Overlaying the Just-in-Time with Kanban System on an American Production Environment

by

Patrick Robert Philipoom

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Management Science

APPROVED:

L. P. Rees, Chairman

E. R. Clayton

P. Y. Huang

L. J. Moore

B. W. Taylor, III

February, 1986

Blacksburg, Virginia

Overlaying the Just-in-Time with Kanban System on an American Production Environment

by

Patrick Robert Philipoom

L. P. Rees, Chairman

Management Science

(ABSTRACT)

During the past several years, the publicized successes of Japanese production management techniques have created an interest in the potential of these techniques for application in an American manufacturing environment. One such Japanese technique that has been the focus of much attention from American manufacturers and production managers is the "just-in-time (JIT)" technique implemented with "Kanbans."[1] However, the applications of the JIT technique in Japan that have been reported have been for large scale assembly line operations that, in general, encompass the unique physical and philosophical characteristics typical of Japanese production systems.

The factors that contribute to the success of the JIT system in Japan are frequently not exhibited in manufacturing systems in the United States, especially in American systems

---

[1] Toyota uses a system of cards, called Kanbans, to control inventory and schedule production in their automotive assembly plants.

that combine assembly and shop-type operations and encompass a high degree of system variability. As such, it is questionable whether the JIT technique can be successfully adapted to American manufacturing systems that do not display the characteristics of Japanese production operations. Nevertheless, a number of American manufacturing companies, in hope of achieving at least some of the Japanese success in inventory control, quality control and production scheduling, have begun implementing the JIT technique in their own unique production environment. The purpose of this dissertation is to investigate implementing JIT in a non-Japanese production environment and to show how JIT can be adapted so that it can have a broader range of applicability, especially under the particular set of conditions that are very likely to exist in many American production environments.

ACKNOWLEDGEMENTS

I would like to thank Dr. Loren P. Rees, my chairman and mentor, for the endless support, effort and guidance that he has provided throughout my graduate studies at Virginia Tech - especially during the dissertation stage of my doctoral program.

I would like to thank Dr. Bernard W. Taylor, III, Head of the Department of Management Science, for his provision of financial aid and resources during my advanced studies and especially for his help in developing three journal articles from the chapters of this dissertation.

I would like to thank Dr. Philip Y. Huang for making available his expert insight in Just-in-Time production systems.

I would like to thank Dr. Edward R. Clayton and Dr. Laurence J. Moore for their assistance with the dissertation and their support and encouragement throughout my graduate studies at Virginia Tech.

I would like to thank Dr. Roberta S. Russell for kindly agreeing to attend the defense of my dissertation in place of Dr. Huang.

I would like to thank
for their cheerful dispositions and their help with the typing of this dissertation.

Finally, I would like to thank my parents,

for their continual love and support throughout this and all of my endeavors.

TABLE OF CONTENTS

1.0 INTRODUCTION

During the past several years, the publicized successes of Japanese production management techniques have created an interest in the potential of these techniques for application in an American manufacturing environment. One such Japanese technique that has been the focus of much attention from American manufacturers and production managers is the "just-in-time (JIT)" technique implemented with "Kanbans." However, the applications of the JIT technique in Japan that have been reported have been for large scale assembly line operations that, in general, encompass the unique physical and philosophical characteristics typical of Japanese production systems.

The factors that contribute to the success of the JIT system in Japan are frequently not exhibited in manufacturing systems in the United States, especially in American systems that combine assembly and shop-type operations and encompass a high degree of system variability. As such, it is questionable whether the JIT technique can be successfully adapted to American manufacturing systems that do not display the characteristics of Japanese production operations. Nevertheless, a number of American manufacturing companies, in hope of achieving at least some of the Japanese success in inventory control, quality control and production scheduling, have begun implementing the JIT technique in their own unique production environment.

The purpose of this dissertation is to investigate implementing JIT in a non-Japanese production environment and to show how JIT can be adapted so that it will have a broader range of applicability, especially under the particular set of conditions that are very likely to exist in many American production environments. The objectives are to determine:

1. how JIT can be used even when the firm cannot reduce the setup times for all machines sufficiently to allow single container lotsizes;

2. the factors that are present in an American production environment that influence the number of Kanbans in a shop, and how to set the initial number of Kanbans in such an environment; and

3. how the number of Kanbans at a workcenter should be dynamically adjusted when conditions in the shop are not stable.

Before the investigation conducted in this dissertation can be presented, certain terms and concepts must be classified and/or defined; this is done in the remainder of this chapter. First, two classifications of production shops are given to help facilitate the discussion of how the sample shop used in this investigation (which is described in Chapter 2) differs from the shops in which JIT has been used by the Japanese. Next, two alternative production systems, order point and material requirements planning, are discussed. The Kanban system is essentially an order point system, whereas material requirements planning is a system widely used in the United States for dependent demand.

Having developed a framework as to different types of shops and different types of systems, the philosophical and cultural differences between Japanese and American production systems are examined. These include such factors as respect for humanity, lifetime employment, frozen demand schedules and workforce attributes. Then, the JIT system developed by Toyota, the JIT with Kanban system is described, and the role that the Kanban system plays within the overall JIT system is delineated. The JIT with Kanban system is also compared to material requirements planning.

Finally, a review of the quantitative research in JIT is presented. The current literature is summarized and its limitations are discussed. This provides a focus for the motivation and scope of the investigation which follows.

1.1 TYPES OF PRODUCTION SHOPS

There are many ways to classify production shops. Groover [13] suggests two schemes: by the production volume and by the layout of the plant. Each type of shop under these classifications is briefly summarized and references are provided for the interested reader.

1.1.1 CLASSIFICATIONS BY PRODUCTION VOLUME

Classification of production shops by production volume is based on the fact that as the volume of production increases, benefits accrue by using more specialized machines and layouts.

1.1.1.1 Job Shop Production

Job shop production [55, pp. 27-48] involves the manufacture of small lots with low production volumes. Shops within this category produce orders to meet specific customer requests, which are often one-time orders. A job shop must have general purpose production equipment and highly skilled workers because of the variety of products it manufactures. Such a shop is typically laid out to allow for great flexibility in machining requirements.

1.1.1.2 Batch Production

Batch production [55, pp. 49-76] is used for manufacturing medium-sized lots for orders that tend to be somewhat repetitive and continuous. There is usually less variety than with job shop production in the types of orders received. This homogeneity allows the shop to use more specialized equipment, which reduces much of the labor cost in producing an item and enables the production rate to be increased. With batch production, a batch is produced and placed in inventory; when the inventory is depleted or reduced to low levels, another batch is produced.

1.1.1.3 Mass Production

This category of shop is used for high-volume production where demand is so high that the product is continuously being produced. With mass production [55, pp. 83-106], equipment is especially designed for the production of one product, thereby eliminating much of the labor and material handling costs generally incurred in the two previous types of shop discussed.

There are two types of mass production: quantity and flow production. Quantity production is generally invoked with a product which requires a limited number of operations - making a bolt for example. Flow production refers to the production of an item which requires a long sequence of operations - such as manufacturing an automobile.

1.1.2 CLASSIFICATIONS BY PLANT LAYOUT

Production volume has a significant impact on the choice of plant layout. Production volume, however, is not the only factor. The physical properties of the product and material handling concerns also affect the plant layout choice.

1.1.2.1 Fixed Position Layout

A fixed position layout [&MT., pp. 132-134] is used when a product is so large that it is not feasible to move it through the shop. The product remains stationary, and all equipment is brought to it. This layout is used, for example, in the manufacture of large aircraft and ships.

1.1.2.2 Process Layout

In order to allow for greater flexibility in machining requirements for products, equipment is grouped by function in a process layout [55, pp. 27-48]. Each function (or process) is a separate department. Jobs must be routed from department to department, throughout the shop, to be completed. This allows for great flexibility in machining requirements for products.

1.1.2.3 Product Flow Layout

In the product flow layout [55, pp. 83-106], jobs are moved along a production line, often on a conveyor belt, with machines adjacent to the line. This configuration is used for high-volume production and is highly inflexible: once the shop is laid out in this fashion, only an extremely narrow range of products can be produced. Nonetheless, this layout will allow the shop to produce the few products processed highly efficiently.

1.2 ALTERNATIVE PRODUCTION SYSTEMS

1.2.1 ORDER POINT SYSTEMS

Material requirements planning was perhaps the first production-inventory system used, albeit informally [47]. But as products became more sophisticated during the first half of this century, companies were no longer able to perform all the calculations this approach requires. To handle this emerging problem, order point systems were used. In an order point system, the components necessary for production are kept in inventory. When the inventory is reduced to a specified level, a pre-determined quantity of the item is ordered. A reorder point is set for each item by determining the lead time demand plus a safety stock to handle variability in lead time and demand. An order quantity is often set using the Economic Order Quantity (EOQ) or some sort of probabilistic approach [14] [40]. Another approach to setting order quantities is the ABC classification method in which the manager is urged to concentrate on reducing inventory for the relatively small number of components which compose the bulk of the company's inventory costs [67].

The most fundamental problem with the order point approach is that it assumes that there exists an independent relationship (i.e., there is no relationship) between the demand for the various components needed for production. There is actually a dependent relationship between the demand for components used in the production of a final product since, for example, every time a car is ordered four tires, two axles, one steering wheel and so forth are ordered. If the demand for components were independent there would be no relationship between these items. One implication of this assumed independence is that large amounts of safety stock must be carried for each item since if only one of the items needed to produce the final product is out of stock, production must halt. As the number of components used to make a product increases, the probability of any component being out of stock must decrease to keep the overall probability of stopping production constant, thus raising the level of safety stock required for each component [62].

Another problem with order point systems is that they are inappropriate for "lumpy" demand which causes large fluctuations in daily demand. EOQ and many other techniques used in lotsizing assume there is a continuous rate of demand for the item - which is generally not a valid assumption. Thus, the lotsizes determined by these methods are based on incorrect model of inventory costs. Finally, order point systems do not explicitly consider capacity requirements or the timing of requirements necessary for demand to be met just as needed [42].

1.2.2 MATERIAL REQUIREMENTS PLANNING

The objective of Material Requirements Planning (MRP) [42] is to determine the requirements for meeting demand and to generate the information needed for correct inventory actions. MRP works backwards from a schedule of demand for final products to determine the requirements for components in terms of the quantity required and the date by which each process must be completed.

There are three categories of inputs to an MRP system [42]: a master production schedule, a bill-of-material file and an inventory records file. The master production schedule is a list of quantities and need dates for the final products. The bill-of-material file is a list of the components needed to make those products and data about the production process necessary for each component. The inventory records file contains the projected inventory balance for each component and final product. In an MRP system, it is imperative that the projected inventory balance be extremely accurate. Many go as far as suggesting that all storerooms be kept under lock and key to insure the accuracy of the inventory information [47].

Once the master production schedule is set, the gross requirements for components are generated using information in the bill-of-material file. A net requirements list is then generated by subtracting the projected inventory balance found in the inventory records file from the gross requirements. Then working backwards from the final product, the date when a process must be completed so the following process will not be impeded is determined. The output for an MRP system is a schedule indicating when each production process must begin and the lotsizes that are required.

MRP systems have three major advantages over order point systems [65]. First, using an order point system, safety stock must be kept for each component to handle forecast errors. This springs directly from the fact that an order point system acts as if the demand for components of a final product are independent. In MRP, safety stock for forecast errors need only be carried for the final product - thus, in-process inventory for components can be significantly reduced. A second advantage to MRP is that it is designed to handle lumpy demand. Order point systems operate best when demand is smooth [62]. Finally, MRP explicitly considers time phasing of production. Order point systems are oblivious to time phasing concerns.

MRP, however, is not without problems. MRP has been criticized for contributing to long production cycles times because it builds excess safety lead times into the production schedule. There is also a tendency to use large buffer inventories in case there are problems such as machines breaking down or there are many defects. All this leads to excessive in-process inventories [56]. A second fault of MRP is that it does not deal with capacity planning; it is simply assumed that the plant has the capacity to meet the schedule established by MRP. Closed-loop MRP, sometimes called manufacturing resource planning (MRP-II), is an attempt to correct this fault by starting with a basic MRP system and including capacity planning in a closed-loop or iterative fashion [49] [65]. In a recent study by Anderson, Schroeder, Tupy and White [1], it was found that only 9.3 percent of the firms that have implemented an MRP system are considered class A users where scheduling and planning are performed in a closed-loop system, inventory is under control and little expediting is required. Sixty two percent of the users surveyed are considered class C or D users where capacity planning is done informally and expediting is used to control the flow of work. Other research [38] indicates that of the companies surveyed, only 44 percent of the companies using MRP have found it to be cost effective and [58] only 50 percent of the companies were satisfied with MRP.

1.3 JUST-IN-TIME

The concept of just-in-time (JIT) began over 20 years ago in the Japanese shipbuilding industry [57]. Steelmakers had overexpanded making it possible for shipbuilders to get fast deliveries on steel orders. Taking advantage of this, shipbuilders dropped their inventories from 1 month's supply to a supply of 3 days. JIT ideas spread to other Japanese companies who began demanding similar service from their suppliers. Later, these companies began applying the same concepts in their internal operations [57].

The objective of JIT systems is "...to produce the necessary units in the necessary quantities at the necessary time [36]." JIT in a broad sense refers to "...all the activities of manufacturing which make the just-in-time movement of material possible" [15] and not just the movement or transport of material activities associated with the system. JIT systems are characterized by a fanatic obsession to reduce in-process inventories. The Japanese consider in-process inventory to be a waste of resources - but even worse, they claim that "inventory is the root of all evil [16]." It has been shown that by reducing in-process inventory (by reducing lotsizes) significant improvements in quality, worker motivation and productivity can be obtained [57].

JIT received widespread attention during the oil shock of 1973; when most Japanese companies lost money, Toyota showed a huge profit using their just-in-time with Kanban system [36]. In 1980, Toyota turned over their inventory every 4 days and reduced their break-even point to 64 percent of sales. It was determined that Japan's cost advantage for a comparable car was $1,700 during that time. The cost difference over United States' firms was attributed mainly to "adversarial labor relations, excessive inventories, lagging productivity, and inferior quality performance [7]."

During the 1970's, many Japanese manufacturers switched to using JIT systems. Now, during the 1980's, many American firms are embracing JIT techniques. The list of American companies using JIT includes: Chrysler, American Motors, Hewlett Packard, Apple, Burroughs, Black and Decker, Bendix, GM, Ford, Kawasaki and Harley Davidson [6] [57] [59]. Using JIT concepts, Harley Davidson reduced their break-even point by 32 percent, reduced defects by 24 percent, reduced in-process inventory from $23 million to just over $8.5 million and significantly reduced stockouts [59]. Such performances have led the American Production and Inventory Control Society to call for a Zero Inventory Crusade. (Note that the terms Zero Inventory and JIT may be used interchangeably [66]).

1.3.1 CULTURAL INFLUENCES IN JAPAN

1.3.1.1 Respect for Humanity

The Japanese use a management by consensus approach [22] [32] [43] [45] [53]. They spend a great deal of time getting everyone involved in the decision making process. Although much time is spent obtaining a consensus, once it is reached the plan is implemented more rapidly since everyone is committed to the plan. The Japanese also place a strong emphasis on keeping the lines of communications open within the company. An example of this which has received considerable attention recently is the quality circle [8] [9] [50]. A quality circle is a small group of workers in a shop or department that meet voluntarily to discuss how to improve product quality. Some of these groups take a broader perspective and include other considerations such as productivity improvement; these groups are sometimes called participatory circles. The scope of these groups is limited to improvements within their own group or department.

Another aspect of respect for humanity is "...allying human energy with meaningful, effective operations by abolishing wasteful operations [36 p.125]." It is felt if a worker sees the job as important, motivation and self-esteem will be high. If the worker feels the job is wasteful or insignificant, morale and self-esteem will suffer.

1.3.1.2 Lifetime Employment

Major Japanese companies have a tradition of lifetime employment for their workers. Only in extreme circumstances will these companies lay off workers. This has many implications for the production system. First, the workers tend to identify more with the company and see a link between their success and the company's success - improving worker morale. Also, since the turnover rate is low, the benefits of employee training programs accrue over a longer period of time.

1.3.1.3 Vendor Relationships

Since Japan is a small country geographically, vendors are often located near the companies they serve. This is important because in JIT systems vendors are generally expected to make three or four deliveries a day on demand. Most United States companies receive weekly or monthly deliveries from their suppliers. Many United States companies implementing JIT systems have asked their vendors to relocate closer to their factories.

Another difference in vendor relationships is the way American and Japanese companies manage their vendors [12] [56]. American companies tend to view vendors as adversaries. They generally use multiple sources for items and tend to play the vendors against each other. The Japanese deal with vendors in a totally different manner. Vendors are considered as co-workers and are treated as an extension of the factory. The companies tend to have long term relationships with their vendors.

1.3.2 ASSUMPTIONS NECESSARY FOR JIT SYSTEMS

1.3.2.1 Small Setup Times

The major thrust behind JIT systems is to reduce inventory by producing small lotsizes. To do this economically, it is necessary to reduce setup times. The Japanese try to reduce setup times to under ten minutes. On achieving this they try to reduce setup times to under one minute; this is called a "one touch setup". Frequently it takes American companies from hours to entire days to perform setups on similar machines [37].

The Japanese attack the setup problem in several ways [36]. They do not use specialists to perform setups. Line workers are expected to operate and setup a variety of machines. When a setup is required, the workers get together to setup the machine in a parallel fashion, removing the bottleneck the setup creates. Another way they reduce setup times is by designing their own machines using in-house engineering staffs. The machines are designed specifically to reduce the time needed for setups. A third way they reduce setup times is by separating the actions required in setting up a machine into two categories: the external setup and the internal setup. The external setup is the part of the setup which can be done while the machine is running. The internal setup is the part of the setup which can only be done when the machine is stopped. For example, getting the tools required to perform a setup is part of the external setup since it can be done while the machine is running. Conversely, changing a drill bit can only be done only after the drill press is stopped - so it is part of the internal setup. What the Japanese try to do is to make as much of the setup as possible part of the external setup and make sure all the external setup is performed before the machine finishes a job. This reduces machine downtime during a setup since none of the external setup is performed while the machine is idle.

1.3.2.2 Frozen Demand Schedule

JIT systems, according to the literature, require that production schedules be essentially frozen for about a month [15] [36]. Once the production schedule is set, the lines are balanced. Deviations from this production schedule may cause the lines to get out of balance - causing the production to back up. Toyota has found that its JIT system can handle demand fluctuations of up to 10 percent by adjusting the length of the workday. The workers stay until the work is done. JIT systems purportedly cannot handle larger fluctuations in demand. Many companies smooth demand by filling orders from the previous week. The actual daily production schedule using this approach is the average daily demand for each product from the previous week. Another approach is to use small buffer inventories of popular product lines to smooth production. Companies draw from this inventory or build it up (by producing less or more than the actual demand for that day respectively) to smooth the production schedule. This practice, however, should be kept to a minimum since it defeats the purpose of Just-in-Time production.

1.3.2.3 Workforce Attributes

The Japanese place a great emphasis on training their employees. This is vital in JIT systems because workers must be cross-utilized extensively and variability in processing times must be kept to a minimum [21]. As mentioned earlier, workers must be able to perform setups on a wide variety of machines. JIT systems can be thought of as conveyor lines running throughout the factory. When one workcenter takes longer than its allotted time processing an item, the entire line must stop until it finishes.

1.3.2.4 Quality Control

Quality control is essential to a JIT system [24] [57]. Since there is no buffer inventory, when defects are found the workers must go back to the preceding stage to have the defect-producing worker fix it - stopping the entire line. The Japanese do not use statistical quality control with their JIT systems; they use what they call total quality control which aims for zero defects. At many of these factories defects are measured in parts per million. At United States factories defects are generally measured in parts per hundred.

Quality control is the responsibility of the worker on the production line in JIT systems; companies with JIT systems generally have small quality control staffs. They also generally do not have rework lines which in many United States factories take up from 15 to 40 percent of total machine capacity in the plant [57]. It is the responsibility of the worker making the defect to correct his or her mistake. This is good from a behavioral view because the worker and the rest of the factory get almost immediate feedback about defects [56]. In an MRP system, defects are thrown into a rework bin and a replacement is drawn from inventory. This does not provide the worker with the kind of feedback that a JIT system supplies.

Another aspect of the quality control effort is preventive maintenance, which the Japanese practice religiously. Not only does this help reduce the number of defects, but it reduces the amount of machine downtime. In a JIT system, machine downtime must be kept to a minimum since there is no inventory available to keep the line going while the machine is being repaired.

1.4 TOYOTA'S KANBAN SYSTEM

The Kanban system is a subsystem of Toyota's JIT production system. Monden [36, p. 4] describes the role of the Kanban system in Toyota's production system in the following manner:

> Many people call the Toyota production system a Kanban system: this is incorrect. The Toyota production system is the way to make products, whereas the kanban system is the way to manage the just-in-time production method. In short, the Kanban system is an information system to harmoniously control the production quantities in every process.

The Kanban system is just a system of cards to control production and inventory. There are primarily two types of Kanbans: production Kanbans and withdrawal Kanbans. Another key element in Kanban is the container. A container is the basic unit of production, roughly analogous to a lotsize in

MRP. The size of a container (the number in the lot) is kept constant within components used for a particular product.

Kanban can be understood by examining a single workcenter (stage N) within a factory [21] [35]. All the containers at stage N that are currently being processed or are in finished inventory must have a production Kanban attached. Similarly, all containers in raw material (or pre-process) inventory must have a withdrawal Kanban attached. For the succeeding stage to get its pre-process materials from stage N, a withdrawal Kanban must be presented at stage N, where a container is pulled from finished inventory. The production Kanban is removed from the container and is replaced with the succeeding stage's withdrawal Kanban; the container and withdrawal Kanban go to the succeeding stage. Thus, the withdrawal Kanban authorizes the withdrawal of a container by the succeeding stage. The production Kanban is sent back to the beginning of stage N - authorizing the removal of a container of stage N's pre-process inventory with a withdrawal Kanban attached. The withdrawal Kanban is removed from the container and replaced by the production Kanban which just arrived at the beginning of stage N. This production Kanban authorizes processing of the container. The withdrawal Kanban is sent to the preceding stage which initiates replenishing the pre- process inventory at stage N.

In MRP, each workcenter is given a schedule of what to produce, in what quantity and when it is needed. In Kanban, this information is transmitted by giving the final workcenter the Kanbans needed for meeting that day's demand. The other workcenters get their production instructions by receiving withdrawal Kanbans throughout the day. This is why Kanban is considered a "pull" system - products are pulled through the factory "just in time" to meet demand.

To summarize how Kanban works at Toyota, there are four basic rules [15]: 1) every container holding parts must have a Kanban attached; 2) containers are never transported between workcenters unless a withdrawal Kanban is attached; 3) production is always done in standard container sizes and; 4) production must not be performed unless authorized by a production Kanban. Note that a production Kanban will always stay within a workcenter. A withdrawal Kanban travels between two workcenters.

There are other types of Kanbans, such as signal Kanbans and vendor Kanbans [36]. Signal Kanbans are used when setup times cannot be reduced to acceptable levels forcing Toyota to produce lotsizes of several containers at a time. Vendor Kanbans are essentially withdrawal Kanbans that travel between the vendor and the stage which uses the vendor's product as their raw material. This illustrates a point previously mentioned: vendors are treated as co-workers and are considered an extension of the factory.

What Kanban attempts to do is to apply the principles of a flow shop to the production of small batches. This is a c

compromise between the process layout and the product flow layout called the group technology layout [13] [51]. This layout places a series of machines together which perform different functions - creating something similar to an assembly line for a particular family of products. This allows a series of processes to be totally automated while retaining much of the flexibility of the process layout. The Japanese have made large gains in productivity and have significantly reduced lead times using group technology.

JIT systems are essentially order point systems. So why does Kanban work when order point systems do not [12]? With its sequencing of products and small lotsizes, Toyota is able to achieve a fairly continuous demand for all the component parts that order point systems assume. This also creates a steady flow through all the workcenters which simplifies capacity planning to the extent that it can be performed by the foreman at each workcenter. Also, since lead times are short, the scheduling of jobs (or time phasing of requirements as it is called in MRP) is not crucial.

1.5 COMPARISON OF KANBAN WITH MRP

A primary objective of Kanban and MRP is to help firms better manage their in-process inventory. The way the two systems view in-process inventory is totally different. The Japanese view in-process inventory as a waste of resources.

They try to totally eliminate it. In MRP, large in-process inventory is used to safeguard the factory from machine breakdowns, late shipments, defects and so forth. This is why MRP is sometimes called a "just-in-case" system [41]. The tools the two systems use could hardly differ more [12]. MRP uses a computer to crunch through a long sequence of computations - generating reams of output. Kanban controls inventory by having workers exchange cards. In MRP, mathematical and statistical techniques are generally used to calculate lotsizes. In Kanban, inventory levels are set using a trial and error approach. Kanbans are taken away from workcenters (reducing inventory) until the shop starts to back up.

Using Kanban leads to significant reductions in inventory investment, leadtime and stockouts - which MRP promises but frequently does not deliver [15] [36] [57] [59]. JIT systems have a proven advantage in quality control [56]. Kanban affords the user a better method of keeping track of in-process inventory. Every item of inventory has a Kanban attached. One of the major problems with MRP is keeping an accurate accounting of in-process inventory [11] [27]. Another advantage of Kanbans is that losses from obsolescence are generally less than in an MRP system [36]. This is because there will be much less inventory on-hand in a Kanban system when model changes are made.

A more subtle advantage of Kanban is that it exposes problems in the manufacturing process which can lead to improvements. For example, suppose a machine has a large setup time. The workcenter will have to use Signal Kanbans to produce lotsizes of several containers. The increased inventory at the workcenter will be extremely noticeable, signaling there is a problem with this machine. This will put pressure on everyone to resolve this problem. In MRP the answer would probably be to increase the lead time at that machine and perhaps increase the lotsize. Thus, the tendency in MRP is to work around problems rather than to solve them.

MRP is believed to apply to a wider range of manufacturing environments. The assumptions for using Kanban are much more restrictive. Achieving a near frozen production schedule is hard to do in practice. For example, product options for cars can cause problems when attempting to smooth the production schedule. A particular make of car can have many combinations of options - each combination is a different product. To handle this, Toyota offers a limited number of option packages. This not only reduces the number of different products, but it stabilizes daily demand. Of course, this does have marketing drawbacks since Toyota can not cater as much to the individual customer. The assumptions with regard to implementing Kanban (reducing setup times, reducing the variance of processing times and cross-utilization of workers) are achievable, but may require substantial amounts of time and money. However, one only needs to look at almost any issue of <u>Production and Inventory Management</u> ([2] [4] [28] for example) to see that implementing MRP is probably at least as difficult and expensive as implementing Kanban.

