Hierarchical Production Planning in Make to Order System Based on Work Load Control Method

Ehsan Fattahi^{1,*}, Mahta Khodadad²

¹Mazandaran University of Science and Technology, Iran

²Islamic Azad University, Firouzkoh Branch, Iran

*Corresponding Author: ehsan_f60@yahoo.com

Copyright © 2015 Horizon Research Publishing All rights reserved.

Abstract Subject to competition markets, continues change in expectations, customer requisition and decrease of accessible production resource, the production system of make to order is the appropriate selection for response to today's markets specifications and in order to provide the large spectrum of customer's orders. In this research expend the control and planning production system (work load control) in make to order production system to manage the delivery time. For this, we recommend the hierarchical production planning structure. The major target is managing the activity in make to order system in order to achieve the short released time and entrance orders competition fee. To achieve this target we present the appropriate decision models with attention to restriction of production system. Finally due to consequent below: Present the expended release model to plain the production line and decrease the delivery time; Present the orders timing method in various stations. This method do subject to existences of previous level. Synchronistic performed the two methods state better performance in compare to others researches.

Keywords Make to Order System (MTO), Make to Stock System (MTS), Work Load Control (WLC), Hierarchical Production Planning (HPP)

1. Introduction

"Production Management Systems" (PMS) involves the need for a systematic look at the production. In absence of such a holistic look, production will be doomed to failure. Management of production, which is in a higher level than the production planning, is a way to implement a production schedule by managing all its components. In the above expression, the word "Management" is based on the fact that production planning is not instrumental alone, unless each component of the production schedule is managed and arrives at a result. Finally, the word "Production" is based on the experience that attention to production as an effective competitive weapon can anticipate firms relative to

competitors [1]. With this backdrop, we can say that PMS seeks with a holistic perspective and a proper planning to identify all effective elements and complexities of the production process to deliver a product on the market that increases customer satisfaction.

Today, manufacturing factories, considering the role of production in the competitive market system, find themselves in an environment totally changed. An example of this change can be observed in different industries of consumer goods and capital goods such as automotive, electronics, and appliances. A manager who is facing these rapid changes will have to adopt new approaches to address this competitive environment. Old strategy of integrated production planning has lost credibility and given itself up to the flexibility of new approaches. New approaches in production management have reduced the time required to designing and marketing products and the time between orders and their delivery to customers. Key features of new production environments include [2]:

- a. Increase of product variety
- b. Severe reduction of the product life cycle
- c. Change of social expectations
- d. Change of patterns of costs

Considering these characteristics and to integrate themselves with new conditions, manufacturing firms must plan and manage different levels of their schedules so that they get the flexibility to confront unpredictable market changes.

In this paper, we used the term "Custom Order Decoupling Point" (CODP) to express the difference between various production environments. CODP is a point where all the following production processes are assigned to meet customer demands. In other words, all production processes prior to this point are the basic warehouse, and the subsequent processes are the basic order. This is also called "Inventory/order interface point" (I/O). The more this point is close to the end of production line, the delivery time will be reduced and the ability to meet customer demands and the stock of raw materials and semi-finished products in workstations before this point will be increased. However,

the more it's close to the beginning of production line, the stock of raw materials and semi-finished products is reduced and the product delivery time will be increased.

So far, several models of hierarchical production planning (HPP) in different manufacturing environments to improve the performance of manufacturing management were proposed. Kingsman et al. [13,14] have proposed a structure to manage orders which must be received by the make-to-order (MTO) system. In this case, some of the incoming orders are acceptance, and some of them are rejected. This stage in the MTO literature is called the order entry stage or customer inquiry stage. In this model, entry process is divided in two phases: the client application and the order entry system. Cakravastia et al. [15] was assuming that the order has been accepted by the customer, price and delivery time should be determined through talking between the customer and suppliers. Caloso et al. [16,17] provided the decision making structure for negotiation between customers and suppliers in the supply chain. In these articles, three correct number linear models have been proposed to select the suppliers and orders which have highest profits for the company. Breithoupt et al. [18] determined the sensitivity of the HPP structure to parameters such as workload high limits and time horizon and identified critical parameters. Then, a new method for calculating these parameters were presented to improve the output of the HPP structure. Rabbani et al. [19,20] proposed the HPP structure for combined MTS/MTO environments. As a novelty of this research in comparison to the literature review, the HPP structure was proposed in four levels.

2. Make to Order (MTO) System

In these environments, CODP precedes the production process of semi-finished products. Composites are designed and manufactured according to customer demand. Product selection in MTO production environments is made by customers. The implementation of an environment which engages in production only after the customer orders are received where all workstations produce only the raw materials required for the completion of customer orders, and where the stock of raw materials in all workstations are very low even zero, a coordination is vital between different levels of the factory, reduced time of supply, high flexibility of the production line, and use of complicated and expensive equipment. The planning of such production systems is a difficult and complicated task [3].

3. Hierarchical Approach of Production Planning in Production Management System

Given the major defects of the integrated approach, this perspective will not be a good solution to manage production in today's organizations. In general, planning and control of

production and raw materials is not a single activity that can be managed by the integrated approach. By cons, it depends on a series of decisions about various topics in the production system. These decisions are taken generally in different levels of management according to the hierarchy of production activities, so that any compliance with this hierarchical structure of decision-making will tighten the horizon of decisions but will increase the details of the data needed to make decision. To better manage various activities of production, a close relationship between the decisions taken at successive levels is necessary.

Hierarchical Production Planning (HPP) is an effective method to perform this multilevel process of hierarchical decision-making. In this hierarchical approach, the main problem of decision-making is decoupled into sub-problems at different levels of decision-making; then these sub-problems are solved in a given sequence according to the different levels of hierarchical structure. Also, to maintain the integrity of the main problem and the interconnection of these sub-problems, the response at each level of the hierarchical structure will be considered as a limitation in the decision-making model at a lower level [4].

In general, a bottom-top feedback mechanism is used in a HPP structure to improve decisions made at higher levels [5].

Figure 1. Hierarchical Structure of Decision-Making in Production Management

4. Structure of the Proposed System of Planning and Production Control (WLC) in MTO Production Environments

Work load control (WLC) is an improved approach used to have a successful system of planning and production control in MTO production environments with respect to the hierarchical approach. In other words, we propose in this paper a hierarchical structure of WLC in order to better control the activities that affect the delivery time of customer orders.

The proposed structure for hierarchical production planning (WLC) in MTO production environments provides

four different levels of planning.

4.1. First Level - (Customer Enquiry Stage)

In this stage, given the demands of the customer, we check if MTO firm is able to meet the demands of a customer given the limitations of the system, new orders, and degree of signification of the customer. The output of this level of WLC is to accept or reject orders.

4.2. Second Level - (Order Entry Stage)

In this stage, if the customer's orders are not rejected at the first level, three final decisions will be made:

- Determine the price and delivery time of new orders (delivery is determined only if it is negotiable).
- Plan the ability to accept new orders.
- Choose a group of contractors who are able to provide necessary raw materials for new orders in a timely manner and a reasonable price.

4.3. Third Level – (Release of Orders into the Shop Floor)

This stage includes only accepted orders for which raw materials are available (output of the second level). In this stage, given the delivery time of orders and also the density of the work in different workstations, delivery of new orders in the system will be permitted.

4.4. Fourth Level – (Prioritizing Orders in Workstations)

In this stage, orders awaiting delivery in workstations are given priority so that their delivery time is minimized.

For a better communication and improved efficiency of the production system, feedbacks are taken into account from lower to higher levels.

5. Distinguishing Features of the Research

The distinctive features of this research compared to other researches carried out in this area are as follows:

- Consider the significance level of customers to decide whether to accept or reject new orders by appropriate analytical tools.
- Introduce a new structure of decision-making to accept or reject orders at the first level and determine the delivery time and the price of incoming orders at the second level, by an integer mathematical model.
- Consider all components of the supply chain in the delivery and price control including customer MTO firm and contractors to reject or accept orders received at the MTO system.
- Offer alternatives to determine the delivery time of entry orders and thereafter, different prices for these

orders, and facilitate the negotiation process between customer and MTO system are the most difficult stages of decision-making to reject or accept orders.

- Select contractors to perform works related to new orders received in the system at a given price and delivery time by an integer mathematical model of planning.
- Extend the process of order release into the shop floor to access the shorter delivery time and improve the efficiency of production line by reducing the density of works according to WLC perspective.

6. Delivery Time Components in MTO Systems

Delivery time components in MTO production environments include the following:

- Negotiation Time (NT): The time between the arrival of an order (customer request) and the decision on its acceptance or rejection.
- Raw Material Lead Time (MLT): The time between putting in an order for raw materials needed and the time of its acceptance.
- Product Delivery Date (PDT): The time of entry of raw materials into the system, namely, when orders are ready to be released. These kinds of orders are placed at the warehouse specified for orders to be released. The waiting period between the entry of raw materials and the time of their delivery into the production system is referred to as waiting period in the warehouse of orders to be released.
- Order completion date in the shop floor (SFTT): The time between the release of orders into the production system and delivery of customer orders.

