Int. J. Production Economics 137 (2012) 272–283

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

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

The capacity investment decision for make-to-order production systems with demand rate control

J. Will M. Bertrand, Henny P.G. van Ooijen *

Department of IE&IS, Eindhoven University of Technology, Den Dolech 2 Eindhoven, NL-5612, AZ, The Netherlands

A R T I C L E   I N F O

Article history:
Received 2 December 2010
Accepted 8 February 2012
Available online 20 February 2012

Keywords:
Capacity investment
Make-to-order
Demand rate control
Lead time sensitivity

A B S T R A C T

In this paper we study the capacity investment decision for make-to-order manufacturing firms that utilize a fixed capacity, operate in a stochastic, stationary market, and can influence their demand rate by increasing or decreasing their sales effort. We consider manufacturing situations that differ in sales contribution, in market elasticity to sales effort, work-in-process costs, and demand sensitivity to lead time.

If demand is insensitive to lead time we find that for situations with a low sales contribution and high work-in-process costs (for example the manufacturing of capacity equipment that is at the end of the innovative life cycle, such as food processing machines and textile printing machines), using a dynamic demand rate policy can bring substantial improvements in profit. Moreover, when using the optimal demand rate policy, the profit is quite insensitive to the initial capacity investment. If demand is sensitive to lead time, using a dynamic demand rate policy brings substantial increases in profit in all situations considered. The profit again is quite insensitive to the initial capacity investment.

Consequently, without much loss in profit, for all cases the capacity investment decision can be based on the stochastic model with stationary demand, neglecting the possibility of influencing the demand rate. However, the profit that results from this investment, and the return-on-investment, should be determined from a model that includes the optimal demand rate policy, since the stationary stochastic model can significantly underestimate the profit and could lead to the abandonment of the investment.

1. Introduction

In this paper we study the capacity investment decision for make-to-order manufacturing firms that utilize a fixed capacity, face a stationary demand distribution, and can influence their demand rate by increasing or decreasing their sales effort. In particular, we study how the ability to vary the demand rate in response to the work-in-process level affects profit and the capacity investment decision.

Capacity investment decisions are typically considered at the strategic or tactical level and are mostly studied under simplifying assumptions regarding control at the operational level. Specifically, in these models it is generally assumed that at the operational level demand rates are stationary. This assumption is also prevalent in much research on operational decision-making. For instance, when studying the scheduling of production orders in production systems, it is often assumed that demand is an exogenous variable. A few studies exist that study the demand–production interface, e.g., Palaka et al. (1998) who study the situation where demand rate is affected by the lead times quoted, which can be considered a sales instrument (see also Dewan and Mendelson (1990) and Ray and Jewkes (2003)). However, all these studies look for the optimal steady-state values and neglect the possibility that, at the operational level, sales efforts can dynamically respond to past sales and work-in-process levels. The aim of this paper is to incorporate this possibility in the analysis of the initial capacity decision problem, and to see for what production systems it makes a significant difference to do so.

Firms that exercise a constant sales effort in a stationary market will face a stationary demand rate. Demand per unit time can then be characterized as a stochastic variable with a constant mean. For such systems the optimal demand rate, and thus the optimal sales effort to be deployed, can be determined by maximizing the profit per unit time, where profit is defined as the revenue per unit time minus the sum of sales costs and variable production costs per time unit. Variable production costs include the costs of work-in-process.

For manufacturing systems with stochastic demand and/or production, the demand level must be set below the capacity level, since otherwise capacity utilization would be equal to or

* Corresponding author.
E-mail address: h.p.g.v.ooijen@tue.nl (H.P.G. van Ooijen).

0925-5273/$ - see front matter
doi:10.1016/j.ijpe.2012.02.010

greater than one, and work-in-process, and thus work-in-process costs, would be infinite. Buss et al. (1994) analyze this problem and give results for the optimal capacity and the optimal stationary demand rate in the presence of work-in-process costs due to stochastic order arrivals and stochastic processing times. Palaka et al. (1998) extend this problem by explicitly including lead-times in the model.

In this paper, we assume that the firm, after installation of the capacity, can respond dynamically to the work-in-process level by changing the sales effort in order to change the demand rate. This complicates the analysis but enables us to provide a number of useful insights. We will first focus on environments where customers are lead time insensitive and later check our results for environments where customers are lead time sensitive.

First, we want to know for what type of manufacturing system varying the demand rate can give a substantial increase in profit. Second, we want to determine for what type of manufacturing system the possibility of varying the demand rate should be taken into account when making the initial capacity decision.

