A new decision-making structure for the order entry stage in make-to-order environments

M. Ebadian, M. Rabbani*, F. Jolai, S.A. Torabi, R. Tavakkoli-Moghaddam

Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran

Received 6 April 2005; accepted 5 January 2007
Available online 4 March 2007

Abstract

In this paper, we propose a new comprehensive decision-making structure for the order entry stage in make-to-order (MTO) environments. The aim of the proposed structure is to manage the arriving orders so that the MTO system just proceeds to produce those arriving orders which are feasible and profitable for the system. The appropriate decisions on the arriving orders are taken based on two criteria including price and delivery time. The arriving orders have either fixed or negotiable delivery times. The proposed structure has five major steps. At the two first steps, the new arriving orders either are rejected or appropriate decisions to meet their delivery time are made. At the next step, the optimal prices along with delivery times (if negotiable) of non-rejected orders are determined by a mixed-integer programming model. In the case of the final approval by the customers at the fourth step, another mixed-integer programming model is launched to select a set of suppliers and subcontractors that are able to provide required raw material and workload of the new the accepted orders. Moreover, numerical experiments are presented and a simple example is illustrated to show the applicability of the proposed mathematical models.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Make-to-order; Decision-making structure; Price; Delivery time; Mixed-integer programming model.

1. Introduction

Manufacturing companies often use different policies in order to meet their demands. Some response to their demands through finished products inventories that are called make-to-stock (MTS). Others fulfill an order only when it enters the system. These kinds of production systems which are called make-to-order (MTO), have to supply a wide variety of products, usually in small quantities. In such systems, the arrival of customers’ orders is stochastic over time. The arriving orders usually require different routings and processing times through production facilities (Haskose et al., 2004). Among the arriving orders, due to different constraints, the system is able to fulfill only some of them and the rest are rejected. Under this condition, MTO companies have to accept an optimal combination of the arriving orders so that their profit and share in the competitive market increase. The main criteria to select this optimal combination in such environments are minimum price, short delivery time and high quality. The conflicts between these criteria make the production
planning process too complicated. Production of high-quality orders usually increases their prices and delivery times. Also, fulfillment of orders with short delivery time needs extra resources which increase their prices (Womack et al., 1991). Therefore, MTO systems, which are not able to predict the arrival time of their orders and have to deliver the arriving orders quickly, need a decision-making structure that helps them manage the arriving orders to meet these main criteria.

In recent years, much research on MTO environments has emphasized on delivery time criterion. This research is more about releasing orders to the shop floor and dispatching them through different workstations to decrease their delivery times. Although appropriate release time and accurate schedule of orders on the shop floor can lead to reduction of their manufacturing lead times and work-in-process (WIP), there is no guarantee that their delivery times are reduced (Kingsman et al., 1989). Instances of this research are presented in papers of Philipoom and Fry (1992), Belt (1978) and Irastorza and Dean (1974). Complete review of such research has been offered by Wisner (1995). To decrease the delivery time, MTO companies need a planning stage to manage the arriving orders by considering the existing constraints before the job release and the dispatching stages. The decision problem of this stage is to examine whether the MTO system is able to deliver the arriving orders to customers with a low and competitive price and short delivery time or not. This examination is done regarding existing constraints. In the literature of hierarchical production planning in MTO environments, this stage is considered medium-term planning and is referred to the order entry stage (Hendry et al., 1998; Kingsman, 2000; Breithoupt et al., 2002; Hendry and Stevenson, 2006).

The purpose of this paper is to improve the production planning framework in MTO environments by proposing a new comprehensive decision structure for the order entry stage. The proposed structure leads to better management of the arriving orders by rejecting undesirable orders and determining reasonable prices and delivery times for the accepted orders. Since the decisions on the arriving orders are taken due to the current conditions of the system, the structure also leads to better planning of the job release and the dispatching stages.

The structure of this paper is as follows: In the next section, the previous research on the order entry stage is presented. In Section 3, the proposed decision-making structure for the order entry stage is illustrated. This section consists of illustration of different steps of the structure and the developed mixed-integer programming models. Numerical experiments are presented in Section 4. In Section 5, the advantages of the proposed structure and future research directions are discussed. Finally, to show the applicability of the proposed structure, an appropriate example is presented as an Appendix A.

2. Literature review

The literature review on the production planning of MTO systems reveals that the number of research regarding the order entry stage is far less than the job release and the dispatching stages. Hendry and Kingsman (1989) first considered the order entry stage in the production planning structure of MTO systems. They emphasized that the key factor in success of MTO systems is the order delivery time management. Hence, before the job release and the dispatching stages, it is necessary to launch the capacity planning and control at the order entry stage. In this regard, a hierarchical input–output control was suggested across different stages of production planning including the order entry stage. The goal of input–output control proposed first by Wight (1970) is synchronous control of inputs (i.e., customers' demands) and outputs of the system (i.e., the capacity). Hendry and Kingsman (1993) also suggested a method by means of an input–output control approach in order to accept or reject the new arriving orders. Since the delivery and manufacturing lead times depend on the total backlog (TB) and planned backlog (PB) (Plossl and Wight, 1973), they are controlled by means of defining predetermined minimum and maximum values for TB and PB. If the arrival of a new order causes TB or PB to violate the predetermined values, either the capacity is increased or the new order is rejected. Easton and Moodie (1999) introduced a technique that simultaneously optimizes pricing and lead time decisions for MTO firms with contingent orders.