Total Time Manufacturing (TMT) includes SFTT and PDT. Finally, the Due Date (DD) which includes the total TMT, MLT and NT.

Equation (1-3) shows how to calculate the TMT and DD.

$$TMT = SFTT + PDT$$

$$DD = NT + MLT + PDT + SFTT$$

$$... = NT + MLT + TMT$$

7. Variables Examined in this Study

7.1. Dependent Variables

- Costs of a production system include the costs of production variables in normal time, overtime, sub-contracting, and the costs of postponed orders where delivery delay is permissible.
- Different costs of manufacturing an order in addition to purchase costs of raw materials and workload of contractors, and also the penalty payable for delivery earlier or later than desirable by contractors.

7.2. Independent Variables

- Total capacity of workstation r including normal time, overtime and sub-contracting assigned to order i in period t .
- A certain capacity of workstation r assigned to order i during overtime.
- Degree of delay in delivery of order i in due date.
- Order i completion Date.
- A certain capacity of workstation r during overtime is assigned to order i .

8. First Level of the Proposed WLC Approach: Customer Inquiry Stage

At this level of the proposed structure, we will discuss the question of accepting or rejecting new orders. This decision is based on three main criteria including total capacity available during planning, customer significance level for the organization, and order characteristics such as the workload required on the production line.

This model is divided into two levels depending on order delivery time. The first level consists of two steps: In the first step, incoming orders are grouped according to the number of customers who made an order. In the second step, the ability of production system and all orders are generally compared to see whether new orders can be completed within the planning horizon. If the current capacity can meet new orders, the orders enter into the second level; otherwise, other decisions are made. The steps mentioned in the first level are the same for both types of orders with negotiable and fixed delivery time.

8.1. First Step - Grouping Incoming Orders According to Customer Significance for the Organization

Group 1 - The organization completes customer orders in the best condition.

- Fixed and long relationship with the organization.

Group 2 - The competition between different manufacturing organizations to boost customer satisfaction.

- Improving the functioning of the organization leads to increased sales.

Group 3 - Casual orders

- Weak relationship and lower ability to deliver orders at due date.

Group 4 - Irregular and small amounts of purchases

8.2. Second Step - Capacity Estimated Calculation

In this step, each time a new order comes in, the sufficiency of current capacity is considered within the planning horizon at the first level. To this end, estimated calculation is performed as follows:

- If an order is highly significant, the following relation is used to estimate the capacity thereof:

$$\sum_{i \in O_r} (TWK_{ir} \times p_i) \leq C_{rt};$$

$$\forall r, t = (t_{now}, \dots, T)$$

i : Order index

r : Workstation index

O_r : All orders needed in workstation r

t : Planning period index

t_{now} : Time now

T : Length of planning horizon at the first level

C_{rt} : Maximum capacity of workstation r in period t including normal time, overtime, and sub-contracting.

TWK_{ij} : Workload required by the order i in workstation j .

P_i : Possibility to accept order i put in by a customer. During the entry of a new order, three kinds of order exist in the output system: orders approved by the customer and all of which are released in the production system (RO); orders approved by the customer which are not yet released in the production system and are planned to be backlogged (PO). The third kind of order is contingent order (CO) in which delivery time and price are determined by the production system, whose approval or cancellation is not yet decided by the customer. The new order is of the third kind. Contingent orders may be finally canceled by the customer. P_i -value for orders approved by the customer (RO and PO) is 1. For other orders (CO) which are not yet canceled or accepted by the customer, this value will be 0 or 1 using previous data and experiences of marketing and sales unit.

If for the new order, which is highly significant, there is not sufficient capacity in some workstations, the following alternatives are suggested for decision making about the order:

- Increase the capacity of production system to prepare a new order: due to the significance of the new order in workplaces where there is not sufficient capacity within the planning horizon T , basic decisions are taken by management to increase the capacity.
- Delay certain orders in the system that are of medium significance: in this alternative, we examine whether the delay in preparation of some orders of medium significance gives us the possibility to complete them at the planning period.
- Cancel orders that are unimportant (or of low significance) and are not yet approved by the customer (contingent orders). This way we will have the capacity to complete the new order.
- Reject the new order: the new order can be rejected if its completion requires considerable additional capacity which is unavailable.

- If an order is of small or medium significance, the following equation is used to calculate the estimated capacity:

$$\sum_{i \in O_r} (TWK_{ir} \times p_i) \leq C_{rt} \times (1 - \alpha_{rt});$$

$$\forall r, t = (t_{now}, \dots, T)$$

α_{rt} is a percentage of total C_{rt} capacity that is considered for highly significant orders in future. This value is determined by the management of organization according to future orders forecasted by sales and marketing unit. If the above equation is not established, decision making on this order will include the following cases:

- Delay new orders of medium significance: This alternative can be used only when the delay of an order allows its fabrication within the planning period T .
- Rejecting a new order of low or medium significance: In case the completion of a new order requires considerable additional capacity which is unavailable, this order may be rejected.
- Delay some PO and RO with medium significance: At manager's discretion, some PO and RO with medium significance can be postponed to create sufficient capacity for new orders.

9. Second Level of the Proposed WLC Approach Order Entry Stage

The second level of the proposed structure includes determining the delivery (if unspecified) and price of orders arrived in the production system according to various limitations and inside and outside of organization. Two main criteria for deciding the acceptance or rejection of new orders is their price and delivery time. These two criteria are highly interdependent. If the delivery time is short and compact, in addition to normal production time, production system shall adopt other policies, including overtime, delaying other orders in the system, employment of new staff and sub-contracting to complete the order. All these policies impose costs on the firm and affect the cost price of new orders.

The second level consists of four steps. If the order is not rejected at the first level and if the delivery of the new order is fixed, we use backward method in the first step of the second level to calculate the latest or earliest release time inside the production system. These times will be compared to the average delivery time and the workload required by contractors to make appropriate decisions.

In case delivery time is negotiable, there will be suitable alternatives to create different delivery times. If an order is accepted in the first stage, we use a mathematical planning model in the second stage that we call IP1 to calculate the price and delivery time.

As for orders with stable delivery time, cost price will be announced to the customer, and for orders with negotiable delivery time, the delivery time will be made known to the customer in addition to the price. In the third stage, the

customer compares these data with those from other MTO firms and takes into account its own criteria to determine the rejection or acceptance of the order. In the fourth stage, a good group of contractors will be selected by a mathematical model of planning (IP2) to perform the workload required for the delivery of accepted orders.

Since the implementation of the second and fourth stages is time consuming because of the use of integer mathematical models, these stages are completed according to the following conditions:

- A significant order enters the system.
- When the number of new orders arrived in the system exceeds a certain limit.
- When the difference between the time the first order comes in exceeds a certain limit after completion of the second stage.

9.1. Deciding on Rejection or Acceptance of Orders with a Stable Delivery Time

If a new order has a stable delivery time, the status of the new order will be determined in four stages:

- Create different alternatives to calculate the price of the new order
- Resolve the integer linear model (IP1) to determine the price of the new order
- Decide to accept or reject the order by the customer
- Select contractors to provide raw materials and required workload to complete the new order.

9.1.1. Creating Different Alternatives to Calculate the Price of the New Order

After considering the estimated capacity to determine if an order can be completed within the planning horizon of the first level, we engage in first stage of the second level. To run the integer planning model (IP1) and determine the price of the new order, it is necessary to specify the input parameters of the problem including Order Completion Date (OCD) at each workstation, Earliest Release Date (ERD) and Latest Release Date (LRD). If the order is sent into the system after LRD, it will not be possible to complete the delivery at due date.

Since delivery time is not known, the backward method is used to calculate the OCD and ERD. In previous research, the following formula was used by Kingsman *et al* [5]. to calculate OCD and ERD:

$$OCD_{i,\mu(i,n_i)} = dd_i$$

$$OCD_{i,\mu(i,n_i-1)} = OCD_{i,\mu(i,n_i)} - TWK_{i,\mu(i,n_i)} - W_p$$

$$OCD_{i,\mu(i,r)} = OCD_{i,\mu(i,r+1)} - TWK_{i,\mu(i,r+1)} - W_p$$

$$LRD_i = OCD_{i,\mu(i,1)} - TWK_{i,\mu(i,1)} - W_p$$

$$ERD_i = LRD_i - pool\ delay$$

n_i : Number of workstations on the operational route of order i

$\mu(i,r)$: Workstation r on the operational route of order i

dd_i : Delivery time of order i

C_{rt} : Order i completion date in resource r

$Pool\ delay$: Total workload of all PO orders

W_p : Waiting time in each workstation for one work of priority p

p : p includes both normal and high priorities.

For orders of medium and low significance, normal priority is taken into account, but for orders of high significance, in addition to the normal priority, if delivery of the order is not possible at dd , we can complete the order by reducing the (high priority) waiting period. Due to the capacity limitation of the temporal warehouse of workstations, a percentage of orders of high significance can be given high priority. Wp can be calculated by the following equation :

$$Wp = \frac{T - (\sum_{r=1}^R \sum_{i \in O_r} TWK_{ir})}{R'}$$

In the above equation, R' is the average number of workstations to complete an order.

