Recent studies that consider optimisation of the sequence in which customer orders are processed in a detailed manner, while leaving the admission control decisions out are the following. Ata and Olsen (2013) consider two customer classes and the problem of offering an incentive compatible menu of price/leadtime pairs in which both components are dynamically set. They assume that production decisions are made at discrete points in time and prove that a discrete-time version of the $Gc\mu$ rule is asymptotically optimal. Öner-Közen and Minner (2017) use a discrete time MDP for optimising dynamic dispatching decisions based on the information on the remaining time of orders until their due date. The tardiness penalty combines a fixed and a variable cost. Orders arrive with exogenously determined stochastic customer-required leadtimes and all orders are accepted unless the system reaches its capacity. We use their model as a building block and extend it to simultaneously consider order selection via dynamic price/leadtime quotation.

3. Markov decision model

We model an MTO firm as a discrete time queuing system with a single server and a limited waiting space. We assume that time is divided into sufficiently small discrete periods so that only one of the following three possible events can happen in each period: (i) Arrival of a prospective customer with probability γ, (ii) completion of the order in processing, if there is any, (i.e. departure) with probability β or (iii) no customer arrival or order completion with probability $\theta = 1 - \gamma - \beta$.

Customers are divided into two types. Leadtime sensitive customers (LS) are willing to pay more for receiving shorter leadtimes. Price sensitive customers (PS) are willing to wait more for paying less. Upon arrival of a prospective customer, the customer type is realised. In practice, firms can infer the customer type from available information about the customer. Plambeck (2004) states that in the automotive industry, consumers are typically leadtime sensitive while corporate fleet buyers are typically price sensitive. Furthermore, she gives the example of Timbuk2 who serves end-users via the Internet (leadtime sensitive), in addition, retailers (price sensitive). In these examples, the customer type is realised as soon as marketing knows whether a retailer, a fleet buyer or a consumer is dealt with. The customers from different geographical regions may also have clearly different preferences (Duenyas 1995). In this case, the information on the region of a customer may reveal its type.

Given that an arrival occurs, the probability that it is a leadtime sensitive customer is ζ. The decision about what price/leadtime pair (p, L) to quote is made accordingly. There are lower and upper limits to quotable price and leadtime values. Upper limits represent the maximum amount a customer is willing to pay ($p_{\max}$) and the maximum amount of time he is prepared to wait ($L_{\max}$) for the product. Lower limits, on the other hand, represent the minimum possible price and leadtime that can be quoted for the product, which are denoted $p_{\min}$ and $L_{\min}$ respectively. For example, $p_{\min}$ can be the expected cost and $L_{\min}$ can be the expected time for processing an order.

Customers demand improved service at a lower price. We model the probability that a customer accepts a given quote using a function that decreases both in price and leadtime. By using the S-shaped logistical response function, following Easton and Moodie (1999) and Watanapa and Techanitisawad (2005), we model the probability of acceptance of a (p, L) quote as follows.

$$P^i(p, L) = \begin{cases} \left[1 + \xi_0 e^{\xi_L^i (L - L_{\min}) + \xi_p^i (p - p_{\min})}\right]^{-1} & p_{\min} \le p \le p_{\max},\ L_{\min} \le L \le L_{\max} \\ 0 & \text{otherwise} \end{cases} \quad i = LS, PS \tag{1}$$

Customers know ranges of possible price and leadtime offer ($p_{\min}$ and $L_{\min}$) that they can get on the market. The probability that the quote is accepted by customer type i decreases with factor ξ_p^i as the difference between offered price and $p_{\min}$ increases, and with factor ξ_L^i as the difference between quoted leadtime and $L_{\min}$ increases. The values of these sensitivity parameters can be determined by using historical data. We assume $\xi_p^{PS} \ge \xi_p^{LS}$ and $\xi_L^{PS} \le \xi_L^{LS}$. Note that for $\xi_0 > 0$, even when the firm offers $p_{\min}$ in combination with $L_{\min}$, the probability that the customer accepts the quote is not equal to 1. This accounts for possibly other factors influencing the customers' choice.