In recent years, the workload control approach (WLC) has been considered by researchers to investigate the order entry stage. This approach is an extended input–output control approach, in which the amount of input workload and the capacity are controlled simultaneously. In this regard, three relatively different approaches have been introduced by Bertand and Wortmann.
A comprehensive review of the WLC approach has been done by Breithaupt et al. (2002). By means of simulation, Hendry et al. (1998) showed that in comparison with systems without control, the use of WLC at the order entry stage leads to decrease in manufacturing lead times of orders. Kingsman and Hendry (2002), using simulation of a real job shop at Lancaster University, showed that the use of input–output control has a positive effect on performance measures such as lead time, queuing time, and the capacity utilization in comparison with the use of input control alone and also a system without control. Kingsman (2000) also modeled WLC in a mathematical form to assist in providing procedures for implementing input and output control. It enables dynamic capacity planning to be carried out at the order entry stage. Haskose et al. (2004) modeled the WLC approach as a queuing network with limited buffer capacities for each workstation. Then, the obtained results have been used to analyze issues in WLC. Hendry and Stevenson (2006) developed an aggregate load-oriented WLC concept for the MTO industry referred to the Lancaster University Management School (LUMC) approach. A key characteristic of the LUMC approach is that it controls a hierarchy of backlogs beginning when a potential customer first makes an inquiry. Henrich et al. (2006) investigated the impact of grouping machines on the effective of the WLC approach within MTO job shops. They provided a starting point for a profound decision on grouping machines as well as on the control of the respective capacity groups. With respect to the importance of the WLC approach for job shop practices, Land (2006) analyzed the sensitivity of the WLC approach to improve its performance. Land (2006) provided a number of new insights that improve the basis for setting parameters in WLC. The results have important consequences for implementations of WLC concepts in practice.

In recent research, suppliers have been also considered as a party of the MTO system that affect prices and delivery times of orders. Cakravastia and Nakamura (2002) developed a model for price and due date negotiations between a manufacturer and its multiple suppliers to fulfill a single order from a customer. Calosso et al. (2003, 2004) discussed in detail the structure of a standardized negotiation process occurring in a multi-enterprise setting and presented three mixed-integer linear programming models that may be used by different parties involved.

The main disadvantage of previous research is the lack of a comprehensive structure for the acceptance/rejection decision on the arriving orders and also selection of the best prices and delivery times for the accepted orders considering all relevant elements of the supply chain, such as suppliers, subcontractors, the MTO company, and customers. Hence, in this paper, a comprehensive decision-making structure is proposed to manage the arriving orders at the order entry stage of MTO environments by taking into account all affected parties of the supply chain. The proposed structure makes decision on the arriving orders regarding two criteria including price and delivery time. It is assumed that the production system produces high-quality orders.

Four main decisions are taken on the new arriving orders by the proposed structure:

(a) Identifying and rejecting undesirable orders: Some arriving orders not only do not increase the profit and market share of the company, but also bring chaos for the production system. The main chaos caused by undesirable orders is to delay the accepted orders due to the lack of the capacity taken by undesirable orders. Such a decision also causes the presented optimization models neither to be infeasible nor run for each arriving order.

(b) Computing prices and delivery times (if delivery times of the new orders are negotiable) of non-rejected orders: According to the results of the previous decision, if the new arriving orders are not rejected, their prices and delivery times are computed subject to the existing constraints.

(c) Quoting prices and delivery times to the customers and negotiating with them. The final outcome of the negotiation process is to accept or reject the arriving orders.

(d) Selecting the best set of suppliers and subcontractors to provide the required material and workload of the new accepted orders: If the proposed prices and delivery times of the new arriving orders are accepted after the negotiation, the company takes the last decision which consists of providing required material and workload to produce the accepted orders.

The above-mentioned decisions are taken in the structure, respectively. In brief, by means of the
proposed structure, the MTO system proceeds to produce only those kinds of the arriving orders which do not cause serious modifications in the production schedule of the previous accepted orders and also bring significant profit and market share for the MTO system. Therefore, the proposed structure can manage the arriving orders at the order entry stage of MTO systems efficiently.

3. The proposed decision-making structure

In this section, the new decision-making structure is presented for a typical MTO system. It is assumed that the production line of the considered system is job shop. In other words, each arriving order has a predetermined unique process route that must be adhered to fully; that is, no alternative routings are allowed. Each order includes different subassemblies which are produced by common raw material and components. No subassemblies, raw material and components are held in the system as inventory, i.e., required material and subassemblies are replenished based on the requirements of the new accepted orders. The proposed decision-making structure to manage the arriving orders is depicted in Fig. 1.

In order to make feasible decisions on the arriving orders, all affected parties of the supply chain which their decisions and performances have significant

![Fig. 1. The proposed structure at the order entry stage.](image-url)
effect on prices and delivery times of the new arriving orders are considered in the structure. These parties consist of customers, the MTO company, suppliers and subcontractors (Fig. 2). Customers offer their orders to the MTO company with a set of features such as routing process, price, and delivery time. The MTO company produces the accepted orders based on their demanded features and also tries to deliver them at their promised due dates. Due to the capacity limitation, the MTO company often has to provide some percentage of its required capacity and raw material through its subcontractors and suppliers to produce the new accepted orders. Subcontractors supply the demanded capacity in the form of semi-finished products and subassemblies, which are called workload in this paper, and suppliers provide raw material and components. The structure of inbound side of the considered supply chain is a single tier, parallel suppliers, and subcontractors. Each party has different objectives, which make acceptance/rejection decision on the arriving orders too complicated. Customers expect the MTO company delivers their orders with low prices and short delivery times. On the other hand, the subcontractors’ objective is to provide workload with high price. It is assumed that the subcontractors deliver the required workload of the MTO company on time. Suppliers are willing to sell raw material and components with high price and long delivery time. Moreover, the objective of the MTO company is to earn the maximum profit as well as market share by producing the best set of the arriving orders. To reach such an objective, two mixed-integer programming models are presented in the proposed structure. At first, prices and delivery times of the new arriving orders are determined by the first mixed-integer programming model (IP1 model). Then, an optimal combination of suppliers and subcontractors are selected by another mixed-integer programming model (IP2 model).

