ABSTRACT

Kanban systems differ from traditional inventory oriented pull systems in that they permit consideration of the interaction between production lead time and inventory levels. In this paper, the effect of batch sizing policy on production lead time and hence on inventory levels and cell performance is studied. Single and Dual card Kanban cells and Two Stage Kanban Systems are modelled as Markovian processes and the effect of batch sizing on expected inventory and backorder costs is studied. It is shown that batching policy has a significant impact on costs. The effect of varying the card count in the cell is also studied. The analytical results agree qualitatively with results reported in simulation studies.
Introduction

While manufacturing control in U.S. companies has tended to evolve away from "pull" systems, some Japanese companies have developed types of pull control systems that appear to be very effective in overcoming some of the problems associated with such systems. Sugimori et al. (1977), Monden (1983), Schoenberger (1983), Hall (1983) and others have described the mechanics of Kanban systems. Huang, Rees and Taylor (1983), Shen (1986) and Kimura and Terada (1981) have analyzed the performance of such systems using simulation models. The literature suggests that despite the pull nature of these systems, they are able to function effectively, and have become identified with "just-in-time" methods. Karmarkar (1986a,b) discusses some of the factors underlying this performance and relates Kanban systems to older forms of "pull" systems. However, there appears to be little in the way of analytical studies of Kanban systems that relate their performance to design parameters.

Recent research in modelling batch manufacturing facilities (Karmarkar, 1983; Karmarkar, Kekre and Kekre, 1985; Karmarkar, Kekre, Kekre and Freeman, 1985) has shown that there is a significant relationship between batch sizing policies and production lead times. Controlling lead times is an important aspect of the Kanban system as described by Sugimori et al., who give a relationship between production lead time and inventory levels for a Kanban system. However, their expression assumes deterministic demand, and is essentially a restatement of Little's law. Based on the results of investigations in other areas, it seems likely that batch sizes play an important role in lead time control. The reduction of lead time in turn,
allows inventories to be reduced, without incurring high stockout costs. In effect, lead time reduction is crucial in operating a pull systems in a varying demand environment, without incurring high levels of either safety stocks or stockouts. The effect of batching policies in batch production facilities has also been studied by Zipkin (1986) using methods similar to the studies cited earlier. The effect of production lead times on safety stocks has been discussed by Magee (1956).

In this paper we model single card, dual card, and two stage Kanban systems. Our use of these terms is roughly in correspondence with the usage in Schoenberger. Approximate Markovian models of system behavior are developed to study the effect of batch sizing and card count on the expected costs of inventory holding and backorders or shortages which cause delays in filling demand. It is shown that batch sizing has a significant effect on production lead times and hence on costs, in a manner analogous to results obtained for other multiproduct batch manufacturing facilities. The results are qualitatively compared with simulation studies of batching in Kanban cells done by Shen (1986), and are seen to be in correspondence.
A Kanban system is a control mechanism for a manufacturing facility (or cell) consisting of a production process and an inventory storage location. The facility, which may process one item or many, may itself draw upon other inventories or production facilities. We will here assume that the inputs to the production process, whether raw material or labor, are always available.

The essential properties of the Kanban scheme are:

1. Production is carried out in multiples of a minimum quantity or batch. Each batch, whether completed or authorized for production must correspond to an identifier, which is typically a card or container. The word "kanban" itself means "action-plate" and refers to a card.

2. The number of cards (or containers) in the system is fixed, and hence the total quantity of on-hand and on-order inventory in the facility is fixed.

3. Production is only initiated when finished inventory is removed from the cell, thereby releasing a card (or container).

The triggering or release mechanism in 3 by which the card acts as a work order released by the occurrence of demand, identifies the Kanban system as a "pull" approach. It is also apparent that the fixed number of cards which gives a fixed total on-hand and on-order quantity, makes the Kanban scheme analogous to the Base Stock system described by Kimball (1956). See Karmarkar (1986a,b) for a discussion of these issues.
In principle, the Kanban scheme can be used to control not only production in manufacturing cells, but also transportation between cell or facilities. In this case, the cards controlling transportation, travel between facilities; material cannot be transported until the using facility draws on incoming inventory, thereby releasing a card which authorizes the movement of more material. To distinguish the physical transportation operations from production processes, the terms production kanban and conveyance or transportation kanban are sometimes used for the kanban cards used in the two types of application. It should be recognized, however, that there is essentially no logical difference in the two applications of the scheme.