In order to relate our research to previous work, we model the manufacturing system and its market setting following Buss et al. (1994). These authors extend the classical deterministic capacity investment model by modeling the production system as a single server queuing system with exponential customer order inter-arrival times and exponential processing times. They assume revenues per period that are concave in demand rate, capacity costs per period that are a power function of the capacity (rate), and work-in-process carrying costs that are linear in work-in-process. Furthermore, they assume that after the capacity decision is taken, both the production rate and the demand rate are stationary variables. Their problem was to find optimal values for the demand rate and capacity as a function of the system parameters. We extend their model by including the possibility of varying the demand rate after the capacity has been installed and thus the production rate has been set.

We first develop a value iteration method to determine, for a given capacity, the optimal dynamic demand rates as a function of the evolution of the work-in-process. Then we develop a Markov model to determine, for a given capacity, revenue and cost setting, the optimal demand rate policy and the average profit per unit time. This Markov model is also used to determine the average profit per time unit obtained under the optimal stationary demand rate.

Next, we select sets of model parameter values that represent manufacturing systems with different market elasticity, sales contribution, and work-in-process costs. For each set of parameter values we use our models to determine the average profit per time unit under the optimal stationary and dynamic demand rate policy, in order to identify manufacturing situations where being able to vary the demand rate leads to substantial increases in profit. We then investigate whether anticipating the possibility of varying the demand rate after installation of the fixed capacity improves the capacity investment decision. It can be shown that the optimal capacity level under a dynamic demand rate lies between the optimal capacity level under stationary stochastic demand and production rates, and the optimal capacity level under deterministic demand and production rates (the latter resulting in zero work-in-process costs). For each set of parameter values, we vary the capacity level in this range to find the optimal capacity. From these results, we provide insights into the type of manufacturing system for which it is important to consider the availability of dynamic demand rates when making the capacity investment decision.

The rest of the paper is organized as follows. In Section 2 we give a short overview of related literature. In Section 3, we formulate the economic model of the production system and its market setting. Section 4 presents the value iteration method and the Markov model, and gives an example of the optimal demand rate policy for a given manufacturing situation. In Section 5, we present the different manufacturing situations we study. The results obtained for each situation under the optimal demand rate policy as well as under the optimal stationary demand rate are discussed in Section 6. In Section 7 we will check the results of this research for situations where customer demand is lead time sensitive. Finally, in Section 8, the conclusions of the paper are given.

2. Literature review

In the operations management literature, research on optimal demand rate setting is scarce. In systems with admission control customers arrive according to an arrival process with a fixed rate. However, upon arrival, a customer may either be permitted or denied access to the system. Stidham (1985) reviews open-loop systems where each arriving job is accepted with probability p, and closed-loop systems where arriving jobs are admitted, based on the observed queue length. The main emphasis is on the difference between socially optimal and individually optimal (equilibrium) policies, and on the use of dynamic programming inductive analysis to show that an optimal control is monotonic or characterized by one or more "critical numbers". Xu and Shantikumar (1993) determine the optimal admission policy for a first-come, first-served M/M/m queuing system to maximize the expected discounted profit. They derive an easily computable approximation for the optimal threshold value (for the number of customers in front of the customer at the end of the queue) that triggers the last customer to renege his decision to enter the queue.

Our research does not focus on admission of customer orders from a fixed externally given stream of customer orders, but on influencing the stream of customer orders at its source, thus influencing the arrival rate. Hillier (1963) studied the problem of determining the proper balance between the amount of service and the amount of waiting for that service. He used an economic model for determining the level of service which minimizes the total of the expected cost of service and the expected cost of waiting for that service. Buss et al. (1994) determine for a single machine production system the arrival rate (or demand rate) and the capacity level (or production rate) that maximize the profit if contributions are concave to scale, capacity costs are convex to scale and work-in-process costs are linear to scale. They assume stationary production rates and order arrival rates. Palaka et al. (1998) extend this model to include lead-time setting as a sales instrument.

So and Song (1988) study a production system where the demand rate, or order arrival rate, is sensitive to both price and lead-time, and thus can be influenced. They determine the joint optimal selection of the stationary values of price, lead-time and capacity for the production system. They do not consider the possibility of dynamically influencing the demand rate by varying the price and lead-time over time. The work of So and Song can be viewed as an extension of the research of Buss et al., since it explicitly models the sales instruments price and lead time. Tijms and van der Duyn Schouten (1978) study a system that at each point in time can be in one of two possible states and where, at any moment, the system can be switched from one state to another. The arrival rate and the service rate both depend on the state of the system. They derive a formula for the long-run average expected cost per unit time if holding costs are state dependent and switch-over costs are fixed.

There is also extensive literature on the production–marketing interface where the interaction between demand and production is discussed in various ways (e.g., Upasani and Uzsoy (2008)). Another stream of research related to this topic is the research