As shown in Fig. 1, based on the arriving orders with fixed or negotiable delivery times, the decision route is different. To make appropriate decisions on the arriving orders, they are classified into two groups based on their importance. In practice, companies may have different criteria to assign different priorities to the arriving orders. In this paper, two important criteria (i.e., profit and market share) are considered to prioritize the arriving orders as follows:

- High priority orders that can make significant profit and increase the company’s market share.
- Low priority orders that can only increase the company’s market share.

Only the low priority orders can be delayed and the high priority orders must be delivered on time. The proposed structure has five major steps. In the two first steps, the main objective is to identify and reject undesirable orders. If the delivery time is fixed and the new order is not rejected, the price of the order is determined by the IP1 model at the third step. In the case of the final approval by the customer at the fourth step, the IP2 model is run to select the best combination of suppliers and subcontractors at the last step. If the delivery time is negotiable, different delivery times are generated at the second step. Then, different combinations of delivery times and prices are generated by the IP1 model. In the case of choosing the final delivery time and price through negotiation between the customer and the company, the best set of subcontractors and suppliers is selected.
The two first steps of the proposed structure are implemented for each arriving order. To avoid several runs of the proposed models and also increase their applicability, other steps are only implemented when one of the following conditions occurs:

- When a high priority order enters the system.
- As soon as the number of the new arriving orders reaches a predetermined value specified by the production planners.
- As soon as time gap between the arrival time of the first arriving order after the last run of the IP1 model and the present time becomes greater than a predetermined time.

The predetermined values can be specified based on the past experiences.

### 3.1. Orders with fixed due dates

If the new arriving order has a fixed due date, the following steps are implemented:

#### 3.1.1. Calculating rough-cut capacity

In this step, the possibility of meeting the due dates of the new orders is estimated within the planning horizon $T$. In this regard, the following stages must be done:

(a) If the order has high priority, the following equation is used to compare the required capacity of all existing orders with the maximum available capacity:

$$\sum_{i \in O_r} (TWK_{ir} \times p_i) \leq C_{rt}, \quad \forall r, t = (t_{now}, \ldots, T). \tag{1}$$

where $i$ is the order index ($i = 1, \ldots, I$); $r$ the resource index ($r = 1, \ldots, R$); $O_r$ the set of orders to be processed on resource $r$; $t$ the period index ($t = 1, \ldots, T$); $t_{now}$ the current period; $C_{rt}$ the maximum available capacity of resource $r$ during period $t$ (including regular time, overtime and subcontracting); $TWK_{ir}$ the required processing time of order $i$ on resource $r$; and $p_i$ the acceptance probability of order $i$.

There are three types of orders in MTO systems: Orders that have been already confirmed and released to the shop floor (type I), orders that have been confirmed but have not been yet released (type II). These kinds of orders are waiting in a place called jobs pool. The third type of orders is twofold: Orders which their prices and due dates have been determined by the company but have not been yet confirmed by the customers (i.e., the company is waiting for the final decision of the customers) and also orders which have just entered the system and are still under assessment by the company (type III). The values of $p_i$ for types I and II are equal to one. For the orders belonging to type III, the value of $p_i$ should be set realistically by consideration of attributes including delivery time and features of the new order determined by the customers, the estimated price to cover variable costs, the available capacity in the system, the priority of the new order considered by the company, and the bidding behavior of each competing firm. The values of $p_i$ can be computed with the same approach presented by Easton and Moodie (1999).

If there is not enough capacity in some resources (i.e., Eq. (1) is not respected), the following alternatives are presented to make proper decisions on the order:

- Increasing the capacity of the resources that do not have enough capacity to fulfill the new order.
- Rejecting some low priority orders belonging to type III, so that there is enough capacity to fulfill the new order.
- Delaying some low priority accepted orders.
- Rejecting the new order with considerable required capacity. This order can be an undesirable order. It causes chaos in the production planning that leads to delay of the accepted orders.

(b) If the order has low priority, the following equation is used to compare the required capacity of all the existing orders with the maximum available capacity:

$$\sum_{i \in O_r} (TWK_{ir} \times p_i) \leq C_{rt} \times (1 - \alpha_{rt}), \quad \forall r, t = (t_{now}, \ldots, T) \tag{2}$$

$\alpha_{rt}$ indicates a percentage of $C_{rt}$ that is put aside for high priority future arriving orders at period $t$. To estimate the values of $\alpha_{rt}$, the approach suggested by Sridharan (1998) can be used. Sridharan proposed a method to determine the reserved capacity for the high priority orders. Although this reserved capacity is not in the form of percentage, this method can be used to find an estimated quantity for $\alpha_{rt}$. Capacity rationing has been found to result in consistently improved performance compared to the
case where no rationing is implemented. If Eq. (2) is not respected, the decisions for the new orders consist of rejecting or delaying the new order or delaying some low priority accepted orders. On the other hand, in the case of sufficient capacity for the new arriving order, the next step of the structure is implemented. The outline of the rough-cut capacity planning is depicted in Fig. 3. Thus, with the arrival of each new order, the management is able to take the necessary decisions such as rejecting the order, increasing the capacity or postponing the orders before running the IP1 model. It is noted that rough-cut capacity planning has not been used in the previous contributions to reject or accept the arriving orders.

