 1,2,3,... \tag{31}$$

Objective function 28 minimizes the total cost of holding semi-finished products in the buffer, cost of lost customers, cost of delay in fulfilling customer orders, cost of holding completed products by MTS, backorder cost and the cost of establishing T customization production lines. Constraint 29 denotes the service level of the product where the expected customization time $\frac{\alpha(1-\theta)+\mu(m-g)}{T\mu\alpha}$ must be greater than or equal with a portion τ of the expected order fulfillment time. Constraints 30 and 31 denote the range of model variables.

In Section 5, a numerical example is presented to depict the applicability of presented models and a vast sensitivity analysis is performed to understand how the system behaves versus various parameter values.

5. Numerical example

In order to represent the applicability of the presented model and the solution method, a numerical example with parameter analysis is presented in this section. We consider a production line with $m=5$ stations where 20% of product completion is processed in each station. The whole production line with one machine in each station manufactures its products based on an exponential distribution with the parameter of $\mu=1$. The customers of this production line enter to the system following an Poisson distribution with the rate of $\lambda=0.9$.

The line manager wants to choose a strategy between Full MTO, Full MTS and MTS/MTO manufacturing according to the presented scenarios in Section 4. According to system cost parameters, it is desired to find a place for the semi-finished products buffer (finding θ as OPP) and the number of production lines, which can be added as product customization lines to improve the service level of the system (finding T). The system lets at most $N=10$ customers to be in the line (including the one being served) and there are just $S=4$ places to buffer the semi-finished products. The customers of system are impatient and may enter to the system with the probability of P_n as presented by Bhat (2008) and after entering may renege of being served following an exponential distribution with parameter $\beta=0.2$.

There is a setup time for each machine if it wants to customize a product due to a specific order, which follows an exponential distribution with the parameter of $\alpha=40$. Customers of the system have a due date of $DD=0.1$ of time unit and for delayed product customization there is a penalty of $C_D=5$, which is a reward if the product is completed earlier than the due date. Customers are so sensitive to not having a semi-finished product in the line for customization, but there is an opportunity of backordering an order with the cost of $C_B=20$. If a customer prefers to not getting the service from the line when there is not a semi-finished product in the buffer, there would be a customer loss cost of $C_{LO}=6$. Moreover, buffering semi-finished products in the line has a cost of $C_K=0.5$ per item per time unit.

As discussed in the model presentation, there are some specific costs for each scenario. The cost of idleness for each machine in scenario 1 is $C_I=0.5$ and the cost of holding finished products in scenario 2 is $C_H=0.5$. There is a constant fraction of completion delay for the system, which is chosen as $\tau=0.04$ according to customers and market characteristics.

As a logical assumption, the holding cost of semi-finished products is related to the completion percentage of the product (Jewkes & Alfa, 2009), so the product value follows the equation $V(\theta)=\theta\%$ for each semi-finished product. The cost of constructing a new line after the semi-finished products buffer is a decreasing function of product completion, as well. We can use a function format of $a*(1-bV(\theta))$ as the decreasing function of θ, and for the current production line the construction cost is $C_T=0.2*(1\text{-}(0.7*\mathrm{V}(\theta)))$. It is notable that the available space of the line just can be applicable for constructing 5 customization lines after the OPP place ($T_{\max}=5$).

The results of solving this numerical example by MATLAB 7 are presented in Table 5. As it is seen in Table 5, presented model works well for the considered production line. In this example we also want to consider Full MTO and Full MTS options besides the MTS/MTO manufacturing strategy, which is equal to $\theta=0\%$ for Full MTO and $\theta=100\%$ for Full MTS. Since our formulation doesn't work for θ values of 0% and 100%, here $\theta=1\%$ is considered to calculate the total cost of Full MTO strategy and $\theta=99\%$ is used to calculate the total cost of Full MTS strategy. It is noteworthy that the only limitation of our model caused some of our solutions be out of acceptable solution space, which are shown by *NS* in Table 5.

Table 5. Results of numerical example.