Order completion date in a workstation is determined by reducing the waiting period (Wp) and workload required by the order (TWK) from order completion date in the next workstation. This formula can determine relatively accurate dates for ERD and OCD only when there is sufficient capacity in each workstation.

In addition to the above equation, the following is presented for the first time given the capacity of individual workstations:

$$OCD_{i,\mu(i,n_i)} = ST_{i,\mu(i,n_i)} + TWK_{i,\mu(i,n_i)} \times P_i + B_n = dd_i$$

$$OCD_{i,\mu(i,n_i-1)} = ST_{i,\mu(i,n_i-1)} + TWK_{i,\mu(i,n_i-1)} \times P_i + B_{(n-1)};$$

$$OCD_{i,\mu(i,n_i-1)} \leq ST_{i,\mu(i,n_i)}$$

$$OCD_{i,\mu(i,r)} = ST_{i,\mu(i,r)} + TWK_{i,\mu(i,r)} \times P_i + B_r;$$

$$OCD_{i,\mu(i,r)} \leq ST_{i,\mu(i,r+1)}$$

$$OCD_{i,\mu(i,1)} = ST_{i,\mu(i,1)} + TWK_{i,\mu(i,1)} \times P_i + B_1;$$

$$OCD_{i,\mu(i,1)} \leq ST_{i,\mu(i,2)}$$

$$LRD_i = ST_{i,\mu(i,1)}$$

$$ERD_i = LRD_i - pool\ delay$$

ST_{r_i} : Is the earliest time when workstation r has the

sufficient capacity to complete the order i . This time, in the example of the previous page, is the third day of the ninth week. Order completion date in the previous workstation is less than or equal to that date, which is written as an "unequal" in front of each equation.

B_r : Is the waiting period at workstation r before the necessary operation is accomplished in this station on the new order. Queuing at each workstation is separated into two parts:

- Queue of order with normal priority: In this status, the new order is placed at the end of the queue in each workstation.
- Queue of order with high priority: This queue consists of two parts: One is for orders of medium and low significance, and the other for orders of high significance. The first part precedes the second. The stage where a high priority is given to new orders is described below.

Because of capacity problems, calculated values of OCD and ERD will be more accurate and the possibility of a reliable delivery time will be increased. In addition to capacity, the calculation time necessary for the orders according to the possibility of their acceptance (P) is a further advantage of the proposed formula.

Prior to the implementation of IP1 model (second stage), it is necessary to compare the Raw Material Arrival Date (MAD) by contractors and LRD and ERD values to ensure that it is possible, considering the MAD, to complete the new order at due date. At first and second level of the proposed WLC approach in MTO systems, medium term planning system is made and we do not need more accurate data. We only calculate an average value of MAD according to the performance of contractors in the past. Where raw materials arrive on time into MTO system, there is no need to prioritize or make other additional works including overtime and sub-contracting, and the order will be placed at the end of the queue of orders to be released. Otherwise, if raw materials do not arrive to the firm at due date, the following solutions are proposed for better decision-making and management of new orders. To this end, t_{now} , LRD , ERD and MAD values will be compared. The following status will occur when comparing these values:

A- $LRD_i < t_{now}$: In this case, it is not possible to complete the order at due date, and it is better to reject the order.

B- $ERD_i < t_{now} < LRD_i$: Under such conditions, it is not possible to complete the order in a normal status. The order ($MAD-ERD$) will be completed with delay (assuming that this order is at the end of the queue of orders to be released). Here are the alternatives proposed to deliver the order at due date:

- Increasing the priority in the queue of orders to be released: As indicated in the above equations, new orders are queued in order of arrival at the end of the queue; in other words, these orders will be given a normal priority. Reducing the waiting period for backlogged orders is one of the methods to access a

shorter delivery time. This reduction will allow this relation $ERD = MAD$ to be established. This reduction also increases the priority of orders waiting to be released. In other words, in this alternative, the order is given a higher priority at least to the point of $MAD - ERD$. The new value is equal to Pool Delay:

$$Pooldelay' = LRD_i - MAD$$

- 2) Change in OCD values: Change of OCD in different workstations is another way to reduce delivery time. This change can be performed by three ways:
 - a. Changes in the values of OCD for a new order by increasing the capacity of workstations: In this status, all workstations that are necessary to accommodate the new order are considered and, in case of increased capacity, new OCD values will be calculated. The change in OCD values should be calculated as the new ERD is equal to MAD. This alternative does not change the OCD for other orders in the system.
 - b. Changes in the values of OCD for a new order through changing the OCD for other orders: Orders where ERD is greater than MAD, it is likely to change the values of OCD, without their delivery time is altered. By changing the OCD values for other orders, capacity for the new order will be increased; with this new capacity, OCD values of the new order will be reduced and, subsequently, ERD will also be reduced. Raw materials required for orders where ERD is greater than MAD will be entered into the system before due date and, therefore, we need to keep the raw materials in the relevant warehouse. This alternative will reduce the waiting period for raw materials needed by such orders and, therefore, the cost of raw materials warehoused will be reduced, which is one of the goals of MTO systems.
 - c. Changes in OCD values using the capacity assigned to other orders: by assigning the capacity for orders in the production system, it is possible to achieve a new order where ERD value is less than that of MAD, thereafter, reduce the values of OCD in individual workstations and accelerate the delivery. This alternative changes the OCD values where additional capacity for the new order is created. Using the capacity assigned for these orders will delay their delivery. This additional capacity will be calculated so that the new ERD is at least equal to MAD. This alternative is used for new orders of high significance.
- 3) High priority in the queue between different sources of production: Giving high priority to a new order, it will be possible to reduce the waiting period and thus delivery time. High priority for orders of low significance include placing the new order at the end

of the queue of the first part, and for orders of high significance include placing the new order at the end of the queue of the second part. Reducing waiting period in different workstations shall be such that the relation $ERD = MAD$ is established.

- 4) Lateness: If possible, orders can be completed behind schedule. This alternative is only possible for orders of less significance.
- 5) If it is not possible to use such alternative and if there is no other suitable solution for the delivery of the new order at due date, the order will be rejected at the discretion of management.

It is also possible to use a combination of these alternatives to complete the order at due date.

$$C- t_{now} < MAD_i < ERD_i$$

Raw materials are delivered on time and the order is placed at the end of the queue for orders waiting to be released.

D- $ERD_i < MAD_i < LRD_i$: This status resembles the status (B).

$$E- MAD_i > LRD_i$$

In this status, MAD is later than LRD. Here, it is difficult to prepare the order on schedule and it is highly possible that the order is postponed. All such alternatives of the second part are also available in this status, with the difference that the level of reduction of delivery time will be so that MAD and LRD values are equal.

9.1.2. Resolving the Integer Linear Model (IP1) to Determine the Price of the New Order

At the first level of hierarchical production planning model, integer linear model is used to calculate the price of incoming orders where the delivery time is specified. If the order is not rejected by the MTO system in the second stage, in this stage, an integer linear model is used to calculate various costs imposed on the system because of the new order. Final price of the order is calculated by totaling costs. As in the previous stage various OCD and ERD values were calculated by different alternatives, therefore, IP1 model was applied in exchange for each set of values (OCD and ERD). In addition, the lowest price of the order will be announced to the customer. The optimal value (OCD and ERD) is chosen for each optimal response at IP1 model. Indexes, parameters and variables of the decision taken by the IP1 model are listed:

Indexes:

i : Order index ($i = 1, \dots, n$)

r : Workstation index ($r = 1, \dots, R$)

t : Time index ($t = 1, \dots, T$)

Input parameters:

cr_{irt} : Manufacturing cost of order i in workstation r in normal time and period t

co_{irt} : Manufacturing cost of order i in workstation r during overtime in period t

cs_{irt} : Manufacturing cost of order i in workstation r by contractor in period t

ct_i : Cost of delaying the order i per unit of time

iwp_r : Total workload for orders behind the schedule

iw_{irt} : Total workload of order i in workstation r with a ST time in period t (coming into the system)

ow_{irt} : Total workload of order i in workstation r with an OCD in period t (coming out of the system)

CR_{rt} : Maximum capacity available in workstation r in normal time and period t

α_{rt} : A percentage of CR_{rt} total capacity is taken into account for orders of high significance.

CO_{rt} : Maximum capacity available in workstation r in period t during overtime

CS_{rt} : Maximum capacity available in workstation r in period t by contractors of the firm

$OS(i)$: All orders that are completed on time.

M : A very large number.

Decision variables:

Y_{irt} : Total capacity in workplace r including the normal time, overtime and sub-contracting assigned to order i in period t .

O_{irt} : A certain capacity of workstation r assigned to order i during overtime.

S_{irt} : A certain capacity of workstation r assigned to order i by contractors.