If the customer accepts the quote, the order joins the production system and a revenue $R = p$ is earned; otherwise, the customer is lost. Whenever a prospective customer accepts the quote, it is referred to as 'order'. We incorporate an upper bound K on the number of pending orders that the system can accommodate. This upper bound represents the financial or space-related limitations that may apply for a firm. For example, consider a firm that orders raw material for processing customer orders. The raw material can consume available physical space in the production facility and/or budget. A prospective customer is automatically rejected when K is reached. The due date of an order is equal to its arrival time plus the quoted leadtime. When the remaining time of an order until the due date becomes negative (i.e. the order is late), a tardiness penalty is incurred. The tardiness penalty is the sum of a fixed and a variable cost. We assume that the customers want to get the product as soon as completed, therefore there is no penalty for completing orders before their due dates. A variable processing cost is incurred in each period during which an order is processed.

The orders that join the production system are put into a pool and the decision which of the pending orders to process next is made upon completion of the currently processed order based on the information about the remaining time of orders until their due date. The pending orders are numbered according to their arrival sequence, so that number one is always taken for processing, when the special case of FCFS applies.

The problem is modelled as a discrete-time MDP where every period is a decision epoch. An infinite planning horizon is considered. The objective is maximising the long-run average profit per time unit.

3.1 *State and action space*

We define states of the MDP such that they describe the system at a period after the event occurrence and before the decision. The state description includes an indicator variable to distinguish between three possible events since the decisions made at periods with arrival differ from those made at periods of completion and at all other periods there is no decision to make. Furthermore, the decisions at arrival periods are distinguished between customer types. Thus, the state of the MDP is denoted as $(ind, n, r_1, r_2, \ldots, r_n)$ where $ind = 0$ represents a period with no order arrival or completion, $ind = 1$ with an order completion, $ind = LS$ with a prospective LS customer arrival, and $ind = PS$ with a prospective PS customer arrival. n denotes the number of orders in the system. It is the number of orders a completed order leaves behind when $ind = 1$. When $ind = LS, PS$, it is the number of orders that an arriving prospective customer *finds* in the system. So, after receiving the quote if the customer places an order, the total number of orders in that period becomes $n + 1$. r_i, $i = 1, 2, \ldots, n$ denote the remaining time of order i until the due date. r_1 gives the value for the order currently being served, if there is any. We denote the state space by S and the set of states in which the indicator variable is equal to $ind = 0, 1, LS, PS$ by S^{ind}.

Define $K(s)$ as the set of all possible actions in state s. For $ind = 1$, $K(s) = \{1, \ldots, n\}$, i.e. one of the n orders left behind the completed order is selected for processing next. For $ind = LS, PS$, $K(s) = \{(p, L)|p \in \{p_{\min}, \ldots, p_{\max}\}, L \in \{L_{\min}, \ldots, L_{\max} + 1\}\}$. Due to (1), making quote $(p, L_{\max} + 1)$ for any $p \in \{p_{\min}, \ldots, p_{\max}\}$ means that the customer is rejected. The values in the set of quotable leadtimes are integer numbers, since time is discrete. The set of alternative prices can contain real numbers. There is no decision, $K(s) = \{\}$, in dummy decision periods ($ind = 0$).

3.2 *Transition probabilities*

We present the transition probabilities for three different cases.

Case 0: $ind = 0$

Consider a period t, in which no order arrival or completion occurred. There is no decision to make. We define $p^0_{s,s'}$ as the probability that the system will be in state s' at the next decision epoch if the current state is s.

- For $s = (0, n, r_1 + 1, \ldots, r_n + 1)$, $\forall n \in \{1, \ldots, K - 1\}$,

$$p^0_{s,s'} = \begin{cases} \theta & s' = (0, n, r_1, r_2, \ldots, r_n) \\ \beta & s' = (1, n-1, r_2, \ldots, r_n) \\ \gamma \cdot \zeta & s' = (LS, n, r_1, r_2, \ldots, r_n) \\ \gamma \cdot (1-\zeta) & s' = (PS, n, r_1, r_2, \ldots, r_n) \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

  If the current state is s and, at the next decision epoch $(t + 1)$, no order arrival or completion occurs (θ), the system moves to $s' = (0, n, r_1, r_2, \ldots, r_n)$. The number of orders remains the same, the remaining times of the orders until their due date decrease by one. In s', the indicator also takes the value zero, since no order arrival or completion occurred in $t + 1$. If, at the next decision epoch, an order completion occurs (β), the system moves to $s' = (1, n - 1, r_2, \ldots, r_n)$. Since the completion of the order in processing has happened, s' includes the remaining times until the due date, which have decreased by one, only for $n - 1$ orders left behind, i.e. $r_2, \ldots, r_n$.
  If at the next decision epoch, a leadtime sensitive customer arrives $(\gamma \cdot \zeta)$, the system moves to $s' = (LS, n, r_1, r_2, \ldots, r_n)$. The number of orders remains n although an arrival occurred, since for $ind = LS$ n indicates the number of orders that the customer *finds* in the system. At this point neither the quote nor the decision by the customer whether to place an order or not is made. The logic is the same for PS.
- For $s = (0, 0, null)$, an order completion in the next decision epoch is not possible.