T	first Scenario					second Scenario				
	θ^*	g^*	Total Cost			θ^*	g^*	Total Cost		
			Full MTO	MTS/MTO Manufacturing	Full MTS			Full MTO	MTS/MTO Manufacturing	Full MTS
1	80%	4	6.79	*4.58*	10.34	60%	3	6.05	*4.72*	10.52
2	60%	3	6.39	***4.56***	10.87	60%	3	4.40	*3.91*	11.06
3	80%	4	7.37	*5.13*	NS	40%	2	4.01	***3.85***	NS
4	80%	4	8.88	*5.62*	NS	40%	2	4.09	*4.07*	NS
5	80%	4	10.63	*6.15*	NS	20%	1	*4.38*	4.46	NS

Each T value has an optimal total cost which is shown by italic underlined font, and the best answer of each scenario is shown by bold font in Table 5. For example in second Scenario, manufacturing based on the MTS/MTO strategy for $T=5$ is optimal (total cost=4.46) in a condition that we just complete 20% of the product at the first station. But this is not the general optimal case for $T=5$, and we can adopt the full MTO strategy with lower manufacturing cost of 4.38.

Table 5 shows the optimum completion percentage of each semi-finished product with the number of product completion lines after OPP. As it is seen in the first scenario, two product completion lines is the best option based on cost function value. If we consider two completion lines after OPP, the best percentage of product completion before OPP is 60%, which means to place OPP after the third station. Following the above mentioned condition, adopting the MTS/MTO manufacturing strategy with two completion lines and producing 60% completed products has a cost of 4.56 monetary unit in steady state, which is lower than Full MTO and Full MTS strategies due to customer and market characteristics.

In the first scenario, the total cost and completion percentage are both convex in T. The reason of initial decreasing of total cost in T can be explained by the improvement of customer service level. Due to increase in T from one to two, the customization speed of semi-finished products will increase and this will reduce the cost of delay in fulfilling orders and backorders which happened by lower speed of product completion. But, when we increase T from two to three we can see that the total cost of system is increased.

It is notable that even in this case, service level decrease costs by improving the customization speed, but establishing a new completion line has its own cost, which was represented by a decreasing formula of θ. In the case of moving from two customization lines to three customization lines, the related cost of establishing a new production line is more than the reduction of customer service costs. So, we can see that the total system cost starts to increase by establishing more than two product customization lines.

The behavior of θ is also convex by increasing T. It is logical that by increasing T the speed of customization gets faster and the system prefers to do a higher amount of work within the part of the system which performs with a better speed. So, when T increases from one to two, the production line prefers to perform less work before the OPP where just one line is available, and do the more amount of product completion after OPP where two lines are available.

But, with faster lines the customers will get faster service and the idle time of the system increases in the line. This is the reason that the total cost increases when T moves from two to three. More explicitly, when the system has three product customization lines, the completion of customer orders gets done very fast and there will be a huge amount of idle time for machines in three lines. Due to the trade-off between the benefits of faster product completion and the cost of higher idle times, the convex behavior of θ against T is sensible.

As shown in Figure 6, the minimum cost of second scenario is related to a production line with three completion lines after OPP and producing 40% semi-finished products. More explicitly, the OPP place in the second scenario is after the second station and the remaining process is done in three lines, each has three stations. According to customer and market characteristics, the cost of optimal answer in the second scenario is 3.85 in steady state. Comparing costs of two offered scenarios for this production line, it is shown that the second scenario has a lower cost and its better for the system to produce completed MTS products when there is no customer in the system, and bear the holding cost of finished products instead of bearing the cost of idle machines.

