THE OPTIMAL ARRIVAL RATE CONTROL POLICY IN PRODUCTION SYSTEMS WITH WORKLOAD DEPENDENT PROCESSING RATES.

H.P.G. van Ooijen and J.W.M. Bertrand

Sub-department of Operations Planning and Control, Department of Technology Management, Technische Universiteit Eindhoven

Abstract

Recently a number of operations management reports have been published which provide evidence for the existence of a relationship between work-in-process and productivity. In this paper we investigate job-shops where management is aware of this relationship and also has the possibility to influence the arrival rate. By influencing the arrival rate the level of work-in-process and thus the productivity is influenced. For a number of cost settings we obtain the optimal profits using markov theory and the OptQuest® software system. It turns out that manipulating the arrival stream by management leads to substantially higher profits for all cost settings.

Keywords: productivity, workload, arrival rate switching

1. Introduction

Recently a number of operations management reports have been published which provide evidence for the existence of a relationship between work-in-process and productivity (Schmenner [1988], Holström [1994] and Liebermann and Demeester [1999]). Work-in-process is measured as the number of work orders on the shop floor and productivity is measured as output per employee. There seems to be a certain level of work-in-process at which, given a number of resources, the output is at its maximum. If the work-in-process comes above or below this level the productivity decreases and thus the output decreases. There is a psychological argument for this relationship between workload and productivity. The argument builds on the assumption that performing operations requires human perception, human information processing, human decision-making and human actions to take place. When the number of work orders on the shop floor increases, the work pressure increases, which increases the level of arousal of the shop floor personnel. This in turn has an impact on the perception, information processing, decision making and actions of the shop floor operators and its shop floor management, which all have an impact on the performance (see Wickens (1992), chapter 10, Stress and Human Error).

If a company operates in a dynamic market, which is often the case, then it might be relevant, given the above mentioned relationship, to control the workload in the shop. By controlling the workload in the shop one can try to have a level of work-in-process that gives the highest productivity. Manipulating the order arrival rate might influence the level of work-in-process. Increasing (decreasing) the order arrival rate, in general, will lead to an increase (decrease) in throughput and thus to an increase (decrease) in income. However, it also leads to an increase (decrease) of the work-in-process (costs) and, in general, switching the arrival rate cannot be done without costs. So, in practice management has to make an economic trade-off. In this paper we investigate this trade-off and we try to determine the relationships between revenue, switching costs and costs of work-in-process. The remainder of this paper is as follows. In Section 2 we will discuss the relevant literature. In section 3 we will present the job-shop model, the relationship between work-in-process and productivity, the arrival rate switching policy and the economic model. Thereafter, in Section 4 the results for a
different number of parameter settings will be presented and discussed. The paper is completed with
the conclusions in Section 5.

2. Literature review

In his survey study on differences in productivity between a large number of manufacturing firms, and
within firms over time, Schmenner (1988) found a significant relationship between productivity,
measured as output per employee, and production lead times; productivity turned out to be inversely
related to lead time. Since in manufacturing firms long lead times go with high workloads, this
suggests that a high workload might lead to a low productivity.

The paper of Holmström (1994) shows that on the basis of industrial statistics collected by the
Statistical Office of the United Nations Secretariat, a strong correlation exists between productivity
improvement and the speed of industry networks in a number of industries. These results support the
study by Schmenner.

Lieberman and Demeester (1999) analysed historical data of 52 Japanese automotive industries to
evaluate the inventory-productivity relationship (inventory including work-in-process). They found
that firms increased their productivity ranks during periods of substantial inventory reduction.
Moreover, on average, each 10% reduction in inventory led to about 1% gain in labour productivity,
with a lag of about one year.

Based on this literature we conclude that in production systems where the productivity, determined by
the processing rates, depends on the work load it might be beneficial to control the work load by
directly controlling the arrival rate of the orders. A simple way to achieve this is to switch from a high
arrival rate to a low arrival rate if the workload exceeds a certain value, and to switch back to the high
arrival rate if the workload drops between a certain value. However, switching between arrival rates in
general will lead to costs. These costs might be costs for loss of goodwill, costs associated with the
marketing efforts that are necessary to attract customers and so on. Therefore, the manufacturer will
want to restrict the frequency of switching between demand levels and the magnitude of the difference
between the order arrival rates.