In this paper, we examine three configurations of facilities operated by Kanban methods. In the Single Card system, the items within the production cell are controlled by the Kanban scheme. However, transportation away from the cell is not controlled by Kanbans. To use the terminology given above, the system has production kanbans but no conveyance kanbans. The substantive implication is that while the quantity in the cell is constrained by the number of cards (kanbans) in the cell, there is no limit on the demand for output from the cell. As a consequence, if material is not available in the output storage locations, there is no upper bound on the backorders or unfilled demand that can accumulate.

In the Dual Card configuration, the facility as well as transportation out of the facility are controlled by production and withdrawal kanbans respectively. In this configuration, the total number of backorders possible is limited by the number of conveyance kanbans.
Finally, we consider a Two Stage Kanban system, consisting of two cells in series with one drawing upon the other. This is the smallest example of a multi-stage Kanban controlled process, and begins to consider the interactions that can occur between Kanban systems in a production network.

We model these three configurations using Markovian models. The state of the system is represented in terms of the number of cards (kanbans) on order, the number of cards and batches in finished inventory, and the number of batches on backorder. The limitations on production and withdrawal kanbans can be used to simplify state representation. These models are heuristic in several senses. First, the assumption is made that demand to the facility follows a Poisson process. Secondly, it is assumed that free cards queue for production, and that the production process has an exponential distribution. Thirdly, the cell is taken to be processing a single item class. Note that the distribution of processing times implicitly derives from the heterogeneity of items within this class, rather than from randomness in processing times per se. Finally, it is assumed that as batch sizes are changed, the average demand arrival and production rates per batch change, but that the Markovian assumptions still hold. These assumptions are made primarily because of the tractability of the Markovian model. More realistic assumptions would allow considerably more generality in the demand arrival and production processes, would account for multiple items or item classes, and would explicitly recognize the effect of the batching or aggregation process (for example see Groenevelt, 1986; Groenevelt and Karmarkar, 1986).
While these assumptions are quite severe, there are certain mitigating factors. The primary intent of this investigation is to develop a qualitative analysis of Kanban systems that can provide insights into their parametric behavior in a tractable way. A similar approach has been very successful in modelling other manufacturing and computer systems, where Markovian assumptions have not only been valuable in providing qualitative analysis (Jackson, 1963; Denning and Buzen, 1978) but have even proved to be robust enough to give good numerical results in modelling actual scenarios (Solberg, 1977; Suri, 1983; Karmarkar, Kekre, Kekre and Freeman, 1985).

The following sections model Single Card, Dual Card and Two Stage Kanban systems using Markovian models. The models are used to relate the choice of system parameters (batch size and card count) to expected costs of operating the system. The costs that are considered include inventory holding costs and backorder or shortage costs. The inventory holding costs depend on the production lead times in the cell and also act as a surrogate for other lead time related costs.

**A Single Card Kanban System**

Consider a cell producing a single item in batches. Each batch corresponds to a card or kanban so that the number of batches in process (on order) and in inventory is fixed. Let

\[ N = \text{Total number of cards and batches (in inventory and in process)} \]
\[ I = \text{Average number of batches in inventory} \]
\[ Q = \text{Batch Size} \]
\[ D = \text{Total demand / production (units / year)} \]
\[ \tau = \text{Setup time} \]
\[ P = \text{Processing rate} \]
\[ \bar{x} = \frac{1}{\mu} = \text{Average time to produce a batch} = \tau + \frac{Q}{P} \]
\[ \lambda = \text{Average arrival rate of batched demand} = \frac{D}{Q} \]
\[ \rho = \frac{\lambda}{\mu} = \frac{Dr}{Q} + \frac{D}{P} \]
\[ u = \frac{D}{P} \]

We assume that demand in batches of size \( Q \) arrives according to a Poisson process and that the production time for a batch is exponentially distributed. The inventory position or backlog at the cell can be described by a semi-infinite birth death process, as shown in Figure 1.