1.6 REVIEW OF QUANTITATIVE RESEARCH IN KANBAN

Osaka and Terada [25] compared the amplification of production and inventory fluctuations in a pull system (such as Kanban) and a push system using analytical models of the two systems. They found that when container sizes in a pull system were small, amplification of production fluctuation did not occur in the succeeding workcenter. However, using larger container sizes, amplification of production fluctuation in the succeeding workcenter was quite large. The other factor found to cause amplification in a pull system was large lead times for jobs at a workcenter. Amplification of production and inventory fluctuations in push systems was found to be caused only by errors in forecasting demand.

Huang, Rees and Taylor [20] [21] examined how adaptable a JIT with Kanban system is to a United States production environment via a Q-GERT simulation model of a multi-stage flow shop. Their findings show that variability in processing times and demand rates results in an increase in overtime in their sample shop. It was found that by increasing the number of Kanbans, these effects could be reduced. However, it was shown that an increase in Kanbans would not reduce the effects of bottlenecks in the production line. Their conclusions are that a company considering making the change to a Kanban system should be prepared for a lengthy transition period. They estimate that it would take at least one year to train workers and standardize machine processing times and setup times.

Krajewski, King, Ritzmann and Wong [26] compared Kanban to MRP using their Manufacturing Simulation System (MASS) package. They found that Kanban does well (better than MRP) in the more favorable manufacturing environments they studied. However, they found that an order point system worked equally well in the same environments. MRP, however, appeared to be the more robust system. Their conclusions were that Kanban is a good way to implement small lot production and to expose environmental problems which, if fixed, would lead to more efficient production.

Rees, Huang and Taylor [52] also compared MRP to Kanban for a shop with both serial and assembly operations using two Q-GERT simulation models. (However, they only examined JIT without group technology applied.) They found in their sample shop that MRP could be converted to a Kanban operation by reducing leadtimes and setup times, but that more savings would be obtained if the shop were left as an MRP shop and the same leadtime and setup reductions were effected.

1.7 MOTIVATION AND SCOPE OF INVESTIGATION

The current quantitative literature on JIT can be divided into two categories. The first category examines the operation of a JIT system in a flow-shop environment which does not possess the unique aspects generally found in Japanese production shops. The second category compares JIT with MRP within specific sample shops. The thrust of the current literature is to transfer a JIT system as is, into a different environment and see how well it works. The investigation in this dissertation differs from those in the current literature in that the focus here is on how to adapt a JIT system so that it works well in suboptimal environments rather than simply to transfer a JIT system into a different environment and see if it works. For example, the literature indicates that machine setup times must be drastically reduced in order to use the standard JIT with Kanbans system. Part of this investigation focuses on how to adapt a JIT system so that it will work even if setup times cannot be reduced enough to allow for standard JIT with Kanban operations.

As stated above, the purpose of this dissertation is to explore adapting JIT so that it will work when less than ideal settings and conditions are in effect. These suboptimal conditions are very likely to exist in American production environments that are considering implementation of the JIT system with Kanbans. Specifically, this dissertation examines three important problems faced by an American firm implementing a JIT system with Kanban: what should be done if setup times cannot be reduced to a level that allows small lot production (one container lots) at all workcenters; what factors in an American production environment influence the number of Kanbans required at a workcenter, and how should the number of Kanbans at a workcenter be determined; and, how should the number of Kanbans at a workcenter be dynamically adjusted when conditions in the shop are not stable.

In particular, Chapter 3 addresses the problem of using JIT with Kanbans when the firm cannot reduce all machine setups to acceptable levels. The use of Signal Kanbans, which facilitates lot production within a Kanban framework to handle this problem, is explored. The investigation examines setting lotsizes for the Signal Kanbans and the feasibility of using Signal Kanbans versus standard Kanbans in an American production environment.

Chapter 4 investigates how various factors that are present in the typical American production environment influence the number of Kanbans that are necessary for smooth operation of the shop. First, a descriptive model of the relationship between the various shop factors and the number of Kanbans is developed and tested. Second, a methodology for determining the initial number of Kanbans is presented and tested. Finally, the issue of the extent to which shop conditions can deteriorate and JIT still work is addressed.

Chapter 5 addresses whether JIT will work if the shop environment is not as stable as in Japanese shops. A methodology for dynamically adjusting the number of Kanbans in response to unstable shop conditions is presented. The methodology is tested over a wide range of conditions in the sample shop.

In summary, this dissertation investigates the implementation of JIT in a non-Japanese production environment and how JIT should be adapted so that it can have a broader range of applicability. This research should be of interest to American production managers considering the implementation of JIT in their factories. Each of the three parts of this investigation addresses important concerns facing a firm implementing JIT in suboptimal environments.

2.0 THE EXAMPLE SHOP

A simulation model of an example shop is used throughout this investigation to verify the analytical results developed in each of the next three chapters. The example shop, the product structure and the Q-GERT model used throughout the dissertation are described in this chapter.

2.1 THE PRODUCT STRUCTURE AND EXAMPLE SHOP

The product structure and example shop used throughout this investigation are illustrated in Figure 2-1 and Figure 2-2, respectively. The shop has the characteristics of both a job shop and a flow shop. This configuration was chosen because most American shops usually combine assembly and serial operations. Almost all quantitative JIT research to date has focused on pure flow shops. The choice of shop configuration furthers one of the major aims of this research: to investigate JIT in an American production environment.

Workcenters 1, 2 and 4 have the characteristics of a job shop. A variety of operations are performed at these workcenters. Workcenters 3, 5 and 6 are assembly workcenters. Workcenter 3 combines two B's and three H's to make an E, three A's to make a J, two C's and three K's to make an I and,

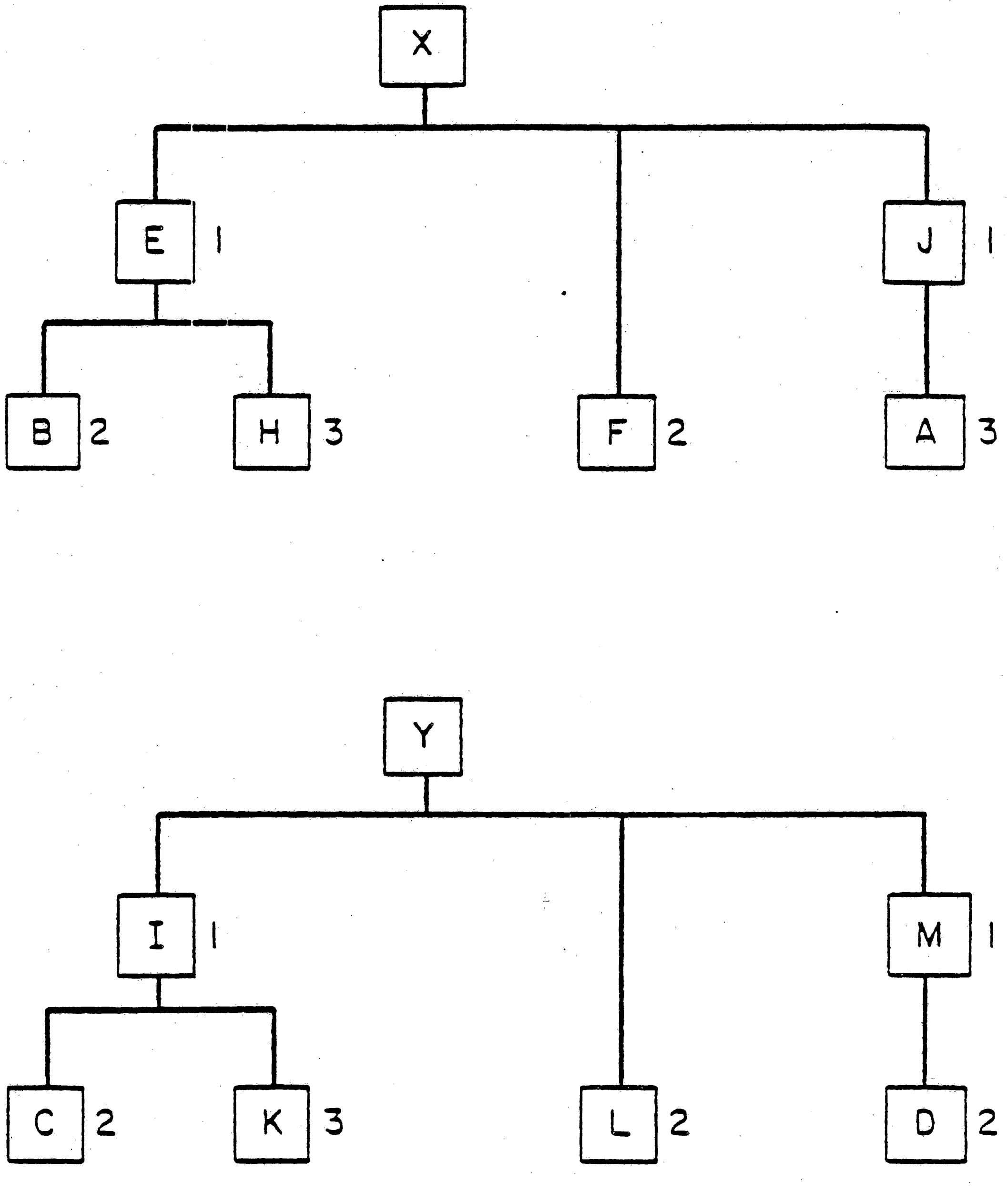


Figure 2-1. The Product Structure.

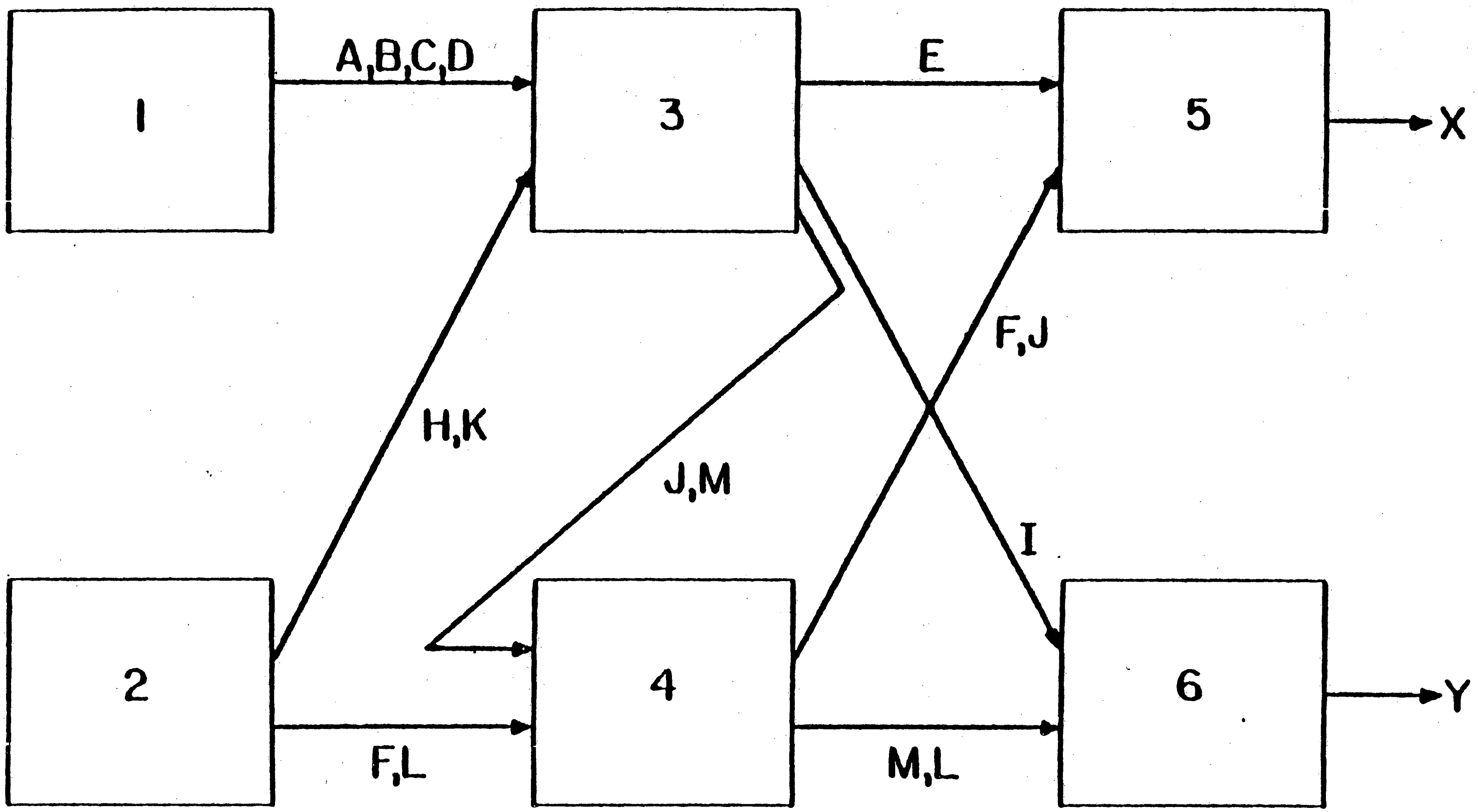


Figure 2-2. The Example Shop and Item Flows.

two D's to make an M. Workcenters 5 and 6 are different from the other workcenters in several respects. They only produce one product each, so they never incur setups. Also, these two workcenters have only one machine whereas the other four workcenters each have two machines. Workcenter 5 assembles one E, two F's and one J to make an X. Workcenter 6 assembles one I, two L's and one M to make a Y. Further details of the shop such as demand for products and costs of production vary from chapter to chapter and hence are provided when appropriate.

2.2 THE SIMULATION MODEL

The example shop is investigated using a Q-GERT model. Q-GERT [48] is a network oriented simulation language written in ANSI FORTRAN. Sixty user written routines (approximately 2000 lines of FORTRAN) are used in addition to the network model. The model keeps track of inventory, setup and backorder statistics. Moreover, the total cost of production is computed and displayed along with its components: inventory costs, backorder costs, and shortage costs. The basics of the model can be explained by describing one assembly workcenter (workcenter 6) and one fabrication workcenter (workcenter 4) since the rest of the workcenters are similarly constructed.

2.2.1 THE ASSEMBLY WORKCENTER

The simulation model for assembly workcenter 6 is illustrated in Figure 2-3. Orders are generated by user function 70, which is called whenever Node 180 is realized. The user function places a transaction representing an order (or a withdrawal Kanban) into Queue Node 80. Finished goods inventory is stored in Queue Node 86. Whenever Queue Nodes 80 and 86 contain transactions at the same time, Assembly Node 88 is realized. This is analogous to an order and a final product being matched. Upon matching,a transaction is sent to Node 89 and then to Queue Node 2, the production ordering post. This signifies that the workcenter should start production for Y as soon as the machine is available.

Whenever Nodes 64, 74 and 84 are realized, the full requirement of I, L and M, respectively, to make a product Y has arrived. It takes two transactions to realize Node 74 since it takes two L's to make a Y. Only one transaction is needed to realize Nodes 64 and 84 since product Y only requires one I and one M. Upon being realized, Nodes 64, 74 and 84 send transactions to Queue Nodes 81, 82 and 83, respectively. When Queue Nodes 81, 82 and 83 all contain transactions (which means that all the pre-process materials needed to make a Y are present) and Queue Node 2 contains a transaction (signalling that an order has been placed for a Y) the production order and the pre-process materials are

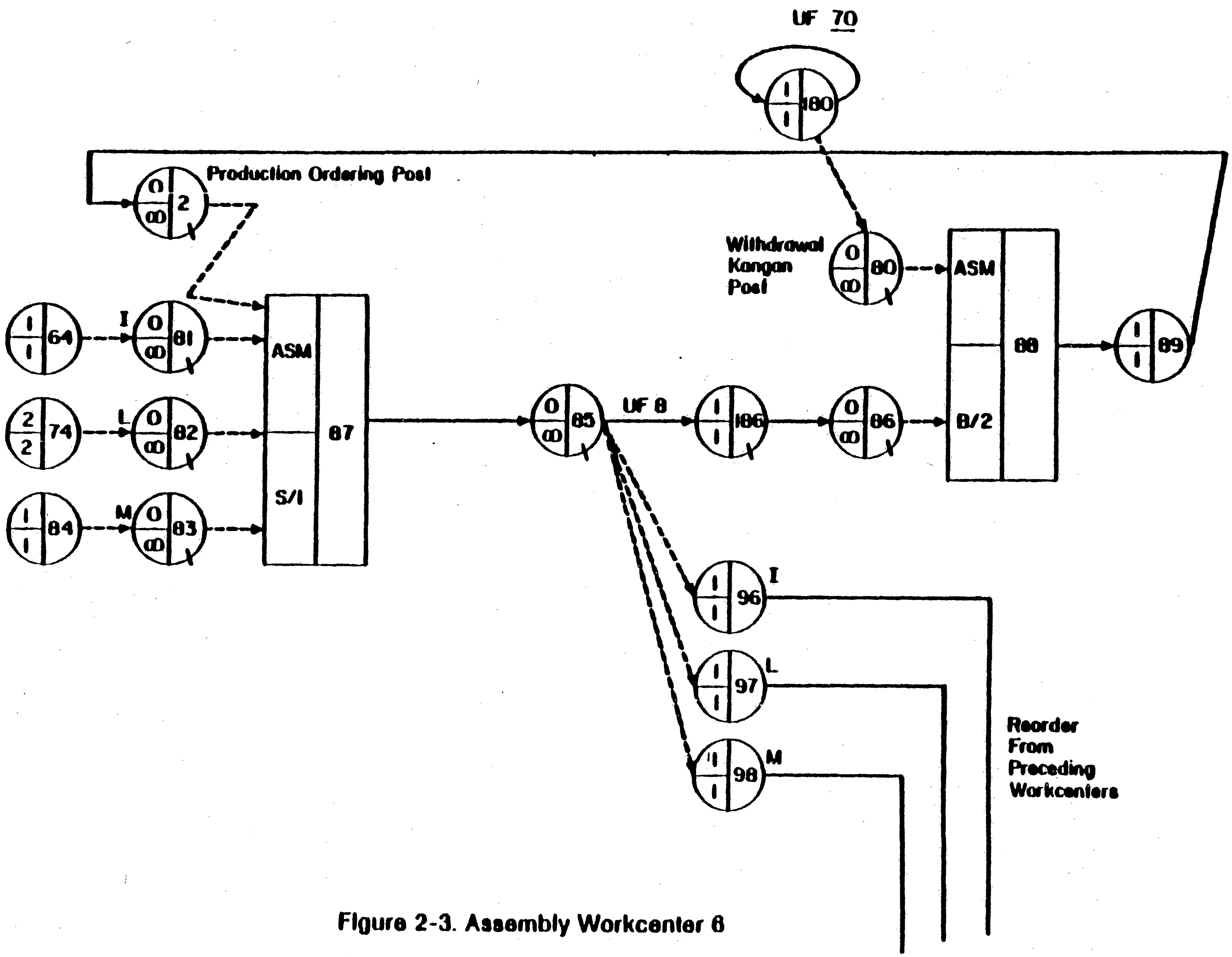


Figure 2-3. Assembly Workcenter 6

matched by Assembly Node 87 which causes a transaction to be placed in Queue Node 85 in front of the machine.

When the machine becomes available, User Function 8 is called. This causes a transaction to be placed in Nodes 96, 97, and 98 using Subroutine PTIN. From here the transactions go to the appropriate preceding workcenters. This represents the workcenter ordering additional containers of pre-process materials that are depleted by starting production. Then, a service time is assigned for the production activity which spans Nodes 85 and 186. When Node 186 is realized, a transaction is sent to Queue Node 86 signifying that a new Y has been produced and put into inventory.

2.2.2 THE JOB WORKCENTER

Workcenter 4 performs job-shop type operations on components F, J, L or M. The simulation model for job workcenter 4 is illustrated in Figure 2-4. Orders for finished goods at workcenter 4 come from workcenters 5 and 6. Each of these is represented by a transaction being placed in Queue Node 40. Transactions representing the workcenter's finished goods inventory are stored in Queue Node 46. Attribute 1 identifies the type of job the transaction represents, i.e, whether the component to be processed is an F, J, L or M. When transactions that have the same attribute 1 value are in Queue Nodes 40 and 46 at the same time, they are matched

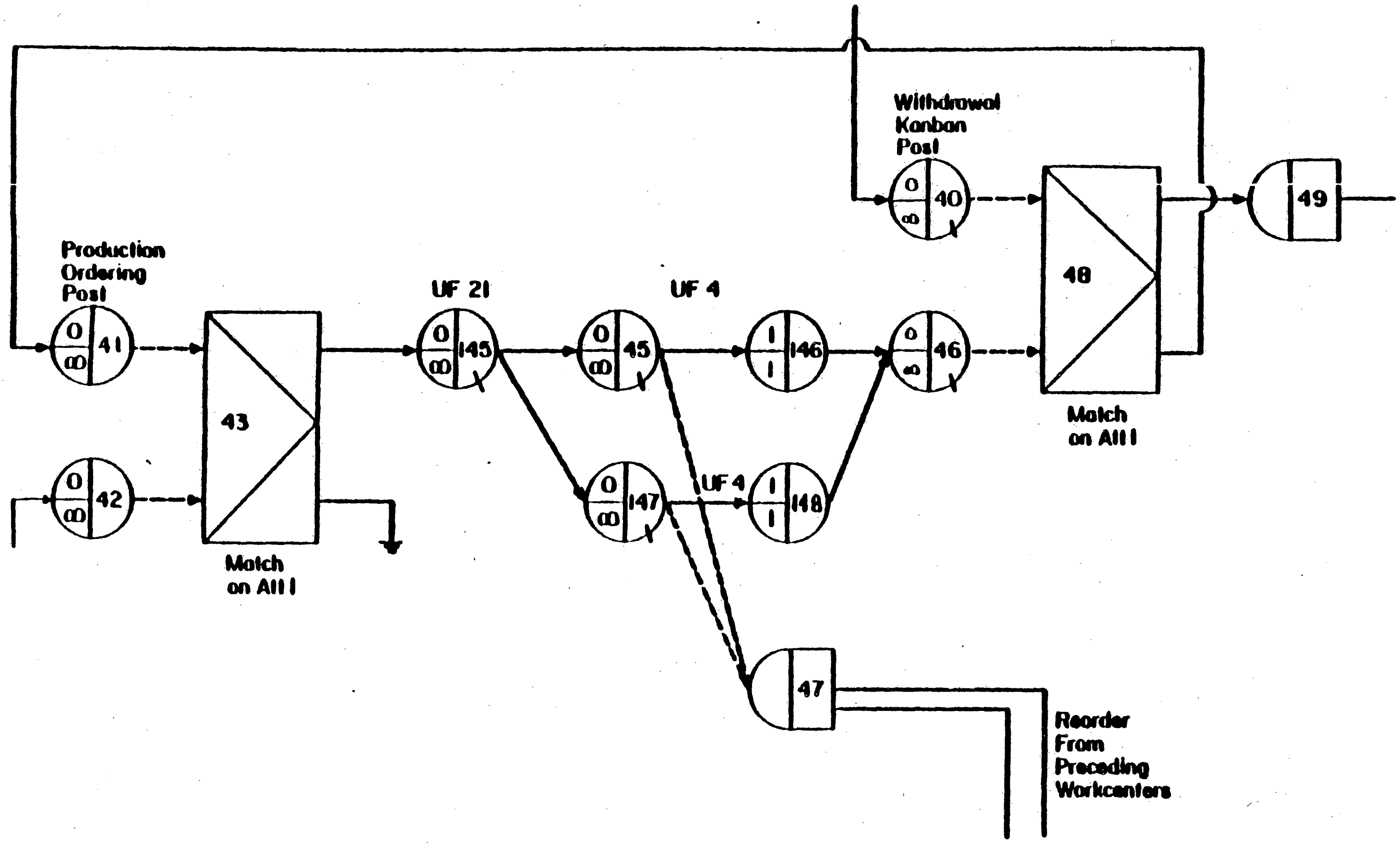


Figure 2-4. Job Workcenter 4

by Match Node 48. A transaction representing the filled order is sent to the ordering workcenter. Another transaction is sent to Queue Node 41, the production ordering post. The transactions in Queue Node 42 represent the pre-process materials for workcenter 4. When transactions that have the same attribute 1 value are in Queue Nodes 41 and 42 at the same time, a transaction is placed in Node 145. When Node 145 is realized, User Function 21 is called. Here it is determined which of the workcenter's two servers will get the job finished quicker using a minimum throughput rule. If the first server is chosen, the transaction is sent to Queue Node 45 by placing a 1 in attribute 2. Likewise, a 2 is placed in attribute 2 if the transaction is to be sent to Queue Node 147. Node 145 branches on attribute 2, so the transaction is sent to the proper queue. When the server becomes available, User Function 4 is called. This causes the pre-processed material that is being used to be replaced by putting a transaction in Node 47. From there, an order is sent to the appropriate preceding workcenter. Then, User Function 4 assigns the service time from the appropriate distribution. When the activity is completed, Node 146 or 148 (depending on which server handles the transaction) is realized. Then, a transaction is sent to Queue Node 46. This represents the finished goods inventory at workcenter 4.

2.3 SUMMARY

The example shop, the product structure and the Q-GERT simulation model used throughout the next three chapters have been presented. Shop parameters, such as processing times and setup times, vary throughout the investigation and are described in the following chapters for each set of simulation runs.

3.0 DETERMINING WORKCENTER LOTSIZES FOR SIGNAL KANBANS

Japanese experiences have implied that small setup times relative to processing times are essential in a JIT production system with Kanbans. Without such small setup times large bottlenecks at workcenters result and the production operation becomes hopelessly delayed. The inability to significantly reduce setup times prohibits many American firms from attempting to use the JIT technique with Kanbans especially if the production system is a non-repetitive job shop-type operation rather than an assembly line-type operation. American firms are often unable to make the necessarily large investment in new machinery or extensive worker training required to reduce setup times. However, the Toyota Company [36] [37] has developed a novel means for employing the JIT technique with Kanbans for an operation that encompasses several workcenters that have large setup times relative to processing times (such as a forging process or punch process). In such a system a special type of Kanban, referred to as a "signal Kanban" is used in conjunction with the workcenter or process that has relatively large setup times. In effect, a signal Kanban triggers the production of larger than normal lots at workcenters with large setup times within a JIT framework, wherein standard Kanbans at normal workcenters concurrently trigger the production of containers encompassing only a very small number of units, hopefully only one.