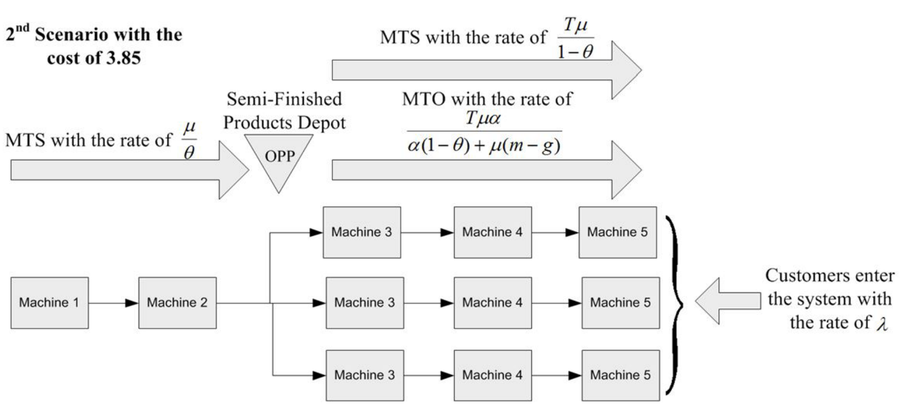


Figure 6. Production line related to second scenario in the numerical example (MTS = Make to Stock; MTO = Make to Order).

In the second scenario, the behavior of total cost against increasing T is the same as the first scenario and can be explained with the same logic. But, the behavior of θ against increasing T is not the same as the first scenario. In the second scenario there is no idle cost for the system and when the system is empty of customers, it starts to produce based on MTS. So, as having more lines let the system to have faster production rates, the system prefers to do less work with low rates and higher amount of work by higher rate. Here, there is not a preventive factor such as idle machines that impose a cost on the system and θ decreases as T gets higher values.

In this section, some other system behaviors are analyzed by various values of system parameters to better show the applicability of model and clarify the concept. Moreover, the logical behavior of system performance measures can be considered as the validity of the presented calculations and the proposed model.

5.1. Total cost variations against λ fluctuations

It is a basic rule in queuing theory to consider the rate of customer arrivals less than the service rate of the system to make the system stable; otherwise the queue will be long and longer to infinity. In current study, there is no need to consider such a rule due to the fact that our system has a finite queue and also the service rate of the system is more than μ due to the several customization lines. In order to analyze the system cost behavior against various values of customer arrival rates, the numerical example is run for different values of λ from 0.1 to 0.7 by the increment of 0.2 in λ value.

According to the convergence of the total cost of the system by increasing λ, we didn't go further than 0.7. But as we noted, here we can even consider λ values greater than 1. The cost behavior of production line under scenario 1 is presented in Figure 7 for possible values of T and $\theta = 60\%$.