LT_i : Delay of delivery of order i

FT_i : Order i completion date

$$Y_{i,\mu(i,n_i),t} > 0$$

Otherwise $x_{it} = 0$

$$MIN \quad Z = \sum_{i \in O_r} \sum_{r=1}^R \sum_{t=1}^T [cr_{irt} \times (Y_{irt} - O_{irt} - S_{irt}) + co_{irt} \times O_{irt} + cs_{irt} \times S_{irt}] + \sum_{i \notin OS(i)} ct_i \times LT_i$$

ST .

$$(1) \quad \sum_{i \in O_r} (Y_{irt} - O_{irt} - S_{irt}) \leq CR_{rt} \times (1 - \alpha_{rt}); \forall r, t$$

$$(2) \quad \sum_{i \in O_r} O_{irt} \leq CO_{rt}; \forall r, t$$

$$(3) \quad \sum_{i \in O_r} S_{irt} \leq CS_{rt}; \forall r, t$$

$$(4) \quad iwp_r + \sum_{i \in O_r} \sum_{t=1}^T iw_{irt} \times p_i \leq \sum_{i \in O_r} \sum_{t=1}^T Y_{irt}; \forall r$$

$$(5) \quad \sum_{\substack{i \in OS(i), \\ i \in O_r}} \sum_{k=1}^t ow_{irk} \times p_i = \sum_{\substack{i \in OS(i), \\ i \in O_r}} \sum_{k=1}^t Y_{irk}; \forall r, t$$

$$(6) \quad \sum_{k=1}^t iw_{irk} \times p_i \leq \sum_{k=1}^t Y_{irk}; \forall r, i \in OS(i), i \in O_r, t \in (1, \dots, dd_i)$$

$$(7) \quad Y_{i,\mu(i,n_i),t} \leq M \times X_{it}; \forall t, i \notin OS(i),$$

$$(8) \quad -FT_i + t \leq M \times (1 - X_{it}); \forall t, i \notin OS(i)$$

$$(9) \quad LT_i \geq (FT_i - dd_i); \forall i \notin OS(i)$$

$$(10) \quad LT_i \leq T - dd_i; \forall i \notin OS(i)$$

$$(11) \quad \sum_{k=1}^t ow_{irk} \times p_i = \sum_{k=1}^{t+(T-dd_i)} Y_{irk}; \forall r, i \notin OS(i), i \in O_r, t \in (1, \dots, dd_i)$$

$$(12) \quad Y_{irt}, O_{irt}, S_{irt} > 0; \forall r, t, i \in O_r$$

$$(13) \quad LT_i, FT_i \geq 0 \text{ \& } X_{it} \in \{0,1\}; \forall t, i \notin OS(i)$$

The objective function of the model consists of minimizing the production system's costs including the costs of production variables in normal time, overtime and sub-contracting, and also those for postponed delivery of orders that may be completed behind schedule. By resolving the IP1 model, main components of the new order's cost price will be determined including all costs of production variables and postponed delivery of orders that may be completed behind schedule due to the addition of new orders. Given the necessary workload that must be acquired from contractors and the benefit that is determined by management, we can calculate the cost price of the new order. Limitation (1): Limited capacity in normal time; limitation (2 and 3): Limitation of maximum time for overtime and sub-contracting; limitation (4): This limitation guarantees that all input works are completed within the planning horizon T . Only the limitation (4) does not guarantee if orders can be completed at due date, because OCD takes place at different periods within the planning horizon of the first level. Limitations (5 and 6): These limitations are about orders to be completed at due date. Application of limitation (5) helps the completion of orders in different workstations at their relevant OCD. This limitation increase the possibility of order completion at due date. Planning at first level is performed weekly. An OCD may be on third day of the week. Application of this limitation (5) simply guarantees that the order will be completed at the end of a week. To make sure that the order will be completed exactly on the third day of the week, limitation (6) is applied. This limitation is considered only for orders of high significance that must be completed at due date. In other words, the limitation (6) guarantees that the remaining work related to the order shall be completed in period t on due date. In limitation (7), OCD is calculated at the final workstation. This time is saved by taking the binary value of $(X_{it} = 1) X_{it}$. By application of this limitation (8), OCD is calculated. In limitation (9), this time is compared with the promised delivery time to calculate the delivery delay. Limitation (10) ensures that the maximum delay of any order is to the extent that the order will be produced within the planning horizon. Limitation (11) is similar to limitation (5) with this difference that, here, OCD may be delayed (LT_i). In other words, limitation (11) ensures that the amount of work remaining until period t for orders that are delayed should be completed in the time interval between t and $ddi+LT_i$. This limitation causes that sufficient capacity for timely delivery of orders is released. Limitations (12 – 13) are nonnegative limitations to make sure that decision variables are correct.

9.1.3. Deciding to Accept or Reject An Order by the Customer

A new order's price is determined by implementing IP1 model. This price is announced to the customer. If the customer approves the order to be completed at MTO firm, the firm will perform necessary actions in the next stage by providing necessary workload and raw materials through contractors.

9.1.4. Selecting Contractors to Provide Raw Materials and Required Workload to Complete New Orders

The question of selecting a set of contractors to provide necessary workload for accepted new orders is the last stage of the suggested structure. Criteria for contractor selection are price and delivery time. As mentioned before, in exchange for various alternatives, each of which contains their own ERD and OCD, IP1 model is implemented and, finally, the lowest price offered is chosen. IP1 implementation has an ERD and OCD value in different workstations in exchange of which the lowest price is obtained.

Given the output of the IP1, if MTO firm is obliged to use contractors to provide new orders' workload and ensure their delivery time, it is necessary that some of the existing contractors are chosen.

ST_{ir} times at the second stage of the proposed structure are calculated.

ST_{ir} is the latest time that necessary operation is commenced in workstation r regarding order i .

If necessary workload for order i which is provided by contractors enters production system after this time, it will not be possible to get a pre-calculated OCD. Since it is possible that some specified or unspecified orders are rejected from the time these values are calculated to the implementation of IP2 model, ST_{ir} values should be recalculated and updated.

In this stage, the objective is to determine a set of contractors who offer the least delivery time difference with respect to ST_{ir} values, and suggest a lower price to provide necessary workload. IP2 model is suggested to find this set of contractors.

According to requirements of orders whose fabrication has been newly accepted in the system, the necessary amount of raw materials and workload are provided by contractors. Therefore, the amount of raw materials and workload are already known. In other words, IP2 is an allocation model in which given the criteria of price and delivery time some of the contractors available in the present set are selected. Unlike IP1 model, IP2 model is applicable only for orders that are accepted by the customer and their production has been ensured. Orders that have not been yet approved by their customer can cause raw materials and workload to be ordered more than needed. This also may impose many costs on MTO systems. After the new order price is announced to the customer, a time shall be spent by the customer to perform the negotiating process and confirmation of the order. During this time, some of the orders that were not still approved by the customer in the IP1 model may be approved before a new order is confirmed by the customer in IP2 model. It is assumed in IP2 model that workload required by an order in a workstation can be provided only by one contractor.

Indexes:

s : Contractors index ($s = 1, \dots, S$)

Input parameters:

ST_{ir} : The latest time that necessary operation is commenced in workstation r regarding the order i ; otherwise, the order will be postponed.

P_{irs} : Suggested price of contractor s to deliver the required workload for the order i in workstation r .

MAD_{irs} : Delivery time of the required workload for the order i in workstation r by contractor s .

S_{irst} : Maximum amount of workload for the order i in workstation r that contractor s can provide in period t .

β_{ir} : Penalty for the arrival of necessary workload of order i in workstation r prior to schedule.

β'_{ir} : Penalty for the arrival of necessary workload of order i in workstation r behind schedule. Considering that the late

arrival of raw materials and, consequently, late delivery of orders may impose many costs on the firm, the amount of the delay should be so that raw materials enter the firm on or prior to schedule ($\beta_{ir} \gg \beta'_{ir}$).

$NO(i)$: Set of new orders accepted.

$L'(i)$: Set of contractors who can deliver the workload required by order i in workstation r before their S_{tir} .

$S(r_i)$: Set of contractors who can deliver the workload required by order i in workstation r .

Decision variables:

$X_{ir} = 1$, if workload for order i in workstation r is provided by contractor s .