This altered approach developed by the Japanese for employing the JIT system using both standard and signal Kanbans offers an alternative for American firms who desire to implement a JIT system with Kanbans, but, are unable to reduce setup times at selected workcenters to a feasible level, at least initially. However, this altered version of the traditional JIT system requires that appropriate lotsizes be determined to be used in conjunction with the signal Kanbans. Failure to achieve effective lotsizes can result in large backorders and the inability to meet demand. For reasons that will be presented later in this chapter, the traditional multi-product EOQ lotsizing approach does not always work for signal Kanbans. As such, the purpose of this chapter is to demonstrate the use of integer mathematical programming for determining the optimal lotsizes to be used in conjunction with signal Kanbans in a JIT system. Specifically two integer programming model versions are developed and solved, one of which minimizes inventory at a workcenter while the other model version minimizes inventory and setup costs. Both models encompass constraints that prohibit backorders and require demand to be met. Preceding the development of these two integer programming models, the multi-product EOQ lotsizing approach is tested using a Q-GERT simulation model of an example shop. This same simulation model and example is subsequently used to test the two integer programming lotsize models, and, to explore some of the characteristics that an American firm might encounter in the implementation of a JIT system with both signal and standard Kanbans.

3.1 THE JUST-IN-TIME SYSTEM WITH SIGNAL KANBANS

The Japanese control the stage-to-stage authorization of container production with two "Kanbans," which are simply cards. One card, called a production Kanban, accompanies the containers as they are being produced, as shown in Figure 3-1a. Looking specifically at stage N in Figure 3-1a, when the production of a container is completed and demand from the succeeding stage (N-1) occurs (as indicated by a withdrawal Kanban from stage N-1), the production Kanban is removed from the container and is returned to the production-ordering Kanban post at the same stage (N). The withdrawal Kanban from stage N-1 actually replaces the production Kanban on the container, and it accompanies the container to stage N-1. For production activity to take place at stage N, both a production Kanban and a container of the required parts accompanied by a withdrawal Kanban must be present at that stage. The production Kanban subsequently replaces the withdrawal Kanban, and the withdrawal Kanban is sent back to stage N+1 where it authorizes stage N+1 (the preceding stage) to produce another container now required

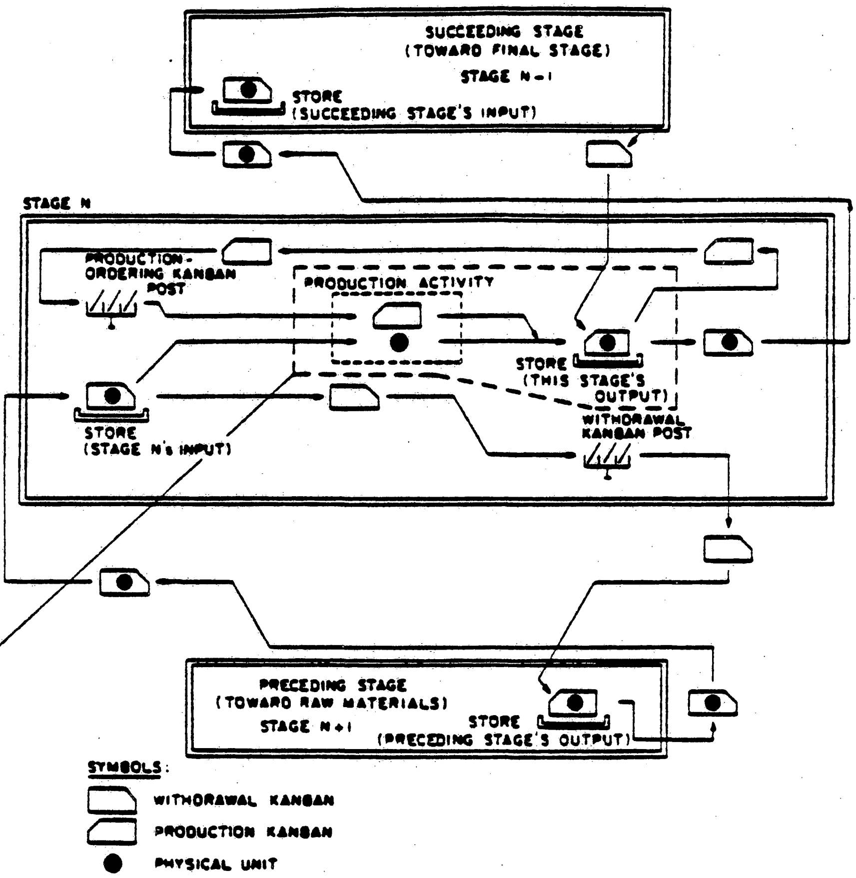


Figure 3-1a. The Normal Kanban Operation.

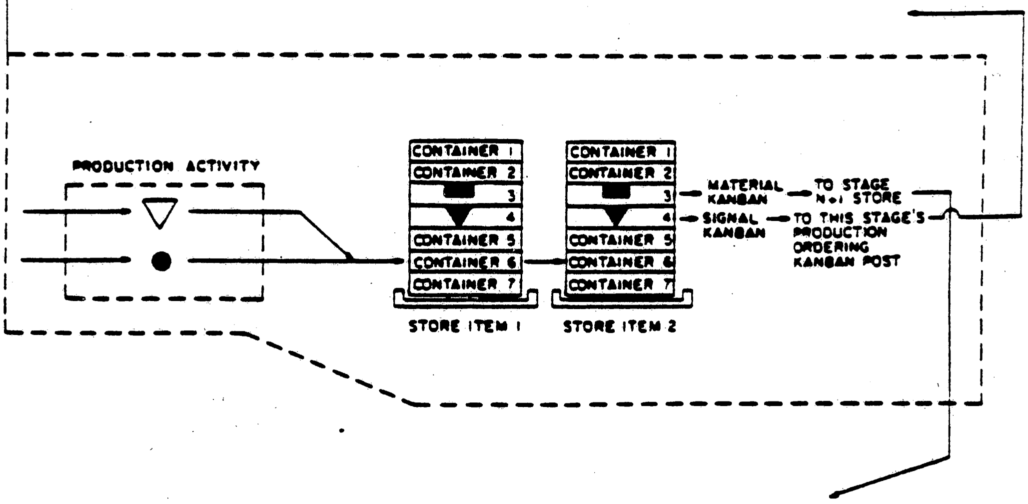


Figure 3-1b. The Kanban Operation with a Signal Kanban.

at stage N. This creates a continuous cycle of container movement between the stages.

In effect, the Kanbans "pull" the containers through the production system "just-in-time" to meet demand at each production stage, thus minimizing in-process inventories. In this process two Kanban swaps are made; one immediately prior to the production activity and one immediately following the production activity at each stage. The production Kanban never leaves its "home" stage while the withdrawal Kanban moves between stages. The production Kanban acts as an intraprocess control apparatus and the withdrawal Kanban serves as the interprocess control apparatus.

The Japanese success with the JIT system with Kanbans is attributable to several factors, most prominent of which are the small setup time, highly skilled and trained Japanese worker, and job automation. These factors enable Japanese companies to achieve small and almost constant processing times with very little variability. Such scheduling systems are further enhanced by extremely close cooperation between the supplier, manufacturer and customer that minimizes variability in input rates and demand schedules.

When the production system consists almost exclusively of workcenters at which small setup times (relative to processing times) cannot be achieved, such as exists in many American job shops, large bottlenecks occur at these workcenters in the JIT system, which in effect, causes the system to collapse. However, in batch-mode production that encompasses a number of workcenters with small setup times and a limited number of workcenters with large setup times, the JIT system can still be employed with the use of "signal" Kanbans.

The JIT system with a signal Kanban is illustrated in Figure 3-1b, which is actually an overlay of the corresponding section in Figure 3-1a. (The signal Kanban system employed in this chapter is more efficient than several other alternative signal Kanban systems available. The interested reader can see Monden [36] for an alternative approach.) The workcenter depicted in Figure 3-1a produces two items and the setup times for the two items are large relative to the items' processing times. In this scenario, both stages N-1 and N+1 operate with normal Kanbans while the workcenter at stage N (in Figure 3-1b) employs a signal Kanban. For example purposes this workcenter is shown with two "stores" for items 1 and 2, respectively, each containing a lotsize of seven containers.

Since stage N-1 is a standard Kanban workcenter, when a container of item 1 is needed at stage N-1, a withdrawal Kanban is sent to stage N and the top container (labelled number 1) is taken. Now six containers exist at the store for item 1 in stage N. In the standard Kanban process when a withdrawal is made at N, a production Kanban would be sent back to the beginning of stage N to trigger the production of another container of item 1, (which would create setups as production switches for each container from one item to the other). However, when using a signal Kanban no production Kanban is sent back to the beginning of stage N -- yet.

As stage N-1 continues to send withdrawal Kanbans for item 1 to stage N, the supply of containers from the store of item 1 will continue to be depleted until the container with the "material" Kanban denoted by the rectangular designation (▭) is reached. When this material Kanban is reached a production Kanban is still not released, but this particular Kanban is a warning that the lot is running low, and, a withdrawal Kanban is sent back from stage N to stage N+1 for a container of the required item produced at stage N+1. Although, when this container reaches stage N there will be no production Kanban to trigger production, there will be a container ready for the imminent arrival of the signal Kanban which initiates production of a lot.

The next container withdrawn from the store at stage N contains the signal Kanban designated by an inverted triangle (∇). As this container is withdrawn, the signal Kanban is sent to the beginning of stage N to "signal" the initiation of the production of a lot. Since required items were previously ordered from stage N+1, production can start immediately. However, note that there is only one container of in-process items at the beginning of stage N. The signal Kanban also authorizes the production of multiple containers, in this case a lotsize of seven containers, at stage N+1. As soon as production begins at stage N, a withdrawal Kanban is sent to stage N+1 for the next container and the normal Kanban process operates between stages N and N+1 until the lotsize of seven containers is completed.

As a result of this large lot, it is unnecessary to make a production run for each withdrawal Kanban from N-1, thus eliminating setups. Although the use of such buffer inventory lots is inconsistent with the Japanese philosophy of inventory reduction embodied in the JIT system, it is an operational compromise that enables a company that does not totally meet the requirements for JIT implementation to reap some of the benefits of the JIT system.

A crucial aspect in the implementation of signal Kanbans is the determination of the lotsize used in conjunction with the signal Kanban. A lotsize larger than necessary will needlessly increase inventory costs, thus offsetting the objective of employing the JIT system, while a lotsize too small will incur excessive setup costs and may create backorders that will ultimately cause a JIT system to completely collapse. As such, the purpose of this chapter is to demonstrate how an integer mathematical programming model can be developed that will determine the optimal lotsize to be used in conjunction with a signal Kanban in a JIT system. However, prior to proceeding with this model development, a case scenario of an example shop using a JIT system will be described in order to test the classical EOQ approach to lotsizing, and, to subsequently test the mathematical programming modeling approach.

3.2 CASE EXAMPLE AND SIMULATION MODEL OF A SHOP OPERATION

The case example that will be employed to analyze the lotsizing models encompasses a production operation with six workcenters that produce two products, X and Y. The product structure for each product is shown in Figure 2-1. The configuration of the six workcenters in the example shop is illustrated in Figure 2-2. This shop has been given the characteristics of both a job shop and an assembly shop in order to explore the feasibility of the signal Kanban system in the broadest context possible. At workcenters 1, 2 and 4, which have the characteristics of a job shop, a variety of operations is performed. However, workcenters 5 and 6 differ from the other workcenters in several respects. They only produce one product apiece so they never incur setups, and, these two workcenters have only one machine whereas the other four workcenters each have two machines.

Of the four workcenters that perform operations requiring setups only workcenter 1 has high setup times relative to container processing times. Workcenter 1 requires 0.5 hours to setup for any of the four items, A, B, C and D, produced at that workcenter. Container processing time is deterministic at workcenter 1 and is approximately 9 minutes (0.15 hours) per container on either of the two (identical) machines. The daily demand for end items X and Y and all in-process items (A through L) as well as model cost parameters are shown in Table 3-1. The setup times for all shop workcenters as well as container processing times for all items are shown in Table 3-2.

As a prelude to the analysis of the JIT system with signal Kanbans, a simulation experiment was conducted using this simulation model to attempt to ascertain if the shop would work with standard Kanbans. (This required some adjustments in the basic shop simulation model.) As part of this simulation analysis, the shop was run using the JIT system with standard Kanbans and it did not work. Extremely large backorders resulted, especially at workcenter 1, that created excessively long delays and demand not being met. After extended run time the shop effectively "collapsed."

Given that the JIT system with standard Kanbans will not work with our example shop, a JIT system with signal Kanbans appears as a workable alternative. In this modified JIT system the signal Kanban will be employed in conjunction with workcenter 1 where large setup times are experienced. However, a crucial decision accompanying the use of signal Kanbans is the lotsizes to be employed, and specifically in the case of our example, the lotsizes at workcenter 1. As such, the next portion of this chapter will be directed at

Table 3-1. Cost and Demand Parameters for Example Shop

Item	Container Cost ($)	Daily Demand (Containers)
A	1500	30
B	1500	20
C	1500	20
D	1500	20
E	4125	10
F	375	20
H	375	30
I	3000	10
J	4500	10
K	375	30
L	375	20
X	9375	10
Y	7875	10

Holding cost = 25% per year of container cost
Backorder cost = $5.63 per container per hour
Setup cost = $65 per hour

Table 3-2. Workcenter Processing and Setup Times.

Workcenter	Container Processing Time (Hours)	Setup Time (Hours)
1	.150	.50
2	.135	.01
3	.225	.01
4	.170	.01
5	.760	N/A
6	.760	N/A

the development of an integer linear programming model for determining the optimal lotsizes to be used in conjunction with signal Kanbans.

3.3 AN INTEGER PROGRAMMING MODEL FOR DETERMINING LOTSIZES

3.3.1 THE MULTI-PRODUCT EOQ MODEL

Prior to actually developing the mathematical programming model for determining lotsizes for signal Kanbans in a JIT system the feasibility of using EOQ analysis for determining lotsizes will be explored. A slightly modified version of the following multi-product EOQ formula for single machines with no backorders [5] will be used to determine the lotsizes at workcenter 1 for two machines:

$$Q_j = \frac{R_j}{\sqrt{\dfrac{\sum_{j=1}^{n} C_{H_j} R_j (1 - d_j PT_j)}{\sum_{j=1}^{n} C_j}}}$$

where, C_{H_j} is the holding cost for item j

C_j is the order (setup) cost for item j

d_j is the demand rate for item j

PT_j is the processing time per unit of item j

R_j is the annual demand for item j

j = item number (at workcenter 1, j = A, B, C, D)

n = number of items processed at workcenter 1.

This particular EOQ formulation is employed in order to eliminate backorders which, if allowed, would create delays throughout the shop that used items produced at workcenter 1. The economic order quantity (lotsizes) for each item produced at workcenter 1 is, as follows: $Q_A = 56$ containers and $Q_B = Q_C = Q_D = 38$ containers. The signal Kanbans used in conjunction with each of the four items produced at workcenter 1 are processed on a first-come, first-served basis.

The simulation model of the shop with these EOQ lotsizes resulted in 40 total containers backordered after 21 weeks. At this point in time with this number of backorders, the shop was two full days behind schedule (for meeting demand) which was viewed to be unacceptable. Thus, the determination of signal Kanban lotsizes using the classical EOQ formula was not felt to be a feasible procedure.

This was a result that could have been logically anticipated. First, note that the EOQ formulation contains setup costs but not setup times. Next let us define the "workcenter production cycle" as the minimum time period required to produce the EOQ determined lotsize of each item processed or assembled at the workcenter. At workcenter 1, the quickest way to produce the four items, A, B, C and D, on the two available machines is to produce A on one machine and then either B, C or D on the same machine. If we select B such that the first machine produces A and then B, this machine will require two setups per cycle and the time to process A and B on machine 1 will be,

$$t_{\text{machine 1}} = \sum_{j=1}^{n} (Q_j^* \text{ x processing time + setup time})$$

$$= (56)(.150) + (38)(.150) + 2(0.5)$$

$$= 15.1 \text{ hours}$$

The processing time for items C and D on machine 2, computed similarly, is $t_{\text{machine 2}}$ = 12.4 hours. However, demand for item A (from Table 3-1) is 3.75 containers per hour. Since Q_A, the EOQ determined lotsize, is 50 containers, the items will be depleted in 13.33 hours, which means that this workcenter will never be able to meet demand and will always be behind schedule.

Faced with such results, the natural inclination might be to attempt to adjust the multi-product EOQ model to explicitly include feasibility considerations, such as setup times. This is not as simple as it might initially appear since the number of setups in a cycle will depend on the machine selection rule and the particular sequence of jobs arriving at each machine. The sequence of jobs, in turn, depends on end-item demand and the particular synchronization of the rest of the shop up to the present instant in time.

Rather than attempt to modify the multi-product EOQ model, an integer mathematical programming model will be developed in which the model constraints guarantee no backorders for a workcenter employing signal Kanbans. The math programming model will be developed using two alternative objective functions, one which minimizes in-process inventory (reflecting the Japanese philosophy of driving inventory levels to zero), and, an objective function that minimizes total (i.e., inventory plus setup) costs.

3.3.2 MODEL CONSTRAINTS

When developing a model that will insure that no backorders are incurred at workcenters in which signal Kanbans are used, the shop workcenters may be "decoupled" so each can be considered independently of the other. In a shop employing the JIT system with one or more signal Kanban workcenters, in general, it is not possible to decouple the workcenters and consider each separately if any of the workcenters incurs backorders. This is because under the JIT system, a workcenter must wait for component parts from preceeding workcenters, and, cannot begin processing until authorization for outbound work-in-process is received. All workcenters are integrally tied to each other and if one is delayed and falls behind schedule, other workcenters may be affected.

However, it is our purpose to develop a mathematical programming model with constraints that prohibit backorders altogether. If there are no backorders in the entire shop and in-process items (or raw materials) are available "in time" at those workcenters needing them, any given workcenter will have all of its needed pre-process items arriving "just-in-time" for processing, and, will have the outbound in-process items ready "just-in-time" to be forwarded to workcenters demanding these items. Consequently, all workcenters will operate as though they are isolated from each other; pre-process items will always be available when needed, and demand for a workcenter's processed items will simply be a delayed image of the end-item demand. Therefore, constraint equations that guarantee that a signal Kanban workcenter will incur no backorders need not encompass variables that describe or define characteristics of any other workcenter as long as we do not allow all other workcenters to have backorders. Thus, developing the constraint equations will be greatly simplified since it will not be necessary to consider any other workcenter except the one for which we are developing constraints.

There are three classes of constraints that must be satisfied to insure that a workcenter with signal Kanbans will have no backorders. First, the workcenter production cycle time for each machine must be no smaller than the time it takes to complete all work for that machine, i.e., the cycle time for each machine must be greater than or equal to the production time, including setups. This constraint is formulated mathematically for a workcenter with signal Kanbans as,

$$t_i \geq \sum_{j=1}^{n} (q_{ij}PT_j + Y_{ij}S) \quad \text{for } i = 1 \text{ to } m \text{ machines}$$

where, n = number of total items to be produced at a workcenter

q_{ij} = lotsize in containers for item j processed on machine i

t_i = the production cycle time for the ith machine at the signal Kanban workcenter

PT_j = the processing time per unit of item j

S_j = the setup time (assumed to be the same for all items at the workcenter in our example)

$$Y_{ij} = \begin{cases} 1 \text{ if } q_{ij} > 0 \\ 0 \text{ if } q_{ij} = 0 \end{cases}$$

Strictly speaking, this equation only holds if lot production is performed for more than one product on a machine. If only one product is being processed on a machine, then the product should be processed with standard Kanbans since no setups are incurred.

The second constraint forces each item produced at a workcenter to be produced on one machine; this will reduce time-consuming and costly setups. That is, once it is decided to produce a lot of an item on a machine, then all the demand for that item during that machine's production cycle time will be produced on that machine. This constraint is written mathematically as follows:

$$\sum_{i=1}^{m} Y_{ij} = 1 \quad \text{for } j = 1 \text{ to } n \text{ items}$$

$$q_{ij} \leq MY_{ij} \quad \text{for all } i \text{ and } j,$$

where all q_{ij} are integer,

Y_{ij} is binary (as above), and

M is a very large positive constant.

The third and final constraint states that for each machine, demand for each item produced on that machine during its production cycle must exactly equal the lotsize for that item. This constraint is expressed mathematically as a pair of constraints for each item and machine:

$$\left.\begin{aligned} Q_j &\leq d_j t_i + (1 - Y_{ij})M \\ Q_j &\geq d_j t_i - (1 - Y_{ij})M \end{aligned}\right\} \quad \text{for all } i \text{ and all } j.$$

where $Q_j = \sum_{i=1}^{m} q_{ij}$ (the sum of lotsizes for all machines at a workcenter), for all j,

and d_j = the demand (containers) per unit time for item j.

Consider the following two cases with respect to the above two equations.

(1) if item j is produced on machine i, then $q_{ij} > 0$ and hence $Y_{ij} = 1$.

In this case, the equations may be rewritten as

$Q_j \leq d_j t_i$

$Q_j \geq d_j t_i$,

which, taken together, are equivalent to

$Q_j = d_j t_i$.

Therefore, if item j is produced on machine i, an equality constraint holds.

(2) if item j is not produced on machine i, then $q_{ij} = 0$ and hence $Y_{ij} = 0$.

In this case, the equations may be rewritten as

$Q_j \leq d_j t_i + M$

$Q_j \geq d_j t_i - M$.

Since M is a huge positive constant, these constraints are always trivially satisfied for any values of Q_j d_j, and t_i.

Therefore, if item j is not produced on machine i, there is no third, final constraint to be satisfied.

When the three constraints presented in this section are simultaneously satisfied, no backorders will occur at a signal Kanban workcenter.

3.3.3 THE INVENTORY MINIMIZATION MODEL

The constraints developed in the preceeding section, while guaranteeing that demand will be met with no backorders, do not necessarily result in a cost effective lotsize. There are several objectives that can be employed that will result in an effective or "optimal" lotsize, and in the following discussion we will offer two of them, inventory minimization and cost minimization.

The first objective to be examined is the minimization of inventory which is consistent with the philosophy inherent in the JIT technique to reduce inventory to the absolute minimum level. The approach presented here is a generalization of the approach reported by Monden [36] that is used by Toyota. This objective function combined with our previously defined model constraints results in the following integer mathematical programming model.

$$\text{minimize } Z = \sum_{j=1}^{n} Q_j$$

subject to

$$t_i \geq \sum_{j=1}^{n} (q_{ij}PT_j + Y_{ij}S), \qquad \text{for } i = 1 \text{ to } m.$$

$$\sum_{i=1}^{m} Y_{ij} = 1, \qquad \text{for } j = 1 \text{ to } n.$$

$$q_{ij} \leq MY_{ij}, \qquad \text{for } i = 1 \text{ to } m \text{ and } j = 1 \text{ to } n$$

$$Q_j \leq d_j t_i + (1 - Y_{ij})M,$$

$$\text{for } i = 1 \text{ to } m \text{ and } j = 1 \text{ to } n$$

$$Q_j \geq d_j t_i - (1 - Y_{ij})M,$$
$$\text{for } i = 1 \text{ to } m \text{ and } j = 1 \text{ to } n$$

$$Q_j = \sum_{i=1}^{m} q_{ij}, \quad \text{for } j = 1 \text{ to } n.$$

Y_{ij} is binary (0 or 1)

for $i = 1$ to m and $j = 1$ to n

q_{ij} and Q_j are integer

for $i = 1$ to m and $j = 1$ to n

$t_i \geq 0$ for $i = 1$ to m.

Recall the example shop we described previously with four items produced at workcenter 1 (A, B, C and D) on two machines. Workcenter 1 was the only one of the six workcenters in the shop that required the use of signal Kanbans; thus, we will develop one integer programming model for this single workcenter. For convenience and standardization of notation we will redefine items A, B, C and D as 1, 2, 3 and 4 respectively.

minimize $Z = Q_1 + Q_2 + Q_3 + Q_4$

subject to

$$t_1 \geq \sum_{j=1}^{4} (0.15q_{1j} + 0.50Y_{1j})$$

$$t_2 \geq \sum_{j=1}^{4} (0.15q_{2j} + 0.50Y_{2j})$$

$Y_{11} + Y_{21} = 1$
$Y_{12} + Y_{22} = 1$
$Y_{13} + Y_{23} = 1$
$Y_{14} + Y_{24} = 1$
$q_{11} \le MY_{11}$
$q_{12} \le MY_{12}$
$q_{13} \le MY_{13}$
$q_{14} \le MY_{14}$
$q_{21} \le MY_{21}$
$q_{22} \le MY_{22}$
$q_{23} \le MY_{23}$
$q_{24} \le MY_{24}$
$Q_1 \le 3.75t_1 + (1 - Y_{11})M$
$Q_1 \ge 3.75t_1 - (1 - Y_{11})M$

$Q_1 \le 3.75t_2 + (1 - Y_{21})M$
$Q_1 \ge 3.75t_2 - (1 - Y_{21})M$

$Q_2 \le 2.50t_1 + (1 - Y_{12})M$
$Q_2 \ge 2.50t_1 - (1 - Y_{12})M$

$Q_2 \le 2.50t_2 + (1 - Y_{22})M$
$Q_2 \ge 2.50t_2 - (1 - Y_{22})M$

$Q_3 \le 2.50t_1 + (1 - Y_{13})M$
$Q_3 \ge 2.50t_1 - (1 - Y_{13})M$

$Q_3 \le 2.50t_2 + (1 - Y_{23})M$
$Q_3 \ge 2.50t_2 - (1 - Y_{23})M$

$Q_4 \le 2.50t_1 + (1 - Y_{14})M$
$Q_4 \ge 2.50t_1 - (1 - Y_{14})M$

$Q_4 \le 2.50t_2 + (1 - Y_{24})M$
$Q_4 \ge 2.50t_2 - (1 - Y_{24})M$

$$Q_1 = q_{11} + q_{21}$$
$$Q_2 = q_{12} + q_{22}$$
$$Q_3 = q_{13} + q_{23}$$
$$Q_4 = q_{14} + q_{24}$$

$Y_{ij} = 0$ or 1 for $i = 1$ to 2 and $j = 1$ to 4

q_{ij}, Q_j integer for $i = 1$ to 2 and $j = 1$ to 4

$t_1, t_2 \geq 0$.