**Figure 1: Markovian Model of a Single Card Cell with \( N \) Cards**

Let \( p(i) \) = probability that system is in state \( i \) (\( i = \ldots, -2, -1, 0, 1, 2, \ldots, N \)).

Then the steady state balance equations for the system are:
\( \mu p(i) = \lambda p(i+1), \text{ for all } i \leq N-1 \)

\[ \sum_{i} p(i) = 1 \]

Solving these equations gives the well known probability distribution for the system:

\[ p(i) = p_{N-i}^{N-i}(1-p) \text{ for all } i \]

The cell as modelled here, has two controllable parameters \( N \) and \( Q \). Consider the expected costs associated with inventory and backorder given

\[ b = \text{Cost of backorder ( \$ / unit time)} \]
\[ h = \text{Cost of inventory ( \$ / unit time)} \]
\[ \bar{B} = \text{Average number of batches backordered} \]

We assume that in-process batches have a cost equal to that of finished batches. This assumption is made to simplify computations but can easily be relaxed. The costs under consideration can be written as

\[ C(N,Q) = b \bar{B}Q + hNQ \]

where \( \bar{B} = \sum_{i=1}^{N+i} ip(1-p) \)

If \( N \) is treated as a continuous variable, for purposes of qualitative analysis,

\[ \frac{\partial C}{\partial N} = b \log p \bar{B}Q + hQ \]
and after some rearrangement
\[
\frac{\partial C}{\partial N} = b \bar{B} \left[ 1 - \frac{D \tau N(1-\rho) + 1}{Q \rho(1-\rho)} \right] + hN
\]

Clearly \( C(N,Q) \) is convex in \( N \), the first term being convex decreasing and the second linear increasing. The behavior of \( C \) with \( Q \) is a little more complicated, but can be seen to be similar to the relationships discussed by Karmarkar (1983). As \( Q \) tend to its lower limit of \[ \frac{D \tau}{1-(D/P)} \], \( \bar{B} \) increases without bound. As \( Q \) becomes large, \( \bar{B} \) tends to \[ \frac{u^{N+1}}{1-u} \] and \( C \) becomes asymptotically linear, increasing in \( Q \).

Define the dimensionless lot size \( q = (Q/D\tau) \). Then the expected costs (3) can be written in dimensionless form as

\[
(5) \quad c(N,q) = \frac{C(N,Q)}{hD\tau} = \left( \frac{b}{h} \right) \bar{B} q + Nq
\]

If we assume that \( \left( \frac{b}{h} \right) = 10 \) say, we can determine the best value for \( q \) and \( N \) for a given real utilization rate \( u \). It is clear that \( \left( \frac{b}{h} \right) \) is a service level factor that captures the weight given to backorders relative to inventory holding.

The sensitivity of costs to batch size (\( c \) with \( q \), given \( u \) and \( N \)) is shown in Fig 2. Note that for large \( q \), the cost growth is asymptotically linear, and increases with increases in \( u \) and \( N \). This implies that lotsizing is very critical when utilization is high, and when there are more containers in the system. As \( q \) becomes small, the expected costs rise to infinity as \( q \) approaches \( 1/(1-u) \) for all \( N \).
FIG 2: Variation of $c(N,q)$ with $q$ for

$(b/h) = 10$ and $U = 0.6$
These results are qualitatively similar to the relationships obtained by Karmarkar (1983) and Karmarkar, Kekre and Kekre (1985) for lead times and work in process. While the mechanisms are similar, it should be recognized that in this case, safety stock and backorder costs are also being included.

**A Dual Card Kanban System**

For the dual card cell, let $M$ be the number of "transportation" or "conveyance" kanbans which are needed to draw from finished inventory at the cell. The state space for the dual card model is now a truncated birth death process, as shown in Figure 3.

