G Model
JMSY-166; No. of Pages 9

ARTICLE IN PRESS

Journal of Manufacturing Systems xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical Paper

Adaptive lead time quotation in a pull production system with lead time responsive demand

Barış Selçuk*

Department of Industrial Engineering, College of Engineering, Bahçeşehir University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 22 September 2011
Received in revised form 27 June 2012
Accepted 23 July 2012
Available online xxx

Keywords:
Kanban
Lead time
Markov chain
Quasi-Birth-Death
Tardiness

ABSTRACT

In this paper, a single-stage single-class kanban controlled manufacturing system is considered, where the problem is quoting accurate lead times for orders that are generated by an MRP system. Order release mechanism as in classical MRP is sensitive to the changes in the lead time according to the concept of lead time syndrome. The objective is to establish a cost effective lead time quoting procedure. The problem is modeled as a two-dimensional Markov chain, and it is solved explicitly by using matrix geometric techniques. Comparative analysis is done between static and dynamic lead time quoting procedures, and significant cost benefits of the dynamic procedure are shown under various scenarios. Guidelines for setting design parameters such as the number of kanbans and the frequency of updating the lead time are provided through numerical tests.

1. Introduction and literature review

The terms push and pull have been widely used to characterize the underlying decision making process in a wide variety of manufacturing and distribution systems. In general, pull systems use local information about the status of production and inventories to authorize the order releases, while push systems use global information including the firm and forecasted customer orders, lead times, capacities, etc.

Kanban is the most widely known pull type production control system. As the finished goods inventory is depleted and containers are emptied, kanban cards attached to each container are released to authorize production in order to replenish the consumed finished goods. Research on kanban type production control systems is huge. Akturk and Erhun [1], Kumar and Panneerselvam [16] and Junior and Filho [11] include a comprehensive review of various approaches and techniques for the analysis of kanban systems. Some recent examples include Erhun et al. [4], Koukoumialos and Liberopoulos [14], Krieg and Kuhn [15], and Iwase and Ohno [9] for performance evaluation and optimization of kanban systems. Majority of the literature aims to establish valid relationships between some design parameters (number of kanbans, kanban sizes, order cycle, and safety stock level) and the performance measures (total cost, throughput, inventory and WIP levels, utilization and fill rate) and to find (near) optimal values of design parameters under various production and demand settings. One important assumption that is commonly made in the literature is that demand for the product(s) is generated by some external stationary process, and information flow is limited to one direction in terms of orders from a downstream stage to an upstream stage. In this paper, we question this assumption by modeling a kanban system within an information-feedback framework. There have been various examples of information-feedback framework modeled in a production planning and control setting such as Takahashi and Nakamura [28], Pandey et al. [22], and Masin and Prabhu [17]. To the best of our knowledge, this paper is the first to present an exact analysis of adaptive lead times in a Kanban/MRP setting.

In a supply chain context it is possible to have both push and pull type systems tied together with flows of information in both directions, for example, capacity and lead time information from upstream to downstream and demand information from downstream to upstream. Joint consideration of push and pull systems has been an interesting research topic in the literature (e.g., [3,5,8,20]) such that the main motivation was to eliminate the drawbacks of both systems by designing an effective integration. The technique utilized by Flapper et al. [5] and Nagendra and Das [20] was to have MRP as responsible for the order generation and Kanban/JIT for the execution of production and material flow between stages of production. In Nagendra and Das [20], the lead time parameter is dynamically adjusted based on the production characteristics and utilization. The new lead time parameter is then feedback to both MRP and Kanban systems. Such an information-feedback system is similar to the setting we analyzed in this paper. However, as suggested by J.W. Forrester in his pioneering work

* Tel.: +90 212 3810890; fax: +90 212 3810550.
E-mail address: baris.selcuk@bahcesehir.edu.tr

0278-6125/$ – see front matter © 2012 The Society of Manufacturing Engineers. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jmsy.2012.07.017

[6] "... a complex information-feedback system designed by happenstance or in accordance with what may be intuitively obvious will usually be unstable and ineffective." Thus, within the context of Kanban/MRP integration, it is worthwhile to analyze the cyclic effects of adjusting lead times, which is called the lead time syndrome in the literature. Such cyclic effects are ignored in the well established due date setting literature where the main concern is to set optimal due dates for exogenously generated orders (e.g., [2,7,10]).

In this paper, we consider a situation where a kanban controlled manufacturing system is responsible for providing accurate supply related information such as lead times to an MRP system that releases production orders. The lead times are quoted based on the status information of the production system such as WIP and backlog levels, and MRP system plans and releases orders according to the quoted lead times. It has been argued in the literature that once the lead times are updated dynamically, this will cause the orders to be released in a more lumpy manner, and thus, lead time syndrome occurs. In response to larger work-in-process (WIP) levels, longer lead times are quoted, then orders are released earlier to cover increased expected demand during the longer lead time. This leads to longer queues in the shop floor, which in turn, lead to further increases in the quoted lead time. The reverse case holds for decreasing lead times. For some conceptual discussions on the lead time syndrome we refer to Mather and Plossl [18], Plossl [23], and Kingsmanet al. [13]. A simulation study by Selçuk et al. [27] provides experimental results for the existence of the lead time syndrome in a push type system. Further, Selçuk et al. [26], through an exact analytical model, shows that the lead time syndrome increases the variability in the manufacturing system, and leads to higher average WIP levels, and thus, longer flow times are experienced. These studies emphasize the fact that once different lead times are quoted, the order release mechanism will be affected in a way to increase the variability in the system. Therefore, it will be interesting to see how this phenomenon can be incorporated into establishing an effective lead time quoting procedure. Especially, in such an information-feedback system, there are two important factors that needs to be carefully considered. These are (1) the frequency with which the updating is performed and (2) the response probability such that the order release mechanism reacts to changing lead times. In this paper, we explicitly model these factors separately, and we aim to provide quantitative insights on their roles.

The time between a customer places an order and the customer receives the order in a kanban controlled system has been considered under different names in the literature. Kim and Tang [12] uses *response time*. Yavuz and Satir [29] uses *mean total demand satisfaction lead time*. Kumar and Panneerselvam [16] also identifies the response time as an important performance measure for kanban controlled systems. Kanban systems, by limiting WIP levels, are powerful in controlling manufacturing lead times between certain limits. However, short manufacturing lead times with low variability do not necessarily induce short and less variable response times [31]. This argument is supported by a simulation study of Yücesan and De Groote [30]. Kim and Tang [12] is the first to determine the optimal number and size of kanbans that minimize the average manufacturing lead time, and at the same time, satisfy a service level constraint on response time. Although these studies provide valuable insights on the impacts of various design variables on the response time, they consider the response time as a statistical measure (not as a quoted parameter), and assume that the demand process is independent of the response time. In this paper, different from the existing literature, we present an exact model of adaptive kanban response time (quoted lead time) such that the demand (orders released by MRP system) is sensitive to the quoted lead time.

In summary, we aim to answer the following research questions:

1 How to quote dynamically adjusted planned lead times in a kanban/MRP integration?
2 What is the effect of lead time syndrome?
3 What are the effects of design parameters such as the number of kanbans, the update frequency and the response probability?

This paper is organized as follows. In the following section, the detailed problem description is provided together with an explicit explanation of the lead time syndrome. Then, in Section 3 the problem is modeled by a Quasi-Birth-Death process, and an exact solution of this model is given. Section 4 presents various quantitative results generated by the solution of the model under different cases. In Section 5 the paper is concluded with some future research directions.