The solution to this example model is,

$Q_1 = q_{11} = 60$ containers

$q_{21} = 0$

$Q_2 = q_{12} = 40$ containers

$q_{22} = 0$

$Q_3 = q_{23} = 10$ containers

$q_{13} = 0$

$Q_4 = q_{24} = 10$ containers

$q_{14} = 0$

$t_1 = 16$ hours

$t_2 = 4$ hours

Using our original example shop notation, this solution indicates that the signal Kanban lotsize for item A at workcenter 1 is 60 containers while the lotsize is 40 for B and 10 containers each for C and D. Note that these lotsizes are significantly higher than one might consider for a normal Kanban workcenter. This solution also indicates that items A and B will be processed on machine 1 with a production cycle time of 16 hours for these two items. Items C and D will be processed on machine 2 with a production cycle time of 4 hours. Machine 1 satisfies 2-days' demand during its two-day cycle, and Machine 2 satisfies one-half day's demand during its one-half day cycle.

3.3.4 THE COST MINIMIZATION MODEL

The mathematical programming model developed in the preceeding section minimized inventory and ignored the direct minimization of cost. This second version of the mathematical model will consider cost minimization as an objective and is based upon a "rotation cycle policy" [23]. This model encompasses both holding and setup cost in the determination of an optimal Q^*_j. The objective function now becomes

$$\min \sum_{j=1}^{n} C_j \frac{R_j}{Q_j} + C_{H_j}\left[\frac{Q_j}{2} (1 - d_j PT_j) \right]$$

This function stipulates that the sum of holding and setup costs for all items at the workcenter is minimized. To this new objective function we add all the constraints developed to prevent backorders for the previous mathematical programming model that minimized inventory.

The optimal solution to this model for our example shop is $Q_A = 60$, $Q_B = 40$ (both processed on machine 1), $Q_C = Q_D =$ 38 containers (both processed on machine 2). Machine 1's production cycle time is 16 hours and machine 2's is 15.2 hours.

The two mathematical programming models were subsequently solved for a variety of different setup times at workcenter 1. The different lotsizes obtained from these setup times for workcenter 1 are shown graphically in Figure 3-2.

The results in Figure 3-2 for our example shop show that the minimum inventory integer programming model generates consistently lower lotsizes than the minimum cost model for all setup times. This is not unexpected since the objective of the minimum inventory model is to minimize lotsizes. In order to analyze the implications of these models on shop costs it is necessary to employ the simulation model developed previously.

3.4 SIMULATION ANALYSIS OF MODEL LOTSIZES

Recall that total cost in our shop is the sum of inventory, backorder and setup costs. Since both mathematical programming models eliminate backorders, total cost is simply the sum of inventory and setup costs. When the various lotsizes shown in Figure 3-2 are substituted into the simulation model of the example shop the total cost values at workcenter 1 shown in Figure 3-3 are generated. Notice that while the "minimum cost" model produced the higher lotsizes in Figure 3-2 this same model resulted in the lower cost in Figure 3-3. This result occurs because the larger lotsizes

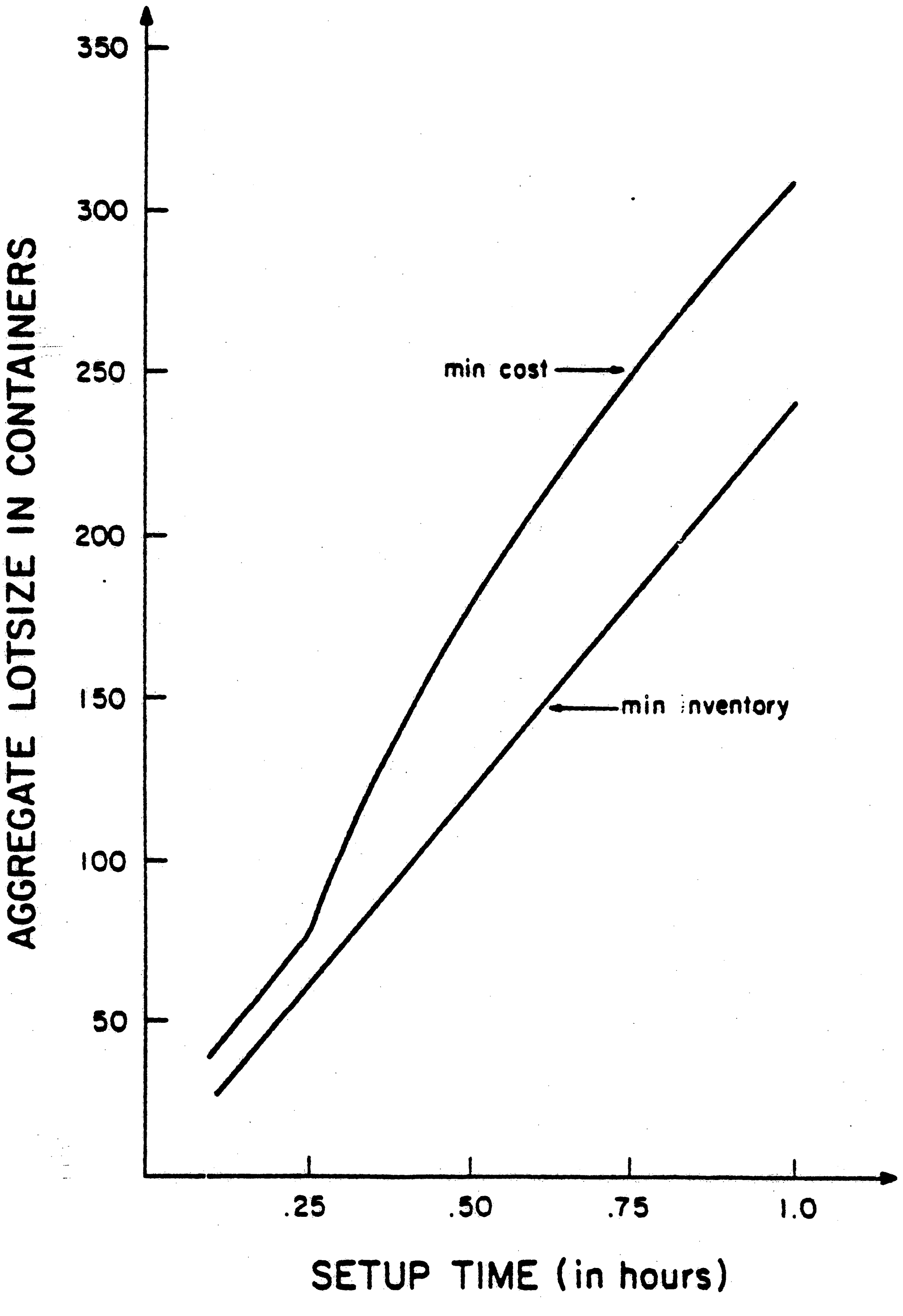


Figure 3-2. The Effect of Setup Time on Lotsize.

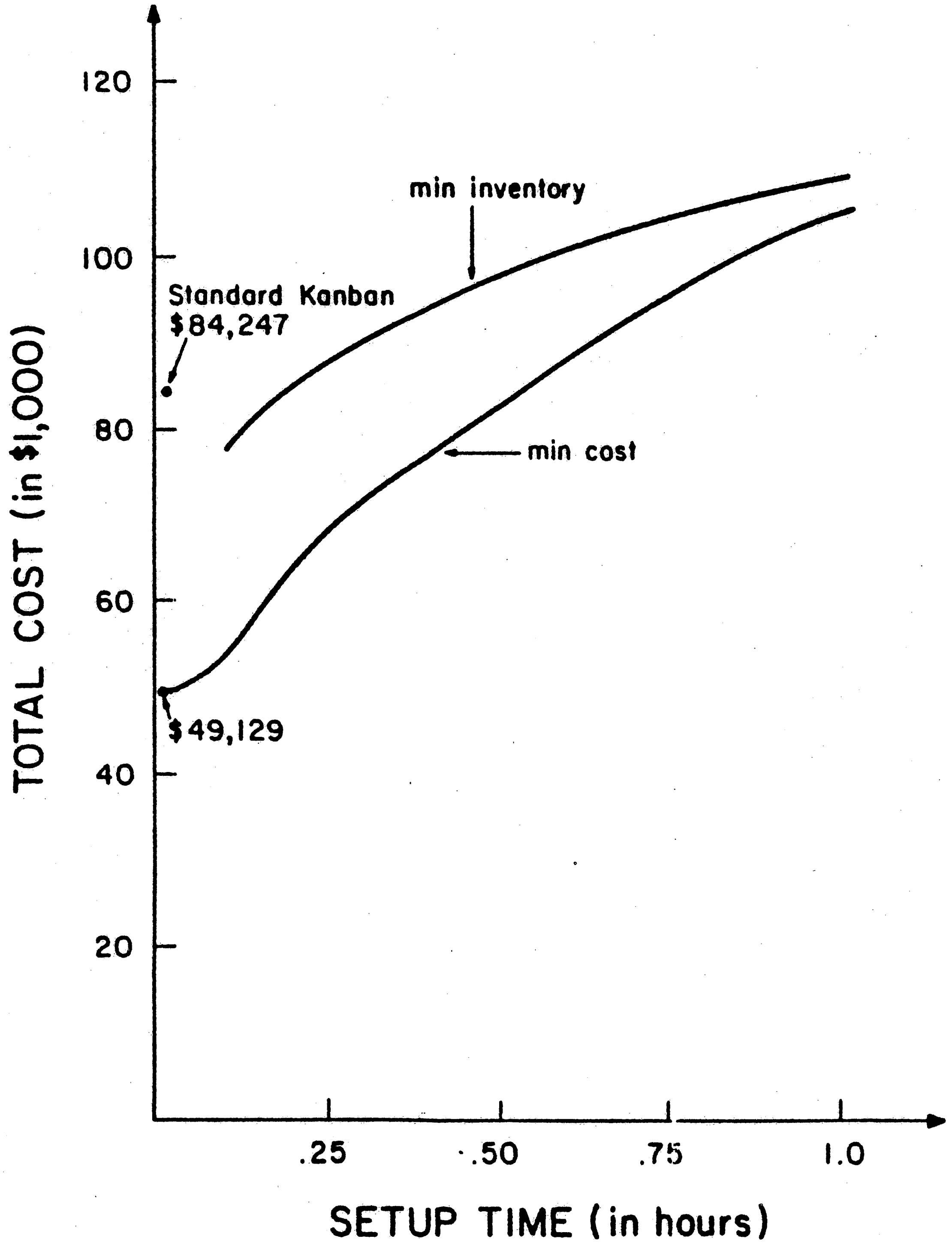


Figure 3-3. The Effect of Setup Time on Total Cost.

generated by the "minimum cost" model reduce the number of setups, and in this particular example shop, costs are very sensitive to setups.

While it is inappropriate to develop any conclusions from only one data point (i.e., the example shop) these results do have several implications for the interested manager contemplating the installation of a JIT system with signal Kanbans. Although the Japanese philosophy inherent in the JIT system emphasizes the minimization of inventory (frequently exclusive of cost considerations), it may be prudent for the manager to consider cost more prominently. From Monden's book [36, pp. 215-216] it appears that Toyota employs the minimum inventory model approach we have presented in this study. However, the results presented in Figure 3-3 indicate that a minimum cost approach is superior, at least for this particular example. In other words, there may be some penalty associated with a philosophy that focuses on inventory reduction without at least a glance at the cost implications. In order to explore this question further we will again employ our simulation model to analyze the impact of container processing times on the example shop.

3.5 SIMULATION ANALYSIS OF CONTAINER PROCESSING TIMES

In this experiment we will explore the impact of varying container processing times (with constant setup times) on workcenter lotsizes and costs. First, various lotsizes are generated from the two integer programming models. These results are graphically displayed in Figure 3-4. Notice from Figure 3-4 that, as in Figure 3-2, lotsizes for the minimum inventory model are consistently smaller than those for the minimum cost model when container processing times are varied for our example shop. It is interesting to note in Figure 3-4 the rather dramatic impact a slight reduction in container processing times has on lotsizes. At least for this example shop, for container processing times greater than approximately 0.15 hours, a slight decrease in processing times results in a very large decrease in lotsizes.

Employing the lotsizes obtained from the two integer programming models (with different container processing times) in the simulation model of the example shop generates the cost results shown graphically in Figure 3- 5. The minimum cost model dominates the minimum inventory model over the range of container processing times explored for our example.

It is interesting to note that contrary to expectations, for the "minimum inventory" model, costs increase as processing times decrease. The explanation for this occurrence is that, as container processing times are reduced, smaller lotsizes result which creates more setups. Thus, the cost curve shown for the "minimum inventory" model in Figure 3-5 is a result of high setup costs in our shop. With the cost parameters used in this shop example, the increased cost due

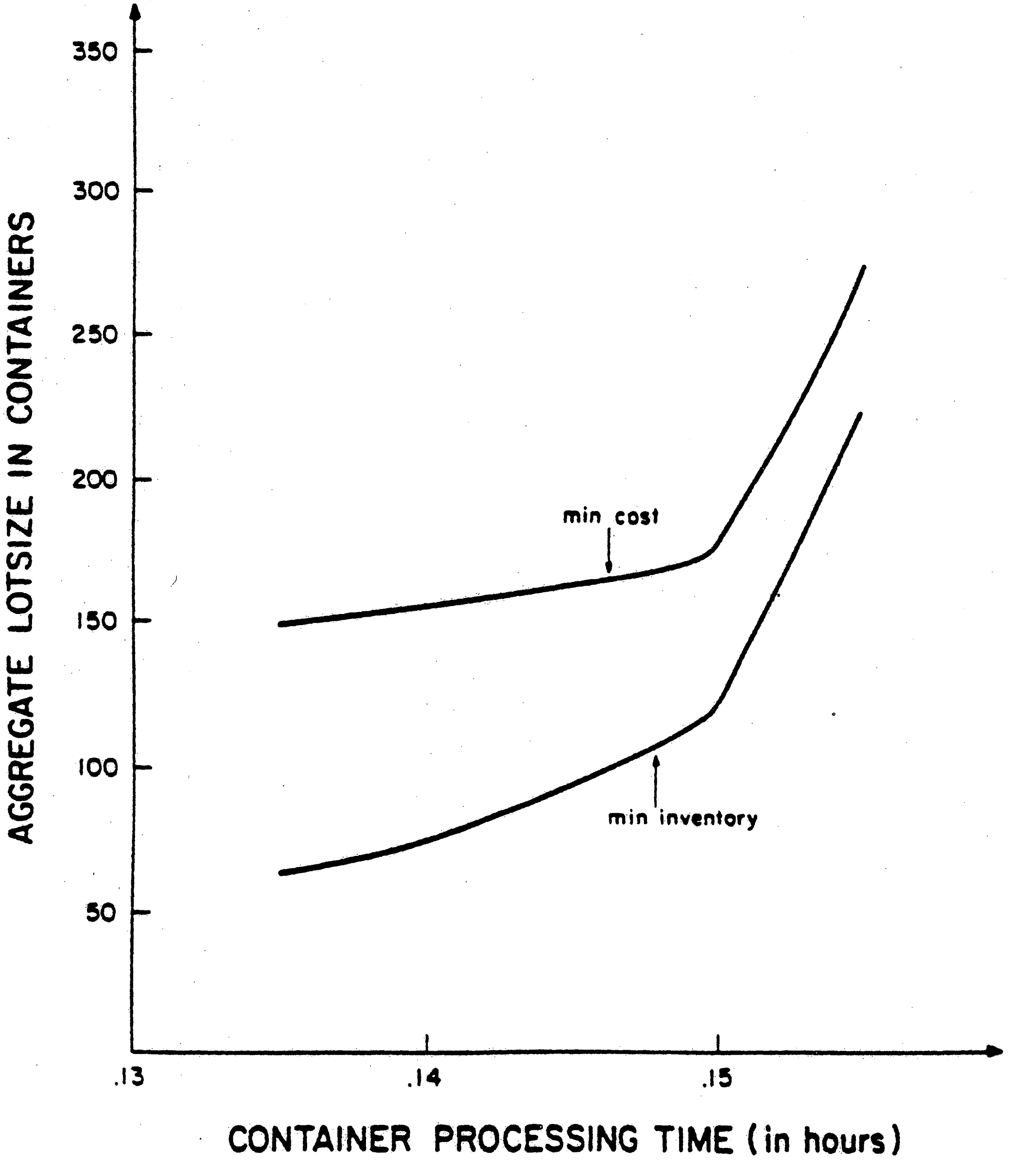


Figure 3-4. The Effect of Container Processing Time on Lotsize.

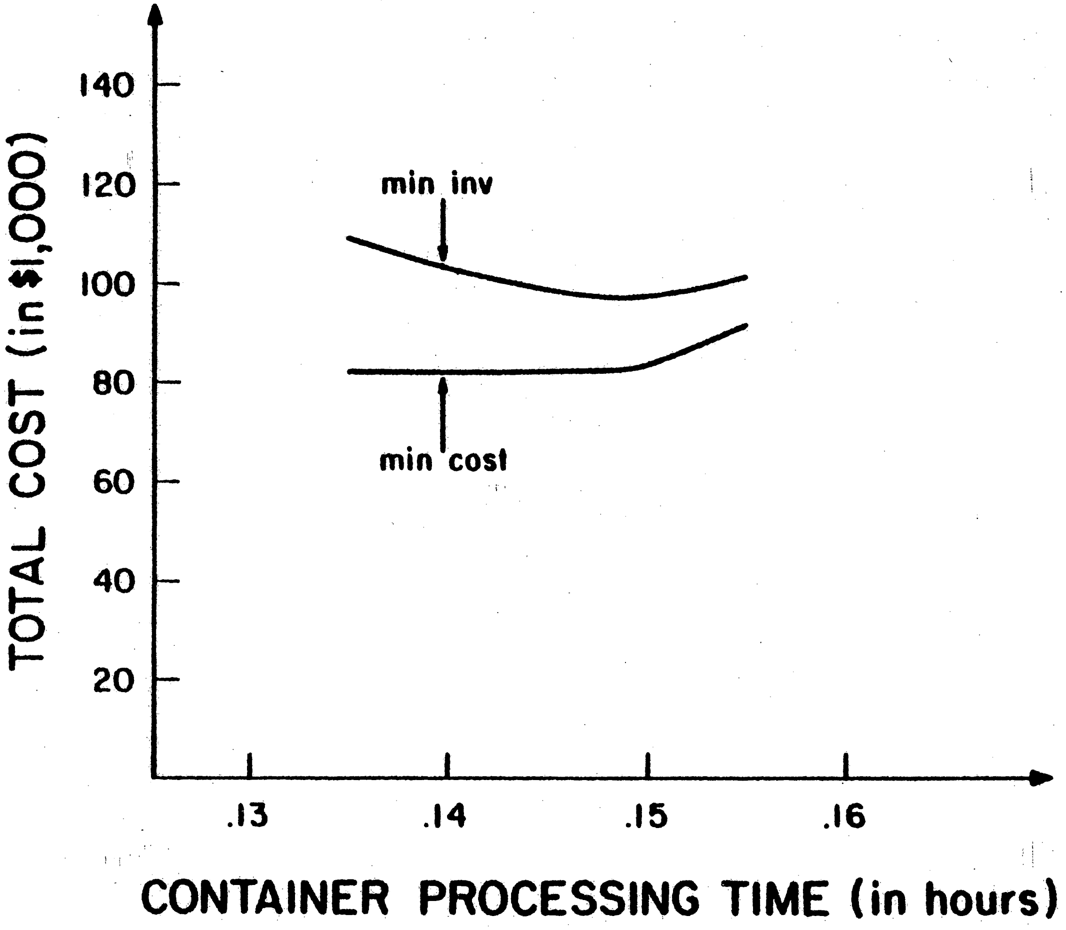


Figure 3-5. The Effect of Container Processing Time on Total Cost.

to setups is much greater than the cost savings for reducing inventory. Although this result is example dependent, it still indicates that it may be important for a manager to consider the cost implications of signal Kanban lotsizes prior to implementing a JIT system.

3.5.1 COMPARING SIGNAL AND STANDARD KANBAN SYSTEMS

In our final simulation model experiment we will compare the JIT system employed in our example shop using standard Kanbans and signal Kanbans at workcenter 1.

In order to operate the example shop using standard Kanbans, setup times at workcenter 1 must be .025 hours or less (a value determined by simulation model experimentation). The total cost at workcenter 1 for the shop with standard Kanbans (and setup times of .025 hours) is $84,247, a point designated in Figure 3-3. However, notice in Figure 3-3 that the total cost incurred at workcenter 1 using signal Kanbans with this same setup time is significantly lower ($49,129). The total cost with signal Kanbans at workcenter 1 increases as setup times increase, but remains below the standard Kanban value up until setup times of about 0.5 hours.

Although this result is based on just a single example, it does have significant implications. There is a tendency among managers implementing a JIT system to attempt to reduce setup times to the point where a standard Kanban system can be used. However, observing Figure 3-3, it can be seen that, at least for this example shop, if setup times are reduced to a point where standard Kanbans are feasible then signal Kanbans may be even more cost efficient than standard Kanbans. As such, it might be that a standard Kanban system should not always be the ultimate objective, and, that a signal Kanban system should be retained as setup times are reduced. Carrying this logic one step further, under certain circumstances and operational shop scenarios a JIT system with signal Kanbans at all workcenters might be more cost effective than a JIT system with standard Kanbans. At the very least, these results indicate that the implementation of a JIT system encompasses a number of interrelated and complex considerations that should be fully explored before the final system is operationalized.

3.6 SUMMARY

The purpose of this chapter has been to develop a model for determining the appropriate lotsizes to use in conjunction with signal Kanbans in a shop using the Just-in-Time technique with Kanbans. It was demonstrated via a simulation model of an example shop that the classical multi-product EOQ model does not always work in a JIT shop. As a result, two integer mathematical programming models for determining signal Kanban lotsizes and eliminating backorders were developed. One model version employed an objective function that minimized inventory while the other model version minimized cost. The impact of these results on total workcenter costs was determined using the simulation model of the example shop. These simulation results, while not providing any universal guidelines or conclusions, do offer several implications for the manager considering the implementation of a JIT system. In general, these implications are that it may be prudent to consider inventory and setup costs rather than simply seeking to reduce inventory to its minimum level in a JIT system; and, under certain conditions a signal Kanban system may be more cost effective than a feasible standard Kanban system.

The next chapter focuses on determining for an American environment which factors influence the number of Kanbans at a workcenter and how the initial number of Kanbans should be determined in such a setting.

4.0 AN ANALYSIS OF FACTORS INFLUENCING THE NUMBER OF KANBANS REQUIRED AT A WORKCENTER

The purpose of this chapter is to analyze two aspects of the JIT technique that confront a production manager implementing the JIT technique for the first time in a uniquely American production environment. First, the factors that influence the number of Kanbans required at workcenters in a JIT system will be identified and discussed. The impact of these factors on the efficiency of a JIT system will be demonstrated via a simulation model of an example production operation. Second, a methodology will be presented for determining the initial number of Kanbans to use in a JIT system. This methodology will also be demonstrated via a simulation model of a production system that does not generally exhibit the characteristics of a Japanese production operation.

4.1 A JAPANESE APPROACH TO DETERMINING THE NUMBER OF KANBANS

Prior to analyzing the factors that influence the number of Kanbans used in a JIT system it will be beneficial to observe how the Japanese set the number of Kanbans at a workcenter. The Toyota Motor Company is one Japanese firm that has had particular success using the JIT system; thus, we will use their procedure for setting the number of Kanbans, as follows [36]:

$$\text{Number of Kanbans} \geq \frac{(\text{demand})(\text{lead time})(1 + \text{safety factor})}{\text{container capacity}} \qquad (4\text{-}1)$$

where lead time = processing time + waiting time + conveyance time + Kanban collecting time [36, p. 70]

Kanban collection time in this formula is the time during which Kanbans are waiting to be picked up or returned to the beginning of a production operation.

If the safety factor in (4-1) is set equal to zero and if demand is expressed in containers per unit time then the formula becomes:

$$\text{Number of Kanbans} \geq (\text{demand})(\text{lead time}) \qquad (4\text{-}2)$$

Thus, the number of Kanbans (according to the Toyota Company) is at least the lead time demand expressed in terms of containers. This formulation has intuitive appeal in that it allows only enough buffering to compensate for lead time demand. We will employ this formulation in the next section of the chapter as a basis for determining the factors that influence the number of Kanbans, given significant variability in the production system.

The Japanese have significantly reduced the variability in their production system. The slight variability in the system that does exist is compensated for by the use of the safety factor in equation (4-1), which is typically determined by the workcenter foreman. The Japanese can employ such a simplified and direct approach because of the foreman's experience combined with the small amount of variation in the production parameters.

However, if the production system is not under such rigid control then setting a safety factor or estimating lead time or lead time demand becomes much more complicated, and, the straightforward Japanese approach for determining the required number of Kanbans becomes less effective and predictive. In the following section we will employ the Japanese formulation for determining Kanbans shown in (4-2) as a basis to examine the effect of stochastic variation on a JIT system, and, identify the factors that influence the number of Kanbans to use in a system with variation.

4.2 FACTORS INFLUENCING THE NUMBER OF KANBANS TO USE IN A JIT SYSTEM WITH VARIATION

In order to identify the factors that influence the number of Kanbans to use at a workcenter we will take a simplified approach and look only at a single workcenter using a single Kanban (and container) during one time period. Furthermore, we will assume that the workcenter encompasses only one machine; the shop produces only one product; that waiting time, conveyance time and Kanban collecting time are all zero or negligible relative to processing time; that product demand arrives instantaneously during each cycle period; and, setup times are zero and all processing times are equal. Given this sample scenario we will attempt to determine the factors that create the likelihood of a backorder at the workcenter since it is backorders that necessitate additional Kanbans to use as a buffer until the backordered demand can be met from excess capacity in succeeding periods.