![Figure 3: Markovian Model for a Dual Card Kanban System with $N$ Production Kanbans and $M$ Withdrawal Kanbans](image-url)
The steady state equations for this system are

\[ p(i) = (\lambda/\mu) \ p(i+1) \quad i = -M, -M+1, \ldots, 0, 1, 2, \ldots, (N-1) \]

\[ \sum_{i=-M}^{N} p(i) = 1 \]

The solution of these equations gives

\[ p(N) = \frac{1 - \rho}{\rho \left[ 1 - \rho^{N-M} \right]} \]

\[ p(i) = \frac{\rho^{N-i} (1 - \rho)}{1 - \rho^{N+M}} \quad i = -M, -M+1, \ldots, 0, 1, \ldots, (N-1) \]

As before, we can write expressions for the expected number of backorders and the average inventory level. The expected costs associated with a choice of \( N, M \) and \( Q \) can then be written as

\[ C(N, M, Q) = bBQ + hNQ \]

In the model above, if we make the alternative assumption that in-process inventory is valued at less than finished inventories we would rewrite the cost as

\[ C(N, M, Q) = bBQ + hIQ + h(N-1)Q \]
Where $h$ is the average cost of work-in-process, and $I = \sum_{i=1}^{N} i \cdot p(i)$. The analysis of this dual card system is completely analogous to the preceding case, and we will not elaborate on it further.

A Two Stage Kanban System

Consider a two stage production chain linked by a Kanban type system of control (Figure 4). Stage 2 provides input to Stage 1. At Stage 1, if an empty Stage 1 container ($E_1$) and a full Stage 2 container ($F_2$) is available Stage 1 processing begins, and when complete, the output is stored at $F_1$.

Let

$\lambda$ = Demand rate for Stage 1 output, in containers/unit time
$\mu$ = Processing rate at Stage 1.
$\sigma$ = Processing rate Stage 2.
$I$ = Number of containers in circulation at Stage 1.
$J$ = Number of containers in circulation at Stage 2.
i = Index for full Stage 1 containers at $F_1$, $i=0,1,2,\ldots,I$
j = Index for full Stage 2 containers at $F_2$, $j=0,1,2,\ldots,J$
FIGURE 4: A Two Stage Kanban System
For computational convenience, we assume that when a shortage occurs for Stage 1 output (i.e. there is external demand but there are no full containers at F 1), the demand is lost. There is however a shortage cost incurred, and we assume that this is b(\$/per unit time). The inventory holding costs are \( h_1 \) and \( h_2 \) for Stage 1 and 2 respectively (\$/unit of product/per unit time).

If we now assume that all production and withdrawal times for batches at both stages are exponentially distributed, we can model the production chain as a finite two dimensional Markov chain. The state variables are the number of full Stage 1 containers \( i \) and the number of full Stage 2 containers \( j \). The state transition diagram for the Markov chain is shown in Figure 5.

The two relevant kinds of costs that we consider are the inventory holding costs at Stage 1 and Stage 2, and the shortage costs at Stage 1 when external demand is unfulfilled. The costs are again a function of the steady state probabilities of the Markov chain. Let \( p(i,j) \) denote the steady state probability that there are \( i \) full containers at Stage 1 and \( j \) full containers at Stage 2. These probabilities are determined by solving the flow balance equations discussed in Appendix A. The expected cost for the production chain can be written as:

\[
C \left[ I, J, \lambda, \mu, \sigma \right] = \sum_{i=0}^{I} \sum_{j=0}^{J} c_{ij} p(i,j)
\]

where \( c_{ij} \) is the associated cost of holding \( i \) and \( j \) numbers of containers at Stage 1 and 2 respectively:  
\[
c_{ij} = \begin{cases} 
(iQ_{h_1} + jQ_{h_2}), & \text{for } i \neq 0 \\
(iQ_{h_1} + jQ_{h_2} + b), & \text{for } i = 0 
\end{cases}
\]
FIGURE 5: State Transition Diagram for Markovian Model of Two Stage Kanban System
Note that i=0 implies that there are no full containers available to meet external demand, and there is an associated penalty in terms of a shortage cost. In cases where Stage 1 is feeding an assembly line, this cost can be taken to be the cost of keeping the assembly line idle.