2. Problem description

There is a single-stage single-class kanban controlled manufacturing system that dynamically quotes planned lead times to an MRP system that, accordingly, releases orders. We assume that the orders follow the dynamics of the lead time syndrome. A detailed description of the lead time syndrome is provided in the following section.

2.1. Lead time syndrome

Consider an item for which production orders are released in integer multiples of a lot size x. Let us denote the lead time of the item as L. Let us also define $\lambda(t)$ as the expected release rate of lot size x at time $t \geq 0$. The expected total demand for the item during a period of time between now and a future time $t > 0$ is denoted by $d(0, t)$, and $d(t)$ is the expected demand rate at time $t > 0$. Let us also define the inventory position at time $t \geq 0$ by $IP(t)$. Then,

$$\lambda(0) = \frac{d(0,\ L) - IP(0)}{x}. \tag{1}$$

Consider that after a planned release of an order at time $t > 0$ the next order is planned to be released at time $t + \Delta t$. Then,

$$\begin{aligned}\lambda(t+\Delta t)\cdot x &= d(t+\Delta t,\ t+\Delta t+L) - IP(t+\Delta t)\\ &= d(t+\Delta t,\ t+\Delta t+L) - (d(t,\ t+L) - d(t,\ t+\Delta t))\\ &= d(t+L,\ t+L+\Delta t),\end{aligned}$$

which yields

$$\lim_{\Delta t\to 0}\lambda(t+\Delta t) = \lambda(t) = \frac{d(t+L)}{x},\ t > 0. \tag{2}$$

This means that the planned release rate at time $t > 0$ chases the demand rate with a time lag equal to the lead time.

Eqs. (1) and (2) demonstrate the fact that an update in the lead time at time $t = 0$ creates a significant effect in the current release rate, and the effect on the planned release rates in the future depends on the stationarity of the demand process. When L is increased by ΔL then, $\lambda(0)$ is increased by $\Delta\lambda(0) = d(L, L + \Delta L)/x$, and $\lambda(t)$, $t > 0$, is changed by $\Delta\lambda(t) = d(t + L + \Delta L) - d(t + L)/x$. Similarly, when L is decreased by ΔL then, $\lambda(0)$ is decreased by $\Delta\lambda(0) = d(L - \Delta L, L)/x$, and the change in the planned release rate at time $t > 0$ is $\Delta\lambda(t) = d(t + L - \Delta L) - d(t + L)/x$.

Given that the demand is stationary with a fixed mean, $d(t) = d$, then the effect of updating the lead time is formulated by

$$\Delta\lambda(0) = \frac{\Delta L\cdot d}{x}, \tag{3}$$

$$\Delta\lambda(t) = 0,\ t > 0. \tag{4}$$

Please cite this article in press as: Selçuk B. Adaptive lead time quotation in a pull production system with lead time responsive demand. J Manuf Syst (2012), http://dx.doi.org/10.1016/j.jmsy.2012.07.017

Eq. (3) and (4) indicate that for a stationary demand process, updating the lead time does only affect the current release rate, and future planned release rates remain unaffected. However, this is based on the assumption that the lead time is fixed during the rest of the time, which is a common assumption in MRP systems. Eq. (3) implies that whenever the lead time is increased, then additional orders are generated, and whenever the lead time is decreased, then some existing orders are canceled.

2.2. Lead time quotation

There is a fixed number of kanbans in the manufacturing system, each attached to a container of finished goods. If there are available finished goods, then an order is satisfied directly. In case of stock out, the unsatisfied order waits in the backlog queue. Let us denote $b_r(t)$ as the number of orders in the backlog queue at time t. The quoted lead time reflects the perceived duration of time that the next order will be satisfied. The perception is reflected through an update parameter r. There is also L_0 that refers to the minimum allowable lead time. The lead time quoted at time t is

$$L_r(t) = L_0 + \lfloor \frac{b_r(t)}{r} \rfloor. \tag{5}$$

The second term in the right-hand-side of Eq. (5) implies that the lead time is changed by every change of size r in the backlog level. This directly implies that the value of r gives the frequency with which the lead time is updated. The bigger r is, the lower the update frequency is, and vice versa.

We consider two different types of performance measures, *internal* and *external*. The internal performance measure relates to the efficiency of the manufacturing system with respect to the total amount of WIP and finished goods inventory, which is fixed to the number of Kanban cards, K. The external performance measure refers to the performance related to the order deliveries. The delivery performance in our case has three dimensions. These are backlogs, the length of the lead time, and tardy deliveries.

Let $E[B]$ and $E[L]$ are respectively the average backlog level and the average lead time. Similarly, $E[T]$ refers to the expected number of tardy deliveries that are completed (not canceled). Let c_h and c_b be respectively the inventory holding cost and the backlog cost of one container. c_L is the penalty cost for each unit of the quoted lead time, and c_T is the cost of late delivery of an order to the customer. The objective is to minimize the total cost of

$$TC(K, r) = c_h K + c_b E[B] + c_L E[L] + c_T E[T]. \tag{6}$$

A Markov chain model is developed in order to evaluate the performance of the manufacturing system for any given values of K, r, L_0.

3. Model

We assume that the backlog queue discipline is First-Come-First-Serve, and the lot size is equal to the mean demand, $x = d$, which is also equal to the container size. Order arrivals from the customer follows a Poisson process with a fixed rate of λ, and production cycle time for each container is exponentially distributed with a fixed mean of $1/\mu$. We model this queueing system as a two-dimensional Markov process defined by $\{X_r(t), t \geq 0\}$, $X_r(t) = (W(t), L_r(t))$, where $W(t)$ is the total amount of WIP in the system at time t, and $L_r(t)$ is the lead time at time t quoted by Eq. (5).

When $b_r(t) = r \cdot (L_r(t) - L_0) + r - 1$, an arrival triggers an increase in the lead time, and an additional order is released immediately with probability β. When $b_r(t) = r \cdot (L_r(t) - L_0)$, a departure triggers a decrease in the lead time, and an order in the production backlog is canceled immediately with probability β. As long as $b_r(t)$ remains in between, the lead time is kept constant. The transition rate diagram of the process $\{X_r(t), t \geq 0\}$ with $r = 3$ is provided in Fig. 1. We consider a setting (as also suggested by McKay and Wiers [19]) where the order release planner does not only act as an executor but also assumes a role as an anticipator, and this is incorporated into our model through the response probability β.

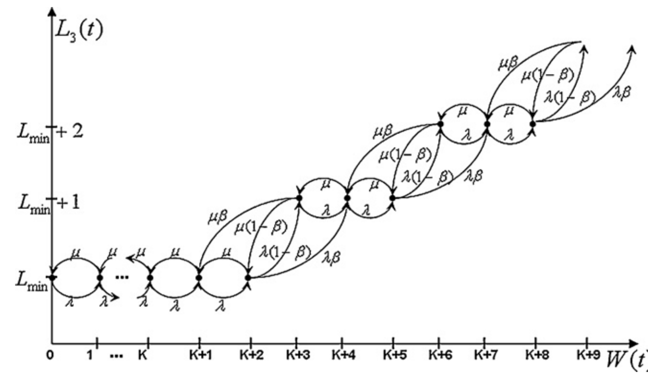

Fig. 1. Transition rate diagram for the process $\{X_3(t), t \geq 0\}$.