In order to facilitate understanding of the forthcoming discussion it is necessary to define several terms, as follows. Container demand cycle (CDC) is the interarrival time of orders for the shop's final product; therefore, CDC = 1/demand. Cycle processing time (CPT) is the time required to complete production at a workcenter during one container demand cycle assuming the workcenter is idle when the cycle's orders arrive. (For example, if an order for 3 identical component parts arrives each CDC at a workcenter, and if no setups are necessary and the part processing time is 19 minutes, then the CPT equals 57 minutes.) Cycle throughput velocity (v) is the average number of items necessary to meet demand per machine per workcenter per time period. (For example, if workcenters A and B each have one machine and if 3 items are demanded at workcenter A and 10 items at workcenter

B per hour, then v_A equals 3 items per machine per hour and v_B equals 10 items per machine per hour.)

Now, recall formula (4-2) that was presented earlier to determine the number of Kanbans:

$$\text{Number of Kanbans} \geq (\text{demand})(\text{lead time}) \qquad (4\text{-}2)$$

Given our previous assumption that our example workcenter requires only one Kanban, (4-2) can be rewritten as,

$$1/\text{demand} \geq \text{lead time} \qquad (4\text{-}3)$$

However, also recall that we assumed that waiting time, conveyance time and Kanban collecting time are all equal to zero or negligible, which means that lead time equals processing time. Thus (4-3) can be reformulated as,

$$1/\text{demand} = \text{CDC} \geq \text{lead time} = \text{CPT} \qquad (4\text{-}4)$$

or,

$$\text{CDC} \geq \text{CPT}. \qquad (4\text{-}5)$$

The cycle processing time (CPT) can also be expressed as,

$$\text{CPT} = \Sigma(\mu_{PT} + \varepsilon_{PT}) \qquad (4\text{-}6)$$

where μ_{PT} = the mean processing time per item

ε_{PT} = the random error in the processing time.

If processing times are assumed to be independent and identically distributed, the mean and standard deviation of the CPT can be written as follows:

$$\mu_{CPT} = v\mu_{PT}$$

$$\sigma_{CPT} = v^{1/2}\sigma_{PT}.$$

Employing these formulas, the following statements can be written describing the probability that CDC will equal CPT:

$$CDC = CPT = \mu_{CPT} + Z_P\sigma_{CPT} \qquad (4\text{-}7)$$

or,

$$Z_P = (CDC - \mu_{CPT})/\sigma_{CPT}$$

$$= (CDC - v\mu_{PT})/(v^{1/2}\sigma_{PT}), \qquad (4\text{-}8)$$

where Z_P is the number of standard deviations that the CPT must be greater than its mean in order for CDC to equal CPT.

As an example, if CDC = one hour, μ_{CPT} = 0.85 hour and σ_{CPT} = 0.1786 hours then Z_P =0.84. If it is assumed that the CPT is normally distributed, then the probability that the CDC will be exceeded is 0.201. In other words, the probability that a backorder will occur (i.e., there is an insufficient number of Kanbans or containers) is 0.201.

Equation (4-8) can also be reformulated in terms of various production parameters, as follows:

$$\text{utilization} \equiv \text{util} = v\mu_{PT}/CDC \qquad (4\text{-}9)$$

$$\text{coefficient of variation} \equiv CV = \sigma_{PT}/\mu_{PT} \qquad (4\text{-}10)$$

$$\mu_{CPT} = v\mu_{PT} = (v\mu_{PT}/CDC)(CDC) = (\text{util})(CDC) \qquad (4\text{-}11)$$

Thus, the standard deviation of CPT can now be written as

$$\sigma_{CPT} = \sigma_{PT}v^{1/2}$$

$$= v^{1/2}(\mu_{PT}\sigma_{PT}/\mu_{PT})(v/v)(CDC/CDC)$$

$$= (\sigma_{PT}/\mu_{PT})(v\mu_{PT}/CDC)(CDC/v^{1/2})$$

$$= (CV)(util)(CDC)/v^{1/2},\ v \neq 0. \qquad (4\text{-}12)$$

Recall from (4-7) that,

$$CDC = \mu_{CPT} + Z_P\sigma_{CPT}$$

which, with (4-11) and (4-12) substituted, becomes

$$CDC = (util)(CDC) + Z_P(CV)(util)(CDC/v^{1/2}) \qquad (4\text{-}13)$$

or,

$$Z_P = (v^{1/2}/CV)[(1/util) - 1],\ 0 < util \leq 1 \text{ and } CV \neq 0 \qquad (4\text{-}14)$$

(A coefficient of variation (CV) equal to zero implies no probability of a backorder.)

From equation (4-14) we can now identify the factors that influence the probability of backorders, and hence, the number of Kanbans needed. The first factor is the throughput velocity, v. Z_P varies directly with the root of the throughput velocity; therefore the probability of a backorder will vary inversely with the throughput velocity. As the throughput velocity decreases, the probability of a backorder will increase, as will the need for additional Kanbans.

A second factor from (4-14) that influences the number of Kanbans needed is the coefficient of variation (CV) of the processing time at the workcenter. Z_P varies inversely with the coefficient of variation, therefore, the probability of backorders will increase as CV increases. Further, the probability of backorders will increase faster for CV increases then it will decrease for throughput velocity increases.

The third factor identified from (4-14) is the utilization of machines at the workcenter. As the utilization rate increases, Z_P will decrease, and thus, the probability of backorders and the number of Kanbans will increase.

A fourth factor that can also influence the number of Kanbans employed at a workcenter is whether or not autocorrelated processing times exist. Recall that in the above derivation, processing times were assumed to be independent and identically distributed. If n items are produced per CDC at a workcenter and the processing times are positively autocorrelated so as to follow an AR(1) process (see Box and Jenkins [3]) rather than being independent, then,

$$\sigma'_{CPT} = \sigma_{CPT}(1 + \gamma_n) \qquad (4\text{-}15)$$

where,

σ'_{CPT} = the standard deviation of cycle processing time with autocorrelation following an AR(1) process.

σ_{CPT} = the standard deviation of cycle processing time with independent and identically distributed processing times

γ_n = a function strictly greater than 0 for all n.

In the case of positive autocorrelation, equation (4-14) becomes,

$$Z'_p = Z_p/(1 + \gamma_n) \qquad (4\text{-}16)$$

where Z'_p has the same meaning as Z_p, but for the autocorrelated case. Thus, if processing times are autocorrelated, Z'_p will decrease by a factor of $(1 + \gamma_n)$ from the case with no correlation, and, the probability that the CDC will be exceeded and there will be backorders increases.

4.2.1 A SIMULATION CASE EXAMPLE

In order to demonstrate the impact of these factors on the number of Kanbans used at a workcenter, an example shop will be simulated with varying operating conditions. The example shop is described in Chapter 2. The product demand is 200 units for X and 100 units for Y. The container size is 20 units for X and 10 units for Y.

In order to provide a more realistic test of the factors we identified via equation (4-14), the assumptions originally specified prior to the development of equations (4-1) through (4-14) will be relaxed. In our example shop all workcenters are linked together and simulated for a number of cycles, two final products are produced, waiting times are not zero and the arrival of demand during each cycle is not instantaneous. Processing times are normally distributed and assumed to follow an AR(1) series with a positive autocorrelation of 0.5. The means of the container processing times for each product at each workcenter and the setup times at each workcenter are shown in Table 4-1.

The example shop was simulated and the results are shown in Table 4-2. Table 4-2 contains the value of Z'_p, computed for each workcenter using equation (4-16), with an ordinal ranking of the workcenter's Z'_p "score," the coefficient of variation (CV), the utilization rate (util), the cycle throughput velocity (v), the number of Kanbans determined from trial and error, and, the average number of backorders incurred at the workcenter. (The average number of backorders at a workcenter is the average number of withdrawal Kanbans which cannot be matched with an item of finished inventory.)

The objective of this simulation analysis is primarily to determine if the methodology used to determine the probability of backorders (and, thus, if there are enough Kanbans), developed in the previous section, is effective given a more realistic and complex production environment than the one used to develop equations (4-14) and (4-16). In our example, the value of Z'_p was first computed for each workcenter using the indicated values for CV, util and v, and then ranked

Table 4-1. Parameters for the Simulation Example.

Workcenter	No. of Machines	Container Processing Time (Hours)		Setup Time (Hours)
		μ_X	μ_Y	
1	1	0.124	0.062	0.005
2	1	0.138	0.069	0.005
3	1	0.328	0.164	0.020
4	1	0.1768	0.0884	0.012
5	1	0.648	N/A	N/A
6	1	N/A	0.736	N/A

Table 4-2. Simulation Results with Autocorrelated Processing Times.

	Factors Influencing the Number of Kanbans					Simulation Results		
Workcenter	CV	util	V	Z	Z rank	Average Backorders	Number of Kanbans	Performance Rank
1	.3	.726	9	2.33	1	.0342	1	1
4	.3	.722	6	2.056	2	.0673	1	2
5	.3	.799	1	.890	3	.0079	2	3
2	.3	.889	10	.800	4	.0587	2	4
3	.3	.920	4	.388	5	.1060	2	5
6	.3	.917	1	.293	6	.4807	3	6

ordinally from 1 to 6 with the best performance being ranked first. These values and rankings are shown in the first part of Table 4-2. Note that a high value for Z'_p indicates a low probability of backorders. For example, in Table 4-2, a value of 2.33 for Z'_p for workcenter 1 indicates a lower probability of a backorder than a Z'_p value of 2.056 for workcenter 4; hence, workcenter 1 is ranked higher than workcenter 4. These Z'_p values (and the rankings) were all determined manually using equation (4-16).

Alternatively, the second part of Table 4-2 consists of the results obtained using the simulation model. The number of Kanbans used per workcenter was determined by trial and error with the basic objective of reducing backorders to an average of .5 or less per hour. The performance ranking for the simulation results was the number of Kanbans required per workcenter to meet this arbitrarily selected guideline. Thus, workcenters 1 and 4, which each requires only one Kanban, are ranked first and second. The differential ranking between these two workcenters is based on fewer average backorders at workcenter 1 than workcenter 4. The differential rankings for workcenters 5, 2 and 3, each of which requires two Kanbans is determined the same way.

The significant aspect of the results in Table 4-2 is that the performance rankings of the workcenters determined manually using the Z'_p formulation and with the simulation model, are in complete agreement. In other words, the use of Z'_p appears to be an entirely effective means for analyzing the influence of the production factors noted on the number of Kanbans required. These results also appear to verify that the coefficient of variation (CV), utilization (util) and throughput velocity ($\dot{v}$) directly influence the number of Kanbans required at a workcenter.

4.3 A SIMULATION APPROACH FOR DETERMINING THE INITIAL NUMBER OF KANBANS AT A WORKCENTER

The first part of this chapter consisted of an analysis of the factors in a production system that will influence the number of Kanbans that must be used at a workcenter. In this second part of the chapter a methodology using simulation will be demonstrated for determining the number of Kanbans to use at a workcenter given a realistic and variable production environment. Such a production environment would include multiple machines at workcenters, stochastic processing times and setup times, and, complex item sequences through the production system.

As a prelude to the development of the methodology for determining the number of Kanbans at a workcenter we will make the simplifying but important assumption that the production system will operate with no backorders. That is, it will be assumed that the objective is to specify just enough Kanbans throughout the shop so that backorders will never be incurred at any workcenter. This is a reasonable objective since systems using the JIT technique are typically very sensitive to backorders, especially in large shops where many workcenters are cascaded. This assumption will enable all workcenters to be "decoupled" from each other since the items produced at each workcenter will always be available at other workcenters when they are needed.

Recall the previous formulation (4-2) for determining the number of Kanbans at a workcenter used by the Toyota Company:

$$\text{Number of Kanbans} \geq (\text{demand})(\text{lead time}) \qquad (4\text{-}2)$$

We will employ basically the same formulation for determining the number of Kanbans, as follows:

$$n \geq D_m L_{0.95} \qquad (4\text{-}17)$$

where, n = the number of sets of Kanbans for an item at a workcenter

D_m = the maximum demand for that item's final product in containers

$L_{0.95}$ = the maximum lead time at that workcenter determined to be the 95th lead time value from 100 simulated lead times ranked in order of magnitude from lowest to highest. (Lead time is determined in this manner because it was concluded from experimentation that the absolute maximum lead time allowed too many extra Kanbans floating in the system whereas the 95th highest lead time did not.)

(Note that in (4-17) "n" equals the number of sets of Kanbans. The total number of actual Kanbans is computed by multiplying the sets of Kanbans by the number of containers of an item that are needed to make a container of final product. For example, assume at a workcenter that two sets of Kanbans are necessary for product X as determined by equation (4-17). If at the workcenter 2 B's and 3 A's are required to produce an X then the total number of actual Kanbans needed is 4 for B and 6 for A or 10 overall.)

The essential difference between (4-2) and (4-17) is that our latter formulation assumes a production operation wherein the variability in production parameters makes it difficult to estimate D_m and L. Alternatively, in the first (Japanese) formulation the lesser degree of variability would result in relatively accurate estimates of demand and lead time. As a result of these differences, we will employ a simulation approach to determine the parameters on which the number of Kanbans is based in (4-17).

As already indicated, the maximum forecasted demand is employed as the estimate for D_m. This is somewhat different from the Japanese approach where actual demand is simply estimated and multipled by one plus a safety factor set by the foreman as shown in (4-1). The determination of lead time

($L_{0.95}$ in equation (4-17)) is generally more complex. In some instances it may be possible to use analytical results for time in system to determine lead time if the workcenter can be represented by a simple queueing model such as an M/M/1 or M/M/c model. However, it is more likely that lead time cannot be determined analytically in which case a simulation approach is required.

4.3.1 THE SIMULATION MODEL

The simulation methodology for determining lead time will be described via an example which is a slightly modified version of the production operation described earlier. Demand for X and Y will be changed to 190 and 100 units per day, respectively, in order to complicate the system by throwing demand out of synchronization. The setup and processing times as well as the number of machines at each workcenter are shown in Table 4-3. Notice that the processing times for the containers of product X items are twice as large as the containers for items used for product Y at workcenters 1 through 4 since because X has a container size of 20 units whereas Y has a container size of 10 units.

The simulation methodology will be demonstrated via two examples, one for workcenter 6 and one for workcenter 4. We will begin with workcenter 6 because it is relatively easy

Table 4-3. Simulation Example Parameters.

Workcenter	No. of Machines	Container Processing Time (Hours)		Setup Time (Hours)
		μ_X	μ_Y	
1	2	0.166	0.083	0.007
2	2	0.184	0.092	0.007
3	2	0.452	0.226	0.028
4	2	0.250	0.125	0.017
5	1	0.682	N/A	N/A
6	1	N/A	0.736	N/A

to model since it has only one demand stream. In other words, workcenter 6 satisfies demand for product Y only by assembling parts into 10 containers of Y every eight-hour day. As a result, container demand is deterministic with an interarrival time of 0.8 hours. Assume service times follow a normal distribution and recall that there is only one machine at the workcenter.

If we assume that there is sufficient capacity and there is a sufficient number of Kanbans to prevent backorders at each workcenter, then workcenter 6 will never experience a shortage of in-process items. These assumptions enable us to decouple workcenter 6 from the rest of the operation as discussed above and consider it as an isolated case with demand that has a deterministic interarrival pattern and a normal service distribution with a single server, which is actually the description of a D/N/1 queueing model.

In order to obtain an estimate for lead time, a D/N/1 queue is simulated. Time in the system, which is our estimate of lead time, is recorded for every fiftieth arrival. (The fiftieth arrival is collected because the lead times are autocorrelated in the sense that if one job has a long lead time the probability increases that the next job will have a longer than normal lead time. By collecting only the fiftieth observation a sequence of "independent observations is obtained.) 100 observations are obtained in this manner.

These 100 lead time observations are ranked from smallest to largest and the 95th highest value is used as our estimate of lead time as explained previously. This value is then used (together with an estimate of maximum forecasted demand, D_m) to compute the number of Kanbans required at workcenter 6 according to equation (4-17).

The second simulation example for workcenter 4 is more complex than workcenter 6 since workcenter 4 must process components to satisfy demand at two workcenters, 5 and 6. Demand from workcenter 5 for component items to make X (i.e., 2 F's and 1 J) arrives at the same rate as the demand for end-item X (i.e., 190/20 = 9.5 containers) due to the "pull" nature of the JIT shop described earlier. Therefore, the interarrival time for orders at workcenter 4 from workcenter 5 is 0.842 hours (i.e., 8 hours/9.5 containers). Similarly, demand from workcenter 6 for component items to produce Y (i.e., 2 L's and 1 M) arrives every 0.80 hours at workcenter 4. Thus, the demand for orders arriving at workcenter 4 is actually a superposition of two demand streams. There are also 2 machines at workcenter 4.

If sufficient Kanbans are provided to always meet demand at workcenters 5 and 6, then workcenter 4 may be decoupled from them. Likewise, if workcenter 1, 2 and 3 also have sufficient Kanbans to always meet demand, then in-process items will always be available at workcenter 4. This enables us to consider workcenter 4 as an isolated case and model it separately from the remainder of the shop just as we did with workcenter 6.

Workcenter 4 can be treated as a queueing model with interarrivals following the superposition of two deterministic rates, normal service times and two servers. This model was simulated until 100 independent observations could be collected with the 95th highest observation selected as the estimate for lead time. This estimate for lead time was subsequently used in equation (4-17) to determine the number of Kanbans required at workcenter 4.

The simulation methodology described for workcenters 4 and 6 was employed for all six workcenters in the production system described in Figure 2-1. Estimates for L and n were obtained for three different cases of variation in processing times: CV = 0.1, 0.2 and 0.3. These results are shown in Table 4-4.

Next, a simulation model of the example production system described in Figures 2-1 and 2-2 and Table 4-3 was simulated using the number of Kanbans determined by the methodology described above and shown in Table 4-4. Processing times at each workcenter were assumed to be independent and identically distributed with CV = 0.1, 0.2 and 0.3, in that order. The simulated production operation ran with all demand being met and no backorders for all three cases. This indi-

Table 4-4. Number of Kanban Sets for Processing Times Which are Independent and Identically Distributed (IID).

Workcenter	Lead time L	Number of Kanban Sets		Coefficient of Variation for Processing Times
		n_X	n_Y	
1	0.61	1	1	CV = 0.1
2	0.75	1	1	
3	0.78	1	1	
4	0.60	1	1	
5	0.78	1	N/A	
6	0.90	N/A	2	
1	0.64	1	1	CV = 0.2
2	0.74	1	1	
3	0.87	2	2	
4	0.62	1	1	
5	0.97	2	N/A	
6	1.23	N/A	2	
1	0.65	1	1	CV = 0.3
2	0.72	1	1	
3	0.88	2	2	
4	0.61	1	1	
5	1.14	2	N/A	
6	2.17	N/A	3	

cates that, at least for this example production system, the methodology for determining the number of Kanbans is correct.

In order to determine if there were too many Kanbans in the system, i.e., too much "slack," sensitivity analysis was conducted using the simulation model. This analysis consisted of removing a set of Kanbans for an item at a workcenter that had more than one set of Kanbans and repeating the simulation. Results indicated that backorders occurred at each workcenter when Kanbans were removed.

Similar simulation model results were obtained for a second simulation analysis wherein processing times were assumed to be positively autocorrelated.

4.3.2 THE EFFECT OF LESS-THAN-IDEAL PRODUCTION FACTORS

This methodology can be used to explore the effect of less-than-ideal production factors on the number of Kanbans. Consequently, as a final simulation analysis, the influence of the various factors identified in the first part of this chapter on the number of Kanbans used was examined within the context of this example production system and simulation model. Figures 4-1, 4-2 and 4-3, indicate respectively the effects of the coefficient of variation (CV), utilization (util) and positive autocorrelation on the number of Kanbans.

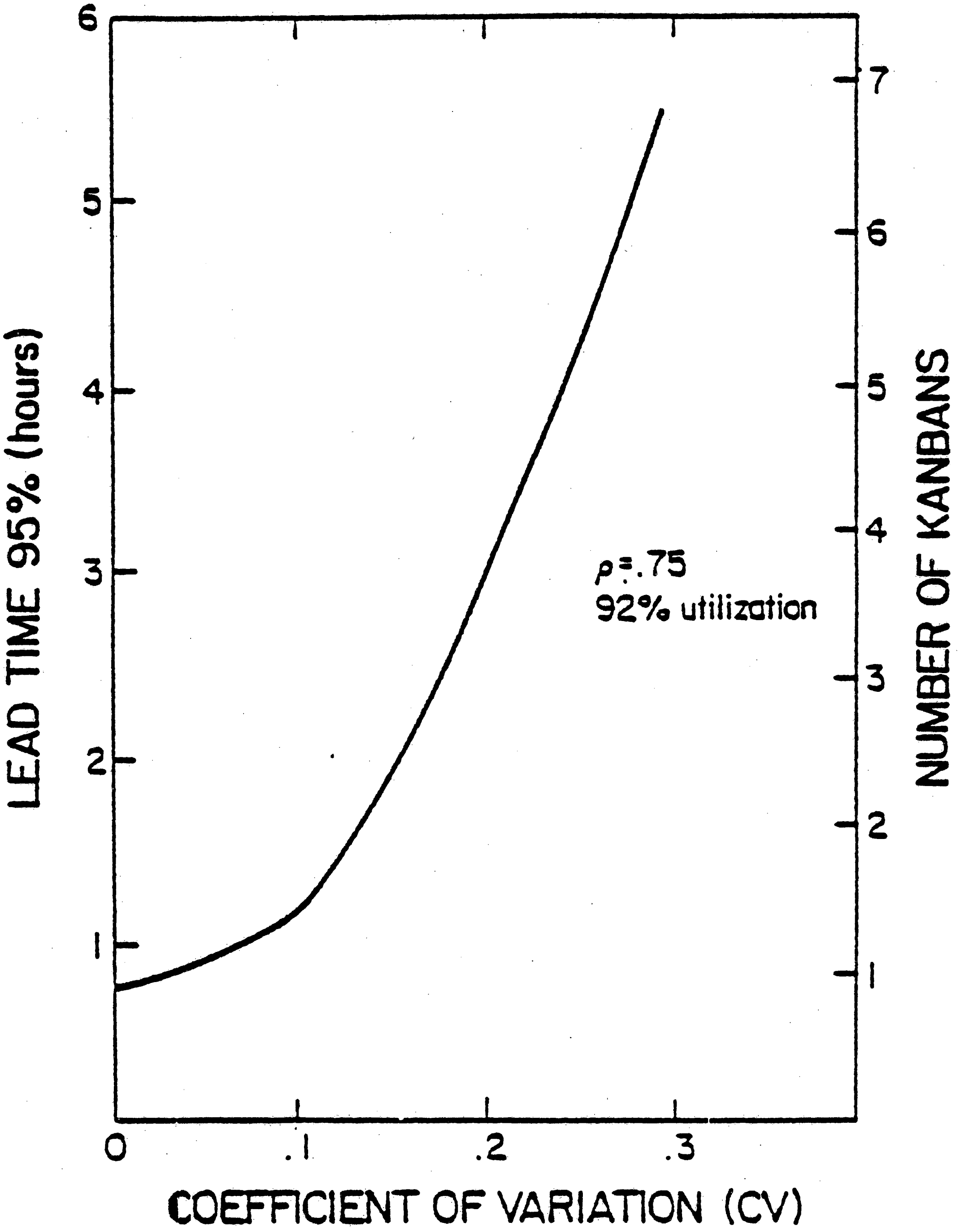


Figure 4-1. The Effect of Variability in Processing Time on the Number of Kanbans at a Workcenter.

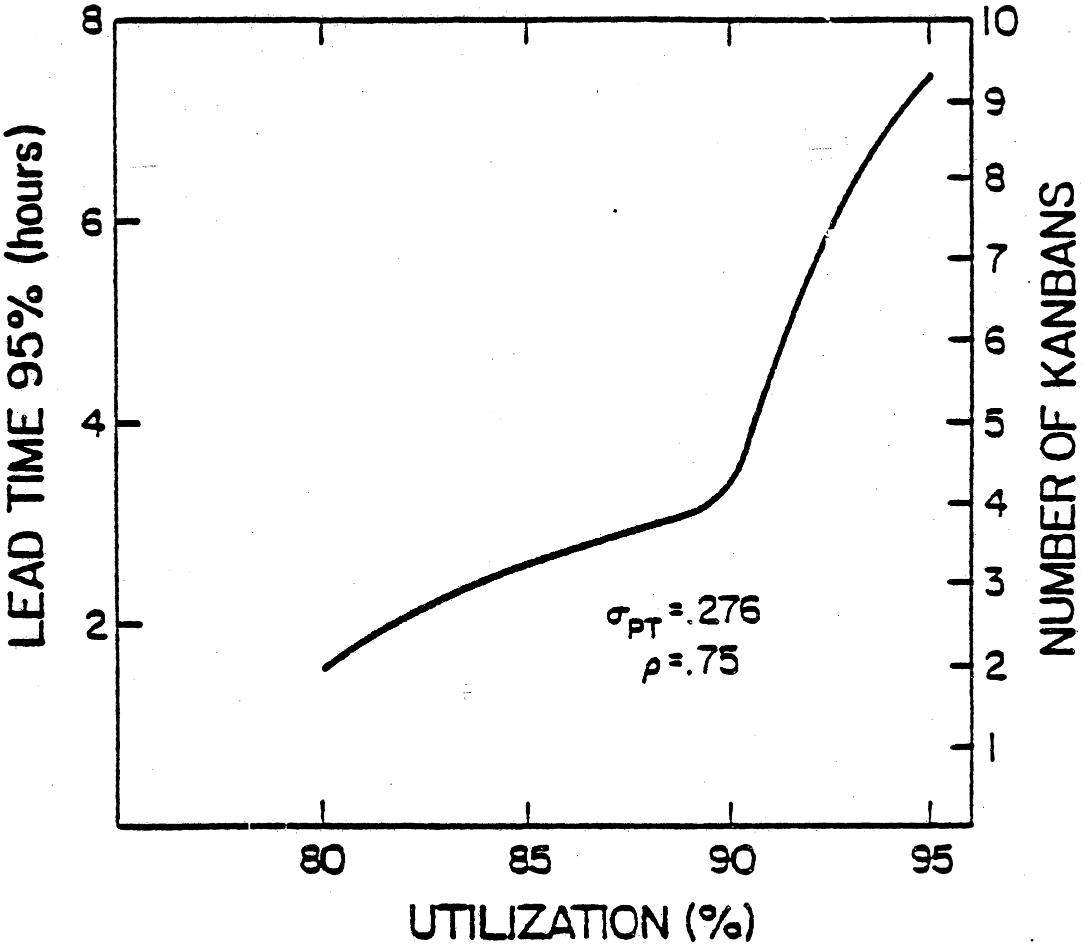


Figure 4-2. The Effect of Utilization on the Number of Kanbans at a Workcenter.