Otherwise, $X_{ir} = 0$

$$\begin{aligned} MIN \quad Z' = & \sum_{i=1}^I \sum_{r=1}^R \sum_{s \in S(r_i)} (P_{irs} \times X_{irs} \times S_{irst}) + \sum_{i \in NO(i)} \sum_{r=1}^R \sum_{s \in L'(i)} \beta_{ir} \times (ST_{ir} - MAD_{irs}) \times X_{irs} \\ & + \sum_{i \in NO(i)} \sum_{r=1}^R \sum_{\ell \notin L'(k_i)} \beta'_{ir} \times (MAD_{irs} - ST_{ir}) \times X_{irs} + \sum_{i=1}^I \sum_{r=1}^R \sum_{t=1}^T \left[cr_{irt} \times (Y_{irt} - O_{irt} - \sum_{s \in S(r_i)} (S_{irst} \times X_{irs})) + co_{irt} \times O_{irt} \right] \end{aligned}$$

$$(1) \quad \sum_{i=1}^I (Y_{irt} - O_{irt} - (\sum_{s \in S(r_i)} (S_{irst} \times X_{irs}))) \leq CR_{rt} \times (1 - \alpha_{rt}); \forall r, t$$

$$(2) \quad \sum_{i=1}^I O_{irt} \leq CO_{rt}; \forall r, t$$

$$(3) \quad iwp_r + \sum_{i=1}^I \sum_{t=1}^T iw_{irt} \leq \sum_{i=1}^I \sum_{t=1}^T Y_{irt}; \forall r$$

$$(4) \quad \sum_{i \in OS(i)} \sum_{k=1}^t ow_{irk} = \sum_{i=1}^I \sum_{k=1}^t Y_{irk}; \forall r, t$$

$$(5) \quad \sum_{k=1}^t iw_{irk} \leq \sum_{k=1}^t Y_{irk}; \forall r, i \in OS(i), t \in (1, \dots, dd_i)$$

$$(6) \quad \sum_{k=1}^t ow_{irk} = \sum_{k=1}^{t+LT_i} Y_{irk}; \forall r, i \notin OS(i), t \in (1, \dots, dd_i)$$

$$(7) \quad \sum_{s \in S(r_i)} X_{irs} = 1; \forall r, i \in NO(i)$$

$$(8) \quad Y_{irt}, O_{irt} \geq 0; X_{irs}; \forall i, r, t, s \in S(r_i)$$

The objective function of IP2 model includes different costs of manufacturing an order plus the costs for purchasing raw materials and workload from contractors and the penalty for delivery prior or behind schedule by contractors. In IP2 model, input parameters include the following: necessary amount of workload and raw materials to be provided by contractors, and also raw materials' delivery time before running the model. As previously mentioned, IP2 model only applies to orders confirmed and approved. Therefore, the value of such orders is 1 in IP2.

Limitations (1 – 6) are similar to IP1 problem limitations, but they are only for approved orders. Limitation (7) guarantees that the necessary workload regarding an order in a workstation may be provided just by one contractor. The application of IP2 model only for approved orders will update data and parameters at the first level. In other words, fifth stage indicates the modality of implementing the rolling horizon approach for the first level of hierarchical production planning structure.

By implementing IP1 and IP2 models, one the one hand, the order is completed with the least possible cost and it is more likely acceptable by customer; one the other hand, a group of contractors are selected who will provide necessary workload and raw materials at the lowest possible cost and at the latest period. MTO manufacturing firms have an appropriate and effective relation with different components of supply chain which plays a major role in the effective performance of MTO systems; it also creates a flexible production schedule which complies with internal and external limitations in order to earn greater profits. This

planning method can facilitate release and scheduling of orders at low levels of planning.

9.2. Deciding to Reject or Accept Orders with Negotiable Delivery Time

In this status, no fixed time for order delivery is offered to the firm by the customer, but it asks the firm to offer a combination of delivery time and price so that it makes its decision according to its own criteria. In this case, similar to fixed delivery time, four main activities are carried out that are as follows:

- B (1) – Production of various alternatives to calculate the delivery time for new orders
- B (2) – Determining delivery time and price of new orders by implementing integer linear model (IP1)
- B (3) – Deciding to reject or accept orders through negotiation with MTO firm and customer
- B (4) – Selecting contractors for the supply of raw materials and workload required by approved orders.

9.2.1. Production of Various Alternatives to Calculate the Delivery Time for New Orders

If an order is not rejected at the first stage, in order to implement IP1 model and calculate delivery time and price of new orders, we need to calculate OCD and ERD values. Because delivery time is not offered directly by the customer, so in this case, different delivery times are produced by using forward method and the following relations:

$$LRD_i = ERD_i + \text{pool delay}$$

$$OCD_{i,\mu(i,1)} = LRD_i + TWK_{i,\mu(i,1)} \times P_i + B_1$$

$$OCD_{i,\mu(i,2)} = ST_{i,\mu(i,2)} + TWK_{i,\mu(i,2)} \times P_i + B_2$$

$$OCD_{i,\mu(i,r)} = ST_{i,\mu(i,r)} + TWK_{i,\mu(i,r)} \times P_i + B_r$$

$$OCD_{i,\mu(i,n_i-1)} = ST_{i,\mu(i,n_i-1)} + TWK_{i,\mu(i,n_i-1)} \times P_i + B_{(n-1)}$$

$$OCD_{i,\mu(i,n_i)} = ST_{i,\mu(i,n_i-1)} + TWK_{i,\mu(i,n_i-1)} \times P_i + B_n$$

$$dd_i = OCD_{i,\mu(i,n_i)}$$

$$ST_{i,\mu(i,2)} \geq OCD_{i,\mu(i,1)}$$

$$ST_{i,\mu(i,r+1)} \geq OCD_{i,\mu(i,r)}$$

$$ST_{i,\mu(i,n_i)} \geq OCD_{i,\mu(i,n_i-1)}$$

Four general alternatives are proposed for calculating different delivery times so that various delivery times are determined by the combination of these alternatives:

- 1) Calculation of different values of ERD: in this status, three different times for ERD are proposed which include the most optimistic, moderate, and the most pessimistic times for the delivery of raw materials by contractors.
- 2) Giving different priorities to an order at the warehouse of backlogged orders: By giving different priorities at the warehouse of backlogged works, different delivery times may be determined similar to order with fixed delivery time. These priorities include the following conditions:
 - a. Normal priority: In this condition, the order is placed at the end of queue for unreleased orders.
 - b. High priority: In this status, upon arrival of raw materials, the order is placed on the release line, and ERD and LRD values become equal ($Pool_delay = 0$).
 - c. Other priorities: In addition to the two above priorities, depending on the order significance, we can assign different priorities at the warehouse of non-released orders.
- 3) Change in OCD values: In this alternative, by producing different OCD values similar to orders of stable delivery time, different delivery times are produced. This change can be done in three ways:
 - a. Change in OCD of new orders by increasing the capacity of workstations where new operations are carried out on new orders.
 - b. Change in OCD of new orders by changing OCD values of other orders.
 - c. Change in OCD of new orders by using the capacity allocated to other orders.
- 4) Change of waiting period priority in each workstation: Like the previous status, depending on the significance of orders, two normal and high priorities are given to relevant orders.

For each alternative, a set of values (dd , OCD , LRD) are provided that constitute the most important inputs of IP1 model. Some of these values can be removed:

- 5) dd values that have a short time interval with the “time now”. Although this delivery time is short, but we need extra activities such as sub-contracting and overtime to meet this need. Also, a short delivery time may produce a problem for production schedule and result in late delivery of some approved orders.
- 6) add values whose due date are beyond planning horizon T .

9.2.2. Determining Delivery Time and Price of New Orders by Resolving the Integer Linear Model (IP1)

IP1 linear model which was discussed in section (A2) is used to determine the price of new orders in which delivery time is negotiable. In this case, for each set of values

(dd, OCD, ERD), IP1 model is applied. Thus, for each dd value, one price is obtained. Finally, different values (dd, P) will be offered to the customer.

9.2.3. Deciding to Reject or Accept Orders through Negotiation with MTO Firm and Customer

By applying IP1 model, different combinations of (dd, P) values are generated for new orders. These combinations are offered to the customer. At the end, both parties agree on one combination of price and delivery time.

9.2.4. Selecting Contractors for the Supply of Raw Materials and Workload Required by Approved Orders

Finally, similar to orders with stable delivery time, and using IP2 integer linear model, a set of best contractors, who offer the lowest cost to the system, is selected to provide necessary workload.

10. Third Level of the Proposed WLC Approach (Release of Orders into the Shop Floor)

In this model, the lateness in delivering orders occurs in one of the following conditions:

- If the amount of workload available in a workstation approaches to zero.
- If all orders in a workstation are non-urgent.
- If one order becomes urgent in the warehouse of orders to be released.

10.1. If the Amount of Workload Available in a Workstation Approaches to Zero

This does not mean that available amount of workload is certainly zero. As soon as the available workload approaches zero, the production system sends signals to inform the site supervisor of this issue. To avoid unemployment of such a workstation, some of the orders that are firstly processed in this workstation are released so that the amount of workload which is available distances itself from zero. To do this, the following steps are performed in sequence:

A (1) – Determine the set of orders in the warehouse of orders to be released where first operation is carried out.

A (2) – Release urgent orders with a higher LRD.

A (3) – If two or more urgent orders have identical LRD, the sequence of releasing orders according to the time spent in the workstation to process them. Orders for which shorter time is spent are released first.

A (4) – Non-urgent orders with a higher LRD:

A (5) – Like the second case, if two or more non-urgent orders have identical LRD, the sequence of their release will depend on the time spent for operations performed on them in the relevant workstation.