This process has a Quasi-Birth-Death (QBD) structure. Let us define this QBD process by $\{Y_r(t), t \geq 0\}$ with $Y_r(t) = (\hat{L}_r(t), \hat{b}_r(t))$, where $\hat{L}_r(t) = L_r(t) - L_0$ and $\hat{b}_r(t) = W(t) - K - r\hat{L}_r(t)$. Level 0 of this process has states $\{(0, -K), (0, -K+1), \ldots, (0, r-2), (0, r-1)\}$. For $l = 1, 2, \ldots$, level l of this QBD process is composed of r states with state space as $\{(l, 0), (l, 1), \ldots, (l, r-2), (l, r-1)\}$. The transition rate diagram in Fig. 1 can be transformed into a simpler representation as shown in Fig. 2.

Define $\delta_{m,n}$ as the Kronecker delta; $\delta_{m,n} = 1$ if $m = n$, and $\delta_{m,n} = 0$ if $m \neq n$. In order to conveniently describe the infinitesimal generator of the process $\{Y_r(t), t \geq 0\}$, we will employ the following notation for arbitrary integer k. $I^{(k)}$ is the k dimensional identity matrix. $T_R^{(k)}$ is the k dimensional right drift matrix, $(T_R^{(k)})_{m,n} = \delta_{m,n-1}$, $m, n = 0, 1, \ldots, k-1$. $T_L^{(k)}$ is the k dimensional left drift matrix, $(T_L^{(k)})_{m,n} = \delta_{m-1,n}$, $m, n = 0, 1, \ldots, k-1$. $e_m^{(k)}$ is the k dimensional unit column vector with 1 on the $(m+1)$th row, $(e_m^{(k)})_n = \delta_{m,n}$, $m, n = 0,$

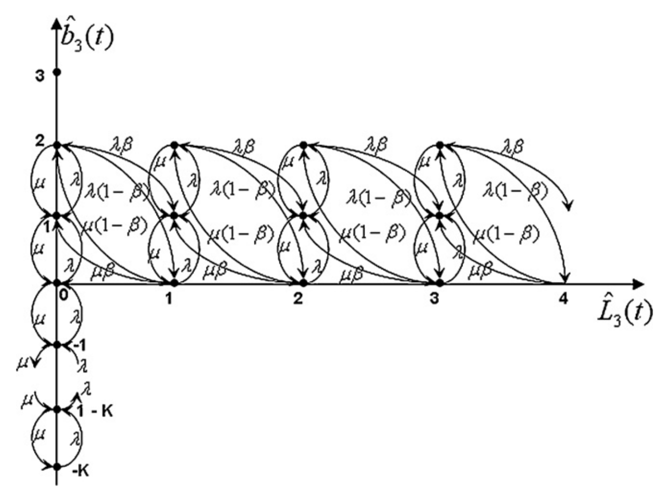

Fig. 2. Transition rate diagram for the processes $\{Y_3(t), t \geq 0\}$.

1, . . ., $k-1$. Given that the states are in lexicographic order, the generator $Q^{(r)}$ of $\{Y_r(t), t \geq 0\}$ is

$$Q^{(r)} = \begin{bmatrix} B_1^{(r)} & B_0^{(r)} & 0 & 0 & \cdots \\ B_2^{(r)} & A_1^{(r)} & A_0^{(r)} & 0 & \cdots \\ 0 & A_2^{(r)} & A_1^{(r)} & A_0^{(r)} & \cdots \\ 0 & 0 & A_2^{(r)} & A_1^{(r)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

$B_1^{(r)}$ is a $K+r$ dimensional square matrix, and it gives the transition rates between the states of level 0. $B_0^{(r)}$ is a $(K+r)\times r$ dimensional transition rate matrix from level 0 to level 1, and similarly $B_2^{(r)}$ is a $r\times(K+r)$ dimensional transition rate matrix from level 1 to level 0. $A_0^{(r)}$, $A_2^{(r)}$ and $A_1^{(r)}$ are r dimensional square matrices. $A_0^{(r)}$ gives the transition rates from level l to level $l+1$, $A_2^{(r)}$ is the transition rate matrix from level $l+1$ to level l, and $A_1^{(r)}$ gives the transition rates within level l, $l=1, 2, \ldots$. These transition matrices are written as

$$B_0^{(r)} = \lambda e_{K+r-1}^{(K+r)}((1-\beta)e_0^{(r)} + \beta e_1^{(r)})^T,$$

$$B_1^{(r)} = \lambda T_R^{(K+r)} + \mu T_L^{(K+r)} - (\lambda+\mu)I^{(K+r)} + \mu e_0^{(K+r)}(e_0^{(K+r)})^T,$$

$$B_2^{(r)} = \mu e_0^{(r)}(\beta e_{K+r-2}^{(K+r)} + (1-\beta)e_{K+r-1}^{(K+r)})^T,$$

$$A_0^{(r)} = \lambda e_{r-1}^{(r)}((1-\beta)e_0^{(r)} + \beta e_1^{(r)})^T,$$

$$A_1^{(r)} = \lambda T_R^{(r)} + \mu T_L^{(r)} - (\lambda+\mu)I^{(r)},$$

$$A_2^{(r)} = \mu e_0^{(r)}(\beta e_{r-2}^{(r)} + (1-\beta)e_{r-1}^{(r)})^T.$$

3.1. Steady state solution

For the derivations done throughout the rest of this paper we assume that the system modeled in the previous section is stable. We refer to Selçuk [25] for the derivation of the stability condition of this process, which is given by $\rho^r + \beta\rho^{r-1} - \beta\rho < 1$ for all $r=2$, 3, . . ., $0 \leq \beta \leq 1$ and $\rho=\lambda/\mu$. It is obvious that once the system is stable in the static case, $\rho<1$, then the stability is also retained in the dynamic case.

Let $p_l^{(r)}$ denote the vector of equilibrium probabilities for level l of the QBD process $\{Y_r(t), t\geq 0\}$. For $l=0$ there is a $K+r$ dimensional steady state probability vector, and for $l=1, 2, \ldots$ there are r dimensional steady state probability vectors. These are denoted as

$$p_0^{(r)} = (p^{(r)}(0,-K),\ p^{(r)}(0,1-K), \ldots,\ p^{(r)}(0,r-2), p^{(r)}(0,r-1)),$$

$$p_l^{(r)} = (p^{(r)}(l,0),\ p^{(r)}(l,1), \ldots,\ p^{(r)}(l,r-2),\ p^{(r)}(l,r-1)), \quad l=0,1,\ldots,$$

where $p^{(r)}(l, j)$ is the steady state probability that the QBD process is in state (l, j). The equilibrium equations for this process are

$$p_0^{(r)}B_1^{(r)} + p_1^{(r)}B_2^{(r)} = 0^{(c+r+1)} \tag{7}$$

$$p_0^{(r)}B_0^{(r)} + p_1^{(r)}A_1^{(r)} + p_2^{(r)}A_2^{(r)} = 0^{(r)}, \tag{8}$$

$$p_{l-1}^{(r)}A_0^{(r)} + p_l^{(r)}A_1^{(r)} + p_{l+1}^{(r)}A_2^{(r)} = 0^{(r)}, \quad l=2,3,\ldots, \tag{9}$$

where $0^{(k)}$ is a k dimensional vector of zeros for arbitrary k. In addition, the normalization equation is

$$\sum_{j=-K}^{r-1} p^{(r)}(0,j) + \sum_{l=1}^{\infty} p_l^{(r)}e^{(r)} = 1, \tag{10}$$

where $e^{(r)}$ is the r dimensional column vector of ones. Solving the system of linear equations given by Eqs. (7)–(10) will provide explicit expressions for the steady state probabilities of the process $\{Y_r(t), t\geq 0\}$.