A number of studies have been published which investigate production situations where the processing
rate depends on the level of work-in-process. However, in all those studies, the processing rate is
controlled by management and is not an autonomous variable (as in our study). Cohen (1976)
investigates a single server with two service rates, where the service rate is changed if the level of
work-in-process exceeds a certain threshold K, and there are costs related to switching and the level of
work-in-process. Purpose of the study is to determine the optimal value for K. Tijms (1977) studies a
server with two constant service rates and fixed switchover costs. The server switches from rate 1 to
rate 2 only when the workload exceeds a level $y_1$ and switches from rate 2 to rate 1 only when the
workload falls to a level $y_2$, where $0 < y_2 < y_1$. The costs of the system consist of a linear holding cost, a
service –cost rate and fixed switchover costs. In this paper an explicit expression for the average cost
of this switchover policy is derived. In Doshi et al. (1978) a system is investigated with two possible
production rates, where the control is based on two critical stock levels, and where the arrival rate and
the distribution of the demand depend on the current production rate. This paper gives the long run
average costs per unit time as a function of the chosen critical levels. Nishimura and Jiang (1995)
consider a server with two service modes: regular speed and high speed, and a service rule that is
characterized by two switchover levels. A key feature is that the server takes vacation for setup before
a new service mode is available. The paper derives for the general model an expression for the generating
function of the equilibrium queue-length distribution in terms of the switchover levels. Bar-Lev et al. (1996)
analyse a stochastic production/inventory problem with state (i.e. work-in-process)
dependent production rates. When the inventory $W(t)$ falls below a critical level $m$, production is
started at a rate of $r(W(t))$, i.e. the production rate dynamically changes as a function of the inventory
level. Production continues until a level $M$ is reached. The two-sided $(m,M)$ policy is optimized using
the expected cost obtained from the stationary distribution of $W$. 
Finally a number of studies have been published which investigate situations where the order arrival rate can be controlled in a static way. Buss et al. (1994) determine for single machine job shop-like production systems, amongst others, the optimal arrival rate that maximizes the profit if there are cost related to capacity and congestion. In this situation the production rate is fixed. So and Song (1998) study a production system where demand is sensitive to both price and lead time and determine the joint optimal selection of price, lead time and capacity. In all these studies the processing times are independent of the workload in the shop.

In this paper we investigate the economic trade-off for a production system where the processing time is a variable that depends on the work-in-process in the shop, and where management has the possibility to influence the order arrival rate. This latter will only be possible with some costs and therefore the arrival rate must not be switched too often. In the next section we describe the relationship between work-in-process and productivity as used in this study. Moreover, we also present the economical model that has been used to investigate the impact of the proceedss per item, the switching costs and the costs of work-in-process.

3. The economical model

We assume that orders arrive at the shop according to a Poisson process that can have to values for the arrival rate: a high arrival rate $\bar{e}_h$ and a low arrival rate $\bar{e}_l$. Management influences the value for the arrival rate: if the arrival rate has a high value and the level of work-in-process exceeds the value $N_{hl}$,
measures are taken such that the arrival rate switches to the low value. On the other hand, if the arrival rate has a low value and the work-in-process drops below the value $N_{lh}$, measures will be taken such that the arrival rate switches to the high value.

The dependency of the productivity on the work-in-process is modeled as a dependency of the departure rate of orders from the shop (which is determined by the production times) on the number of orders in the shop. We assume that the departure rate increases from $\mu(0)$ to $\mu(N_{m})$ in equal steps from $N_{t}=0$ to $N_{t}=N_{m}$, and decreases from $\mu(N_{m})$ to $\mu(N_{f})$ in equal steps from $N_{t}=N_{m+1}$ to $N_{t}=N_{f}$. For $N_{t} > N_{f}$, $\mu = \mu(N_{f})$. For an illustration see Figure 1.

We further assume that management in the shop is aware of the existence of a dependency of the production efficiency on the workload in the shop and has knowledge of both the maximum departure rate of orders $\mu(N_{m})$ and the minimum departure rate, $\mu(N_{f})$, and of the workload for which the maximum departure rate is obtained.