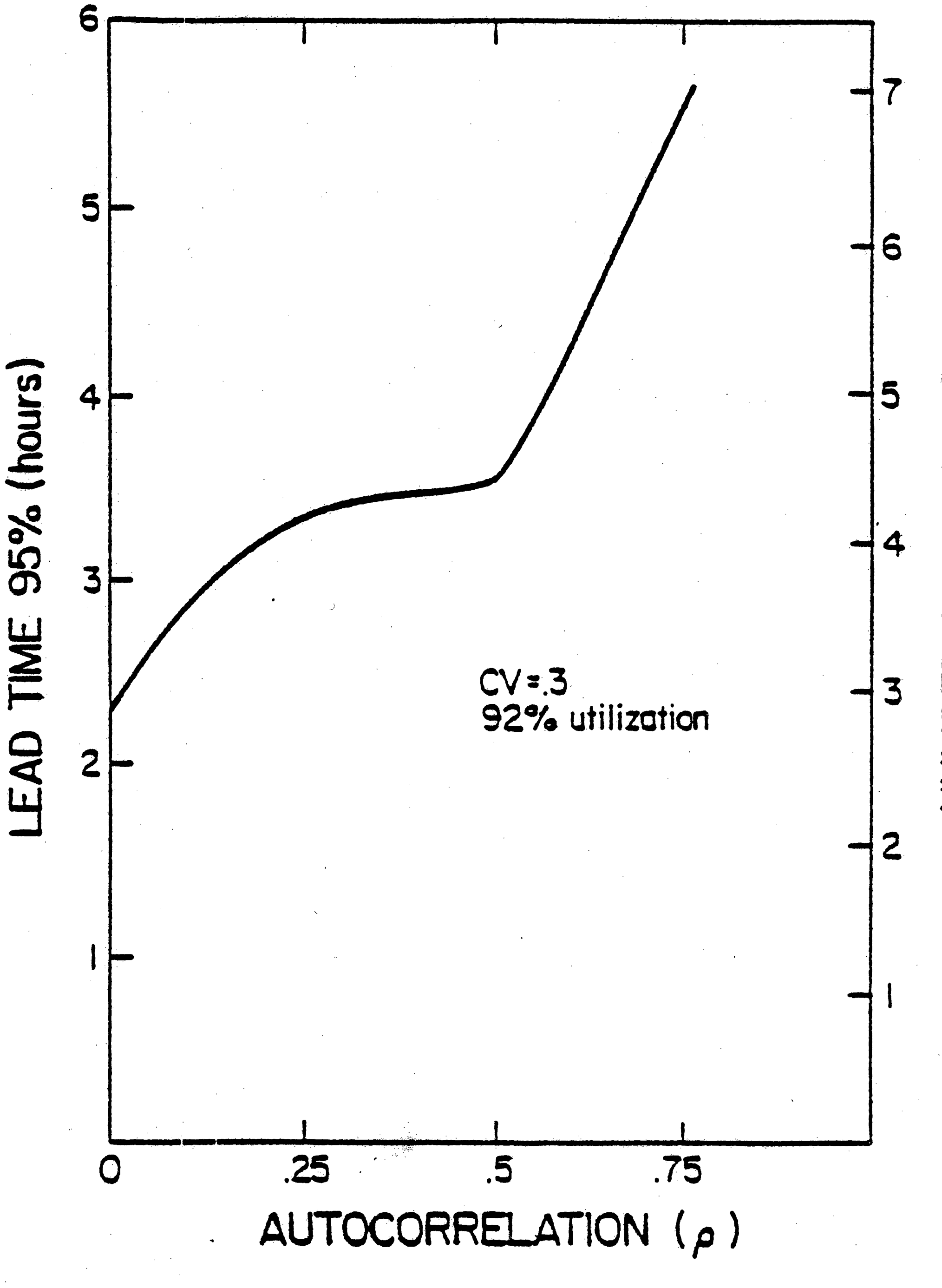


Figure 4-3. The Effect of Correlation of Processing Times on the Number of Kanbans at a Workcenter.

The curves shown in these figures further demonstrate the effects predicted by equations (4-14) and (4-16).

For the production manager contemplating the implementation of the JIT technique with Kanbans, the type of analysis embodied in Figures 4-1 ,4-2 and 4-3 can have a significant impact. For example, Figure 4-1 indicates that if the variability in processing times can be reduced, and processing times standardized, then the number of Kanbans can be reduced. This is a factor that has enabled the Japanese to employ the JIT technique successfully. Alternatively, if a manager is considering use of the JIT technique in a job-shop type production operation, where it is difficult if not impossible to reduce the variability in processing times, then the number of Kanbans required may be so high as to render the JIT system ineffective.

Similar insight into the JIT technique can be gleaned from Figure 4-2. If new, more efficient machines, more effective production processes and better trained workers can be achieved which will reduce utilization, then the number of Kanbans can be reduced. However, if a company cannot afford such alterations in their operation and the utilization rate is very high, then the number of Kanbans required will in all likelihood render the JIT technique ineffective.

4.4 SUMMARY

The first part of this chapter identified those factors in the production environment that influence the number of Kanbans that will be required at workcenters if a JIT system is implemented. Four major factors were identified via an analysis of the Japanese procedure for determining the number of Kanbans required at a workcenter: throughput velocity, the coefficient of variation in processing times, the machine utilization rate and the autocorrelation between processing times. For the American production manager contemplating the possibility of implementing the JIT technique in a production operation that does not display the characteristics of a Japanese operation, these factors can have a significant impact on the effectiveness of the JIT technique.

The second part of the chapter described a simulation approach for determining workcenter lead times, and hence, the number of Kanbans required at a workcenter necessary to prevent backorders. This approach was based on the simplifying (but logical) assumption that workcenters could be decoupled and modeled separately as queueing systems in order to determine lead times. A simulation analysis of an example production operation indicated the procedure was effective in determining the minimum number of Kanbans required at a workcenter. This same simulation model was subsequently used to test the validity of the factors identified in the first part of the chapter as influencing the number of Kanbans required.

In this chapter it has been assumed that shop conditions, while perhaps bad, are at least static. In the next chapter a methodology for dynamically adjusting the number of Kanbans in a shop in which conditions are not static is presented.

5.0 DYNAMICALLY ADJUSTING THE NUMBER OF KANBANS USING ESTIMATED VALUES OF LEADTIME

Typically in a JIT operation the master production schedule is frozen for one month and the number of Kanbans at each workcenter is set based on the average demand for the period [36]. When the monthly demand changes, one would expect that the total number of Kanbans per month would also be changed. Such is not the case at Toyota, as Yasuhiro Monden [36, p. 174] explains: "For example, suppose it is expected that the average daily demand of next month will be two times the demand of the current month.... At Toyota ... the total number of Kanbans ... [is] unchanged." Companies using JIT such as Toyota do not have to routinely adjust the number of Kanbans from month to month for at least three reasons: they have a large market share and hence demand variations from the forecasted value are a small percentage of the total; they have cross-trained workers whom they are able to switch from workcenter to workcenter to mitigate temporary bottlenecks; and their JIT shops are so well run that they can handle day-to-day problems as well as variations in demand. But many firms either using or thinking of using Kanbans do not exhibit these characteristics. They do not have large market shares; workers do not have the training nor (with unions) the opportunity to be switched from workcenter to workcenter as needed; and these firms' JIT operations are still in the lower portion of the learning curve, not to mention being finely tuned and honed. In such firms it is often essential to adjust the number of Kanbans.

Specifically, this chapter explores a workable method for dynamically adjusting the number of Kanbans in a JIT shop. After examining some basic principles of JIT with Kanbans, the methodology is presented. This is followed by three examples. The purpose of the first example is to illustrate in detail the methodology, while the last two examples present solutions to interesting JIT problems using the dynamic adjustment of Kanbans as presented in the methodology.

The Toyota Motor Company sets the number of Kanbans for an item at a workcenter for a period using the following formula [36]:

$$n = [DL(1 + \alpha)], \qquad (5\text{-}1)$$

where n is the number of Kanbans,

D is the average demand expressed in containers

L is the leadtime for the item, which equals processing time + waiting time + conveyance time + Kanban collecting time [36, p.70]

α is a safety factor which is set by the shop supervisor to handle stochastic variation and anomalies and [x] means the smallest integer greater than or equal to x.

This formula (if the safety factor is neglected) indicates that the number of Kanbans depends on demand and the leadtime necessary to produce a container's worth of goods, i.e., the number of Kanbans is the smallest number of containers needed to satisfy demand during leadtime. If considered with the idea of minimizing n, this equation summarizes the Japanese philosophy of carrying only enough inventory to meet demand during leadtime and of reducing leadtime as much as possible.

As mentioned, Toyota does not routinely adjust the number of Kanbans once they are set because of its large market share, cross-trained workers, and well-run shops. With reference to equation (5-1), Toyota does not have to adjust n because its large market share results in D generally being kept within limits, and L is under control (i.e., does not vary widely from its mean) because Toyota's workers are well trained and the shops are well understood and run. Furthermore, even if the product of D and L increases so that more Kanbans are suggested at a workcenter, Toyota still does not increase n. Instead, it takes other measures such as reassigning enough cross-trained workers to that workcenter to reduce L by the amount required to hold n constant.

An important practical question is what should a company do that cannot operate as Toyota does? The purpose of this chapter is to answer this question and determine the conditions under which a company should forego JIT with Kanbans rather than dynamically adjusting n.

Although equation (5-1) states that n may be determined as the product of only two factors, (exogenous) demand and (endogenous) leadtime, it should not be assumed that dynamically adjusting the number of Kanbans is easy. Leadtime is a function of many shop parameters, including utilization, variability in processing times, and throughput rate of jobs at the workcenter. Thus, leadtimes could not be simply forecast statistically using several years of leadtime data because shop parameters themselves change with time. The methodology in the next section explains how the most recent set of shop conditions is used in conjunction with a forecast of demand to set the number of Kanbans for the upcoming period.

5.1 METHODOLOGY FOR DYNAMICALLY ADJUSTING KANBANS

The basic principle underlying the methodology for the dynamic adjustment of the number of Kanbans at a workcenter is to exploit equation (5-1). Periodically (for example, once a month), the number of Kanbans at the workcenter is adjusted based on forecasted demand for the next month and collected observations of leadtime at the workcenter during the past month. In particular, the probability density function (pdf) of leadtime is estimated and is combined with the forecasted demand value to produce the pdf for n, the number of Kanbans. Based on relative values of shortage and holding costs at the workcenter, the supervisor may determine the preferred number of Kanbans to be used the next period.

It was determined through experimentation that leadtime observations are not independent but in fact are strongly positively correlated. Thus it is necessary to also estimate the autocorrelation function of leadtimes so that this factor may be removed when estimating the leadtime pdf.

A "timeline" of our methodology is shown in Figure 5-1. The methodology consists of two measuring periods, one to estimate the autocorrelation function and the other to estimate the pdf. These periods are followed by an "action," i.e., a possible adjustment in the number of Kanbans, and finally a "settle" period which allows the shop to settle down before a new set of observations is made for the next period.

Next, the detailed steps of the methodology are presented for an item at a workcenter.

Step 0. Startup

If this shop is just being converted to Kanban operation or major changes have been made which have perturbed shop

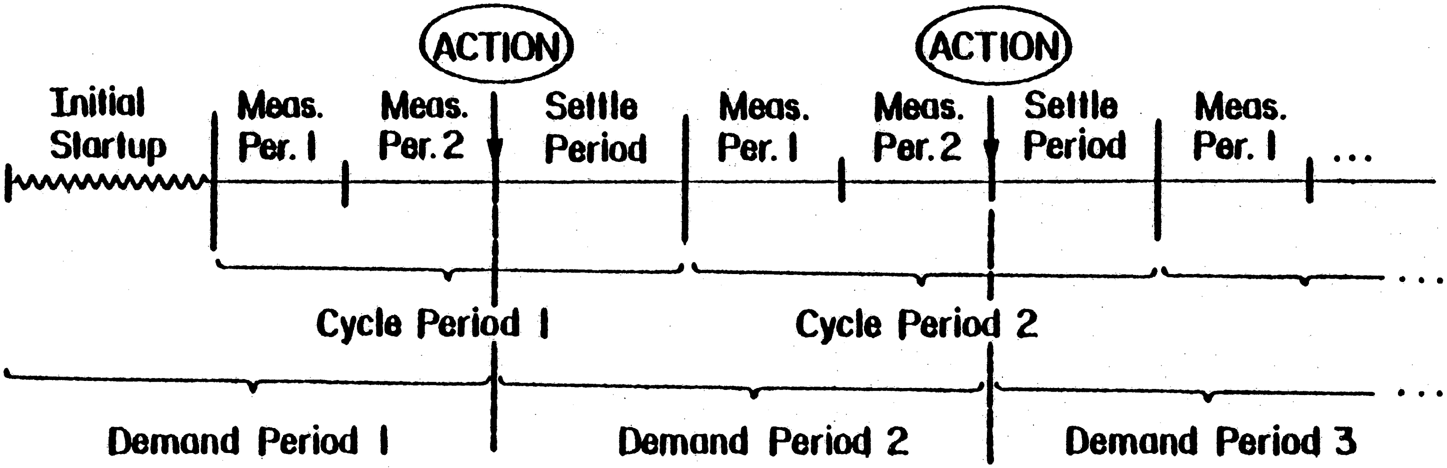


Figure 5-1. A "Timeline" Indicating the Different Activity Periods in the Methodology.

conditions, these transient effects should be permitted to die out before step 1 is attempted.

Step 1. Measuring Period 1

The purpose of measuring period 1 is to obtain autocorrelation estimates of the container leadtime series so that statistically independent observations of leadtime may be obtained during the second measuring period. How independence is achieved is explained in step 2.

Box and Jenkins [3, p. 33] suggest that "in practice, to obtain a useful estimate of the autocorrelation function, we would need at least fifty observations and the estimated autocorrelations r_k [at lag k] would be calculated for $k = 0, 1, \ldots, K$ where K was not larger than say N/4." Since we have observed significant autocorrelations at lags 20 - 25 (under admittedly severe conditions), we recommend that 100 observations be collected during measuring period 1, if feasible, but certainly no fewer than 50.

Once the (100) observations have been collected, the estimated autocorrelation at lag k may be computed using

$$r_k = \frac{\frac{1}{N}\sum_{t=1}^{N-k}(x_{t+k} - x)(x_t - x)}{\frac{1}{N}\sum_{t=1}^{N}(x_t - x)^2}, \quad k = 1,2,\ldots,25, \qquad (5\text{-}2)$$

where x_j is the jth observation of container leadtime.

A correlogram may be drawn, if desired, to show the estimated autocorrelation function.

After equation (5-2) has been calculated for all k listed, the leadtime observations (from measuring period 1) should be discarded.

Step 2. Measuring Period 2

The purpose of measuring period 2 is to estimate the pdf of leadtimes for an item at the workcenter using the results of step 1 to obtain statistically independent observations. Since statistical independence implies no correlation, the correlogram of r_k's produced via equation (5-2) is examined to determine the lag k beyond which all autocorrelations are zero; in practice, the lag k is determined beyond which the autocorrelation remains below 0.05. Then if observations spaced k apart from each other are collected, the observations will be approximately statistically independent. Thus step 2 is conducted to collect observations spaced k apart and estimate the pdf of container lead times.

The histogram historically has been used as an estimate of pdf's, but Law and Kelton [29, pp. 180-181] point out that more "modern" methods exist for this purpose. We recommend that pdf's be estimated using the "conventional" histogram procedure, but recognize that sophisticated users of the methodology will be able to generate better estimates using other methods. In particular, a method developed by Tarter and Kronmal [61] utilizing Fourier series estimates and the orthogonal functions,

$$T_k(x) = (e^{2\pi i x})^k, \qquad i = \sqrt{-1},$$

has been shown to overcome some disadvantages of the histogram as a pdf estimator. Tarter and Kronmal's method has been computerized and the algorithm is described by Tarter, et al. [60]. They furnish an example based on 100 independent and identically distributed observations and state that, based on their experience, "this demonstration is representative, i.e., the nonparametric estimates obtained are typical of what can be expected from samples of this kind."

Law and Kelton state in discussing the histogram procedure that the choice of intervals and interval width is a "vexing" part of the construction and that various widths should be attempted until the histogram "looks smooth." The difficulty that they allude to in this part of the process will be somewhat alleviated by conditions described in step 4, and will be discussed further at that point. For now note that approximately 100 observations of Kanban leadtime should be collected spaced k leadtimes apart and a histogram should be developed from these observations. The histogram will be the estimate of the pdf of L.

Step 3. Forecast Demand

Using standard (company) forecasting procedures, determine an estimate of the next demand period's demand, D, for the item at the workcenter.

Step 4. Determine the pdf for the Number of Kanbans

The purpose of this step is to estimate a probability density function for n given the estimated pdf of leadtime from step 2 and the forecast of demand for the next period from step 3. To accomplish this the pdf of n' is determined, where $n' = DL$, and then the pdf of n (where $n = [DL]$) is found.

$f_{n'}(n')$ is defined as the pdf of the random variable n' and $f_L(L)$ as the pdf of the random variable L. Since D, once estimated, is considered a deterministic constant over the forecasted period, and since it is assumed that D_{t-1} will not differ[2] from D_t so drastically as to appreciably change $f_L(L)$, it can be shown that (see, e.g., [44]),

$$f_{n'}(n') = \frac{1}{\hat{D}} f_L\left(\frac{L}{\hat{D}}\right), \ d \neq 0$$

$$\text{or, } f_{n'}(n') = \frac{1}{\hat{D}} f_L\left(\frac{L}{\hat{D}}\right), \ D > 0. \tag{5-3}$$

Equation (5-3) states that n' has the same general pdf as L, and is in fact just a scaled-down, reshaped version of $f_L(L)$. For example if $\hat{D} = 2$ and L follows the exponential density shown in Figure 5-2a, then n' has the exponential pdf shown in Figure 5-2b. If $\hat{D}$ still equals 2 and L has the empirically determined pdf shown in Figure 5-2d, $f_{n'}(n')$ is as shown in Figure 5-2e.

It remains to determine $f_n(n)$ given $f_{n'}(n')$. The former is just a discretized version of the latter with mass located at n = 1,2,3,..., and, density at each point k equal to $\int_{k-1}^{k} f_{n'}(n')dn'$. That is, the density at each discrete Kanban value will be the area under the $f_{n'}(n')$ curve between the next lower number of Kanbans and this number of Kanbans. For example, the density at n = 3 Kanbans will be the area under the $f_{n'}(n')$ curve between 2 and 3 Kanbans, because any

---

[2] For JIT to work, demand must be fairly constant. Minor fluctuations in demand are handled by adjusting the length of the workday. Thus, even if the daily demand fluctuates, the hourly demand rate should stay constant.

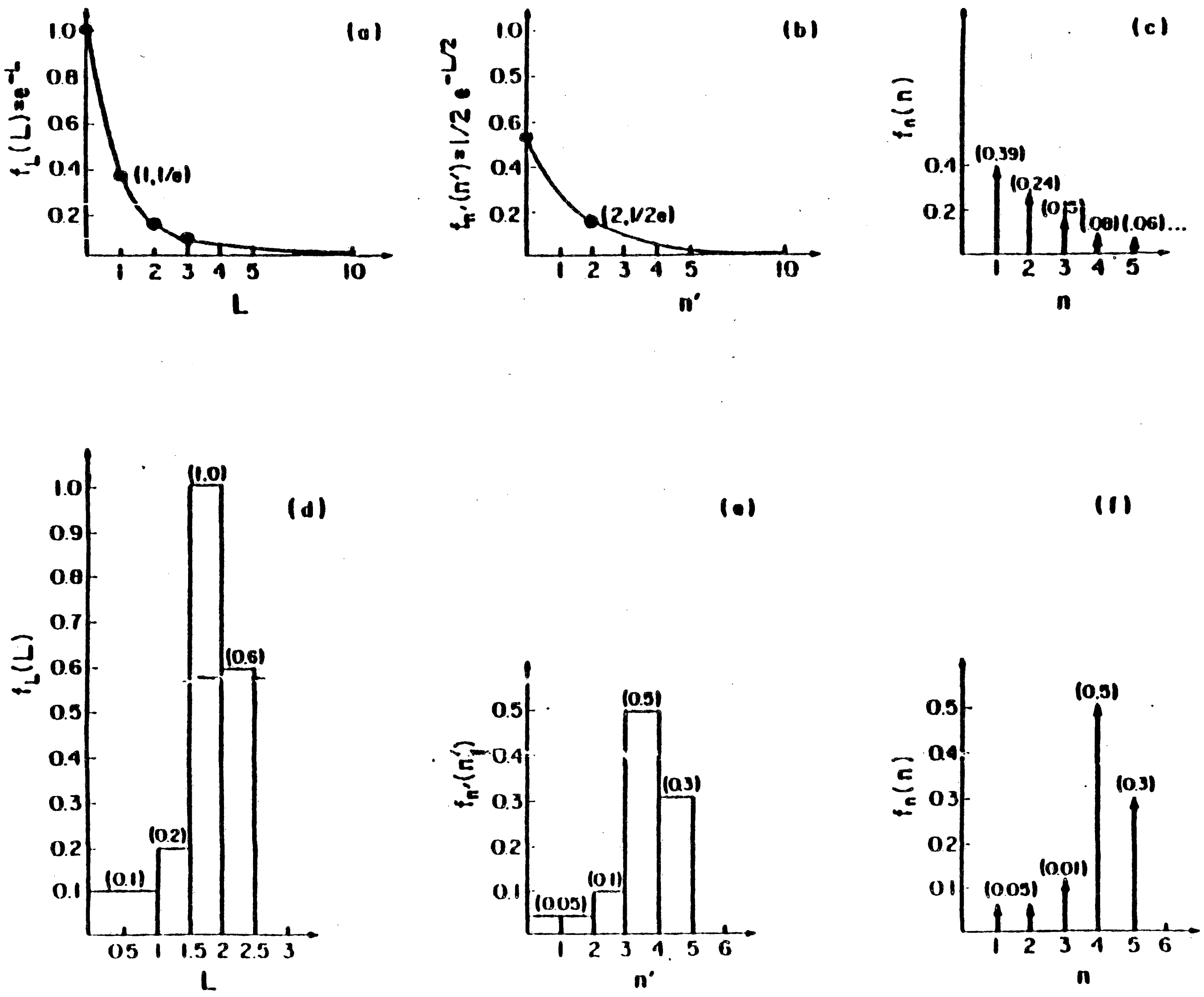


Figure 5-2. PDF's of L, n', n for Two Illustrative Cases.

product of D and L greater than 2 and less than or equal to 3 will require 3 Kanbans according to equation (5-1). The discretized version of Figure 5-2b is shown in Figure 5-2c, and that of Figure 5-2e is shown in Figure 5-2f.

It was mentioned in step 2 that all boundaries to construct the histogram estimating $f_L(L)$ would be conditioned on results discussed in this step (5-4). Since $f_n(n)$ is a discrete distribution taking on values only at the nonnegative integers, the cell boundaries in estimating $f_L(L)$ should be set so that when $f_{n'}(n')$ is constructed, no cell contains any integer as an interior point. This may be accomplished by prohibiting cells for $f_L(L)$ from containing $j/\hat{D}$ as an interior point, where $j = 1, 2, \ldots$. For example, if $\hat{D} = 10$, then the points $L = 0.1, 0.2, 0.3, \ldots$ should not be inside any cell, but rather should be at cell boundaries.

Step 5. Determine the Minimum-Cost Number of Kanbans

In this step the density of n ($f_n(n)$) is utilized to determine the number of Kanbans to use for this item at this workcenter for the next cycle period. To illustrate the procedure, consider the hypothetical pdf $f_n(n)$ given in Figure 5-3. This density states that according to our best estimates, 40% of the time during the next demand period we will need 1 Kanban at this workcenter, another 40% of the time we will need 2, and the rest of the time (20%) we will

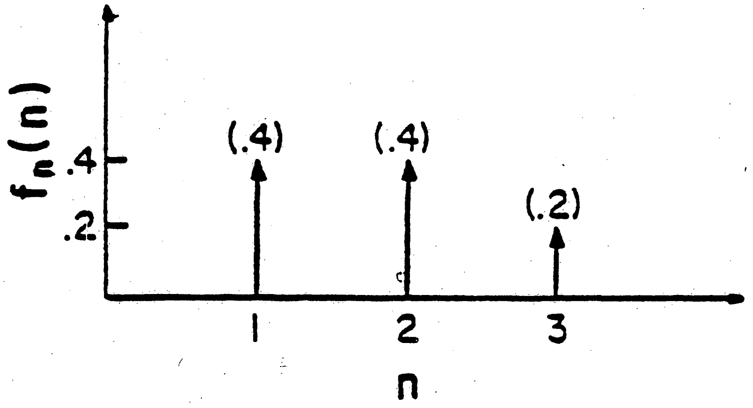


Figure 5-3. A Hypothetical PDF $f_n(n)$ Used to Illustrate Step 5.

need 3 Kanbans. Since we don't know <u>when</u> during the next cycle period we will need these differing number of Kanbans, we select a single Kanban value and implement it for the entire period.

In general there are two costs we may incur if we implement a single Kanban value when we have a pdf whose random variable takes on more than one value (e.g., n = 1, 2, or 3 as in Figure 5-3): a holding cost and a shortage cost. For example, if we implement 2 Kanbans for this item at this workcenter next period, then 40% of the time we will have one Kanban too many and 20% of the time we will have one too few. If we implement 1 Kanban, we will be 1 Kanban short 40% of the time and 2 Kanbans short 20% of the time, but we will not ever hold too many Kanbans. Likewise, if we implement 3 Kanbans, we will be 2 Kanbans "over" 40% of the time and also 1 Kanban over 40% of the time, but we will never be short. If we denote c_s as the cost of a shortage at a workcenter/container/unit time and c_h as the holding cost at a workcenter/container/unit time, then total cost computations may be made for each possible number of Kanbans. These calculations are summarized in Table 5-1.