10.2. If All Orders in a Workstation Are Non-Urgent

If all existing orders are in the queue of workstation for non-urgent orders (their pre-determined OCD at first level is higher than the time of completion in the previous station), urgent orders on which the first operation has been performed in the relevant workstation are released so that their completion is possible at due date. Order releasing procedures are as follows:

B (1) – Determining the set of urgent orders in the warehouse of works to be released the first operation of which is performed in the relevant workstation.

B (2) – Releasing the urgent orders based on LRD priority.

B (3) – By releasing each new order, total workload of the workstation is checked.

10.3. If One Order Becomes Urgent in the Warehouse of Orders to Be Released

As soon as an order becomes urgent in the warehouse of orders to be released, we check if, by releasing the new order, the amount of workload in the workstations, where first operation has been performed, is or is not beyond the permissible limit. If none of these workstations are blocked, the order is released. However, if some of them are blocked, extra capacity required to provide necessary workload for the new order is computed (the difference of total available workload and upper limit); if the capacity required by all blocked workstations can be provided, then the order is released.

As you can see in the proposed WLC model, focus is on two essential cases of on time release to ensure the delivery time computed at first level of the proposed structure and to have a smooth and balanced production system. The only point to be mentioned in the developed WLC model is the way to calculate total workload of each workstation including the set of available workload, workload generated in workstations against the flow and workload generated because of newly released works.

The following steps are suggested in order to make sure if total workload for one workstation is higher than upper permissible limit:

- First step – Calculating the upper limit for the workload of each workstation
- Second step – Calculating total workload in workstations
- Third step – Comparing total workload and upper permissible limit

10.3.1. First Step – Calculating the Upper Limit

To do this, we use the formula provided by Betchet(8):

$$LL_r = 100(PP + LT_r) / PP$$

LL_r : Permissible upper limit for workstation r

PP : Length of planning horizon at the second level

LT_r : Planned lead time for workstation r .

According to the above formula, lead time is determined

by dividing two average amounts of stock in the queue of workstation and the relevant output average. To calculate the lead time in each workstation, first level data are used. At the first level, the amount of orders entering weekly to a workstation (iw_{irt}) and the amount allocated to each individual workstation in different periods are computed. Therefore, the lead time for each workstation in the weekly period t is calculated from the following relation:

$$LT_r = \frac{QUEUE}{Capacity} = \frac{iw_{irt}}{\sum_{i=1}^I Y_{irt}}$$

Lead time has not been calculated in previous research according to the output of first level.

10.3.2. Second Step – Calculating Total Workload in Each Workstation

The following relation indicates how to calculate the workload in workstation i :

$$TL_r = \sum_i TWL_{ir} \times P_{iry}$$

TL_r : Total workload in workstation r

P_{iry} : A percentage of order i released from the current workstation y and entered into workstation r within the planning horizon. If $r = y$, this probability will be 1. In other words, the workstation y is the against-the-flow workstation r for the order i . To calculate p at against-the-flow workstations, the following formula is proposed:

$$P_{jry} = \frac{PQ_y}{TL_y} = \frac{\sum_{i=1}^I Y_{iyt}}{TL_y}$$

PO_y : Planned output value in workstation y within the planning horizon of the second level. This value is equal to the planned capacity in workstation y of the first level of hierarchical structure proposed. If p is higher than 1, this means that the planned capacity in the workstation is more than the total workload and, consequently, total workload in workstation y is completed within the planning horizon and entered into another workstation.

Thus, if order i which is now in queue in workstation y requires m extra workstations before its entry into workstation r , a percentage of this order, which is expected to enter the workstation r within the planning period, will be equal to:

$$P_{iry} = \prod_m \frac{PO_m}{TL_m} \quad m \in \{y, y+1, \dots, r-1, r\}$$

The advantage of calculating p in this research in comparison with other previous researches is that, in many formulas proposed for calculating p , the current status of production system has not been taken into account. In many previous researches, the following formula has been used to

calculate p :

$$P_{jly} = \frac{100}{LL_y}$$

According to the above formula, p -value is only dependent upon LL_y -value which is a fixed value. So, during the calculation of p -value, current status of the workstation is not calculated in terms of capacity and workload available in there. This factor impedes the access to a smooth production system.

10.3.3. Third Step – Comparing the Total Workload with Permissible Upper Limit

After first and second steps, it becomes possible to compare the total workload and permissible upper limit. Before permitting the release of any new order, we need to calculate and compare the total workload and permissible upper limit for workstations which are a part of operation route of the new order. If after the release of new orders into production system, the total workload of any workstations is not beyond the permissible upper limit ($TL_r \leq LL_r$), new orders are permitted to be released; otherwise, ($TL_r > LL_r$) orders will remain at the warehouse of orders to be released.

11. Fourth Level of the Proposed Hierarchical Production Planning Structure: Orders' Priority Stage in Workstations

The focus of this level is on reducing the queuing in individual workstations. The only factor affecting the waiting period in workstations is the method used to prioritize orders inside workstations. Until now, different methods have been used to prioritize works in workstations. In the proposed structure, since values including delivery time, length of the queue in different workstations, and release time are calculated according to the current condition of production system, so at the fourth level, we need not complicated methods to prioritize orders. In previous researches where only second and third levels were used

simultaneously, the use of FIFO, which is one of the simplest methods today, led to acceptable results [11].

11.1. Method for Prioritizing Orders in the Queue of Workstations

Since delivery time in MTO production environments is the most important objective of the system, so, the prioritization of unfulfilled orders being queued in different workstations should be done to begin operation in a manner that their delivery time is ensured. OCD is the most vital characteristic of each order in a workstation and it is obvious that if OCD occurs on time, the timely delivery of orders in each workstation will be ensured. In this level, to prioritize orders in each workstation, we suggest that all orders are given priority according to their OCD. This method of prioritization is called SOCD (shortest order completion date). If two orders have equal OCD in one workstation, the order with higher significance will be given priority.

In the first level of the proposed structure, the degree of hardness of IP1 and IP2 mathematical models was investigated via LINGO 6.0, and to make sure of their output accuracy, outputs generated in the software were analyzed.

12. Evaluation of Hierarchical Production Planning System at the Second Level: Order Entry Stage

At the second level of IP1 and IP2 mathematical planning models, we first saw how computational time in these models increases along with the increase of their dimensions.

Required data on test problems are randomly generated. One week was assumed for time unit and 12 weeks for the length of time horizon. Six types of problems with different dimensions were considered and from each problem 10 samples were generated randomly. Variables used to determine dimensions include the number of orders in the system, resources and orders that can be postponed. For each tested problem, one LINGO model was created and all generated problems were run in a PC. The following tables show the results obtained from IP1 and IP2 models.

Table 1. Structure of IP1 Model in Tested Problem

Number of sample problem	Number of orders in the system	Number of Sources	Number of orders to be delayed	Dimensions of problems		
				Number of integer variables	Number of variables of the continuous number	Number of limitations
1	5	3	3	42	540	384
2	8	5	3	42	1440	653
3	10	3	5	70	1080	526
4	12	6	7	98	2592	1040
5	15	8	9	126	4320	1578
6	18	8	11	154	5184	1806

Table 2. The Structure of IP2 Model in Tested Problems

Number of sample problem	Number of contractors	Number of integer variables	Dimensions of the problem Number of variables of the continuous number	Number of limitations
1	4	60	360	351
2	6	176	960	559
3	12	520	720	425
4	16	1536	1728	978
5	20	1950	2880	1411
6	30	4860	3456	1364

Table 3 shows the mean CPU time required to achieve an optimal solution in IP1 and IP2 models for each set of tested problems. A notable point is that for sample problems with dimensions of $18 \times 8 \times 11$ in IP1 model, there is no possibility to reach an optimal solution in a reasonable CPU time. The computational results also indicate that IP1 model can provide an optimal solution for small-sized and medium-sized problems in a reasonable CPU time. But for some medium-sized problems and all large size problems, it is not possible to access an optimal solution because, in proportion to the increase in the size of problems, the time for an optimal solution increases exponentially. As a general conclusion and according to results, IP1 model is sensitive to the number of orders to be delayed, because the number of integer variables in IP1 model is dependent on the number of orders to be delayed. Unlike the IP1 model, IP2 model is able to compute an optimal solution for all dimensions in a reasonable CPU time, because IP2 model structure is a simple allocation problem in which from the set of the contractors, according to the criteria of price and delivery time, a number of them are selected.

Table 3. CPU Mean Time in Tested Problems

Number of sample problem	Number of tested problems	CPU mean time to get an optimal solution (Min)	
		IP1 Model	IP2 Model
1	10	15	0.8
2	10	43	3.2
3	10	95	3.8
4	10	214	3.6
5	10	442	4.3
6	10	NA	-

Now, a MTO production system with two workstations and four orders (three orders in the production system and one new order with high signification) is considered. In this example, six weeks are assumed for the planning horizon at the first level. All orders have stable delivery time and two orders can be postponed.