### 3.1.2. Generating different alternatives to compute the order price

If the order is not rejected at the first step, the second step is implemented. The main aim of this step is to compute different values for the order completion dates (OCDs) and earliest release date (ERD) of the order as inputs of the IP1 model. Since the due date of the new order is known, the backward method is used to calculate OCDs and ERD. The values of OCDs and ERD are calculated by the following equations proposed by Kingsnam and Hendry (2002):

\[
\text{OCD}_{i,p,i} = \text{dd}_{i},
\]

\[
\text{OCD}_{i,p,i-1} = \text{OCD}_{i,p,i} - \text{TWK}_{i,p,i} - W_p,
\]

\[\vdots\]

\[
\text{OCD}_{i,p,r} = \text{OCD}_{i,p,r+1} - \text{TWK}_{i,p,r} - W_p,
\]

\[\vdots\]

LRD_i = \text{OCD}_{i,p,1} - \text{TWK}_{i,p,1} - W_p,

ERD_i = LRD_i - \text{pool delay},

where \( n_i \) is the number of required resources for order \( i \), \( \mu_{i,r} \) the \( r \)th required resource on the route of order \( i \); \( \text{dd}_{i} \) the due date of order \( i \); \( \text{OCD}_{i,p,i} \) the operation completion date of order \( i \) on resource \( r \); and \( \text{LRD}_{i} \) the latest release date of order \( i \). If the order is released to the shop floor after this time, it is impossible to meet its due date in regular time; \( \text{ERD}_{i} \) the ERD of order \( i \); \( \text{pool delay} \) potential released workload for all the accepted orders with material available waiting in the pool to be released to the shop floor; and \( W_p \) the average waiting time per resource for an order with priority \( p \).

The initial value of \( W_p \) can be computed by the following equation similar to that presented by Kingsman (2000):

\[
W_p = \frac{T - \left( \sum_{r=1}^{R} \sum_{i \in O_r} \text{TWK}_{i,p} \right)}{R'},
\]

where \( T \) and \( \text{TWK}_{i,p} \) are expressed in working days and \( R' \) is the average number of resources per order. As recommended by Kingsman, this initial value needs to be validated by simulation. \( p \) consists of two priorities; normal priority and hot priority. Normal priority is assigned to low priority orders. However, if high priority orders cannot be delivered on time, besides normal priority, hot priority can also be assigned to shorten the delivery time.

Before running the IP1 model, the possibility of meeting the due date should be checked by comparing material arriving date of order \( i \) (MAD_i),
ERD, LRD, and \( t_{\text{now}} \). Since the order entry stage is the medium term planning, the exact data does not need and an estimation of average value of MAD is acceptable based on the past performance of suppliers. If raw material and components are received on time, no prioritization and extra activities (e.g., overtime and subcontracting) are needed and the new order is placed at the end of the pool queue. Otherwise, according to Eq. (3), the due date of the new order can be met by changing the values of pool delay, OCDs, and \( W_p \).

With comparison of MAD, ERD, LRD, and \( t_{\text{now}} \), one of the five following conditions happens:

(a) LRD < \( t_{\text{now}} \): It is impossible to deliver the order on time, so it should be rejected.

(b) ERD < \( t_{\text{now}} \) < LRD: If the new order is placed at the end of the pool, the order is met with delay. To meet the order on time, some alternatives are proposed as follows:

- Increasing the new order priority in the pool. The amount of increase must be at least equal to (MAD - ERD) to reach the equation ERD = MAD.
- Changing the values of OCDs: Besides the previous alternative, it is possible to reach equality between ERD and MAD by changing the values of OCDs. The new values of OCDs are obtained by the following equation:

\[
\text{OCD}'_{i,j,(l,r)} = \text{OCD}_{i,j,(l,r)} - \frac{\text{MAD} - \text{ERD}}{n_i}.
\]

(c) \( t_{\text{now}} \) < MAD < LRD: The raw material and components are received on time and the new order is placed at the end of the pool.

(d) ERD < MAD < LRD: This case is similar to case (b).

(e) MAD > LRD: In this case, the order is probably met with a considerable delay. To decrease the amount of delay, the following alternatives are suggested:

- Changing the values of OCDs so that LRD = MAD (Eq. (6)). So, the order is placed at hot priority. However, this alternative may cut the values of OCDs considerably and hence resources will need high capacity in different periods:

\[
\text{OCD}'_{i,j,(l,r)} = \text{OCD}_{i,j,(l,r)} - \frac{\text{MAD} - \text{LRD}}{n_i}.
\]

- Assigning hot priority to the high priority orders in queues of different resources. This alternative is similar to the fourth alternative of case (b). The queue times must be decreased by means of changing priorities so that LRD = MAD.
- Rejecting the low priority new order that is met with long delay or has a negative effect on production of other orders in the system.
- Delaying the due date of the new order.

The summarization of suggested alternatives is depicted in Fig. 4.

3.1.3. Running the IP1 model to determine the prices of the new orders

As mentioned earlier, the IP1 model is used to determine the prices of the new orders. If the orders are not rejected at the previous steps, different prices are determined by means of different pairs of (ERD, OCDs). Eventually, the minimum computed price of each order is quoted to the customer. It is noted that if there are many pairs, the IP1 model should be run for those pairs that have a considerable difference between their values of ERD and OCDs. MTO companies typically establish their prices using a standard mark-up approach (Easton and Moodie, 1999). Mark-up is the amount that a company adds to its estimated direct costs for the order to cover overhead, unplanned expenses, and desired profit. The final order price is the sum of operating costs plus a standard mark-up.