In order to incorporate the effects of lotsizing (here a lot is one container load), the state transition rates have to be related to container size. We assume that this is the same for both production stages. Let

\[ D = \text{External demand rate for Stage 1 finished goods} \]
\[ Q = \text{Batch size (Container size)} \]
\[ P_1 = \text{Processing rate at Stage 1} \]
\[ P_2 = \text{Processing rate at Stage 2} \]
\[ \tau_1 = \text{Setup time at Stage 1} \]
\[ \tau_2 = \text{Setup time at Stage 2} \]

then

\[ \lambda = \frac{D}{Q} \]
\[ \frac{1}{\mu} = \frac{Q}{P_1} + \tau_1 \]
\[ \frac{1}{\sigma} = \frac{Q}{P_2} + \tau_2 \]

Using Eqns. (9), (10) and (a.2) the expected cost can be related to the lotsize (Q) and can be determined for a configuration of the production chain (given I and J). Figure 6 shows these relationships. It is seen that for large Q, the inventory holding costs at Stage 1 and Stage 2 both grow asymptotically linearly with Q. This is because \( \lambda \), \( \mu \), and \( \sigma \), eventually stay
FIGURE 6: Variation of Costs with Q for Two Stage Kanban System
in the same relative proportion as \( Q \) increases and the transition probabilities stabilize. When \( Q \) decreases, shortage costs rise for any choice of card numbers. This is because of the increase in production leadtimes discussed by Karmarkar (1983) as in the earlier models. Note however, that the rise occurs at \( Q = 0 \) and not at \( Q = D \tau/(1-(D/P)) \) as in the single card Kanban System discussed earlier.

We now examine the behavior of shortage and inventory costs as the number of cards or containers at each stage is varied. The lot size or container size is kept fixed at \( Q = 60 \). The demand rate is \( D = 1000 \) units/time and the shortage cost is \( b = 2000 \) $/unit-time. The other parameters are taken to be: \( h_1 = 1.0, \ h_2 = 0.5, \ P_1 = 7,500, \ P_2 = 5,000, \ \tau_1 = 0.02, \ \tau_2 = 0.01 \). The number of containers (I) at Stage 1 and the number of Stage 2 (J) were each varied individually, while keeping the other constant. The resulting cost variations are shown in Figures 7 and 8.
FIGURE 7: Variation of Costs with Number of Containers

at Stage 1
FIGURE 8: Variation of Costs with Number of Containers at Stage 2
In both cases, the shortage costs decrease. However, as \( I \), the number of Stage 1 containers is increased, the shortage costs go to zero. This is not the case with Stage 2 containers, since inventories at Stage 2 cannot eliminate stockouts at Stage 1. In both cases, increasing the number of containers increases the inventory costs at that stage linearly. However, the effects on the other stage are quite different. Increasing Stage 1 containers lowers inventory costs at Stage 2, since available material at Stage 2 is drawn to produce for Stage 1 finished inventory. However increasing \( J \), the number of Stage 2 containers, also has the effect of increasing material availability to Stage 1, so that Stage 1 holding costs also increase. This asymmetry in consequence arises simply because Stage 1 feeds Stage 2.

The asymptotic limiting behavior of shortage costs as the number \( J \) of Stage 2 containers is increased and that of Stage 2 inventory costs as the number \( I \) of Stage 1 containers is increased, is discussed in Appendix B.

Next, we consider the simultaneous variation of batch (container) size and card (container) numbers, to optimize the behavior of the system. The purpose is to gain an insight into the behavior of the system. We illustrate this on the same two stage example, with two iterations of a search over the number of containers, while optimizing with respect to the batch size \( Q \). First, we fix \( J \), the number of Stage 1 containers at \( J=3 \) and search over \( I \), the number of containers at Stage 1. For each choice of \( I \), the value of container size \( Q^*(I,J) \) is determined. The minimum cost is achieved at \( I=6 \), and \( Q^*(6,3) = 37 \). The results are illustrated in Figure 9(a,b).
FIGURE 9a: Variation in Costs as I is Varied.

\[ C^*(I,3) \]

\( J = 3 \), \( Q^* \) Chosen to Minimize Costs
FIGURE 9(b): Optimal Batch Size $Q^*$ as $I$ is Varied with $J=3$
Then, we vary $J$ while keeping $I$ fixed at 6. The minimum is achieved at $J=3$, with $Q^*(6,3) = 37$. The results are shown in Figure 10(a,b). This can be verified to be the optimal solution for this example. We see from this exercise, that the optimal batch size reduces as the number of containers is increased, as we might expect. However, the relationship is not linear, so that the total quantity represented by the containers does not remain constant, but tends to increase. We also see that the effect of the number of containers on costs is quite pronounced, and is more symmetrical than the effect on cost of $Q$. Of course, too much should not be read into these relationships, since this is just one instance of a highly stylized model.
FIGURE 10a: Variation in Costs as \( J \) is Varied.