Based on definitions of ow_{irt} and iw_{irt} parameters and also the data on table 4, values of these parameters are calculated in table 5. Meanwhile, all iwp_r values are considered zero. Tables 5 – 10 show other data needed in IP1 model.

Table 4. Data on Orders

Orders	ERD_i (wk)	dd_i (wk)	p_i	ct_i	Workstation 1		Workstation 2	
					OCD_{il} (wk)	TWK_{il} (h)	OCD_{il} (wk)	TWK_{il} (h)
A	1	3	1	-	2	20	3	30
B	2	5	6.	-	3	10	5	25
C	2	5	8.	30	3	30	5	40
D	1	4	1	40	2	40	4	35

Table 5. Input Values (iw_{irt}) and Output Data (ow_{irt}) for One Week in IP1 Model

Week	Workstation 1 (iw / ow)				Workstation 2 (iw / ow)			
	A	B	C	D	A	B	C	D
1	/-20			/-40	/-30			/-35
2	20-/	/-10	/-30	40-/		/-25	/-40	
3		10-/	30-/		30-/			
4								35-/
5						25-/	40-/	

Table 6. Data on the Capacity of IP1 model

Week	Workstation 1			Workstation 2		
	Maximum Capacity Normal time	Maximum Capacity Overtime	Maximum Capacity Contractor	Maximum Capacity Normal time	Maximum Capacity Overtime	Maximum capacity Sub-contracting
1	40	20	10	40	20	10
2	40	20	10	40	20	10
3	40	20	10	40	20	10
4	40	20	10	40	20	10
5	40	20	10	40	20	10
6	40	20	10	40	20	10

Table 7. Production Costs at the Normal Time in IP1 model

Order	Workstation 1						Workstation 2					
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
A	50	48	58	60	45	70	60	60	56	65	50	55
B	70	75	68	72	70	65	64	58	68	75	80	70
C	30	35	35	30	32	34	28	25	20	25	20	25
D	80	85	87	80	80	82	60	64	50	55	60	58

Table 8. Production Costs during overtime in IP1 Model

Order	Workstation 1						Workstation 2					
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
A	150	150	155	160	152	158	165	170	160	168	165	170
B	200	204	200	205	210	205	215	210	215	217	220	215
C	120	125	120	123	125	128	115	110	108	110	113	115
D	220	214	215	218	210	220	236	234	235	230	235	230

Table 9. Production Sub-contracting costs in IP1 Model

Order	Workstation 1						Workstation 2					
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
A	350	355	357	360	355	350	355	360	365	367	360	365
B	395	390	390	394	393	398	369	368	370	374	375	370
C	320	320	325	324	320	328	310	318	312	315	312	310
D	400	402	404	410	412	400	402	406	410	415	410	405

Table 10. α_{rt} Values in IP1 Model

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Workstation 1	0	0.1	0.2	0.1	0.3	0
Workstation 2	0.1	0	0.1	0.2	0.3	0.1

The following results are obtained by implementing IP1 model:

$$X_{36}=X_{46}=1 \quad FT_3=FT_4=6, \quad LT_3=1, \quad LT_4=2, \quad Z^*=21824.00$$

To implement IP1 model, it is assumed that second potential order (C) is rejected. Raw materials and workload required by received orders can be provided by six contractors. In addition to IP1 model outputs, other data needed to implement IP1 model are indicated in tables 11–14.

Table 11. S_{ir} Values

Order	Workstation 1	Workstation 1
A	1	2
B	2	3
C	1	2

Table 12. β_{ir} and β_{ir} Values

	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5	Contractor 6
β_{ir}	20	30	50	55	35	30
β_{ir}	2000	3000	5000	5500	3500	3000

Table 13. P_{irs} Values

Order	Workstation 1						Workstation 2					
	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5	Contractor 6	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5	Contractor 6
A	1350	1380	1320	1350	1310	1300	1500	1550	1500	1450	1480	1460
B	1400	1380	1420	1450	1400	1390	1800	1850	1800	1900	1780	1890
D	900	1020	1000	1030	1000	1020	1500	1520	1450	1420	1500	1520

Table 14. MAD_{irs} values

Order	Workstation 1						Workstation 2					
	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5	Contractor 6	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5	Contractor 6
A	2	3	3	2	1	1	4	1	3	1	2	1
B	1	3	2	3	3	2	3	1	2	2	2	3
D	3	1	2	2	2	1	1	2	2	3	1	1

By implementing IP2 model, the following results are achieved:

$$X_{115} = X_{125} = X_{212} = X_{225} = X_{312} = X_{322} = 1$$

$$Z^* = 491541$$

Table 15. Results of simulation of different order release models (numbers shown in the column header indicates the performance criteria of different release models defined above) - Two-day interval during the arrival of two consecutive orders

Release Model	Prioritization method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Release model (3)	SOCD	15.6	9.9	0.7	7	15	9.5	5.1	2.5	2.2
Release model (4)	SOCD	15.3	9.4	0.8	6.7	14.7	9.8	5.1	2.6	2.2
Release model (5)	SOCD	15.1	9.5	0.9	6.8	14.6	9.9	5	2.6	2.0.2
Immediate Release model	FCFS	17.2	17.4	2.7	25.8	19.1	18.1	11.5	4.7	4.2
Immediate Release model	S / OPN	10.1	10.5	1.5	8.6	12.1	12.4	8.4	4.6	4
Immediate Release model	SOCD	15.8	16.5	0.7	9.6	15.2	15.2	9.5	4.4	3.9
slar Model (k = 4)	Pst	16.4	11.9	2.4	12.5	15.8	11	7	3.5	2.2
slar Model (k = 5)	Pst	16.2	12.8	1.8	11.6	15.6	11.9	7.5	3.7	2.4
slar Model (k = 6)	Pst	16.4	13.2	0.9	10.6	15.5	12.5	8	3.8	2.5
wcedd Model (6)	S / OPN	39.4	9.4	24.3	34.7	50.8	8.6	6	3	1.7
wcedd Model (7)	S / OPN	21.4	10.2	7.1	25.4	24.6	9.4	6.5	3.2	1.8
wcedd Model (9)	S / OPN	18.1	13.4	3.2	16.3	18.1	10.9	7.3	3.8	2.4

Table 16. Results of simulation of different order release models (numbers shown in the column header indicates the performance criteria of different release models defined above) – The interval of 2.8 days during the arrival of two consecutive orders

Release Model	Prioritization method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Release model (3)	S OCD	19.5	10.1	2.5	14.8	18.7	11	6.1	3	2.5
Release model (4)	S OCD	19.2	10.3	3	13.1	18.4	11.1	6.1	3.1	2.5
Proposed Release model (5)	S OCD	19.1	10.5	3.1	12.3	18.1	11.3	6	Three. 1	2.5
Immediate Release model	FCFS	21.9	22.4	6.2	37.8	24.6	24.6	15.5	6.1	5.8
Immediate Release model	S / OPN	11.2	11.4	3.4	10.7	15.4	25.6	10.8	6.4	5.7
Immediate Release model	S OCD	20.4	20.4	2.6	25.3	20.3	20.3	13.6	6.2	5.1
slar Model (k = 4)	Pst	21.5	13.1	4.5	26.1	21.3	21.4	7.4	3.8	2.7
slar Model (k = 5)	Pst	21.1	14.4	4.1	24.8	21	13.7	8.1	3.9	2.8
slar Model (k = 6)	Pst	20.9	15.2	3.7	24.3	21.1	15.2	8.7	4.1	3.1
wcedd Model (6)	S / OPN	48.7	10.4	32.4	40.1	65.4	10.3	6	2.9	2.3
wcedd Model (7)	S / OPN	26.1	13	8.4	37.8	29	12.4	7.5	3.5	2.6
wcedd Model (9)	S / OPN	26.1	13.6	8.9	40.2	27.4	11.4	7.2	4	2

13. Evaluation of Hierarchical Production Planning at Third and Fourth Levels

To evaluate release and prioritization of orders in the proposed models at third and fourth levels of hierarchical production planning system, the models were simulated using Arena 7.0 simulation software.

To ensure that the proposed models compared with previous works will provide better performance in the shop floor; results obtained from simulation of the models are presented and compared with three models for orders release into the shop floor. The input data for simulated models include [12]:

- Number of workstations: 6 (Each workstation is has a separate machine)
- Work routes: random and irreversible
- Number of operations required for each order: discrete uniform distribution [1,6]
- Interval between the arrival of two sequential orders: exponential with the arrival rate of 2 orders in one business day
- Operating time: Erlang function 2 with an average of 0.75 days

In this paper, we compare four methods of order release:

- Proposed release model
- Immediate release: in this model, each order upon arrival is released into the shop floor
- Superfluous load avoidance release (SLAR) model [11]
- WCEDD model [10]

The performance of each of the models is checked by four different values of the permissible upper limit (infinite, open, closed, very closed) so that the degree of their dependence to the permissible upper limit is determined. 3, 4, and 5 are selected for the high permissible upper limit of a workload in a workstation, and 6, 7, and 9 for the high permissible upper

limit of total workload in a workstation, respectively for very closed, closed, and open upper limit.