From Eqs. (7) and (8), the following relationships hold.

$$p^{(r)}(0,j) = p^{(r)}(0,0)\rho^j, \quad j=-K, 1-K, \ldots, r-2, \tag{11}$$

$$p^{(r)}(0,r-1) = p^{(r)}(0,0)\cdot\frac{\rho^{r-1}}{1+\beta\rho}, \tag{12}$$

$$p^{(r)}(1,0) = p^{(r)}(0,0)\cdot\frac{\rho^{r}}{1+\beta\rho}, \tag{13}$$

$$p^{(r)}(1,j) = p^{(r)}(1,0)(\rho+\beta)\rho^{j-1}, \quad j=1,2,\ldots,r-2, \tag{14}$$

$$p^{(r)}(1,r-1) = p^{(r)}(1,0)\cdot\frac{(\rho+\beta)\rho^{r-2}}{1+\beta\rho}. \tag{15}$$

Given that the Markov process with generator $Q^{(r)}$ is ergodic, the equilibrium probability vectors are determined by deploying the matrix-geometric form,

$$p_l^{(r)} = p_1^{(r)}(R^{(r)})^l, \quad l=1,2,\ldots, \tag{16}$$

where the r dimensional rate matrix $R^{(r)}$ is the minimal-nonnegative solution of the matrix-quadratic equation,

$$A_0^{(r)} + R^{(r)}A_1^{(r)} + (R^{(r)})^2A_2^{(r)} = 0^{(r\times r)}, \tag{17}$$

where $0^{(r\times r)}$ is the r dimensional square matrix of zeros. In order to solve for the steady state properties of the QBD process $\{Y_r(t), t\geq 0\}$, one must find the rate matrix $R^{(r)}$ that solves Eq. (17) (see Neuts [21] for an overview of matrix geometric methods). Structural results for finding an explicit rate matrix are provided in Ramaswami and Latouche [24] for the QBD processes with transition matrices of rank 1.

Given the Markov process with generator $Q^{(r)}$ is ergodic, by using the matrix algebra provided in Appendix A, the rate matrix $R^{(r)}$ that exactly solves Eq. (17) for every $r=2, 3, \ldots$ is given by

$$R^{(r)} = \begin{bmatrix} 0^{(r)} \\ 0^{(r)} \\ \vdots \\ 0^{(r)} \\ R_{r-1}^{(r)} \end{bmatrix}, \tag{18}$$

where

$$R_{r-1}^{(r)} = \left(\rho,\ \rho(\rho+\beta),\ \rho^2(\rho+\beta),\ \ldots,\ \rho^{r-2}(\rho+\beta),\ \frac{\rho^{r-1}(\rho+\beta)}{1+\beta\rho}\right). \tag{19}$$

The lower diagonal structure of $R^{(r)}$ directly implies that its largest eigenvalue is

$$\eta = \frac{\rho^{r-1}(\rho+\beta)}{1+\beta\rho}, \tag{20}$$

and $\eta<1$. Then the normalization equation is rewritten as,

$$\sum_{j=-K}^{r-1} p^{(r)}(0,j) + p_1^{(r)}\left(I^{(r)} + \frac{R^{(r)}}{1-\eta}\right)e^{(r)} = 1, \tag{21}$$

where the term $(I^{(r)}+(R^{(r)}/1-\eta)e^{(r)}$ is an r dimensional column vector of ones except the last row being equal to $1+(R_{r-1}^{(r)}e^{(r)}/1-\eta)$. Employing the explicit expression of $R_{r-1}^{(r)}$ provided in Eq. (19),

$$1+\frac{R_{r-1}^{(r)}e^{(r)}}{1-\eta} = \frac{(1+\beta\rho)}{(1-\rho)}. \tag{22}$$

Using this result from Eq. (22), together with the findings in Eqs. (11) and (15), in the normalization equation (21) gives

$$p^{(r)}(0,0) = (1-\rho)\rho^K, \quad (23)$$

The steady state probability vectors for all levels of the QBD process $\{Y_r(t), t \geq 0\}$ can be found by using the explicit expression for $p^{(r)}(0,0)$ as in Eq. (23) together with the equilibrium relations stated in Eqs. (11)–(15) and the matrix geometric form in Eq. (16).

4. Performance analysis

By using the results of the previous section we can find exact expressions for $E[B]$ and $E[L]$ as in the following propositions.

Proposition 1. *For given values of ρ, β, r, and K, the expected number of backlogs is given by*

$$E[B] = \frac{\rho^{K+1}}{1-\rho} + \frac{2\beta\rho^{K+r}}{1+\beta\rho-\rho^{r-1}\left(\rho+\beta\right)}$$

Proof See Appendix B.

Proposition 2. *For given values of L_0, ρ, β, r, and K, the expected lead time is given by*

$$E[L] = L_0 + \frac{\left(1+\beta\right)\rho^{K+r}}{1+\beta\rho-\rho^{r-1}\left(\rho+\beta\right)}$$

Proof See Appendix C.

It is analytically not possible to present a closed form expression for the average number of tardy deliveries. The reason is that some of the orders are canceled, and thus, should not be counted as tardy. Besides, the delivered orders are tardy when the actual delivery time exceeds the quoted lead time (which is changing continuously). Therefore, we provide an approximation for $E[T]$.

Consider an arriving order that causes the system to make a transition to quoted lead time L and to backlog level B. The probability that the order will become late is

$$Pr\left\{N(L) \leq B-1\right\} = \sum_{k=0}^{B-1} e^{-\mu L}\frac{(\mu L)^k}{k!},$$

where $N(L)$ is Poisson distributed with rate μL. The above statement holds for all orders that are not canceled under the assumption that only the last customer order in the backlog queue is canceled.

The rate of order cancelations is $\sum_{l=1}^{\infty} p(l,0)\mu\beta$. Similarly, the total arrival rate of orders (including the additional orders due to increase in the lead time) is $\lambda + \sum_{l=0}^{\infty} p(l,r-1)\lambda\beta$. Then, the fraction of orders that are canceled is

$$\frac{\sum_{l=1}^{\infty} p(l,0)\mu\beta}{\lambda+\lambda\beta\sum_{l=0}^{\infty} p(l,r-1)}.$$

Assuming that there are no order cancelations, the expected number of tardy deliveries is given by

$$E[\overline{T}] = \lambda\sum_{l=0}^{\infty}\sum_{j=0}^{r-2}[Pr\{N(L_0+l) \leq rl+j\}p^{(r)}(l,j)] + \lambda(1-\beta)\sum_{l=0}^{\infty}[Pr\{N(L_0+l+1) \leq rl+r-1\}p^{(r)}(l,r-1)] + \lambda\beta\sum_{l=0}^{\infty}[Pr\{N(L_0+l+1) \leq rl+r\}p^{(r)}(l,r-1)].$$

Then,

$$E[T] = E[\overline{T}]\cdot\left[1-\frac{\sum_{l=1}^{\infty} p(l,0)\mu\beta}{\lambda+\lambda\beta\sum_{l=0}^{\infty} p(l,r-1)}\right].$$

Table 1
Experimental design.