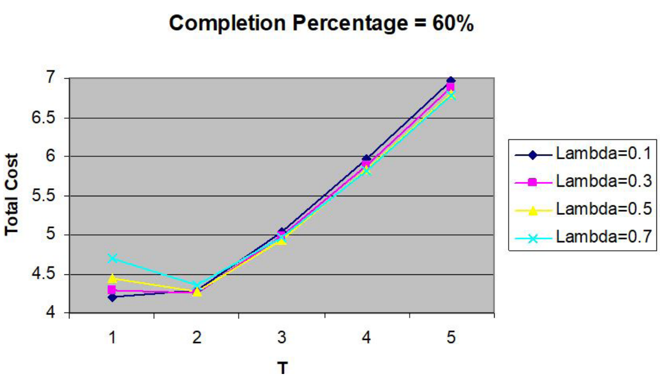


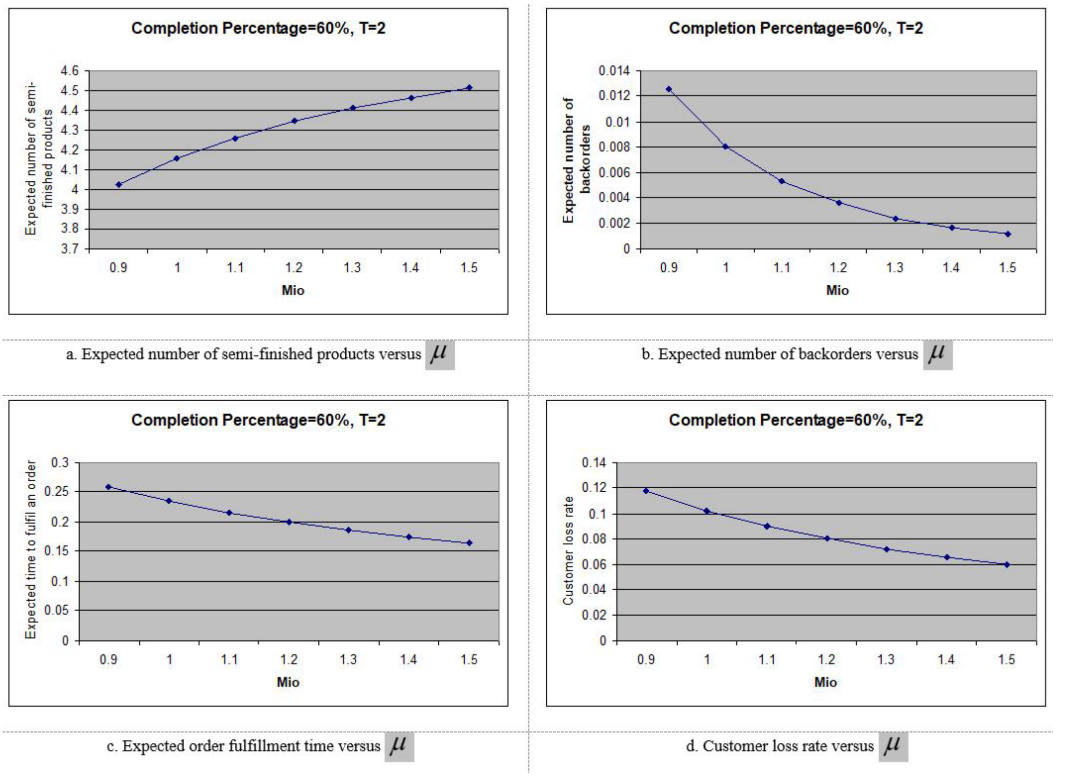
Figure 7. Total cost versus different values of T and λ .

As it is seen in Figure 7, for λ values greater than 0.1 there is a convex behavior for total cost which is vastly explained in the solution of the current example. But for $\lambda = 0.1$, we see a direct increasing trend of the total cost. As the portion of customer arrival rate to system service rate increases, it is logical that the production line needs a faster service rate to increase the service level. In the case of $\lambda = 0.1$, the portion of customer arrival rate to system service rate is very small, so the cost of constructing another line and increasing the value of machine idleness will be greater than the benefit of customer satisfaction which can be obtained by 2 or more customization lines. As a result we can conclude that the line can service the customers with just one customization line and the lowest cost for $\lambda = 0.1$. For higher values of λ, it is necessary to add another customization line.

Another important trend is the higher total costs of lower λ values at larger T values. As it is seen, the total cost of the system for $T = 5$ and $\lambda = 0.1$ is 6.97, but the total cost of the system for $T = 5$ and $\lambda = 0.7$ is 6.78. This behavior can be explained by the same tradeoff between service level and the cost of constructing new lines. Having 5 customization lines for a system with $\lambda = 0.1$ is not profitable, so we can see that higher number of customization lines have lower cost due to more efficient usage of their capacity. It is notable that by increasing T, the difference between the total cost of production line with $\lambda = 0.1$ and $\lambda = 0.7$ gets larger and larger.

5.2. System performance measures versus μ fluctuations

As the customer arrival rate has specific effects on the system cost and its performance, changing production rate can affect the system performance, too. Here, the production rate μ is altered to have a better view of system performance fluctuation versus different values of production rate μ. Different performance measures versus μ are graphed in Figure 8a to 8d.



a. Expected number of semi-finished products versus μ

b. Expected number of backorders versus μ

c. Expected order fulfillment time versus μ

d. Customer loss rate versus μ

Figure 8. System performance measures versus μ.

Figure 8a represents the increasing trend of semi-finished products versus increasing μ. Indeed, increasing μ means a higher number of manufactured products and this issue leads to more entrance of the semi-finished products to the OPP buffer. This higher buffer occupancy is shown as more semi-finished products in steady state. Expected number of backorders has an decreasing trend versus increasing μ, which is shown in Figure 8b. This decreasing trend can be explained by the same logic of semi-finished products trend. When we have higher entrance of products to the buffer, the number of arrived customers who encounter an empty buffer decreases. So, the decrease of expected backorders is a logical result versus the higher values of μ.