Stability requires at least that $\lambda < \mu(N_{f})$. Note that without arrival rate control (thus $\hat{\lambda}=\lambda_{h}=\lambda$) this implies that $\hat{\lambda}< \mu(N_{0})$, which may result in a substantial loss of capacity.

Our departure rate function models the inverse U-shape type of dependency of production efficiency on experienced work pressure as reported in Wickens (1992). In our model we thus assume that the work pressure that an operator experiences is proportional to the workload in the shop and that the experienced work pressure affects the production efficiency. It will be clear that the management of the shop would prefer to operate the shop at the highest possible level of efficiency, that is, realize departure rate $\mu(N_{m})$. However, in our model, management cannot directly influence the departure rate since the departure rate depends on $N_{t}$, which is a stochastic variable. In our model, the parameters that are under direct control of management are the order arrival rates $\lambda_{h}$ and $\lambda$, and the switching values $N_{hl}$ and $N_{lh}$. Now the question is, given the relationship between productivity and work-in-process, proceeds per unit, costs related to switching from a high arrival rate to a low arrival rate, the costs of work-in-process, what are then the optimal switching points $N_{hl}$, $N_{lh}$, and the optimal arrival rates $\lambda_{h}$, $\lambda$, and what is then the profit per unit time?

The profit per unit time can be determined from the revenue per unit time, that can be calculated by multiplying the average number of items produced per unit time with the proceeds of each item, and the costs of switching from a high arrival rate to a low arrival rate, the costs of switching from a low arrival rate to a high arrival rate and the costs per unit time of the work-in-process. The latter can be for instance handling costs, interest costs etc. and will be expressed as a percentage of the proceeds per unit.

In formula:

$$P = mTP + c_{hl}SW_{hl} + c_{lh}SW_{lh} + r.m.WIP \quad (1)$$

where $P$ = profit per unit time-
$m$ = proceeds per unit
$TP$ = average number of items produced per unit of time
$c_{hl}$ = costs related to switching from a high arrival rate to a low arrival rate
$SW_{hl}$ = number of switches from a high arrival rate to a low arrival rate per unit of time
$c_{lh}$ = costs related to switching from a low arrival rate to a high arrival rate
$SW_{lh}$ = number of switches from a low arrival rate to a high arrival rate per unit of time
$r$ = costs of a unit of WIP per unit time, expressed as a percentage of the proceeds per unit
$WIP$ = average level of work-in-process

To find $TP$, $SW_{hl}$, $SW_{lh}$ and WIP we modeled the shop as a two dimensional Markov chain by including the arrival rate as a variable in the birth-death queueing process. To calculate the steady state probability of having $n$ orders in the system, $P_{n}$, we introduce $P(n,1)$ as being the probability of having $n$ customers in the system when the arrival rate is high, and $P(n,2)$ as being the probability of having $n$ customers in the system when the arrival rate is low. Given the arrival rate and departure rate processes as discussed in Section 2 the $P_{n}$’s can be calculated for any feasible combination of parameter values from the following state equilibrium relationships:
Table 1. The different cost scenarios for which the maximum profit per unit of time has been determined.

<table>
<thead>
<tr>
<th>Setting</th>
<th>m</th>
<th>c_{hl}</th>
<th>c_{lh}</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.001</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.01</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.001</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.01</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.01</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.001</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.01</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.0</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1000000</td>
<td>1000000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\[
P(n,1)\left(\lambda_n + \mu(n)\right) = \lambda_h P(n-1,1) + \mu(n+1)P(n+1,1) \quad \text{if } n < N_{hl} - 1
\]
\[
P(N_{hl} - 1,1)\left(\lambda_h + \mu(N_{hl} - 1)\right) = \lambda_h P(N_{hl} - 2,1)
\]
\[
P(N_{lh} + 1,1)\left(\lambda_l + \mu(N_{lh})\right) = \lambda_l P(N_{lh} - 1,1) + \mu(N_{lh} + 1)P(N_{lh} + 1,1) + \mu(N_{lh} + 1)P(N_{lh} + 1,2)
\]
\[
P(N_{lh} + 1,2)\left(\lambda_l + \mu(N_{lh} + 1)\right) = \mu(N_{lh} + 2)P(N_{lh} + 2,2)
\]
\[
P(n,2)\left(\lambda_l + \mu(n)\right) = \lambda_l P(n-1,2) + \mu(n+1)P(n+1,2) \quad \text{if } N_{lh} + 1 < n
\]
\[
\sum_{n=1}^{N_{hl}+1} P(n,1) + \sum_{n=N_{hl}+1}^{\infty} P(n,2) = 1
\]

From the \( P_n \) values the average level of work-in-process (WIP), the average throughput per unit time (TP) and the average switching frequencies per unit time (SW_{hl} and SW_{lh}) can be calculated numerically.
Now the profit per unit time can be calculated using (1).