Factors	Treatments	Number of treatments
Lead time quoting procedure	r = 2, 3, 4, static	4
Response probability, β	1.0, 0.5, 0.0	3
Number of kanbans, K	1, 4	2
Utilization, ρ	0.80, 0.95	2
Costs (c_h, c_b, c_L, c_T)	(1,1,3,20), (10,1,15,20)	2

4.1. Setting

In this paper we aim to make a comparison between static and dynamic cases, to show the conditions under which dynamically quoting lead times might be preferable over the static lead times, and to provide insights about the impact of lead time syndrome. We consider r = 2, 3, and 4, where r = 2 represents the case with the highest update frequency, and r = 4 is the case with the lowest update frequency. In addition, β = 1.0, 0.5, and 0.0 are considered. β = 1.0 represents the case that the planner reacts to every change in the lead time, and β = 0.0 represents the case that orders are independent of the change in the lead time and thus, the lead time syndrome does not exist in this case. We also consider the number of kanbans as a design variable with K = 1 and 4. In addition, different utilization levels of ρ = 0.80 and 0.95 are taken into account. For all cases the value of μ is fixed to 20. Thus, for ρ = 0.80 and 0.95, we set respectively λ = 16 and 19. Cost parameters, which represent different product characteristics and delivery requirements, are also important to provide insights about the performance of the dynamic and the static cases. In general, due to the nature of the problem the main goal is to quote accurate and short lead times. Thus, the costs associated with the length of the lead time and tardiness are expected to be high, and the backlog cost is expected to be low. We choose (c_h, c_b, c_L, c_T) = (1, 1, 3, 20) to represent the case with high tardiness costs and low inventory and lead time costs. (c_h, c_b, c_L, c_T) = (10, 1, 15, 20) represents the case with expensive items with high inventory holding and lead time costs. Table 1 gives the list of all design factors, their treatment levels and the number of treatments for each factor. Applying a full-factorial design, there are in total 80 different treatments. First, both the static lead time and L_0 are set to zero, and we show the effects of the factors given in Table 1. Further, we also find optimal values of static lead time, L_0, K, and r that give the minimum total cost both for static and dynamic cases. In the static case, the optimal lead time is found by marginal analysis. In the dynamic case, for each choice of r the optimal K and L_0 are found by enumeration. As r increases the dynamic case converges to the static case, and search for r that gives minimum cost is stopped. Through these results, we show the relative percentage improvement in total cost by dynamically updating the lead time under different cost parameters and utilization levels.

4.2. Numerical results

Table 2 provides the numerical results for ρ = 0.80 and ρ = 0.95. L_0 = 0 for all dynamic cases, and for the static case, the quoted lead time is also taken as zero. TC_1 refers to the total cost of the case with (c_h, c_b, c_L, c_T) = (1, 1, 3, 20), and TC_2 refers to the total cost of the case with (c_h, c_b, c_L, c_T) = (10, 1, 15, 20).

The insights gathered from the numerical results in Table 2 are as follows. As r increases (in other words as the update frequency decreases) then the average backlog level and the average lead time

Please cite this article in press as: Selçuk B. Adaptive lead time quotation in a pull production system with lead time responsive demand. J Manuf Syst (2012), http://dx.doi.org/10.1016/j.jmsy.2012.07.017

Table 2
Performance measures for $\rho = 0.80$ and $\rho = 0.95$.

		$r = 2$			$r = 3$			$r = 4$			Static
		$\beta = 1.0$	$\beta = 0.5$	$\beta = 0.0$	$\beta = 1.0$	$\beta = 0.5$	$\beta = 0.0$	$\beta = 1.0$	$\beta = 0.5$	$\beta = 0.0$	
$\rho = 0.80$											
$K = 1$	$E[B]$	6.04	4.62	3.20	4.46	3.92	3.20	3.95	3.65	3.20	3.20
	$E[L]$	2.84	2.13	1.42	1.26	1.08	0.84	0.75	0.67	0.56	0.00
	$E[T]$	1.89	2.17	2.56	3.98	4.23	4.61	5.71	5.92	6.25	12.80
	TC_1	53.35	55.50	59.67	88.84	92.71	98.88	121.46	124.99	130.80	260.20
	TC_2	96.48	90.10	85.73	113.01	114.69	117.95	139.42	142.02	146.46	269.20
$K = 4$	$E[B]$	3.09	2.37	1.64	2.29	2.01	1.64	2.02	1.87	1.64	1.64
	$E[L]$	1.46	1.09	0.73	0.65	0.55	0.43	0.38	0.34	0.28	0.00
	$E[T]$	1.11	1.20	1.31	2.18	2.26	2.36	3.05	3.11	3.20	6.55
	TC_1	33.64	33.67	34.04	51.88	52.77	54.11	68.22	69.09	70.46	136.71
	TC_2	87.12	82.78	78.78	95.65	95.42	95.27	108.80	109.20	109.87	172.71
$\rho = 0.95$											
$K = 1$	$E[B]$	35.64	26.84	18.05	26.62	22.95	18.05	23.61	21.39	18.05	18.05
	$E[L]$	17.59	13.19	8.79	8.57	7.34	5.71	5.56	5.01	4.17	0.00
	$E[T]$	0.62	0.73	0.90	1.44	1.56	1.76	2.25	2.37	2.57	18.05
	TC_1	101.74	82.07	63.48	82.04	77.16	71.38	86.22	84.74	83.06	380.05
	TC_2	321.78	249.36	178.00	193.86	174.28	148.91	162.00	153.83	142.11	389.05
$K = 4$	$E[B]$	30.55	23.02	15.48	22.82	19.67	15.48	20.25	18.34	15.48	17.15
	$E[L]$	15.08	11.31	7.54	7.35	6.30	4.90	4.77	4.29	3.58	0.00
	$E[T]$	0.55	0.65	0.77	1.26	1.36	1.51	1.96	2.05	2.21	17.15
	TC_1	90.87	73.86	57.57	74.15	69.74	64.34	77.78	76.27	74.35	328.99
	TC_2	307.82	245.57	184.04	198.30	181.29	159.10	171.03	163.79	153.27	364.99

decrease. The backlog level decreases, because the impact of the lead time syndrome will become less when the update frequency is less. The average lead time is also shorter, because the lead time is not increased until a larger increase in the backlog level is seen. The average number of tardy orders increases, because for the same backlog size a shorter lead time is quoted when the update frequency decreases.

Decreasing the value of β means that the lead time syndrome is experienced in a decreasing degree since now, the planner at the downstream stage less frequently reacts to changes in the lead time. As a result of less responsive planner the average backlog level and the average lead time decreases as β decreases. On the other hand, the average number of tardy deliveries increases.

Increasing the number of kanbans will decrease the average backlog level, the average lead time, and the average tardiness. Because, in this case, more orders are satisfied directly from stock, and the actual delivery times drop as well as the quoted lead times. That is, backlog costs and delivery performance related costs drop as the number of kanbans increases. But the inventory cost increases. As the utilization level increases the impact of K decreases. For higher utilization levels the system operates most of the time at higher backlog levels, and the impact of changing the number of kanbans becomes limited as compared to the case with a lower utilization level.