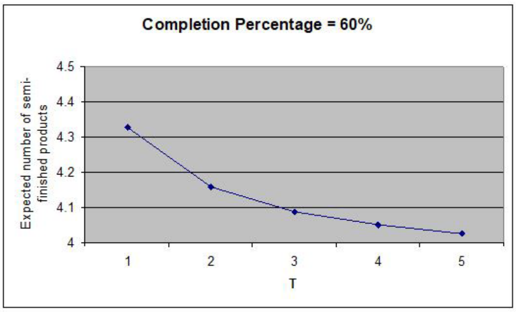
Changes of expected time for fulfilling orders and customer loss rate are shown in Figures 8c and 8d. When the production rate increases, the customer orders are satisfied faster and the queue length gets shorter, which leads to lower waiting and order fulfillment time. The queue length is directly used in the formulas of calculating customer loss rate. As a result, decrease of customer loss rate versus increased μ is a reasonable trend.

5.3. Affect of different numbers of customization lines T on system performance measures

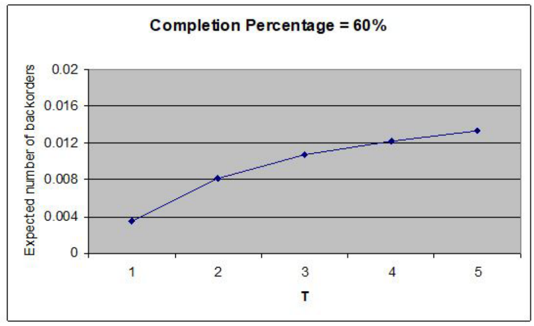
As the production rate and customer arrival rate affected the system performance, number of production lines after OPP can affect the system performance by changing important characteristics of the system, as well. Variations of system performance measures versus the number of product customization lines are depicted in Figures 9a to 9d.

The analysis of system performance measures versus T is similar to the system analysis versus μ fluctuations, due to the fact that the most important effect of increasing T is higher rates of product customization. By these more customization lines, semi-finished products get completed faster and lower amount of semi-finished products will be available at the buffer. As the level of semi-finished products decreases in the buffer, it is logical that more customers will encounter an empty buffer, so the number of backorders increases in steady state.

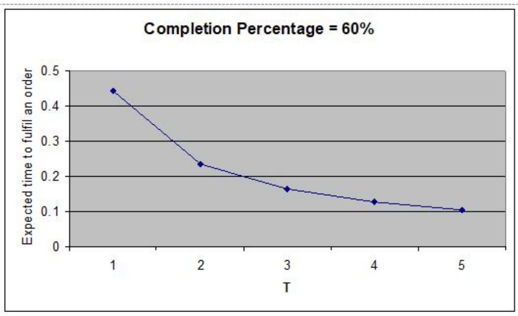
This system behavior versus increasing T is shown in Figures 9a and 9b. Moreover, it is obvious that by more number of customization lines, customer orders get satisfied with a faster rate. This faster rate leads to a shorter customer queue and lower amount of customer loss rate in steady state. These facts are shown in Figures 9c and 9d.



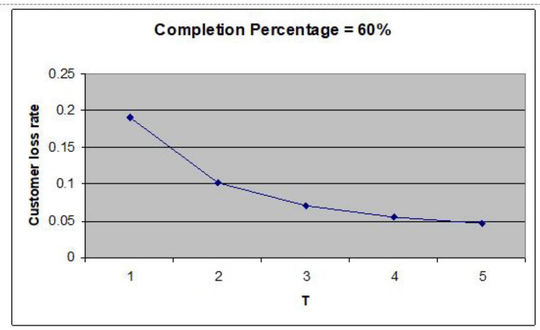
a. Expected number of semi-finished products versus T



b. Expected number of backorders versus T



c. Expected order fulfillment time versus T



d. Customer loss rate versus T

Figure 9. System performance measures versus T .

6. Conclusions

In this article a model is developed to show the application of Make to Stock (MTS)/Make to Order (MTO) manufacturing in a production line to derive the benefits of both MTS and MTO strategies. The model is capable to find the Order Penetration Point (OPP) in the production line, which is the buffer of semi-finished products in the production system. Such as many real production systems, the production processes and customer arrivals have stochastic nature and customers have the possibility of not entering the system due to the long queue or departing the system due to long waiting time.