The breakeven point between 1 and 2 Kanbans occurs (see Table 5-1) when

$$0.8c_s = 0.4\ c_h + 0.2\ c_s$$

Table 5-1. Cost Calculations Using the PDF for the Number of Kanbans of Figure 5-3.

n	Holding Cost	Shortage Cost	Total Cost
4	$0c_h$	$0.4(1)c_s + 0.2(2)c_s$	$0c_h + 0.8c_s$
5	$0.4(1)c_h$	$0.2(1)c_s$	$0.4c_h + 0.2c_s$
6	$0.4(2)c_h + 0.4(1)c_h$	$0c_s$	$1.2c_h + 0c_s$

or when $c_s/c_h = 2/3$, while the break point between 2 and 3 Kanbans occurs when

$$0.4\ c_h + 0.2\ c_s = 1.2\ c_h$$

or when $c_s/c_h = 4$. Thus minimum cost for the pdf of Figure 5-3 implies

Condition	Optimal Number of Kanbans
$c_s/c_h \leq 2/3$	1
$2/3 \leq c_s/c_h \leq 4$	2
$c_s/c_h \geq 4$	3.

Step 5 may be stated as follows: Estimate c_s and c_h and determine the number of Kanbans which minimizes the sum of holding and shortage costs given the pdf for n determined in step 4.

Note that if $c_s/c_h \to \infty$ then step 5 implies that the optimal number of Kanbans is the largest value the pdf takes on (3, for the example), since shortages are so costly that none should ever be incurred. Similarly, if $c_s/c_h \to 0$, then the optimal number of Kanbans is the minimum value the pdf realizes (1, in the example). Since in most JIT shops the cost of a shortage significantly exceeds that of holding at most non-final stage workcenters, we will discuss in detail the special case $c_s/c_h \to \infty$ shortly.

Step 6. Action Step

Set the number of Kanbans for the item at this workcenter to the value determined in step 5. Note that if the number of Kanbans is to be increased, then the additional containers of goods will have to be produced in the upcoming period during what would have been idle (i.e., non-processing) time.

Step 7. Settle Period

The purpose of this step is to ensure that the workcenter has sufficient time to settle down after step 6 has been implemented and before step 1 is repeated. If the number of Kanbans is reduced by step 6, then there is no transient settle period as the excess Kanban cards are merely pulled off processed containers and not recirculated. However, if the number of Kanbans is increased by step 6, then there will be a transient period of expected length,

$$E(\text{length}) = \frac{1}{(1\text{-util})} \left(\sum_i \Delta n_i^+ \, PT_i \right)$$

where Δn_i^+ is the increase in Kanbans for item i at this workcenter, util is the utilization of the workcenter, and PT_i is the average container processing time for item i. (The index i ranges over all items at the workcenter.)

Return to step 1 (i.e., the methodology repeats).

5.1.1 SPECIAL CASE OF THE METHODOLOGY: SHORTAGE COSTS OVERWHELM HOLDING COSTS

It is frequently the case that $c_s/c_h \to \infty$ or at least that $c_s/c_h > M$, where M is the largest breakeven point determined from $f_n(n)$ as in step 5. The ratio c_s/c_h is often large in a JIT shop because when a shortage of processed goods occurs at a workcenter, it can delay work at workcenters located toward the finished product as well as hold up workcenters that furnish materials for this workcenter. (Recall that in a JIT system preceding workcenters cannot make materials for a workcenter until authorized to do so. If there is a shortage at a workcenter, authorizaton at preceding workcenters may be delayed until the shortage is removed and there is a need to replace pre-process goods).

When shortage costs are considerably greater than holding costs the above methodology may be greatly simplified. Note that if $f_L(L)$ has finite support, i.e., is bounded by b above and below by a (even if a and b are unknown), then,

$$f_n(n) = \hat{D} f_{L_{max}}(L_{max}).$$

This phenomenon was described in passing while discussing step 5 and results from the fact that shortages are so expensive that the maximum reasonable number of Kanbans should be chosen.

It is known (see, for example, [31]) that the pdf of the maximum of a sample of size m is related to the density f and the cumulative distribution function (cdf) F from which the sample is taken as follows:

$$f_{x(m)}(x) = f_{xmax}(x) = m(F_x(x))^{m-1} f_x(x) \quad (5\text{-}4)$$

and that the cdf's are related by:

$$F_{x(m)}(x) = (F_x(x))^m \quad (5\text{-}5)$$

Here the subscript x(m) denotes the m-th order statistic. Note that if F has finite support on some interval [a, b], an increase in m in equations (5-4) and (5-5) moves more and more of the probability mass of $f_{xmax}(x)$ in equation (5-4) toward the point b.

Since any "real-world" distribution of leadtime will be bounded below by zero and above by some finite, perhaps unknown number b, $f_L(L)$ has finite support. And since m 0 100 observations are taken during measuring period 2, the density function of L_{max} will be a single-valued random variable with all its mass (p = 1.0) at the point $\hat{L}_{max}$, which is the maximum of the collected leadtimes. $\hat{L}_{max}$ should be at (or very close to) the point b. Thus

$$f_{L_{max}}(L) = \delta(L - \hat{L}_{max}) \; O \; \delta(L - b),$$

where $\delta(x - a)$ denotes an impulse of area one at the point $x = a$.

As a result

$$f_n(n) = \delta(n - [\hat{D}\,\hat{L}_{max}]).$$

This equation says that the density function of the number of Kanbans consists of a single mass at the value $n = [D\,L_{max}]$. Thus to determine the number of Kanbans, one merely needs to forecast D and to find the maximum leadtime over the measurement period. For example, if $D = 5$ and $L_{max} = 0.32$, then the number of Kanbans is

$$n = [\hat{D}\,\hat{L}_{max}] = [(5)(0.32)] = [1.6] = 2.$$

Due to the above simplifications the methodology for the special case becomes:

<u>Step 0. Startup</u> - same as before.

<u>Step 1. Measuring period 1</u> - same as before.

<u>Step 2. Measuring period 2</u> - collect observations as before. Now rather than estimating $f_L(L)$, estimate L_{max}, and use the largest L obtained as this estimate. (Alternatively, a service level may be specified. For example, if a 95% service level is specified and 100 observations are taken, $L_{0.95}$, the

95th percentile of $f_L(L)$, may be estimated using $L_{(95)}$, the 95th order statistic. In this manner one is stipulating that shortages projected for 5% of the time at the workcenter are acceptable, as discussed earlier.)

Step 3. Forecast Demand - same as before.

Step 4. Determine the pdf for the Number of Kanbans - omit.

Step 5. Determine the Minimum-Cost Number of Kanbans - this now becomes the simple calculation,

$$n = [\hat{D}\,\hat{L}_{max}]. \qquad (5\text{-}6)$$

Step 6. Action Step - same as before.

Step 7. Settle Period - same as before.

5.2 CASE EXAMPLES

Three examples will be presented in the remainder of the chapter. The first is furnished to illustrate the details of the methodology. The last two are used to demonstrate that the methodology works well and are also of interest in their own right.

The three examples are furnished in the context of the example shop described in Chapter 2. The container processing times and setup times for each product are shown in Table 5-2. In-process item costs are shown in Table 5-3. The holding cost is 25% of container cost per year, the backorder cost is $5.63 per container per hour, and, the setup cost is

Table 5-2. Container Processing Times and Shop Times for the Example Shop.

Workcenter Number	Container Processing Time (hrs.)		Setup Time
	X	Y	
1	.0936	.0468	.0036
2	.0871	.0436	.0033
3	.2180	.1090	.0088
4	.1453	.0727	.0055
5	.3526	N/A	N/A
6	N/A	.3350	N/A

Table 5-3. Item Costs for the Example Shop.

Item	Unit Cost
A	$150
B	150
C	150
D	150
E	412.5
F	37.5
H	37.5
I	300
J	450
K	37.5
L	37.5
X	937.5
Y	787.5

$30 per hour. The shop is assumed to operate 8 hours per shift, 2 shifts per day, 5 days per week. Demand is 19 containers/shift for final product X and 20 containers/shift for final product Y. The coefficient of variation (CV) of processing times is 0.4 at each workcenter in the shop. With such a large CV it is assumed that the shop is relatively new to JIT and the workers are not yet totally accustomed to procedures; hence the large variances in processing times relative to means. Using these parameters the example shop runs at approximately 84% utilization.

5.2.1 EXAMPLE 1: ILLUSTRATION OF THE METHODOLOGY

The example shop was simulated at 84% utilization (as previously indicated), and also, with processing times increased to yield a 91% utilization case as well. The CV was 0.4 in both cases. It should be noted that the latter case represents a shop pushed close "to the limit"; it has a high utilization rate in addition to a very high variance in processing times.

The methodology is now illustrated step by step using workcenter 6 (which makes final product Y) as an example. Other items at other workcenters are similar.

Step 0. Startup - Through trial and error it was determined that a 1-week startup period was more than sufficient for transient effects to die out.

Step 1. Measuring period 1 - Over 100 observations (in fact 400 over 2 weeks) were utilized to generate correlograms using equation (5-2). These correlograms are shown in Figure 5-4a for the 84% utilization case and Figure 5-4b for 91% utilization. The correlograms are decidedly different and both show appreciable autocorrelations at large lags. (This autocorrelation behavior is due to the large CV in the shop which causes bursts of long leadtimes followed later by bursts of short leadtimes).

Step 2. Measuring period 2 - Utilizing the data that generated the correlograms shown in Figures 5-4a and 5-4b, it can be observed that beyond lag 10 no autocorrelation exceeds 0.05 for the 84% utilization case, and the same is true beyond lag 19 for the 91% case. Thus in measuring period 2, "independent" observations spaced every 10 leadtimes (i.e., 2 per shift) were collected for the lower utilization case, and observations were spaced every 19 (actually rounded to 20) leadtimes (1 per shift) for the other case.

One hundred "independent" observations were collected for each case, which took 25 days in the former case and 50 in

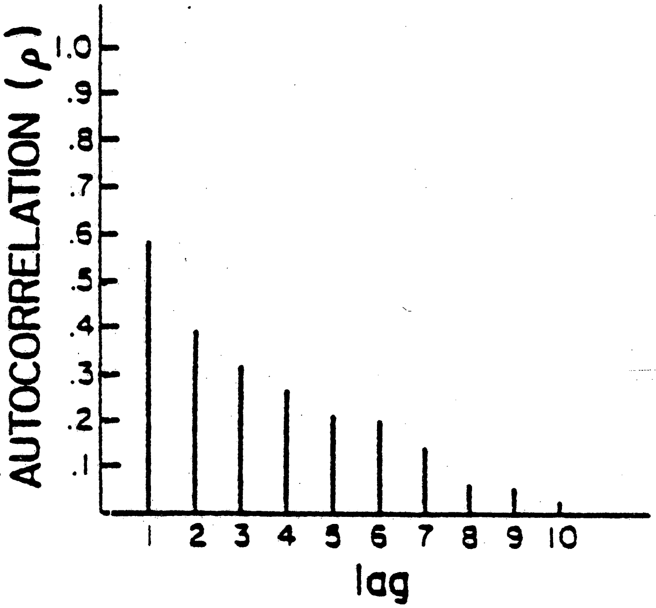


Figure 5-4a. Autocorrelation of Leadtime at a Workcenter (84% util.).

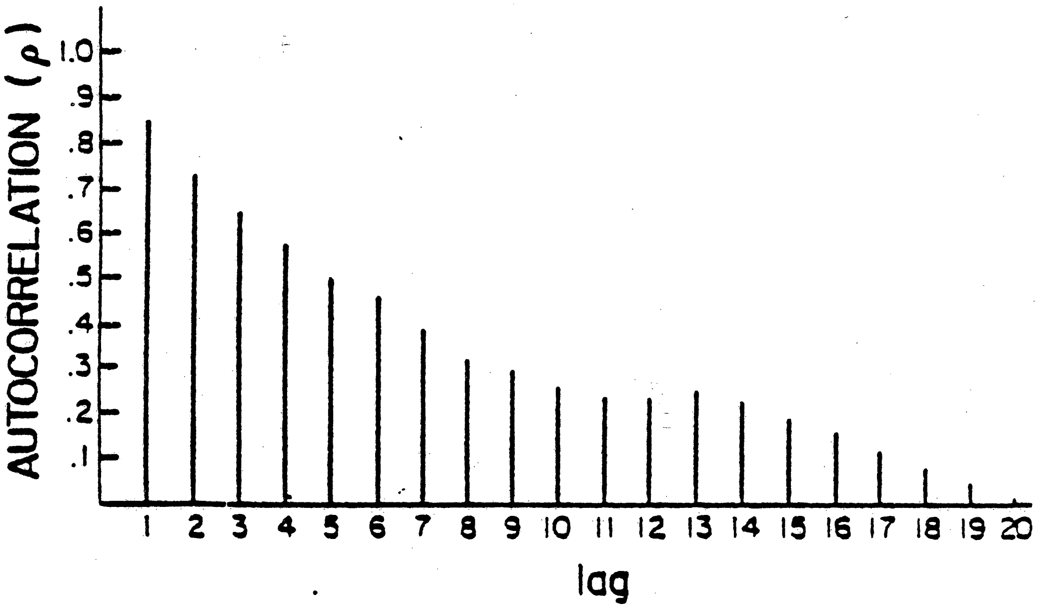


Figure 5-4b. Autocorrelation of Leadtime at Workcenter 6 (91% util.)

the latter. The resulting pdf's for $f_L(L)$ are shown in Figures 5-5a and 5-5b. Note that increasing utilization dramatically changes $f_L(L)$. (For the reader's information, a third case with the shop at 84% utilization and CV = 0.2 was analyzed. This case showed that independent leadtimes could be collected using every 3rd observation, thus greatly shortening the length of measuring period 2.)

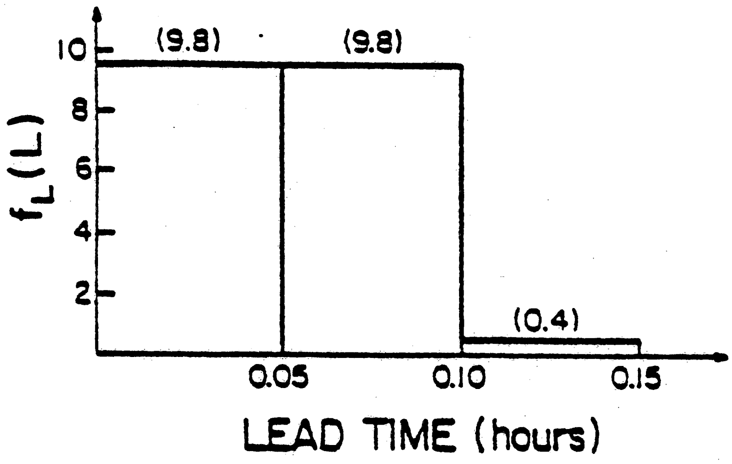
Step 3. Forecast Demand - Since demand is constant for final product Y at workcenter 6, demand for the next demand cycle period will be 20 containers/shift.

Step 4. Determine the pdf for the Number of Kanbans - The density function for n' is obtained via equation (5-3) as,

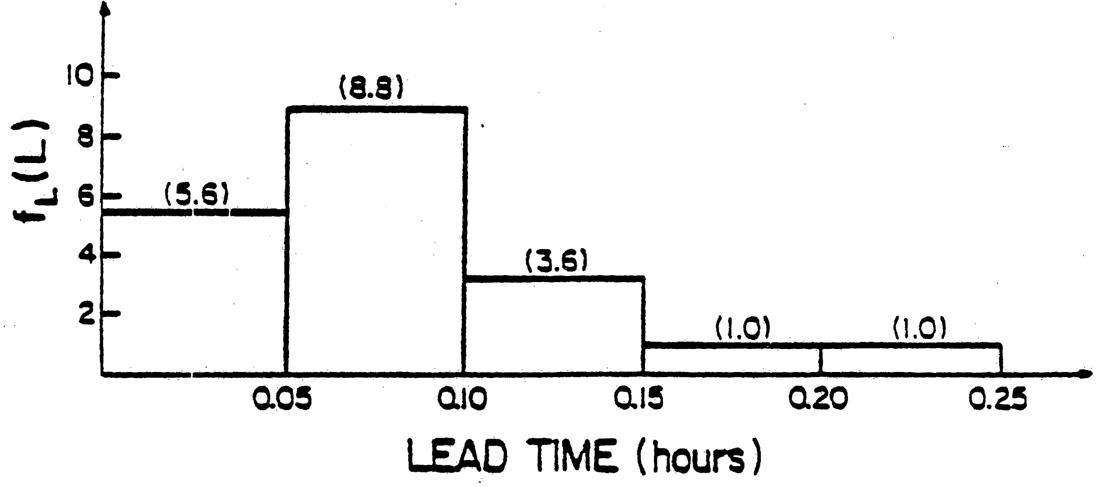
$$f_{n'}(n') = \frac{1}{20} f_L\left(\frac{L}{20}\right).$$

Thus $f_{n'}(n')$ is scaled down vertically (by a factor of 20) and enlarged horizontally (also by a factor of 20) from $f_L(L)$. The densities $f_{n'}(n')$ are shown in Figures 5-6a and 5-6b. The discrete densities $f_n(n)$ are shown in Figures 5-7a and 5-7b.

Step 5. Determine the Minimum-Cost Number of Kanbans - If $c_h = 1$ and $c_s = 10$, then the total cost calculations for the

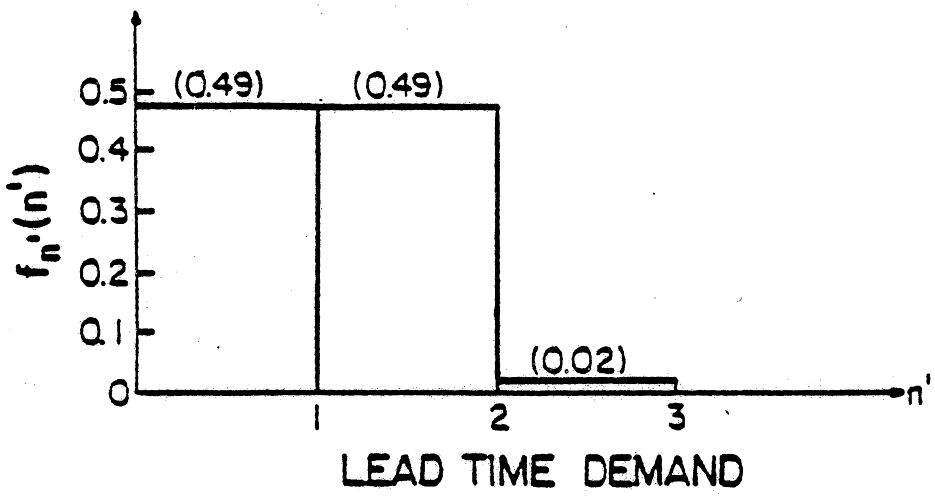


a) 84% Utilization, CV=0.4

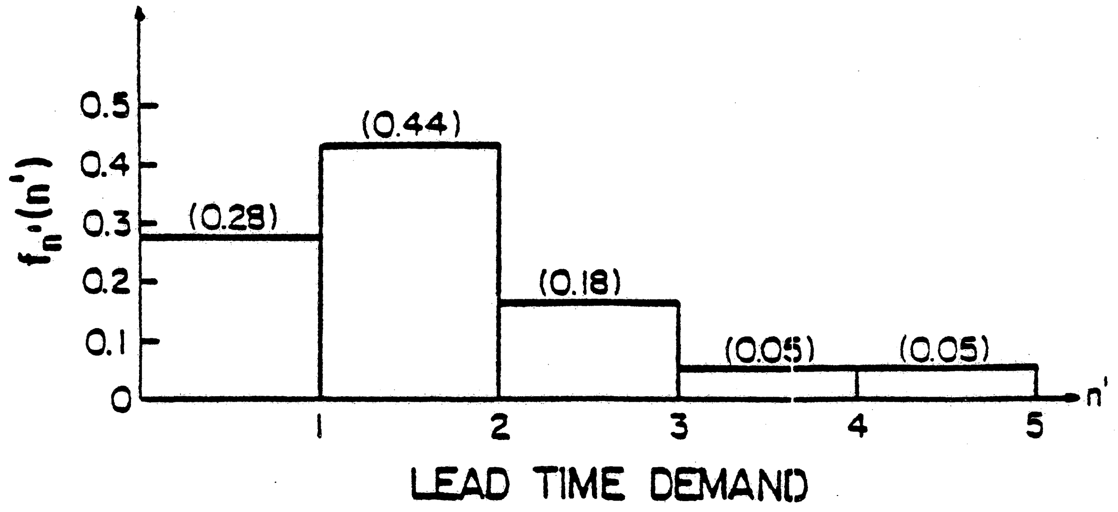


b) 91% Utilization, CV=0.4

Figure 5-5. Histogram Estimates of $f_L(L)$ at Workcenter 6.

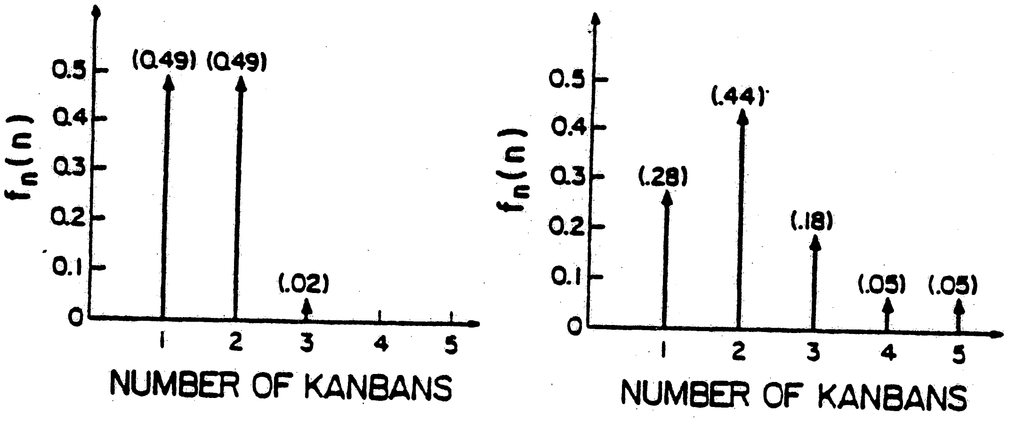


a) 84% Utilization, CV=0.4



b) 91% Utilization, CV=0.4

Figure 5-6. The PDF's of Leadtime Demand, $f_{n'}(n')$, at Workcenter 6.



a) 84% Utilization, CV=0.4

b) 91% Utilization, CV=0.4

Figure 5-7. The PDF's of Number of Kanbans at Workcenter 6.

91% utilization case for n = 1, 2, 3, 4, and 5 Kanbans are as shown in Table 5-4. The minimum-cost number of Kanbans to be used for the next period is 4 for the given values of c_h and c_s.

Step 6. Action Step - Implement 4 Kanbans at workcenter 6 for the next demand period for the 91% case.

Step 7. Settle Period - If there were 3 Kanbans at workcenter 6 during data collection, the expected length of the settle period would be,

$$E(\text{length}) = \frac{\Delta n^{+}(PT_y)}{(1\text{-util})} = \frac{(4 - 3)(0.335 \text{ hours})}{0.09} = 3.72 \text{ hours}$$

It would be reasonable to let the shop settle for one shift, and possibly two, to be safe.

The methodology now returns to step 1 (measuring period 1) with 4 Kanbans circulating at workcenter 6.

5.2.2 EXAMPLE 2: ADJUSTING TO KANBAN MISSPECIFICATIONS

This example explores how well the methodology adjusts to a one-time misspecification in the number of Kanbans. The incorrect number of Kanbans could have been circulated at a

Table 5-4. Minimum Cost Calculations at Workcenter 6 for the 91% Utilization Case Assuming $c_h=1$ and $c_s=10$.

n	Holding Cost	Shortage Cost	Total Cost
1	0	.44(1)(10) + .18(2)(10) + .05(3)(10) + .05(4)(10)	11.50
2	.28(1)(1)	.18(1)(10) + .05(2)(10) + .05(3)(10)	4.38
3	.28(2)(1) + .44(1)(1)	.05(1)(10) + .05(2)(10)	2.50
4	.28(3)(1) + .44(2)(1) + .18(1)(1)	.05(1)(10)	2.40*
5	.28(4)(1) + .44(3)(1) + .18(2)(1) + .05(1)(1)	0	2.85

workcenter, for example, due to a badly missed forecast, or because of a lack of knowledge on how to initialize a system with Kanbans. This latter case not only happens when a shop "opens its doors" for the first time, but also can arise after changes occur in the shop, such as with new or altered products, different routings, etc.

Using previously developed techniques (by the author) it can be determined that the initial total number of Kanbans needed in the sample shop is approximately 35, i.e., about 6 per workcenter on average. However, this initial number of Kanbans was deliberately misspecified and the shop was loaded with 72 Kanbans for this example.

As indicated in Figure 5-8, the shop ran for one period at 72 Kanbans while data were being collected. The methodology predicted that 34 Kanbans should be used in the shop the second period and that number was implemented. The number of Kanbans utilized during the rest of the two-year simulation oscillated between 34 and 38 Kanbans. (This oscillation was caused by differences in leadtime estimates created by variation in processing times from period to period).