In order to maintain the quoted lead time at almost the same level, the update frequency must be decreased when the utilization level increases, and this will be in expense of higher number of tardy deliveries. Under the current choice of cost parameters we see that, for a low utilization level, since the congestion in the shop floor is not high, the lead time syndrome may result in lower total cost when there is low backlog and lead time costs. When there is high lead time cost or when the utilization level is high, the case without the lead time syndrome, $\beta = 0$, is a preferable situation. For a low utilization level, the highest update frequency, $r = 2$, is preferable for all response probabilities. For a high utilization level, the preferable update frequency changes depending on the cost parameters and the response probability. For a high utilization level, the congestion in the shop floor can increase rapidly, and this may lead to longer lead times. Thus, for the case with high lead time costs $r = 4$ is preferable. For the case with low lead time costs, $r = 3$ is preferable for $\beta = 1.0$ and $\beta = 0.5$. Since, the lead time syndrome does not exist for $\beta = 0$, a higher update frequency with $r = 2$ is preferable in this case. That is, when the degrading effect of the lead time syndrome is high as in the case with higher utilization level and higher response probabilities, in order to smooth the effects of the lead time syndrome, a less frequent updating is preferable.

Table 3
Minimum total costs for $\rho = 0.80$ and $\beta = 0.0$.

$(c_h, c_b, c_L, c_T) = (1, 1, 3, 20)$	$r = 11, K = 1, L_0 = 1$	Static, $K = 1, L = 2$
$E[B]$	3.20	3.20
$E[L]$	1.08	2.00
$E[T]$	0.003	0.003
TC	73.0	100.0
$(c_h, c_b, c_L, c_T) = (10, 1, 15, 20)$		
	$r = 14, K = 1, L_0 = 1$	Static, $K = 1, L = 1$
$E[B]$	3.20	3.20
$E[L]$	1.04	1.00
$E[T]$	0.018	0.233
TC	89.7	100.0

Both the update frequency and the response probability have significant effects on how the lead time syndrome drives the system performance. Our results indicate that an increased update frequency is more effective than an increased response probability for decreasing the tardiness. The relationship of $E[L]$ and $E[T]$ for different r and β values for the case of $K = 4$ and $\rho = 0.95$ is visualized in Fig. 3. There is a smoother decrease in tardiness as β increases, however, there are significant shifts in tardiness for different update frequencies. The graph also shows that an efficient frontier between $E[L]$ and $E[T]$ will include those cases with $\beta = 0$, where the lead time syndrome does not occur. This is logical, because the lead time syndrome increases the variability in order releases.

Tables 3 and 4 provide the minimum total costs for $\rho = 0.80$ and $\rho = 0.95$ respectively. The results are for $\beta = 0.0$, and the same insights also hold for other response probabilities. For the ease of

Please cite this article in press as: Selçuk B. Adaptive lead time quotation in a pull production system with lead time responsive demand. J Manuf Syst (2012), http://dx.doi.org/10.1016/j.jmsy.2012.07.017

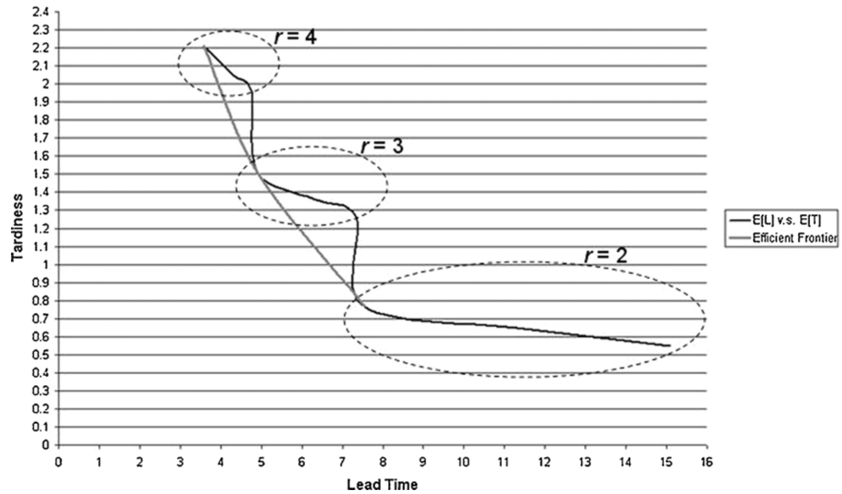

Fig. 3. $E[L]$ vs. $E[T]$ for $K=4$ and $\rho=0.95$.

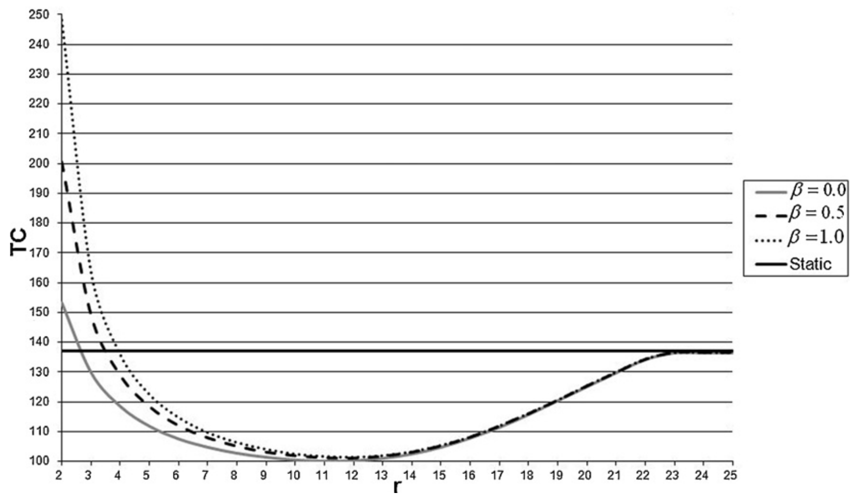

Fig. 4. Total cost vs. r for $\rho=0.95$ and $(c_h, c_b, c_L, c_T)=(10, 1, 15, 20)$.

Table 4
Minimum total costs for $\rho=0.95$ and $\beta=0.0$.

$(c_h, c_b, c_L, c_T)=(1, 1, 3, 20)$		
	$r=13, K=3, L_0=1$	Static, $K=2, L=5$
$E[B]$	16.29	17.15
$E[L]$	1.90	5.00
$E[T]$	0.022	0.114
TC	69.8	100.0
$(c_h, c_b, c_L, c_T)=(10, 1, 15, 20)$		
	$r=15, K=1, L_0=1$	Static, $K=1, L=3$
$E[B]$	18.05	18.05
$E[L]$	1.82	3.00
$E[T]$	0.095	0.897
TC	63.9	100.0

exposition, in each scenario, the larger total cost is represented by 100 in order to show the percentage differences between dynamic and static cases. Depending on the utilization level and the cost parameters, more than 30% saving in total cost is achieved by dynamically updating the lead time. The saving will be higher for a higher utilization level. In almost all cases, dynamically updating the lead time with its optimal design parameters will lead to a lower average backlog level, a shorter average lead time and a smaller average number of tardy deliveries. Therefore, updating the lead time generates significant cost savings in all cases. Tables 3 and 4 also reveal that, in the static case, the response to relatively expensive tardiness cost is by increasing the static lead time (from $L=1$ to $L=2$ for $\rho=0.80$ and from $L=3$ to $L=5$ for $\rho=0.95$), whereas in the dynamic case, this is done by increasing the update frequency. That is, the concept of update frequency can be used as a fine tuning mechanism for different cost parameters in the dynamic case.