A two-dimensional queueing model is developed to obtain the performance measures under two different scenarios. The main difference between two scenarios is the utilization of the customization stations, where the system can only customize semi-finished products based on the specific customer orders or customize semi-finished products based on both customer orders and forecast. The objective function is cost-based and lets the production line manager to find the optimal place of OPP, optimal number of customization lines after the OPP and the best scenario to customize the semi-finished products.

The applicability of presented model and scenarios is shown by a numerical example and the observations show the convexity of total cost in terms of product completion percentage and number of customization lines after the OPP. Also, it is shown that increasing the production rate leads to higher expected number of semi-finished products in the buffer, less expected number of backorders, less expected order fulfillment time and less customer loss rate. Observing the system behavior by increasing the number of customization lines after the OPP shows the reduction of expected number of semi-finished products in the buffer, increase of expected number of backorders, reduction of expected order fulfillment time and customer loss rate.

The current study considers the exponential time between the arrivals of customers and semi-finished products. As a possibility of future research, other applicable distributions can be considered to model this problem. Applying Fuzzy theory is another interesting way of studying the uncertain behavior of customers and production line in the studied system. Moreover, the effect of some other market characteristics such as competing companies or substitute products on OPP or the required production rate can be studied in future research.

References

Almehdawe, E., & Jewkes, E. (2013). Performance analysis and optimization of hybrid manufacturing systems under a batch ordering policy. *International Journal of Production Economics*, *144*(1), 200-208. http://dx.doi.org/10.1016/j.ijpe.2013.02.005.

Bhat, U. N. (2008). *An introduction to queueing theory: modeling and analysis in applications*. Boston: Birkhauser. http://dx.doi.org/10.1007/978-0-8176-4725-4.

Donath, B. (2002). *The IOMA handbook of Logistics and Inventory Management*. USA: John Wiley & Sons.

Ghalehkhondabi, I., Sormaz, D., & Weckman, G. (2016). Multiple customer order decoupling points within a hybrid MTS/MTO manufacturing supply chain with uncertain demands in two consecutive echelons. *OPSEARCH*, *53*(4), 976-997. http://dx.doi.org/10.1007/s12597-016-0265-6.

Herer, Y. T., Tzur, M., & Yucesan, E. (2002). Transshipments: an emerging inventory recourse to achieve supply chain leagility. *International Journal of Production Economics*, *80*(3), 201-212. http://dx.doi.org/10.1016/S0925-5273(02)00254-2.

Hoekstra, S., & Romme, J. (1992). *Integral logistics structures: developing customer oriented goods flow*. London: McGraw-Hill.

Jewkes, E. M., & Alfa, A. S. (2009). A queuing model of delayed product differentiation. *European Journal of Operational Research*, *199*(3), 734-743. http://dx.doi.org/10.1016/j.ejor.2008.08.001.

Kalantari, M., Rabbani, M., & Ebadian, M. (2011). A decision support system for order acceptance/rejection in hybrid MTS/MTO production systems. *Applied Mathematical Modelling*, *35*(3), 1363-1377. http://dx.doi.org/10.1016/j.apm.2010.09.015.

Liu, W., Yang, Y., Xu, H., Liu, X., Wang, Y., & Liang, Z. (2014). A time scheduling model of logistics service supply chain based on the customer order decoupling point: a perspective from the constant service operation time. *The Scientific World Journal*, *2014*, 1-22. PMid:24715818.

Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Baltimore: Johns Hopkins University Press.

Olhager, J., & Prajogo, D. I. (2012). The impact of manufacturing and supply chain improvement initiatives: a survey comparing make-to-order and make-to-stock firms. *Omega*, *40*(2), 159-165. http://dx.doi.org/10.1016/j.omega.2011.05.001.

Olhager, J. (2003). Strategic positioning of the order penetration point. *International Journal of Production Economics*, *85*(3), 319-329. http://dx.doi.org/10.1016/S0925-5273(03)00119-1.