Indices, input parameters and decision variables of the IP1 model are defined as follows:
Calculating different ERDs and OCDs for the new order

- If $LRD < t_{now}$
  - No
  - If $ERD < t_{now}$
    - No
      - If $MAD < ERD$
        - No
          - If $MAD > LRD$
            - Yes
              - Assigning hot priority to the high priority orders in queues of resources
              - Similar to $ERD < t_{now}$
            - No
              - Yes
                - Increasing the priority into the pool
                - Changing the OCDs so that $ERD = MAD$
      - Yes
        - Changing the OCDs so that $LRD = MAD$
  - Yes
    - Rejected the new order

Indices
- $I$ order index ($i = 1, ..., I$)
- $R$ resource index ($r = 1, ..., R$)
- $T$ period index ($t = 1, ..., T$)

Parameters
- $cr_{irt}$ production cost of order $i$ on resource $r$ at period $t$ in regular time
- $co_{irt}$ production cost of order $i$ on resource $r$ at period $t$ in overtime
- $cs_{irt}$ subcontracting cost of order $i$ on resource $r$ at period $t$
- $ct_i$ lateness penalty of order $i$ per unit time
- $iw_{irt}$ workload of order $i$ on resource $r$ awaiting material with ERD at period $t$, i.e., input workload to the shop floor which should be produced ($iw_{irt} = TWK_{ir}$)
- $iwp_r$ total workload that has been remained from the previous planning horizon (workload of the orders in the pool and also the orders which have been released to the shop floor but have not yet gone through resource $r$ and require operation on the resource)
- $ow_{irt}$ total workload of order $i$ on resource $r$ with OCD at period $t$ (i.e., the output of workload required from resource $r$ at period $t$ to ensure that the order $i$ conforms to its planned OCDs and hence meet its planned delivery date ($ow_{irt} = TWK_{ir}$))
- $CR_{rt}$ maximum regular-time capacity of resource $r$ available at period $t$, typically in machine hours
- $x_{rt}$ some percentage of resource $r$ at period $t$ considered for future arriving orders
- $CO_{rt}$ Maximum overtime capacity of resource $r$ available at period $t$, typically in machine hours
- $CS_{rt}$ Maximum subcontracting capacity of resource $r$ subcontracting at period $t$, typically in machine hours
- $OS(i)$ set of orders which must be delivered on time
- $M$ a very large number

Decision variables
- $Y_{irt}$ amount of resource $r$ assigned to order $i$ at period $t$ including regular, overtime, and subcontracting work, typically in machine hours
- $O_{irt}$ amount of resource $r$ assigned to order $i$ at period $t$ during overtime, typically in machine hours
\( S_{i|rt} \) amount of resource \( r \) assigned to order \( i \) at period \( t \) which is supplied by subcontractors, typically in machine hours

\( \text{LT}_i \) lateness amount of order \( i \)

\( \text{FT}_i \) completion date of order \( i \) on the last resource

\[
X_{it} = \begin{cases} 
1 & \text{if } Y_{i,pt(d_i, t)} > 0, \\
0 & \text{otherwise},
\end{cases}
\]

\[
\text{Min } Z = \sum_{i \in O_r} \sum_{r=1}^{R} \sum_{t=1}^{T} [ct_{i|rt} \times (Y_{i|rt} - O_{i|rt} - S_{i|rt})] \\
\times [\text{co}_{i|rt} \times O_{i|rt} + cs_{i|rt} \times S_{i|rt}] \\
+ \sum_{t \not\in OS(t)} ct_i \times \text{LT}_i
\]

s.t.

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

(2) \[
\sum_{i \in O_r} O_{i|rt} \leq CO_{i|rt}, \quad \forall r, t,
\]

(3) \[
\sum_{i \in O_r} S_{i|rt} \leq CS_{i|rt}, \quad \forall r, t,
\]

(4) \[
iwp, + \sum_{i \in O_r} \sum_{t=1}^{T} iw_{i|rt} \times p_i \leq \sum_{i \in O_r} \sum_{t=1}^{T} Y_{i|rt}, \quad \forall r
\]

(5) \[
\sum_{i \in OS(i), k=1}^{t} ow_{i|rk} \times p_i = \sum_{i \in OS(i), k=1}^{t} Y_{i|rk}, \quad \forall r, t,
\]

(6) \[
\sum_{k=1}^{t} iw_{i|rk} \times p_i \leq \sum_{k=1}^{t} Y_{i|rk}, \quad \forall r, i \in OS(i),
\]

(7) \[
Y_{i,pt(d_i, t)} \leq M \times X_{it}, \quad \forall t, i \not\in OS(i),
\]

(8) \[
-FT_i + t \leq M \times (1 - X_{it}), \quad \forall t, i \not\in OS(i),
\]

(9) \[
\text{LT}_i \geq \text{FT}_i - \text{dd}_i, \quad \forall t \not\in OS(i),
\]

(10) \[
\text{LT}_i \leq T - \text{dd}_i, \quad \forall t \not\in OS(i),
\]

(11) \[
\sum_{k=1}^{t} ow_{i|rk} \times p_i = \sum_{k=1}^{t+(T-\text{dd}_i)} Y_{i|rk}, \quad \forall r, i \not\in OS(i),
\]

The objective function of the IP1 model is to minimize the operating costs (i.e., regular time, overtime, subcontracting, and lateness penalty costs). The lateness penalties of the accepted orders which are met with delay due to the arrival of the new order are considered as operating costs of the new order. Constraint (1) is the capacity constraint during regular time. Constraints (2) and (3) represent the limitations of the maximum available overtime and subcontracting capacities. Constraint (4) guarantees that the total workload of all orders into the system over planning horizon, \( T \) will be produced. The left-hand side of constraint (4) is the total workload in the system over \( T \) that has to be carried out on resource \( r \). To meet all orders within \( T \), the sum of actual amount of workload done by resource \( r \) within \( T \) must be equal to or greater than total work in the system. This constraint just provides a simple initial check on the capacity and does not make sure that orders will be met on their delivery times because delivery times are in different periods of the planning horizon. To ensure that the orders are produced based on their due dates, extra constraints are essential.