Results are obtained from 100 performances during 500 business days. All performances were applied independently. The performance includes 10 days as warm-up. Criteria selected for the comparison of different models include 9 cases indicated below:

- Production mean time in the shop floor (SFTT)
- Average of SFTT and waiting period in the stock of orders to be released
- Average lateness in delivering orders
- Number of orders to be postponed
- Total workload of production system including the total workload of all workstations and total workload in the stock of orders to be released
- Total workload in the shop floor including total workload in all workstations
- Combined workload at each workstation including total workload in the relevant workstation and total workload to be transferred into workstations against the flow
- Workload in each workstation
- Standard deviation of workload available. This criterion is used to check the smoothness of the defined production system

Workload has been calculated based on the time.

Three prioritization techniques: FCFS +S/OPN and SOCD proposed in this paper were used to evaluate the immediate release model. In the proposed release and prioritization model, only SOCD method is used to show appropriate combination of second and third levels of hierarchical structure. Since the use of two prioritization methods of PST and S/OPN in SLAR and WCEDD release models has a better impact in the shop floor performance compared with other prioritization methods (11), the two methods were used to prioritize orders in the shop floor. In PST methods, orders are prioritized according to the planned time for operation.

Finally to show the impact of the severity of incoming orders on the performance of release models, we considered two-day intervals during the entry of two consecutive orders. All models were implemented at an interval of 2.8 days during the entry of two consecutive orders. Results are shown respectively in tables 15 and 16.

The results shown in Tables 15 and 16 indicate the proper functioning of the proposed release model to control orders' delivery time compared with other release models. Another notable conclusion is that the proposed model has the ability to match the increasing order entry (comparison of the results obtained in tables 15 and 16 show this). The proposed model has a better performance in comparison with the immediate release model. In the immediate release model where S/OPN prioritization technique is used, the mean value of lateness compared with the proposed model is better, but the value of postponed works is higher. SPT prioritization technique used in SPT immediate release model results in a better performance in the shop floor.

A 40% difference between numerical results of criteria (8) and (9) in the two proposed models and the immediate release model is notable.

Yield comparison of the two proposed models where only high upper limit is used to control and manage delivery time (SLAR and WCEDD models) indicates that only the use of high permissible upper limit is not sufficient to have a better performance of production system in the shop floor. As it is seen in the proposed model, in addition to determining the high upper limit, two release criteria have been presented to prevent unemployment and to achieve smoothing. Another significant point is that WCEDD and SLAR models show a weaker performance of the immediate release model in a number of criteria. This is an issue that many researchers have pointed out that the control of delivery time by defining a high upper limit and in some cases by limiting the amount of workload in the system, production system will not lead to improved functionality, it also causes irregularities in the system [6,7,9].

14. Conclusions

The results of this research can be summarized in the following cases:

- 1) Provide a comprehensive, hierarchical production planning structure in MTO production environments: Although in previous research different levels of hierarchical structure in MTO environments were presented, however, no study is conducted so far in which different levels of decision-making models are provided.
- 2) Provide a new model at the first level of the hierarchical structure to accept or reject orders entering the system. In the proposed model, decisions to accept or reject orders are made in a way that it is possible to determine the accepted orders' delivery time at a competitive price.

- 3) Consider all components of the supply chain in order to determine the delivery time and the price of new orders accepted at the first level. These components include customers, MTO production firms, and various contractors.
- 4) Orders in a MTO system enter the system with distinct characteristics at various random times. Capacity of the system is usually unable to address all of them in the determined planning horizon. Therefore, prior to the implementation of mathematical models, determination of order price and negotiation with customers, some orders that can not be delivered within the planning horizon or those which are allocated a large capacity of the system should be removed from the system.
- 5) Produce various alternatives for calculating the price and different delivery times at the first level. Different values of price and delivery time facilitate negotiation process between customers and MTO firms and puts forward different options for a suitable choice.
- 6) Present a developed integer planning model (IP1) to calculate the price and delivery time of orders entering the system at the first level.
- 7) Present a developed planning model (IP2) at the first level to determine an optimal combination of contractors who offer the lowest price and the shortest delivery time to supply workload required by new orders.
- 8) Provide a developed release model which leads to a smooth production line and reduce delivery time. By comparing the proposed release model with other release models, the final result obtained was a better functionality of the proposed model in a great deal of criteria present in MTO systems.
- 9) Provide an order schedule in different workstations. The method is also conducted according to the outputs of the previous level and orders delivery time.
- 10) One of the strengths of this study was that we considered the rolling horizon topic in production systems within MTO environments. Consideration of the rolling horizon in the proposed structure will increase the applicability of the proposed HPP structure in practice.

15. Suggestions for Future Research

Today, many of production systems work in combined environments of MTS/MTO in which some products are stored according to requested production process, and some others are stored according to the forecasts. So far, only a limited number of researchers have considered the hierarchical production planning (HPP) in such environments. The decision making model presented at the first level can be improved in many ways. Considering the quality criteria in addition to price when selecting contractors will lead to a more complete model at the first

level. One of the results of executing the classic release models and its comparison with our model is that achieving a smooth production system is not possible only by considering a high upper limit as one of the necessities of MTO production environments. Therefore, other criteria are proposed in addition to the upper and lower limits during order release into the shop floor.

REFERENCES

- [1] Higgins, P. D., Tieney, K. and Browne J., 1991. Production management systems: state of the art and perspectives. Proceedings of the CAPE'91 Conference, Edited by G. Doumeings. Amsterdam: North Holland.
- [2] Wortmann, J. C., 1990. Toward one-of-a-kind production: the future of European industry. Proceedings of the International Conference on Advanced in Production Management Systems, Espoo, Finland, 20-22 August.
- [3] Wortmann, J.C., 1992. Production management system for one-of-a-kind product. Computers in Industry, Vol. 19, 79-88.
- [4] Hax, A. C., Meal H. C., 1975. Hierarchical integration of production planning and scheduling. Management Sciences, Vol. 1, 53-69.
- [5] Bitran G. R., Hax, A. C., 1977. On the design of hierarchical production planning systems. Decision Sciences, Vol. 8, 28-55.
- [6] Philipoom, P. R., Malhotra, M. K. and Jensen, J. B., 1993. An evaluation of capacity sensitive order review and release procedures in job shops. Decision Sciences, Vol. 24, 1109-1133.
- [7] Bergamaschi, D., Cigolini, R., Perona, M. and Portioli, A., 1997. Order review and release strategies in a job shop environment: a review and a classification. International Journal of Production Research, Vol. 35, No. 2, 399-420.
- [8] Bechte, W., 1988. Theory and practice of load-oriented manufacturing control. International Journal of Production Research, Vol. 26, 375-395.
- [9] Shimoyashiro, S., Isoda, K. and Awane, H., 1984. Input scheduling and load balance control for a job shop. International Journal of Production Research, Vol. 22, 597-605.
- [10] Melnyk, S. A., Ragatz, G. L., 1988. Order review/release and its impact on the shop floor. Production and Inventory Management. Vol. 29, 13-17.
- [11] Land, M. J., Gaalman, J. C., 1998. The performance of workload control in job shops: Improving the release method. International Journal of Production economics, Vol. 56-57, 347-364.
- [12] Kingsman, B. G., Tatsiopoulos, I. P., Hendry, L. C., 1989. A structural methodology for managing manufacturing lead times in make-to-order companies. European Journal of Operation Research, Vol. 40, 196-209.
- [13] Kingsman, B., Hendry, L., 2002. The relative contribution of input and output controls on the performance of a workload control system in make-to-order companies. Production Planning & Control, Vol. 13, No.7, 579-590.
- [14] Kingsman, B., 2002. Modeling input-output workload for dynamic capacity planning in production planning systems. International Journal of Production Economics. Vol. 68, 73-93.
- [15] Cakravastia, A., Nakamura, N., 2002. Model for negotiation the price and due date for a single order with multiple supplier in a make-to-order environment. International Journal of Production Research, Vol. 40, No.14, 3425-3440.
- [16] Calosso, T., Cantamessa, M., Dang vu, A., 2003. Production planning and order acceptance in business to business electronic commerce. International Journal of Production Economics, Vol. 85, 233-149.
- [17] Calosso, T., Cantamessa, M., Dang vu, A., 2004. Negotiation Support for Make-To-Order Operations in business to business electronic commerce. International Journal of Production Economics, Vol. 85, 233-149.
- [18] Breithoupt, J. W., Land, M. and Nyhuis, P., 2002. The workload control concept: Theory and practical extensions and load oriented order release. Production Planning & Control, Vol. 13, No.7, 625-638.
- [19] Rabbani, M., Tavakoli Moghadam, R., 2006. International conference on industrial engineering and system management , A Proposed Hierarchical Production Planning Structure for Combined MTS/MTO Environments,
- [20] Rabbani, M., 2009, International conference on industrial engineering and system management, complete article, An MADM Framework toward Hierarchical Production Planning in Hybrid MTS/MTO Environments