Rafiei, H., & Rabbani, M. (2011). Order partitioning and order penetration point location in hybrid make-to-stock/make-to-order production contexts. *Computers & Industrial Engineering*, *61*(3), 550-560. http://dx.doi.org/10.1016/j.cie.2011.04.010.

Sharda, B., & Akiya, N. (2012). Selecting make-to-stock and postponement policies for different products in a chemical plant: A case study using discrete event simulation. *International Journal of Production Economics*, *136*(1), 161-171. http://dx.doi.org/10.1016/j.ijpe.2011.10.001.

Sharman, G. (1984). The rediscovery of logistics. *Harvard Business Review*, *62*(5), 71-80.

Shidpour, H., Da Cunha, C., & Bernard, A. (2014). Analyzing single and multiple customer order decoupling point positioning based on customer value: a multi-objective approach. *Procedia CIRP*, *17*, 669-674. http://dx.doi.org/10.1016/j.procir.2014.01.102.

Soman, C. A., van Donk, D. P., & Gaalman, G. (2004). Combined make-to-order and make-to-stock in food production system. *International Journal of Production Economics*, *90*(2), 223-235. http://dx.doi.org/10.1016/S0925-5273(02)00376-6.

Süer, G. A., & Lobo, R. (2011). Comparison of connected vs. disconnected cellular systems: a case study. In V. Modrák, R. Sudhakara Pandian (Eds.), *Operations management research and cellular manufacturing systems: innovative methods and approaches*. Hershey: IGI Global.

Sun, X. Y., Ji, P., Sun, L. Y., & Wang, Y. L. (2008). Positioning multiple decoupling points in a supply network. *International Journal of Production Economics*, *113*(2), 943-956. http://dx.doi.org/10.1016/j.ijpe.2007.11.012.

Teimoury, E., Modarres, M., Khondabi, I. G., & Fathi, M. (2012). A queuing approach for making decisions about order penetration point in multiechelon supply chains. *International Journal of Advanced Manufacturing Technology*, *63*(1-4), 359-371. http://dx.doi.org/10.1007/s00170-012-3913-x.

Wang, J. L. (2009). A supply chain application of fuzzy set theory to inventory control models –DRP system analysis. *Expert Systems with Applications*, *36*(5), 9229-9239. http://dx.doi.org/10.1016/j.eswa.2008.12.047.

Wu, D., & Olson, D. L. (2008). Supply chain risk, simulation, and vendor selection. *International Journal of Production Economics*, *114*(2), 646-655. http://dx.doi.org/10.1016/j.ijpe.2008.02.013.

Wu, M. C., Jiang, J. H., & Chang, W. J. (2008). Scheduling a hybrid MTO/MTS semiconductor fab with machine-dedication features. *International Journal of Production Economics*, *112*(1), 416-426. http://dx.doi.org/10.1016/j.ijpe.2007.04.008.

Zhang, Z. G., Kim, I., Springer, M., Cai, G., & Yu, Y. (2013). Dynamic pooling of make-to-stock and make-to-order operations. *International Journal of Production Economics*, *144*(1), 44-56. http://dx.doi.org/10.1016/j.ijpe.2013.01.012.

Appendix A. Matrix Geometric Method open formula.

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

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

$$\left.\begin{aligned} & B_{n_{k,k}} = \begin{cases} -(\dfrac{\mu}{\theta} + P_n\lambda + n\beta) & k = 0 \\ -(\dfrac{\mu}{\theta} + P_n\lambda + n\beta + \dfrac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}) & 1 \le k \le S-1 \\ -(P_n\lambda + n\beta + \dfrac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)}) & k = S \end{cases} \\ & B_{n_{k,k+1}} = \dfrac{\mu}{\theta} \qquad 0 \le k \le S-1 \end{aligned}\right\} \quad 1 \le n \le N-1 \tag{A.3}$$

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

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

$$C_n = \begin{bmatrix} n\beta & & & & \\ \dfrac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)} & n\beta & & & \\ & \ddots & \ddots & & \\ & & \dfrac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)} & n\beta & \\ & & & \dfrac{T\mu\alpha}{\alpha(1-\theta)+\mu(m-g)} & n\beta \end{bmatrix}, \qquad 1 \le n \le N \tag{A.6}$$