Fig. 4 displays the minimum total costs versus r values, for different values of β and also for the static case. It is clear that the lead time syndrome can be very disruptive, and can lead to a high total cost when there is a high frequency of updating. In such cases, it may even be beneficial to use static lead times. The figure also suggests that, in the dynamic case, if there exists a higher tendency for responding to changes in the lead time (i.e., larger β), then, in order to maintain and control the total cost, the update frequency must be decreased because this will smoothen the disruptive effects of the lead time syndrome.

5. Conclusion and future research

In this paper, an effective lead time quoting procedure is proposed for a pull production control system. The existence of the lead time syndrome has well been proven in the literature, and the concept played a central role in the problem setting considered in this paper. Valuable insights are provided for the impacts of various design and environmental parameters such as update frequency, response probability, the number of kanbans, utilization and cost parameters. The relative improvement by the dynamic lead time setting procedure over the static lead time has been shown through comprehensive numerical analysis.

The main insights gathered from our quantitative results are:

1 Dynamically updating the lead times performs much better than the static lead times. The relative improvement is higher when the production system operates under a higher utilization level.
2 As the lead time syndrome strengthens through larger response probability, the performance degrades with higher costs and more number of tardy orders.
3 Update frequency can be used as a fine tuning parameter to make the lead time quoting procedure flexible against changing environment. Once the lead time syndrome exists in a higher degree (larger β) or the lead time costs are high, the lead time quoting procedure should be adapted by decreasing the update frequency in order to smooth the effects of the lead time syndrome.

We suggest any future research of this study should consider a larger problem set including a multi-stage production system together with a downstream distribution system into the evaluation. Although the problem complexity may easily become very high and may limit achieving closed form solutions, we believe valuable insights remain to be explored about such information-feedback systems within a supply chain context.

Appendix A. Derivation of $R^{(r)}$

The matrix geometric form and the structure of $A_2^{(r)}$ allow us to express the rate matrix as $R^{(r)} = -A_0^{(r)}(A_1^{(r)} + A_0^{(r)}e^{(r)}(\beta e_{r-2}^{(r)} + (1-\beta)e_{r-1}^{(r)})^T)^{-1}$. Further, let us define the r dimensional square matrix

$$A_3^{(r)} = A_1^{(r)} + A_0^{(r)}e^{(r)}(\beta e_{r-2}^{(r)} + (1-\beta)e_{r-1}^{(r)})^T.$$

$R^{(r)}$ has rows of zero except the last one, and its last row can be derived from a linear combination of the first and the second rows of $(A_3^{(r)})^{-1}$. Let us define $(a_{3(0)}^{(r)})^{-1}$ and $(a_{3(1)}^{(r)})^{-1}$ as the first and the second rows of the matrix $(A_3^{(r)})^{-1}$ respectively. Then, the last row of the rate matrix $R^{(r)}$ is

$$R_{r-1}^{(r)} = -\lambda((1-\beta)(a_{3(0)}^{(r)})^{-1} + \beta(a_{3(1)}^{(r)})^{-1}),$$

where $(a_{3(0)}^{(r)})^{-1}$ and $(a_{3(1)}^{(r)})^{-1}$ solve the following matrix equations:

$$(A_3^{(r)})^T((a_{3(0)}^{(r)})^{-1})^T = e_0^{(r)},$$
$$(A_3^{(r)})^T((a_{3(1)}^{(r)})^{-1})^T = e_1^{(r)}.$$

By exploiting the tridiagonal structure of $A_3^{(r)}$ we find

$$(a_{3(0)}^{(r)})^{-1} = -\left(\frac{1}{\mu}, \frac{\rho}{\mu}, \frac{\rho^2}{\mu}, \dots, \frac{\rho^{r-2}}{\mu}, \frac{\rho^{r-1}}{\beta\lambda+\mu}\right),$$
$$(a_{3(1)}^{(r)})^{-1} = -\left(\frac{1}{\mu}, \frac{(1+\rho)}{\mu}, \frac{\rho(1+\rho)}{\mu}, \dots, \frac{\rho^{r-3}(1+\rho)}{\mu}, \frac{\rho^{r-2}(1+\rho)}{\beta\lambda+\mu}\right).$$

Consequently,

$$R_{r-1}^{(r)} = \left(\rho,\ \rho(\rho+\beta),\ \rho^2(\rho+\beta),\ \dots,\ \rho^{r-2}(\rho+\beta),\ \frac{\rho^{r-1}(\rho+\beta)}{1+\rho\beta}\right).$$

Appendix B. Proof of Proposition 1

The expected backlog size is given by

$$E[B] = \sum_{l=0}^{\infty}\sum_{j=0}^{r-1}(rl+j)p^{(r)}(l,j) = r\sum_{l=1}^{\infty} l p_l^{(r)}e^{(r)} + \sum_{j=1}^{r-1} jp^{(r)}(0,j) + \sum_{j=1}^{r-1} j\sum_{l=1}^{\infty} p_l^{(r)}e_j^{(r)},$$

which can be further simplified as

$$E[B] = \frac{rp_1^{(r)}e^{(r)}}{(1-\eta)^2} + \sum_{j=1}^{r-1} jp^{(r)}(0,j) + \sum_{j=1}^{r-1} jp_1^{(r)}\left(I^{(r)} + \frac{R^{(r)}}{1-\eta}\right)e_j^{(r)}. \quad \text{(B.1)}$$

Let us rewrite $E[B] = E[B]_1 + E[B]_2$, where

$$E[B]_1 = \frac{rp_1^{(r)}e^{(r)}}{(1-\eta)^2},$$
$$E[B]_2 = \sum_{j=1}^{r-1} j\left[p^{(r)}(0,j) + p_1^{(r)}\left(I^{(r)} + \frac{R^{(r)}}{1-\eta}\right)e_j^{(r)}\right].$$

We will first write $E[B]_1$ in its explicit form, and then, write $E[B]_2$ similarly in order to provide an explicit expression for $E[B]$. By replacing $p^{(r)}(0,0)$ with the right-hand-side of Eq. (23) and using it in Eqs. (13)–(15), and by replacing η with the right-hand-side of Eq. (20) we get

$$E[B]_1 = \frac{r(1+\beta)\rho^{K+r}}{1+\beta\rho-\rho^{r-1}(\rho+\beta)}. \quad \text{(B.2)}$$

The explicit forms of $p_0^{(r)}$ and $p_1^{(r)}$ can be found directly from Eqs. (13)–(15) and from Eq. (23). Also the r-dimensional square matrix $I^{(r)} + \frac{R^{(r)}}{1-\eta}$ can be rewritten by replacing $R^{(r)}$ and η respectively with the right-hand-sides of Eq. (18) and (20). Then, $E[B]_2$ is

$$E[B]_2 = \frac{\rho^{K+1}(1+\beta\rho)(1-\rho^{r-1})}{(1-\rho)(1+\beta\rho-\rho^{r-1}(\rho+\beta))} - \frac{(1+\beta)(r-1)\rho^{K+r}}{1+\beta\rho-\rho^{r-1}(\rho+\beta)}. \quad \text{(B.3)}$$