This example demonstrates that the shop can adjust very quickly to incorrect Kanban numbers. A further implication for shop supervisors and management is that one need not be overly concerned with getting the number of Kanbans exactly

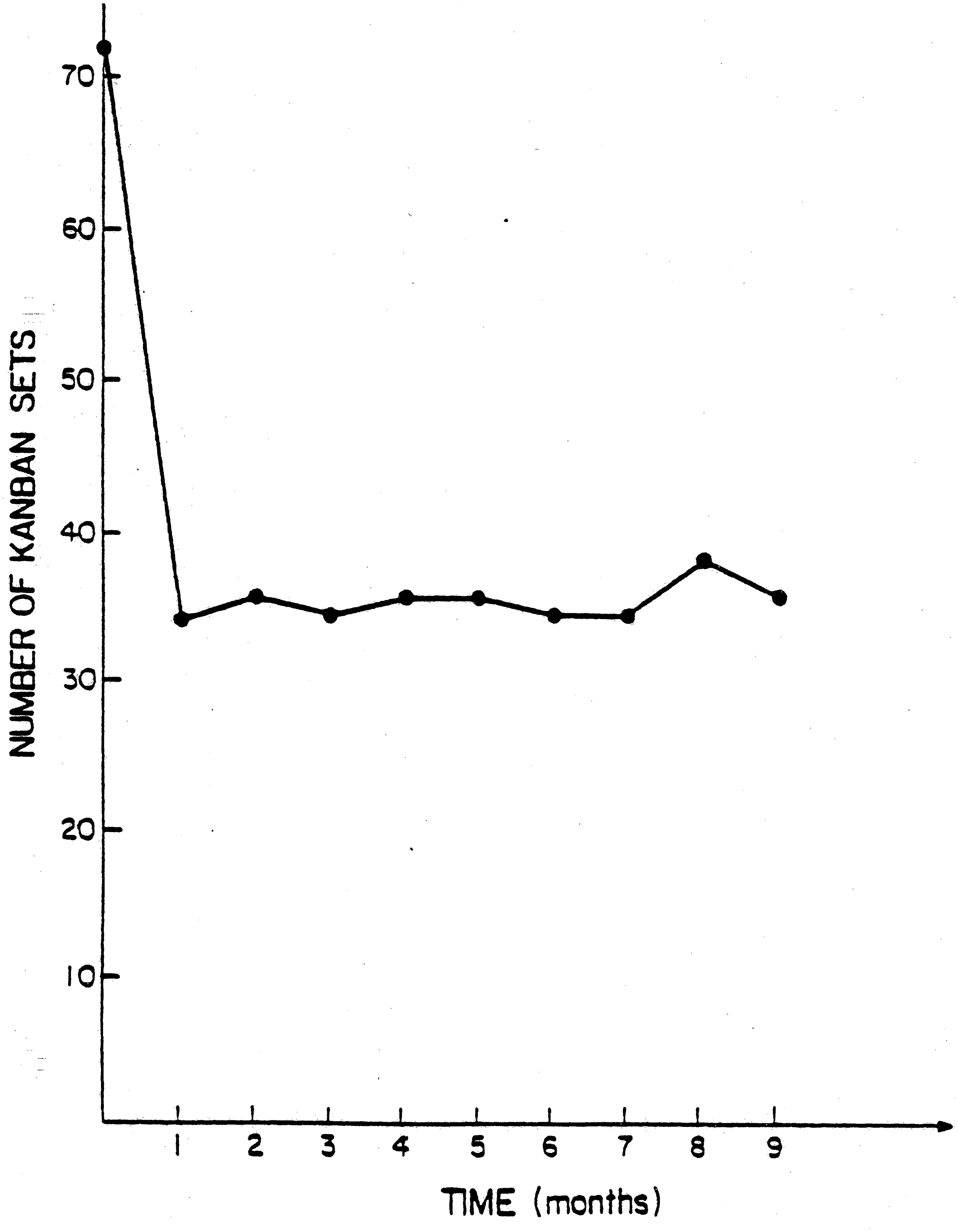


Figure 5-8. Example Illustrating the Effects of Misspecifying Initial Conditions.

"right" during startup or after major changes if one starts with plenty of them. The methodology will adjust to proper levels quickly.

5.2.3 EXAMPLE 3: A TRAINING CASE EXAMPLE

In this example the cost effects of training workers are studied. In the first scenario, there is no training at all, and the shop is run under the conditions described above. This is the baseline case. In the second scenario, workers are trained relatively slowly. The training is two-pronged in its focus: the first aspect is to teach workers to reduce setup times, and the second is to develop in them the ability to meet processing time specifications more consistently. The net effect of this gradual training is that setup times decrease linearly over two years to one-half their original values (which were given in Table 5-2), and, the coefficient of variation (CV) of processing times also decreases linearly over the two years from 0.4 to 0.2. In the third scenario, workers are trained under a one-year concentrated program. The effects are the same as in the second scenario, except that the results take one year instead of two. Thus, CV and setup times reach their terminal values at the end of year 1. They do not change further during the second year.

Cost results for the three scenarios of worker training are shown in Table 5-5. Note that in this table the cost of the worker training itself is not included. Excluding this cost, "fast" training provides approximately $50,000 in savings over the no-training case. The savings come both from reductions in setup costs and inventory costs. Once the workers are trained these savings should accrue indefinitely. The slow training case does provide about $35,000 improvement over the no-training case, but saves less money than the fast-training scenario.

For the interested reader the methodology functioned as follows for the three training cases. In the fast-training case the number of Kanbans was reduced to 24 by the end of the first year, while it took two years for the same result in the slower-training case. The no-training scenario had a "steady-state" value of approximately 35 Kanbans.

5.3 SUMMARY

This chapter has demonstrated how manufacturers that do not enjoy the luxury of cross-trained workers, large market shares, and well-run and understood JIT shops can still function in a Kanban environment. A methodology was developed for dynamically adjusting the number of Kanbans that

Table 5-5. Cost Results for the Three Scenarios of the Worker-Training Example.

Cost	No Training	Slow Training	Fast Training
Inventory	200,700	173,987	162,224
Setup	31,754	23,803	19,804
Backorder	0	0	0
Total	232,454	197,790	182,028

utilized forecasted demand and estimates of container leadtime probability density functions.

Questions of actual implementation of the methodology, such as should all measuring periods be synchronized across the entire shop, were not addressed. It was deemed that they were more properly resolved within the actual constraints of a given shop.

The methodology is "safe" to use when sample sizes recommended at each step are used. PDF estimates are not as critical in this methodology as they might be in other application areas because of the integer-nature of n. Each leadtime observation, for example, needs to be placed in one of a few, wide cells.

Some poorly run shops or other shops with severe internal perturbations will find that so many observations must be collected, that they can only safely adjust the number of Kanbans once every several months. If demand or conditions in the shop change more rapidly than this, these shops should not implement Just-in-Time. Just-in-Time is obviously not a system for everyone. However, as was illustrated in the final two examples, Kanbans can be dynamically adjusted in shops in a sound and successful manner. For the firm still learning about Just-in-Time, or the firm not possessing all the cultural and technical luxuries of a Toyota, this is a potentially significant result.

Chapter 6 concludes this study by summarizing the research performed in light of the overall objectives of this dissertation and by suggesting further research endeavors regarding JIT systems.

6.0 SUMMARY AND CONCLUSIONS

This research has investigated the feasibility of operating a Just-in-Time with Kanbans system within a sub-optimal production environment typical of the environment faced by many American manufacturing concerns. Analytical approaches were developed for dealing with some of the problems caused by this sub-optimal environment. These approaches were tested using a Q-GERT simulation model of a hypothetical shop with both job and assembly operations employing a Just-in-Time with Kanbans system.

The first part of this investigation examined whether a JIT with Kanbans system should be attempted in an environment where setup times cannot be reduced to levels that allow for single container lotsizes at all workstations. There are three important results from this section:

1. An American company which cannot reduce the setup times at all workstations can still utilize JIT with Kanbans by using Signal Kanbans at problem workstations.

2. A lotsizing technique using mathematical programming has been demonstrated.

3. Inventory and setup costs must be weighed against each other rather than simply seeking to reduce inventory to its minimum level. Further, under certain conditions, a Signal Kanban system at a workstation can be more cost-effective than a feasible standard Kanban system.

The second part of this investigation studied both the factors present in the typical American production environment that influence the number of Kanbans at a workstation and how the initial number of Kanbans throughout a shop should be determined. This section of the research makes three important contributions.

1. A descriptive model of the relationship between the number of Kanbans required and the variability in processing time, autocorrelation in processing times, workstation utilization and throughput velocity is derived and successfully verified using the example shop.

2. A method for determining the initial number of Kanbans is demonstrated and successfully tested using the example shop.

3. It is demonstrated that JIT with Kanbans can be used in a manufacturing environment where the machine utilization rate is high and there is a high degree of variability

and autocorrelation in processing times if a company is willing to carry enough inventory. The cost of carrying this large amount of inventory may, however, cause the JIT with Kanbans system to be unprofitable or to be more costly than alternative systems such as MRP.

The first two parts of the investigation when taken together suggest that JIT with Kanbans can be used even if the shop is operating under severe internal conditions such as long setup times, high machine utilization rates, and processing times with a high degree of variability and autocorrelation. However, it has been assumed thus far that even though shop conditions may be less than desirable for JIT operation, they remain stable as do demand rates for final products.

The third part of the investigation examined dynamically adjusting the number of Kanbans for the case where demand and shop conditions are not stable. There are two important results from this section:

1. A workable methodology for dynamically adjusting the number of Kanbans is presented and demonstrated in the example shop.

2. The ability of the shop to adjust to changes corresponds directly to how well the shop is functioning. If shop conditions are good, i.e., variability and

autocorrelation in processing times and utilization rates are low, the shop can be adjusted every several weeks. However, if the shop conditions are poor (i.e., the above factors are at high levels), the adjustment cycle may take several months.

The results from the three parts of the investigation when taken together suggest that in an environment where demand and shop conditions are fairly stable, JIT with Kanbans can still work despite high setup times, high variability in processing times, autocorrelation in processing times, and high utilization rates - conditions that one would expect to find in many American firms considering the implementation of JIT with Kanbans . These conditions can cause a significant increase in the amount of inventory required, but this amount can be determined using the methodology which has been presented. This conclusion in no way implies that JIT with Kanbans should always be used, or that it is more cost-effective than say MRP; it merely says that Kanbans is feasible. Operating under these less than desirable conditions, however, will in all likelihood limit the ability of the shop to adjust to an unstable environment such as when the firm is unable to level the master production schedule, achieve consistent performance from its workforce due to high employee turnover or labor strife, and so forth.

Monden suggests [34] that for the JIT with Kanbans system to be successfully applied, the shop conditions must be "under control" and demand must be relatively stable - conditions which Toyota enjoys. The results from this dissertation suggest that JIT with Kanbans is physically feasible over a considerably broader region than Monden says Toyota uses. "Feasible" here means not only that JIT has potential to work, but rather that it will work <u>and</u> the level of inventory carried can be computed using the methodology developed.

The implications of this suggestion can be further delineated using Figure 6-1. In that figure the horizontal axis corresponds to how good shop conditions are, where by "shop conditions" is meant the level of machine utilization, the amount of variability and autocorrelation in processing times, and the magnitude of throughput velocity. Recall that in the second part of this investigation, these factors were combined analytically into a Z_p-score: a high Z_p-score indicates that shop conditions are "good", while a low Z_p-score suggests that conditions are "poor." The vertical axis in Figure 6-1 represents the stability of demand and shop conditions. Note that this axis reflects the rapidity with which demand and shop conditions are changing, rather than the absolute level of these factors. Monden suggests that the company considering use of JIT with Kanbans operate within the small box-shaped area in the figure. The results presented

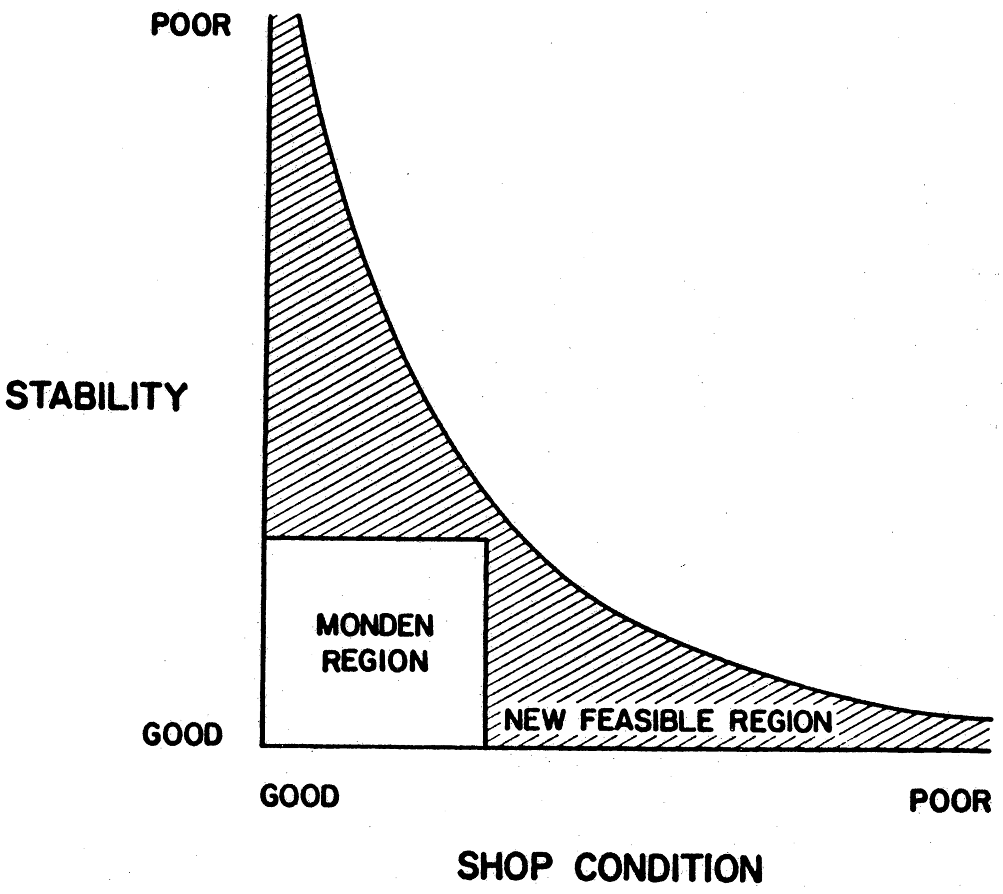


FIGURE 6-1. Feasible Region for JIT with Kanbans.

here indicate that Kanban is feasible and inventory levels are predictable for the entire region under the curve - - a much larger region then the Monden (Toyota) region.

The implication of this finding is twofold. First, a firm can employ JIT with Kanbans even if its demand and shop conditions are somewhat unstable, as long as shop conditions are good in the sense of high Z_p-scores. For example, a shop with high Z_p-scores at its workstations can tolerate master production schedules that cannot be frozen for as long as desired. Such a shop would be able to adjust fairly rapidly to these demand changes (using the methodology presented in the third part of this research) since the adjustment cycle can be relatively short due to the high Z_p-scores. Second, a firm can use JIT with Kanbans even if the manufacturing involves "poor" shop conditions (in the sense of a low Z_p-score) as long as demand and shop conditions are stable.Such a firm would require a long adjustment cycle in dynamically changing the number of Kanbans, but this would not be a problem since demand and shop conditions' stability do not require frequent adjustments. However, if shop conditions are poor and instability is present, JIT with with Kanbans is not advisable because frequent adjustments and long adjustment cycles are required by the methodology. It is impossible to take a long time to adjust and adjust frequently. Hence, the region above the curve in Figure 6-1 is inappropriate for JIT endeavors.

The enlarged JIT with Kanban feasible region is potentially good news to firms considering a change to JIT with Kanbans . They can make the change to JIT with Kanbans now and accrue cost savings as they improve shop conditions and stabilize demand (or as they get closer to the origin in Figure 6-1).

6.1 EXTENSIONS AND FUTURE RESEARCH

These results point the way for future research in Just-in-Time with Kanbans systems. First, while this research has investigated the feasibility of using Kanbans under varying conditions and the setting of inventory levels to handle these conditions, it has not examined in detail the profitability of doing so. It is more than likely that the costs of operating in some areas of the feasible region will be financially prohibitive. Additional research is needed to determine in which portions of Figure 6-1 JIT with Kanbans does have significant advantages over other sustems such as MRP. Studies that have been performed comparing JIT to MRP have taken the whole JIT with Kanban system and tried to apply it blindly in an American environment. The focus of this research has been to adapt JIT to work in an American environment. One path for future research is to compare MRP with a JIT system that incorporates the results from this study.

A second area of research, very pragmatic in nature is to validate this research on actual physical systems. In particular, it would be informative to install and then examine an actual implementation of dynamic Kanban adjustment. "Real world" measurements of processing time autocorrelations and adjustment cycle lengths could be made, thereby establishing the use of the methodology as is, or demonstrating the need for more robust probability density function estimation.

BIBLIOGRAPHY

1. Anderson, John C., Schroeder, Roger G., Tupy, Sharon E., White, Edna M., Material Requirements Planning Systems: The State of the Art, Working Paper, Department of Management Sciences, University of Minnesota, 1981.

2. Blasingame, John W, and Weeks, James K, "Behavioral Dimensions of MRP Change: Assessing Your Organization's Strengths and Weaknesses", Production and Inventory Management, Vol. 22, No. 1, pp. 81-95, January, 1982.

3. Box, G. E. P., and Jenkins, G. M., Time Series Analysis : Forecasting and Control (revised edition), Holden-Day,San Francisco, 1976.

4. Brenizer, Ned W., "The Odyssey of Inventory Management", Production and Inventory Management, Vol. 22, No. 2, pp. 23-36, April, 1982.

5. Buffa, Elwood S. and Taubert, William L., Production-Inventory Systems: Planning and Control, Homewood, Illinois, Richard D. Irwin Inc., 1972.

6. Burgam, Patrick M., "JIT: On the Move and Out of the Aisles", Manufacturing Engineering, Vol 42, No. 6, pp.65-71, June, 1983.

7. Burck, Charles G., "Can Detroit Catch Up?", Fortune, Vol. 105, No. 3, pp. 34-39, July, 1982.

8. Cole, Robert E., "Learning From the Japanese: Prospects and Pitfalls", Management Review, Vol. 69, No. 9, pp. 22-42, September, 1980.

9. _____, "Will Quality Circles Work in the U.S.?", Quality Progress, Vol. 13, No. 7, pp. 30-33, July, 1981.

10. Fitzsimmons, James A. and Sullivan, Robert S., Service Operations Management, McGraw-Hill Book Company, New York, 1982.

11. French, Robert L., "Accurate Work In-Process Inventory - A Critical MRP System Requirement", Production and Inventory Management, Vol. 21, No. 1, pp. 17-23, January, 1981.

12. Goddard, Walter E., "Kanban versus MRP - Which Is Best For You?", Modern Material Handling, Vol. 37, No. 16, pp.40-47, August, 1982.

13. Groover, Mikell P., Automation, Production Systems, and Computer Aided Manufacturing, Englewood Cliffs, N.J., Prentice Hall, 1980.

14. Hadley, G. and Whitin, T. M., Analysis of Inventory Systems, Englewood Cliffs, N.J., Prentice Hall, 1963.

15. Hall, Robert W., Zero Inventories, Homewood, Illinois, Dow-Jones Irwin, 1983.

16. Hayes, Robert H., "Why Japanese Factories Work", Harvard Business Review, Vol. 59, No. 4, pp.57-66, August, 1981.

17. Huang, Philip Y., Philipoom, Patrick R., Rees, Loren P., "Some Queueing Disciplines in a Just-in-Time System with Kanbans," Proceedings of the Annual Meeting of the Southeastern Chapter of The Institute of Management Sciences, Volume 13, pp. 286-291, September, 1983.

18. Huang, Philip Y., Philipoom, Patrick R., Rees, Loren P., Taylor, Bernard W.,III, Server Selection and Queue Disciplines for a Complex Multistage Production Operation using Just-In-Time (With Kanbans), Working Paper, 1984.

19. Huang, Philip Y., Philipoom, Patrick R., Rees, Loren P., Taylor, Bernard W.,III, "Server Selection and Queue Discipline Rules for the Just-in-Time with Kanbans System in a Non-Assembly Line Environment," Proceedings of the Annual Meeting of the American Institute for Decision Sciences, (micro-fiche), November, 1984.

20. Huang, Philip Y., Rees, Loren P., Taylor, Bernard W.,III, "Simulating the Japanese Kanban Technique in an American Production Environment," Proceedings of the Annual Meeting of the American Institute for Decision Sciences, Vol. 2, pp. 113-114, November, 1982.

21. Huang, Philip Y., Rees, Loren P., Taylor, Bernard W.,III, A Simulation Analysis of the Japanese Just-In-Time Technique (With Kanbans) For a Multiline, Multistage Production System, Decision Sciences, Vol. 14, No. 3, pp. 326-343, Summer, 1983.

22. Iwata, Ryashi, Japanese-Style Management: Its Foundations and Prospects, Asian Productivity Organization, Tokyo, 1982.

23. Johnson, Lynwood A. and Montgomery, Douglas C., Operations Research in Production Planning, Scheduling and Inventory Control, New York, John Wiley and Sons, Inc., 1974.

24. Juran, J. M., "Japanese and Western Quality - A Contrast", Quality Progress, Vol. 11, No. 12, pp. 10-18, December, 1979.

25. Kimura, Osamu and Terada, Hirosuke, "Design and Analysis of a Pull System, a Method of Multi-stage Production Control," International Journal of Production Research, Vol. 19, No. 3, pp. 241-253, May, 1979. ✓

26. Krajewski, Lee J., King, Berry E., Ritzmann, Larry P., Wong, Danny S., "Kanban, MRP and Shaping the Production Environment," Proceedings of the Eighteenth Annual Meeting of the American Institute for Decision Sciences, (micro-fiche), November, 1983. ✓

27. Kohn, Robert L., "Keep Accurate Inventory Records? ...Let the Other Guy Do It!", Production and Inventory Management, Vol. 19, No. 3, pp. 92-96, July, 1979.

28. Lantham, David A. Jr., "Are You Among MRP's Walking Wounded?", Production and Inventory Management, Vol. 22, No. 3, pp. 23-36, July, 1982.

29. Law, A. M., and Kelton, W. D. Simulation Modeling and Analysis, McGraw Hill, New York, 1982.

30. Lowerre, William M., "Miracle Requirements Planning", Production and Inventory Management, Vol. 15, No. 1, pp. 50-54, January 1975.

31. Mann, N. R., Schafer, R. E., and Singpurwalla, N. D., Methods for Statistical Analysis of Reliability and Life Data, John Wiley and Sons, New York, 1974.

32. Marshall, Byron K., "Japanese Business Ideology and Labor Policy," Columbia Journal of World Business, Vol. 12, No. 1, pp. 22-29, January, 1977.

33. McClain, John O. and Thomas, Joseph, Operations Management, Production of Goods and Services, Prentice-Hall, Inc., Englewood Cliffs, 1980.

34. Monden, Yasuhiro, "'A Simulation Analysis of the Japanese Just-In-Time Technique (With Kanbans) For a Multiline, Multistage Production System': A Comment", Decision Sciences, Vol. 15, No. 3, pp. 445-447, Summer, 1984.

35. _____, "Adaptable Kanban System Helps Toyota Maintain Just-in-Time Production", IE, Vol. 13, No. 5, pp. 29-46, May, 1981. ✓

36. _____, Toyota Production System, Norcross, Georgia, Industrial Engineering and Management Press, 1983.

37. _____, "How Toyota Shortened Supply Lot Production Time, Waiting Time and Conveyance Time", IE, Vol. 13, No. 9, pp. 22-30, September, 1981.

38. Moorecroft, J. D., "A System Perspective on Material Requirements Planning", Decision Sciences, Vol. 14, No. 1, pp. 1-18, January, 1983.

39. "MRP Symposium", (Production and Inventory Management, Vol. 18, No. 2, pp. 29-39, April, 1978.

40. Naddor, Eliezer, Inventory Systems, Malabar, Florida, Robert E. Krieger Publishing Company, 1982.

41. Nellemann, David O., Smith, Leighton F., "Just-in-Time versus Just-in-Case Production Inventory Systems: Concepts Borrowed Back From Japan", Production and Inventory Management, Vol. 23, No. 2, pp. 12-21, February, 1983.

42. Orlicky, Joseph, Material Requirements Planning, N.Y., McGraw Hill, 1975.

43. Ouchi, William G., Theory Z, Reading, Massachusetts, Addison-Wesley Publishing Company, 1981.

44. Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw Hill, New York, 1965.

45. Pascale, Richard T., "Zen and the Art of Management", Harvard Business Review, Vol. 56, No. 2, pp. 153-162, March, 1977.

46. Philipoom, Patrick R.,"Lotsize Determination for Signal Kanbans," Proceedings of the Southeast Chapter of The Institute of Management Sciences, Vol. 14, pp. 19-22, September, 1984.

47. Plossl, George W., "MRP Yesterday, Today, and Tomorrow," Production and Inventory Management, Vol. 21, No. 3, pp. 1-10, May, 1981.

48. Pritsker, A. Alan B., Modeling and Analysis Using Q-GERT Networks, 2nd ed., New York, John Wiley and Sons, 1979.

49. "Productivity Out of MRP - A New Game Plan!", Modern Material Handling, Vol. 36, No. 1, pp. 65-82, January, 1981.

50. "Quality Circles - What It Takes To Make Them Work For You", Modern Materials Handling, Vol. 37, No. 17, pp. 44-47, September, 1982.

51. Ransom, G. M., Group Technology, Maidenhead, Berkshire, McGraw Hill Book Company (UK) Limited, 1972.

52. Rees, Loren P., Huang, Philip Y., Taylor, Bernard W.,III, A Comparative Analysis of MRP and the Japanese Just-In-Time Concept (With Kanbans) For a Complex Multistage Production Operation, Working Paper, 1984.

53. Reischauer, Edwin O., "The Japanese Way", Across the Board, Vol. 16, No. 12, pp. 34-42, December, 1977.

54. Rice, James W. and Yoshikawa, Takeo, "MRP and Motivation: What We Can Learn From Japan?", Production and Inventory Management, Vol. 21, No. 2, pp. 45-52, April, 1981.

55. Schmenner, Roger W., Plant Tours In Production/Operations Management, Chicago, Science Research Associates, 1984.

56. _____, Production/Operations Management: Concepts and Situations, 2nd ed., Chicago, Science Research Associates, 1984.

57. Schonberger, Richard J., Japanese Manufacturing Techniques, N.Y., The Free Press, 1982. ✓

58. Shaw, R. J. and Regentz, M. O., "How to Prepare Users For a New System", Management Focus, Vol. 27, No. 2, pp. 33-36, March, 1980. ✓

59. Schwind, Gene, "MAN Arrives Just In Time to Save Harley-Davidson", Material Handling Engineering, Vol. 39, No. 8, pp. 28-35, August, 1984.

60. Tarter, M. E., Holcomb, R. L., and Kronmal, R. A., "After the histogram what? A description of new computer methods for estimating the population density," Proceedings, Association for Computing Machinery, 22, Thompson Book Co., Washington, D.C., pp. 511-519, 1967.

61. Tarter, M.E., and Kronmal, R. A., "An Introduction to the Implementation and Theory of Nonparametric Density Estimation," The American Statistician, 30(3), pp. 105-112, August 1976.

62. Tersine, Richard J., Production/Operations Management: Concepts, Structure and Analysis, N.Y., North Holland, 1980.

63. Waters, Craig R., "Profit and Loss," Inc., 7(4), pp. 103-112, April 1985.

64. _____, MRP-II: Unlocking America's Productivity Potential, Boston, CBI Publishing Company, 1984.

65. "'Zero Inventory': Can It Restructure U.S. Industry?", Modern Materials Handling, Vol. 39, No. 1, pp. 47-49, january, 1984. ✓

66. Zimmerman, Gary W., "The ABC's of Vilfredo Pareto", Production and Inventory Management, Vol. 16, No. 3, pp. 1-9, July, 1984.