Constraints (5) and (6) are considered only for those orders which must be delivered on time. Constraint (5) ensures that orders are completed by their OCDs. The left-hand side represents total required output, which should be completed over periods \( k = 1 \) to \( t \). The right-hand side represents the total actual workload done on the resource \( r \) over periods \( k = 1 \) to \( t \). So, by means of Constraint (5) all output in the system will be completed by their OCDs. The equality presents that the production line assigns the capacity to the current orders in the system equal to the required amount. In other words, no inventory is permitted in the system. Although constraint (5) increases the possibility of meeting the order on time, it cannot only guarantee that the orders are certainly completed by their delivery times. OCDs are planned over the time bucket \( t \). Hence, the OCDs may occur at the beginning or at the end of period \( t \) or any other time within the time bucket \( t \). In order to push the model to plan on time orders so that they are completed at an exact time (delivery time), Constraint (6) is essential. Constraint (6)
emphasizes that the required capacity of each on time order over periods \( k = 1 \) to \( t \) (i.e., the left hand side) must be provided over periods \( k = 1 \) to its due date so that the order can be delivered on time.

Constraints (7)–(11) are related to the orders that can be delayed. Constraints (7)–(10) represent the calculation of the lateness value for these types of orders. In Constraint (7), the completion date of the order \( i \) (\( FT_i \)) is determined. When \( Y_{i,t}(i,t) \) becomes positive it means that the time bucket \( t \) is the completion period of the order \( i \). So, to respect Eq. (7), \( X_{it} \) gets the value of one. As a result, the right hand side of Constraint (8) becomes zero and the lower bound of \( FT_i \) is determined. With respect to the other constraints and particularly the lateness penalty, the value of \( LT_i \) is computed by the model. Constraint (10) guarantees that delivery time for delayed orders can not exceed the planning horizon. Constraint (11) is similar to Constraint (5) for on time orders. Based on Constraint (11), the required workload of delayed orders at each period can be supplied with \( (T - dd_i) \) delay, which is the upper limit of \( LT_i \). This constraint leaves the capacity for on time orders over \( t \) to \( t + (LT_i + dd_i) \). Constraints (12) and (13) define non-negativity and type of the decision variables.

### 3.1.4. Acceptance/rejection decision on the new order by the customer

After determining the new order price by the IP1 model, the price is quoted to the customer by the MTO company. The customer selects the best price based on its criteria among some MTO firms. If the order is accepted by the customer, the last step of the proposed structure is implemented.

### 3.1.5. Running the IP2 model to select the best set of suppliers and subcontractors

The decision-making problem of this step is the optimal selection of suppliers and subcontractors to supply the required raw material and workload of the new accepted orders. The current set of suppliers and subcontractors are first selected by considering other major criteria such as quality. The structure of the inbound side of considered supply chain is a single tier parallel supply chain for each required raw material and workload. It is assumed that each required raw material and workload of the accepted orders can be provided by only one selected supplier and subcontractor.

To find the best set of suppliers and subcontractors, the IP2 model is suggested. The MTO system demands raw material and workload based on the requirements of the new accepted orders, so the IP2 model is an assignment model. Criteria of assignment are price and delivery time for suppliers and price for subcontractors. As mentioned earlier, subcontractors can supply required workload on time. Hence, only one criterion (i.e., price) is considered for them. Since the company provides only raw material and required workload of the accepted orders, the IP2 model only considers the accepted orders (types I and II). Consideration of contingent orders in the IP2 model cause the company to provide extra material and workload that can charge redundant expenses to the company. That is why two separated mixed-integer programming models are suggested. There is a time gap between quoting time of the new order price to the customer and the acceptance time of the order by the customer. During this time, some contingent orders are confirmed by the customers. In this case, these orders are considered in the IP2 model. Additional notations of the IP2 model are defined as follows:

### Indices

- \( s \) subcontractor index \((s = 1, \ldots, S)\)
- \( \ell \) supplier index \((\ell = 1, \ldots, L)\)
- \( k \) material index

### Parameters

- \( LT_i \) lateness amount of order \( i \) (i.e., output of the IP1 model)
- \( p_{irs} \) suggested price of subcontractor \( s \) for workload of order \( i \) on resource \( r \)
- \( p_{\ell k} \) suggested price of supplier \( \ell \) for raw material \( k \)
- \( MAD_{\ell k} \) delivery time of raw material \( k \) of order \( i \) by supplier \( \ell \)
- \( S_{rsti} \) maximum workload of order \( i \) on resource \( r \) that can be supplied by subcontractor \( s \) at period \( t \)
\begin{align*}
\text{Min } Z &= \sum_{i=1}^{I} \sum_{r=1}^{R} \sum_{s \in S(r)} P_{irs} X_{irs} \\
&\quad + \left[ \sum_{i \in \text{NO}(i)} \sum_{k=1}^{K} \sum_{t \in L(k)} P_{ikt} X_{ikt} \right] \\
&\quad + \sum_{i \in \text{NO}(i)} \sum_{k=1}^{K} \sum_{t \in L(k)} \beta_i (\text{ERD}_i - \text{MAD}_{ikt}) X_{ikt} \\
&\quad + \sum_{i \in \text{NO}(i)} \sum_{k=1}^{K} \sum_{t \in L(k)} \beta_i^t (\text{MAD}_{ikt} - \text{ERD}_i) X_{ikt} \\
&\quad + \sum_{s \in S(r)} \sum_{i=1}^{I} \sum_{r=1}^{T} \sum_{t=1}^{T} \left( Y_{irt} - O_{irt} \right) \\
&\quad - \sum_{s \in S(r)} (S_{irs} X_{irs}) + c_{ort} O_{irt},
\end{align*}

s.t.