Together with Eq. (B.2) this result yields that the average length of the backlog queue is

$$E[B] = \frac{\rho^{K+1}}{1-\rho} + \frac{2\beta\rho^{K+r}}{1+\beta\rho-\rho^{r-1}\left(\rho+\beta\right)}, \quad \text{(B.4)}$$

and this completes the proof of Proposition 1. □

Appendix C. Proof of Proposition 2

The probability mass function of the random variable L is

$$Pr\{L = L_0 + l\} = \begin{cases} \sum_{j=-K}^{r-1} p^{(r)}(0,j), & l = 0 \\ \sum_{j=0}^{r-1} p^{(r)}(l,j), & l = 1,2,\dots \end{cases}$$

By using the explicit derivation of the rate matrix in Eqs. (18) and (19), and the expressions for $p_0^{(r)}$ and $p_1^{(r)}$ given in Eqs. (11)–(15), the probability mass function of L is rewritten by

$$Pr\left\{L = L_0 + l\right\} = \begin{cases} 1 - \dfrac{(1+\beta)\rho^{K+r}}{1+\beta\rho}, & l = 0 \\ \dfrac{(1+\beta)\rho^{K+r}}{1+\beta\rho}\cdot(1-\eta)(\eta)^{l-1} & l = 1,2,\dots. \end{cases}$$

L can be expressed as L_0 plus a random variable whose probability distribution is based on the update strategy and the system characteristics. $L = L_0$ with probability $1 - (((1+\beta)\rho^{K+r})/(1+\beta\rho))$, and $L = L_0 + 1 + \tilde{L}$ with probability $(((1+\beta)\rho^{K+r})/(1+\beta\rho))$, where the random variable $\tilde{L}$ has a geometric distribution with parameter

Please cite this article in press as: Selçuk B. Adaptive lead time quotation in a pull production system with lead time responsive demand. J Manuf Syst (2012), http://dx.doi.org/10.1016/j.jmsy.2012.07.017

η. The probability mass function of $\tilde{L}$ is $Pr\{\tilde{L}=l\}=(1-\eta)\eta^l$, $l=0,1,\ldots$, and $E[\tilde{L}]=(\eta/1-\eta)=((\rho^{r-1}(\rho+\beta))/(1+\beta\rho-\rho^{r-1}(\rho+\beta)))$. Consequently, the average lead time is given by

$$E[L]=L_0+\frac{(1+\beta)\rho^{K+r}}{1+\beta\rho}\cdot(1+E(\tilde{L}_r))=L_0+\frac{(1+\beta)\rho^{K+r}}{1+\beta\rho-\rho^{r-1}(\rho+\beta)}, \tag{C.1}$$

and this completes the proof of Proposition 2. □

References

[1] Akturk MS, Erhun F. An overview of design and operational issues of kanban systems. International Journal of Production Research 1999;37:3859–81.
[2] Chang F. A study of factors affecting due-date predictability in a simulated dynamic job shop. Journal of Manufacturing Systems 1994;13:389–406.
[3] Cochran J, Kaylani H. Optimal design of a hybrid push/pull serial manufacturing system with multiple part types. International Journal of Production Research 2008;46:949–65.
[4] Erhun F, Akturk MS, Turkcan A. Interaction of design and operational parameters in periodic review kanban systems. International Journal of Production Research 2003;41:3315–38.
[5] Flapper S, Miltenburg G, Wijngaard J. Embedding JIT into MRP. International Journal of Production Research 1991;29:329–41.
[6] Forrester J. Industrial dynamics. (10th ed.) Massachusetts: The M.I.T. Press.; 1980.
[7] Hopp W, Sturgis M. Quoting manufacturing due-dates subject to a service-level constraint. IIE Transactions 2000;32:771–84.
[8] Huq Z, Huq F. Embedding JIT in MRP: the case of job shops. Journal of Manufacturing Systems 1994;13:153–64.
[9] Iwase M, Ohno K. The performance evaluation of a multi-stage JIT production system with stochastic demand and production capacities. European Journal of Operational Research 2011;214:216–22.
[10] Joseph O, Sridharan R. Analysis of dynamic due-date assignment models in a flexible manufacturing system. Journal of Manufacturing Systems 2011;30:28–40.
[11] Junior M, Filho M. Variations of the kanban system: Literature review and classification. International Journal of Production Economics 2010;125:13–21.
[12] Kim I, Tang CS. Lead time and response time in a pull production control system. European Journal of Operational Research 1997;101:474–85.
[13] Kingsman BG, Tatsiopoulos IP, Hendry LC. A structural methodology for managing manufacturing lead times in make-to-order companies. European Journal of Operational Research 1989;40:196–209.
[14] Koukoumialos S, Liberopoulos G. An analytical method for the performance evaluation of echelon kanban control systems. OR Spectrum 2005;27:339–68.
[15] Krieg GN, Kuhn H. Performance evaluation of two-stage multi-product kanban systems. IIE Transactions 2008;40:265–83.
[16] Kumar CS, Panneerselvam R. Literature review of jit—kanban system. International Journal of Advanced Manufacturing Technology 2007;32:393–408.
[17] Masin M, Prabhu V. AWIP: a simulation-based feedback control algorithm for scallable design of self-regulating production control systems. IIE Transactions 2009;41:120–33.
[18] Mather H, Plossl GW. Priority fixation versus throughput planning. Production and Inventory Management, 3rd Quarter 1978:27–51.
[19] McKay K, Wiers V. Unifying the theory and practice of production scheduling. Journal of Manufacturing Systems 1999;18:241–55.
[20] Nagendra P, Das S. MRP/SFX: a kanban-oriented shop floor extension to MRP. Production Planning and Control 1999;10:207–18.
[21] Neuts MF. Matrix-geometric solutions in stochastic models: an algorithmic approach. Baltimore: The Johns Hopkins University Press; 1981.
[22] Pandey A, Masin M, Prabhu V. Adaptive logistic controller for integrated design of distributed supply chains. Journal of Manufacturing Systems 2007;26:108–15.
[23] Plossl GW. Throughput time control. International Journal of Production Research 1988;26:493–9.
[24] Ramaswami V, Latouche G. A general class of Markov processes with explicit matrix-geometric solutions. OR Spectrum 1986;8:209–18.
[25] Selçuk B. Dynamic performance of hierarchical planning systems: modeling and evaluation with dynamic planned lead times. PhD thesis, Eindhoven University of Technology; 2007.
[26] Selçuk B, Adan IJBF, De Kok AG, Fransoo JC. An explicit analysis of the lead time syndrome: Stability condition and performance evaluation. International Journal of Production Research 2009;47:2507–29.
[27] Selçuk B, Fransoo JC, Kok AGD. The effect of updating lead times on the performance of hierarchical planning systems. International Journal of Production Economics 2006;104:427–40.
[28] Takahashi K, Nakamura N. Decentralized reactive kanban system. European Journal of Operational Research 2002;139:262–76.
[29] Yavuz IH, Satir A. A kanban based simulation study of a mixed model just in time manufacture line. International Journal of Production Research 1995;33:1027–48.
[30] Yücesan E, De Groote X. Lead times, order release mechanisms, and customer service. European Journal of Operational Research 2000;120:118–30.
[31] Zijm WHM. Towards intelligent manufacturing planning and control systems. OR Spectrum 2000;22:313–45.

Please cite this article in press as: Selçuk B. Adaptive lead time quotation in a pull production system with lead time responsive demand. J Manuf Syst (2012), http://dx.doi.org/10.1016/j.jmsy.2012.07.017