$$p^0_{s,s'} = \begin{cases} 1-\gamma & s' = (0, 0, null) \\ \gamma\zeta & s' = (LS, 0, null) \\ \gamma(1-\zeta) & s' = (PS, 0, null) \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

- For $s = (0, K, r_1 + 1, \ldots, r_K + 1)$ an arriving customer is automatically rejected.

$$p^0_{s,s'} = \begin{cases} 1-\beta & s' = (0, K, r_1, r_2, \ldots, r_K) \\ \beta & s' = (1, K-1, r_2, \ldots, r_K) \\ 0 & \text{otherwise} \end{cases} . \tag{4}$$

Case 1: $ind = 1$

Consider a period t in which an order completion occurred. A decision about which order to process next is made. We define $p^1_{s,s'}(k)$ as the probability that the system will be in state s' at the next decision epoch if the decision k is made when the current state is s.

- For $s = (1, n, r_1 + 1, \ldots, r_n + 1)$, $\forall n \in \{1, \ldots, K-1\}$ and $\forall k \in K(s) = \{1, \ldots, n\}$,

$$p^1_{s,s'}(k) = \begin{cases} \theta & s' = (0, n, r_k, r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n) \\ \beta & s' = (1, n-1, r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n) \\ \gamma\zeta & s' = (LS, n, r_k, r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n) \\ \gamma(1-\zeta) & s' = (PS, n, r_k, r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n) \\ 0 & \text{otherwise} \end{cases} . \tag{5}$$

  If, at the next decision epoch $(t+1)$, no order arrival or completion occurs (θ), the system moves to $s' = (0, n, r_k, r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n)$. In s', the remaining time of the orders until their due dates are one period less than they were in state s. r_k is represented by the third state variable since order k is selected for processing.
  If an order completion occurs (β) at the next decision epoch, the system moves to $s' = (1, n-1, r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n)$. Since order k has been moved to processing at period t, the completed order is the one with a remaining time of r_k until the due date at period $t+1$.
- For $s = (1, 0, null)$, which is the state where the system is left idle after an order completion, (3) holds.

$$p^1_{s,s'}(k) = p^0_{s,s'} \tag{6}$$

- $s = (1, K, r_1 + 1, \ldots, r_K + 1)$, is not possible, since a completed order cannot leave a full system behind.

Case 2: $ind = LS, PS$

Consider a period t in which a prospective customer arrived. A decision about which (p, L) pair to quote is made. We define $p^2_{s,s'}(k|a=1)$ as the probability that the system will be in state s' at the next decision epoch if the decision k is made when the current state is s and the customer has accepted the quote. This probability is denoted by $p^2_{s,s'}(k|a=0)$ if the customer has rejected quote k.

- For $s = (i, n, r_1 + 1, \ldots, r_n + 1)$, $i = LS, PS$, $\forall n \in \{1, \ldots, K-1\}$ and $\forall k \in K(s) = \{(p, L) | p \in \{p_{\min}, \ldots, p_{\max}\}, L \in \{L_{\min}, \ldots, L_{\max} + 1\}\}$,

$$p^2_{s,s'}(k|a=1) = \begin{cases} \theta & s' = (0, n+1, r_1, r_2, \ldots, r_n, L-1) \\ \beta & s' = (1, n, r_2, \ldots, r_n, L-1) \\ \gamma\zeta & s' = (LS, n+1, r_1, r_2, \ldots, r_n, L-1) \\ \gamma(1-\zeta) & s' = (PS, n+1, r_1, r_2, \ldots, r_n, L-1) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

  If the quote (p, L) is made in state s, the customer accepted the quote, and no arrival or completion occurs (θ) in the next decision epoch $(t+1)$, the system moves to $s' = (0, n+1, r_1, r_2, \ldots, r_n, L-1)$. Since a new order with a leadtime of L is obtained and added to the pool of orders at period t, s' has $n+1$ orders with the newly added order having $L-1$ remaining periods until its due date.
  If the quoted (p, L) pair is not accepted by the customer, transition probabilities are the same as the ones given in (2).

$$p^2_{s,s'}(k|a=0) = p^0_{s,s'} \tag{8}$$

  The transition probabilities are conditioned on the decision of the customer whether to accept the quote or not. The probability that one or the other happens (1) will come into play in the expression for the long-run average profit (13).