1. \( \sum_{i=1}^{I} Y_{irt} - O_{irt} - \left( \sum_{s \in S(r)} S_{irs} X_{irs} \right) \leq CR_{rt}(1 - \alpha_{rt}); \forall r, t, \)

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

3. \( \text{iw}_r + \sum_{i=1}^{I} \sum_{t=1}^{T} Y_{irt} \leq \sum_{i=1}^{I} \sum_{t=1}^{T} Y_{irts}; \forall r, \)

4. \( \sum_{i \in \text{OS}(i)} \sum_{k=1}^{K} \text{ow}_{irk} = \sum_{i=1}^{I} \sum_{k=1}^{K} Y_{irk}; \forall r, t, \)

5. \( \sum_{k=1}^{K} \text{iw}_{irk} \leq \sum_{k=1}^{K} Y_{irk}; \forall r, i \in \text{OS}(i), t \in (1, \ldots, d_d), \)

6. \( \sum_{k=1}^{K} \text{ow}_{irk} = \sum_{k=1}^{K} Y_{irk}; \forall r, i \neq \text{OS}(i), t \in (1, \ldots, d_d), \)

7. \( \sum_{i \in \text{NO}(i)} \sum_{t \in L(k)} X_{ikt} = 1; \forall k, i \in \text{NO}(i), \)

8. \( \sum_{s \in S(r)} X_{irs} = 1; \forall i, r, \)

9. \( Y_{irt}, O_{irt} \geq 0; X_{irs} \); \( \forall i, r, t, s \in S(r), \)

\( X_{ikt} \in \{0, 1\}; \forall k, i \in \text{NO}(i), t \in L(k)). \)

The objective function of the IP2 model, which is the same as the objective function of the IP1 model, is to minimize total operating costs, purchase costs of raw material, purchase costs of workload, and earliness or lateness costs of raw material. Constraints (1)–(6) are similar to the IP1 model constraints. Since the IP2 model is run only for the accepted orders, hence in Constraints (3)–(6), the values of \( p_i \) are equal to one. Constraints (7) and (8) guarantee exactly one supplier and one subcontractor is chosen to supply required raw material and workload of the accepted orders. By means of running the IP2 model, the production plan in a MTO system is updated based on features of the new arriving orders and current conditions of the MTO system. In other words, this step indicates how we can apply the rolling horizon approach at the order entry stage.

### 3.2. Orders with negotiable due date

In this case, instead of setting a fixed due date, the customer asks the MTO company to offer a set of different prices and due dates for the new order.
Then, the customer will select the best combination of price and due date based on its own criteria. Similar to the fixed due date, the five following steps are implemented as illustrated in Fig. 1.

3.2.1. Calculating rough-cut capacity

This step is done exactly similar to the fixed due date case.

3.2.2. Generating different alternatives to compute a variety of delivery times

If the order is not rejected at the first step, the value of delivery time of the new order must be known to run the IP1 model. This value is calculated by the following forward computations:

\[
\begin{align*}
\text{LRD}_i &= \text{ERD}_i + \text{pool delay}, \\
\text{OCD}_{i,p(i,1)} &= \text{LRD}_i + \text{TWK}_{i,p(i,1)} + W_p, \\
&\vdots \\
\text{OCD}_{i,p(i,n_i)} &= \text{OCD}_{i,p(i,n_i-1)} + \text{TWK}_{i,p(i,n_i)} + W_p, \\
\text{dd}_i &= \text{OCD}_{i,p(i,n_i)}. 
\end{align*}
\]  

(7)

By changing the parameters of Eq. (7), different delivery times can be generated. Three following alternatives are proposed to calculate different delivery times:

- Calculating different ERDs: Since the arrival time of raw material and components are not exactly clear, three ERDs are suggested: Optimistic, average, and pessimistic arrival times of raw material.
- Assigning different priorities to the new order into the pool: By means of this alternative, different delivery times can be generated. The order with normal priority is placed at the end of queue while the order with hot priority is released to the shop floor as soon as its raw material and components are received (i.e., pool delay = 0). Besides mentioned priorities, the new order can be placed at different priorities of the pool with respect to its importance rate.
- Assigning normal or hot priority in queues of production resources to generate different delivery times for the new order.

These suggested alternatives are summarized in Fig. 5. By combining the mentioned alternatives, different sets of (dd, ERD, OCD) are generated, which are the main inputs of the IP1 model. The new order price is determined per (dd, ERD, OCD) set. Before running the IP1 model, the following combinations should be eliminated:

- Delivery times of combinations that are more than planning horizon \( T \).
- Combinations in which there is a little time gap between the delivery time and the present time. In this case, the company has to provide considerable overtime and subcontracting to meet such an order and can cause chaos in the system to meet other accepted orders.

3.2.3. Running the IP1 model to determine the prices of the new orders for different delivery times

The IP1 model is run for each set of (dd, ERD, OCD). Finally, a price is obtained for each set.

3.2.4. Acceptance/rejection decision on the new order by negotiation

Different sets of (dd, price) are quoted to the customer. Such information facilitates the negotiation process and helps the MTO company and the customer to reach a reasonable and profitable agreement on the price and the delivery time of the new order.

3.2.5. Running the IP2 model to select the best set of suppliers and subcontractors

With a similar process for fixed due date, the best combination of suppliers and subcontractors is selected by the IP2 model.

By running the IP1 and IP2 models based on the sequence presented in Fig. 1, the MTO company
will have an appropriate relationship with other affected parties to manage the arriving orders. Moreover, the proposed structure for the order entry stage facilitates the other decision stages including the job release and the dispatching stages.

4. Numerical experiments

In this section, to evaluate the performance of the proposed models, we indicate how the computational time increases as the size of the test problems increases. The required data of the test problems are randomly generated. The time unit is assumed 1 week and the planning horizon length is equal to 12 weeks for all test problems. Six sets of test problems with different sizes have been considered, as shown in Tables 1 and 2, and ten problems for each set are randomly generated. For each set of test problems, a Lingo model has been generated using Lingo 6.0 modeling language, and all of the test problems have been solved on a personal computer with an Intel Pentium 4 processor running at 2.8 GHz.