4. The maximum profit

Now we know how to calculate the profit given certain values for \( N_{hl}, N_{lh}, \lambda_h \), and \( \lambda_l \), the question remains how to find the optimal values? This means that we have to maximize (1) over \( N_{hl}, N_{lh}, \lambda_h \) and \( \lambda_l \). Since TP, SW_{hl}, SW_{lh} and WIP depend on the \( P_n \)’s and we do not have explicit expressions for the \( P_n \) values, we decided to use the computer software system OptQuest\textsuperscript{®}. This system automatically search for the optimal solution. Based on tabu search, scatter search, integer programming and neural networks, very complex models can be handled.
Table 2. The optimal profit (OP) per unit of time and the corresponding values for the optimal switching points $N_{hl}$, $N_{lh}$, and the optimal arrival rates $\lambda_h$, $\lambda_l$, for the cost scenarios as given in Table 1, as found by OptQuest.

We assumed the relationship between productivity and work-in-process be given as discussed in Section 3 and we ran the OptQuest® program for a number of settings. For the parameters of the order departure rate function as given in Figure 1, we used the following values: $\mu(0)=0.875$, $N_m=45$, $\mu(N_m)=1$, $N_f=135$. Further we used a number of different scenarios with regard to the costs. The different scenarios are given in Table 1. We scaled the values of the parameters to the proceeds per unit, which therefore has been set equal to 1. If the only costs of having work-in-process are the interest costs then a value of 0.001 for $r$ means that for each unit of work-in-process per unit of time the costs are equal to 0.001m. If for instance the unit of time equals one hour, this corresponds to a yearly interest rate of approximately 9%. For $r = 0.01$ this corresponds to a yearly interest rate of 88%. These values for $r$ are of course very unrealistic but we took extreme values to investigate the sensitivity of the maximum profit for the costs of work in process.

The scenarios 1 till 3 are the scenarios with relatively high costs for switching, both from a high arrival rate to a low arrival rate and from a low arrival rate to a high arrival rate. For the scenarios 10 till 12 these switching costs are relatively low and for the scenarios 4 till 9 they are relatively moderate. In the scenarios 13-15 the costs for switching from a high arrival rate to a low arrival rate are relatively low, whereas the costs for switching from a low arrival rate to a high arrival rate are relatively high. In the scenarios 16-18 this is the other way around.

The software program reported for each setting the maximum profit per unit of time it found and the corresponding values for the optimal switching points $N_{hl}$, $N_{lh}$, and the optimal arrival rates $\lambda_h$, $\lambda_l$. These values are given in Table 2. It is remarkable that the optimal point for switching from a low arrival rate to a high arrival rate is hardly influenced by the different costs and lies at the point where the departure rate has its highest value. Only for the situations were the costs for switching from a low arrival rate to a high arrival rate are low, the optimal point for switching from a low arrival rate to a high arrival rate shifts substantially to the left. If the costs for switching from a low arrival rate to a high arrival rate are relatively low then the optimal low arrival rate takes on a pretty low value for all