- For $s = (i, 0, null)$, $i = LS, PS$,

$$p_{s,s'}(k|a=1) = \begin{cases} \theta & s' = (0,1,L-1) \\ \beta & s' = (1,0,null) \\ \gamma\zeta & s' = (LS,1,L-1) \\ \gamma(1-\zeta) & s' = (PS,1,L-1) \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

If the quoted (p, L) pair is not accepted by the customer, the transition probabilities are the same as in (3).

$$p^2_{s,s'}(k|a=0) = p^0_{s,s'} \tag{10}$$

- $s = (i, K, r_1+1, \ldots, r_K+1)$, $i = LS, PS$, is not possible since a customer who arrives when the system is full is automatically rejected.

3.3 *Cost structure*

We consider the following cost structure to calculate the total tardiness cost incurred whenever a certain state is visited.

$$\begin{aligned} tcost_s = {} & (\text{Number of orders with } r_i = 0) \cdot (\text{fixed cost}) \\ & + (\text{Number of orders with } r_i \leq 0) \cdot (\text{unit variable cost}) \quad \forall s \in S \end{aligned} \tag{11}$$

The fixed cost is due as soon as the remaining time of an order until the due date becomes 0. At this period and at each of the following periods during which the order stays in the system, a unit variable cost is incurred. The state description does not include the remaining time until the due date for the completed order in a completion period. Suppose that the remaining time until the due date is '0' for the order in processing at a certain period and it is completed in the following period. This order experiences one period of lateness but a remaining time of -1 for this order is never represented in the system state.

A processing cost (c) is incurred in each period an order spends in processing. Thus, this cost applies each period unless the system is idle.

$$pcost_s = \begin{cases} c & \text{an order is in process} \\ 0 & \text{otherwise} \end{cases} \qquad \forall s \in S \tag{12}$$

3.4 *Computing the optimal policy*

Define g^* as the maximum long-run average profit per time unit. Let $g(R)$ be the long-run average profit per time unit that results from the actions of a stationary policy R. For obtaining the policy R with a $g(R)$ that is sufficiently close to g^*, we use value-iteration. We compute the value function $V_t(s)$ from the following relationship.

$$V_t(s) = \begin{cases} -tcost_s - pcost_s + \sum_{s' \in S} p^0_{s,s'} V_{t-1}(s'), & \forall s \in S^0 \\ \max_{k \in K(s)} \left\{ -tcost_s - pcost_s + \sum_{s' \in S} p^1_{s,s'}(k) V_{t-1}(s') \right\}, & \forall s \in S^1 \\ \max_{k \in K(s)} \left\{ P^i(k) \left(p - tcost_s - pcost_s + \sum_{s' \in S} p^2_{s,s'}(k|a=1) V_{t-1}(s') \right) \right. & \\ \left. + \left(1 - P^i(k)\right) \left(-tcost_s - pcost_s + \sum_{s' \in S} p^2_{s,s'}(k|a=0) V_{t-1}(s') \right) \right\}, & \forall s \in S^i, \quad i = LS, PS \end{cases} \tag{13}$$

First, for the states where no order arrival or completion occurs, no decision is made. Second, a departure occurs and the decision regarding which order to process next is made. Third, for the states with an order arrival, the decision regarding the price/leadtime quote is made. If the quote is accepted, a revenue that is equal to p is earned (line 3). The direct cost of visiting a state ($tcost_s + pcost_s$) appears in all lines and is incurred regardless of the decision.

Our algorithm stops at iteration t if the following is satisfied:

$$\frac{\max_{s \in S}\{V_t(s) - V_{t-1}(s)\} - \min_{s \in S}\{V_t(s) - V_{t-1}(s)\}}{\min_{s \in S}\{V_t(s) - V_{t-1}(s)\}} \leq \epsilon \tag{14}$$

which guarantees $\frac{g^* - g(R(t))}{g^*} \leq \epsilon$, where $R(t)$ is the policy at iteration t. For further details about the approach (see Tijms 2003, Chapter 6).