Table 3 represents the average CPU time required to obtain an optimal solution of the IP1 and IP2 models for each set of test problems. It is noted that in the IP1 model, for \(18 \times 8 \times 11\) problem instances, it was not possible to find an optimal solution within reasonable CPU time. Computational results indicate that the IP1 model can obtain an optimal solution for small and moderately medium-sized problems within reasonable time. However, it cannot obtain an optimal solution for medium and large-sized problems within reasonable time because solution time grows exponentially with

<table>
<thead>
<tr>
<th>Problem set number</th>
<th>Number of test problems</th>
<th>Average CPU time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IP1 model</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>214</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>442</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA, not available.
increasing the size of the problem. With respect to the numerical experiments, the IP1 model is sensitive to the number of delayed orders in the system. However, the IP2 model can obtain an optimal solution for all sizes within reasonable time. That is because the structure of the IP2 model is a simple assignment problem, in which among pre-selected suppliers and subcontractors, some of them are chosen based on their suggested prices and delivery times.

Therefore, a more efficient heuristic method should be developed to obtain a near-optimal schedule for medium and large-sized problems in the IP1 model within reasonable CPU time. In this regard, one of the modern stochastic search methods such as genetic algorithm (GA) or tabu search (TS) can be used as the solution method for the problem.

5. Conclusions

In MTO systems due to a wide variety of products, low amount of standard products and the lack of possibility of appropriate forecasts, the production planning is more complicated than MTS systems. The arrival times of orders are stochastic over time and each new order often has a significant effect on profit and market share of company. Therefore, MTO companies need an efficient decision-making structure at the order entry stage in which the company can select an appropriate combination of the arriving orders. This combination must lead to the increase of the profit and the market share of the company significantly. In this paper, a new decision-making structure is proposed to select this combination.

The proposed structure contains the following advantages in comparison with the previous research:

- In the proposed structure, all affected parties of the supply chain are taken into account to manage and plan the arriving orders efficiently. These parties consist of customers, the MTO company, subcontractors, and suppliers. The parties are considered in an appropriate hierarchy to take decisions that are acceptable for all parties.
- Since undesirable orders are identified and rejected through some special criteria, it leads to better planning and control for the rest of the arriving orders and also can prevent the IP1 model from infeasibility.
- Generating different alternatives to compute price and delivery time of each arriving order. These alternatives increase the chance of acceptance of the orders because customers face with different values and can select one, which is appropriate for them. In brief, this contribution facilitates the negotiation process for the MTO company and customers that is an important stage in the acceptance/rejection decision (Calosso et al., 2003, 2004).
- The use of two developed mixed-integer programming models in the structure to determine price and delivery time of the accepted orders and also to select the optimal combination of subcontractors and suppliers. These models help the MTO system to manage the arriving orders efficiently.

Moreover, the following directions can be considered for future research:

- So far, extent research has been done on the job release and the dispatching mechanisms as short term planning in MTO environments. Aggregation of the order entry and short-term planning stages within a hierarchical production planning structure in MTO companies will certainly improve the performance of the MTO company. In this regard, the hierarchical production planning structure includes three main stages, i.e., the order entry stage, the order release stage and the dispatching stage. This issue is on our research line.
- Nowadays, besides price and due date criteria, the quality of final products is also important for the customers. Since suppliers and subcontractors have a great effect on the quality of final products, hence quality criterion should be added to the criteria set to select the best set of suppliers and subcontractors.
- The proposed mathematical models can be improved in different ways. For example, the assumption of providing each raw material and workload by only one selected supplier and subcontractor can be neglected in the IP2 model. In such a case, in addition to the selection of the best combination of supply chain partners, the purchased amount of raw material and workload from each partner is also computed.
Appendix A

To show the applicability of the mathematical models, a typical MTO system is considered with two resources (i.e., work centers) and four orders (including three current orders and a new high priority order). The planning horizon length is assumed 6 weeks. All orders have a fixed due date and two of them can be delayed. Table A1 represents the required data for this example. To discuss shortly, most parameters are considered the same for different resources and periods. The maximum regular-time capacity, the overtime capacity and the subcontracting capacity are 40, 20, and 10 h/week, respectively. The value of $a_{rt}$ is set to 0.1 for all resources at all periods. Also, the values of $cr_{irt}$, $co_{irt}$ and $cs_{irt}$ are equal to 50, 150, and 350 per all indices, respectively.

Based on the definition of $iw_{irt}$, $ow_{irt}$ and Table A1, these parameters can be computed as shown in Table A2. Also, the values of $iwp_r$ are set to 0.

By running the IP1 model, the following results are obtained:

$$X_{36} = X_{46} = 1 \Rightarrow FT_3 = FT_4 = 6,$$
$$LT_3 = 1, \ LT_4 = 2, \ Z^* = 20 \ 600.$$

For running the IP2 model, it is assumed that contingent order 2 (C) is rejected by the customer. The required raw material, components and workload of existing orders can be provided by three suppliers and three subcontractors. All existing orders need three common raw material and components. The values of $\beta_i$ are equal to 40, 30, and 20 for the orders respectively. Also, the values of $\beta'_i$ are equal to 4000, 3000 and 2000. Other required data for running the IP2 model are presented in Tables A3 and A4.

By running the IP2 model, the following results are obtained:

Suppliers:

$$X_{111} = X_{122} = X_{132} = X_{213} = X_{223}$$
$$= X_{233} = X_{312} = X_{323} = X_{331} = 1,$$
$$\Rightarrow Z^* = 496, 780.$$

Subcontractors:

$$X_{113} = X_{123} = X_{212} = X_{223}$$
$$= X_{313} = X_{321} = 1.$$
References