<table>
<thead>
<tr>
<th>Scenario</th>
<th>OP</th>
<th>$N_{hl}$</th>
<th>$N_{lh}$</th>
<th>$\lambda_h$</th>
<th>$\lambda_l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.813</td>
<td>44</td>
<td>200</td>
<td>0.821</td>
<td>0.746</td>
</tr>
<tr>
<td>2</td>
<td>0.805</td>
<td>44</td>
<td>195</td>
<td>0.821</td>
<td>0.740</td>
</tr>
<tr>
<td>3</td>
<td>0.731</td>
<td>44</td>
<td>191</td>
<td>0.812</td>
<td>0.727</td>
</tr>
<tr>
<td>4</td>
<td>0.833</td>
<td>44</td>
<td>169</td>
<td>0.842</td>
<td>0.738</td>
</tr>
<tr>
<td>5</td>
<td>0.823</td>
<td>44</td>
<td>164</td>
<td>0.841</td>
<td>0.730</td>
</tr>
<tr>
<td>6</td>
<td>0.736</td>
<td>44</td>
<td>161</td>
<td>0.827</td>
<td>0.702</td>
</tr>
<tr>
<td>7</td>
<td>0.866</td>
<td>44</td>
<td>143</td>
<td>0.867</td>
<td>0.720</td>
</tr>
<tr>
<td>8</td>
<td>0.843</td>
<td>44</td>
<td>140</td>
<td>0.865</td>
<td>0.706</td>
</tr>
<tr>
<td>9</td>
<td>0.739</td>
<td>44</td>
<td>140</td>
<td>0.839</td>
<td>0.640</td>
</tr>
<tr>
<td>10</td>
<td>0.883</td>
<td>37</td>
<td>118</td>
<td>0.897</td>
<td>0.702</td>
</tr>
<tr>
<td>11</td>
<td>0.865</td>
<td>37</td>
<td>117</td>
<td>0.893</td>
<td>0.678</td>
</tr>
<tr>
<td>12</td>
<td>0.740</td>
<td>38</td>
<td>117</td>
<td>0.844</td>
<td>0.519</td>
</tr>
<tr>
<td>13</td>
<td>0.819</td>
<td>44</td>
<td>187</td>
<td>0.828</td>
<td>0.746</td>
</tr>
<tr>
<td>14</td>
<td>0.810</td>
<td>44</td>
<td>180</td>
<td>0.827</td>
<td>0.741</td>
</tr>
<tr>
<td>15</td>
<td>0.733</td>
<td>44</td>
<td>177</td>
<td>0.817</td>
<td>0.730</td>
</tr>
<tr>
<td>16</td>
<td>0.819</td>
<td>27</td>
<td>200</td>
<td>0.827</td>
<td>0.641</td>
</tr>
<tr>
<td>17</td>
<td>0.810</td>
<td>28</td>
<td>200</td>
<td>0.826</td>
<td>0.546</td>
</tr>
<tr>
<td>18</td>
<td>0.733</td>
<td>33</td>
<td>195</td>
<td>0.817</td>
<td>0.271</td>
</tr>
</tbody>
</table>
costs with regard to the work-in-process. The optimal high arrival rates are more or less insensitive to the costs. Only when the switching costs are relatively low they tend to decrease substantially. Although we allowed the high arrival rate to exceed 1, there seems to be no cost setting where such a high arrival rate is beneficial.

If management does not take measures to influence the arrival rate, then the optimal arrival rate gives for \( r=0 \) an optimal profit of 0.749, for \( r=0.001 \) an optimal profit of 0.744 and for \( r=0.01 \) an optimal profit of 0.700, which are all below the corresponding values for the situations where management does react to the level of work-in-process.

5. Conclusions

In this paper we investigated the situation where the productivity depends on the work-in-process. This dependency has been modeled by as a dependency of the departure rate of order from the shop (modeled as a single server station) on the number of orders in the shop. We further assumed that management is aware of the existence of this dependency and has knowledge of both the maximum departure rate of orders and the minimum departure rate of orders, and the workload for which the maximum departure rate is obtained. Management has also the possibility to influence the order arrival rate and thus to influence the level of work-in-process.

To calculate the optimal profit and the values of the switching points and the arrival rates, we modeled the shop as a two dimensional Markov chain and developed an expression for the probability of having \( n \) customers in the shop. Using these probabilities and the OptQuest \textsuperscript{®} program we the optimal values and the corresponding values for the switching points and the arrival rates.

It turned out that the optimal values, with exception of the situations where the costs for switching are relatively low, or the costs of work-in-process are high, are more or less insensitive to the costs.

We an also conclude that if management takes measures to influence the arrival rate in case where the workload is too high or too low, the optimal profit can be substantially increased, compared to the situation where the arrival rate is not influenced; this even holds in the situation with high switching costs.

In further research we will investigate whether costs that are related to the difference between the low arrival rate and the high arrival rate play a role in our setting.

References