\( J = 6 \). \( Q^* \) Chosen to Minimize Costs
FIGURE 10(b): Optimal Batch Size $Q^*$ as $J$ is Varied with $I=6$. 
Discussion and Summary

The purpose of this investigation was to gain insights into the parametric behavior of Kanban systems, and to develop analytical support for some conjectures about this behavior. An alternative approach would have been to use simulation as many others have done. However, the analytical models are actually much simpler to develop, and much easier to analyze both analytically and numerically. In fact, Shen(1986) in a parallel project, developed a SLAM simulation of a two-stage Kanban system followed by an assembly stage; the effort in developing that model was considerable, as was the cost of making runs.

The major results of this investigation are the following:

1. The batch size associated with each card (or the containers size) has a significant effect on the performance of Kanban systems. The results are in correspondence with the models of Karmarkar (1983) and Karmarkar et al (1985), which showed the impact of batching on lead times.

   The effect of batch sizes was also tested by Shen (1986) using simulation. While the system that he examined is somewhat different, the results were qualitatively very similar to those in Figures 2 and 6. In particular, Figure 11 is drawn from Shen, and shows results essentially similar to those in Figure 6.

2. The effect of the number of cards (containers) is also significant. With the batch size fixed, varying the number of cards is analogous to varying the order up to level in a base-stock system (see Karmarkar 1986a), and the results on backorder and inventory holding are as might be expected.
3. There is an interaction between card numbers and batch size. The base-stock analogy above would suggest that doubling the number of cards and halving the batch size would show only a marginal effect on costs, since the pool of on-hand plus on-order stock would stay the same. However, since the batch size also affects lead times, the effect on costs is more complex.

4. In a multistage Kanban system, the parameters at one stage affect the performance at other stages. Increasing the number of cards at one stage appears to increase inventory levels at a subsequent stage and decrease them at a preceding stage. While this effect is intuitively clear, our analysis of these interactions was very limited, and there is room for more work. Furthermore, we fixed the batch sizes at both stages to be the same; this is not in fact necessary. Relaxing this constraint would change the pattern of interactions.

5. It is possible to optimize the behavior of Kanban systems through the control of its parameters. The impact of these parameters on performance is considerable. However, the models developed here are not suitable for representing and optimizing realistic systems.

To elaborate on the last comment, it is apparent that the Markovian model would be stretched beyond its capabilities in attempting to model a realistic system. In particular, a realistic model of a single stage should be able to accommodate multiple items, multiple machines, non-exponential processing
times, and more general demand processes. Such models are studied by Groenevelt (1986) and Groenevelt and Karmarkar (1986a). These models could be used to optimize the behavior of Kanban systems and also to develop "dynamic" systems which react to changing conditions (for example, Groenevelt and Karmarkar, 1986b).
FIGURE 11: Variation of Costs with Batch Size in a Simulation Study of Kanban Systems (Shen, 1986)
Appendix A: Calculating Steady State Probabilities for the Two Stage System.

The steady state probabilities are calculated by solving a system of linear equations representing flow balance in the state transition diagram. Consider a typical interior node \((i, j)\) of the grid in Figure 5. The various possible transitions to and from this node are shown below.

![State Transition Diagram]

and the flow balance equation for \((i, j)\) is
(a.1) \[(\lambda+\mu+\sigma)p(i,j) = \lambda p(i+1,j) + \mu p(i-1,j) + 1 + \sigma p(i,j-1)\]

The probabilities can be determined by solving the system of equations

\[Ap = 0\]

\[(a.2) \sum_{i=0}^{I} \sum_{j=0}^{J} p(i,j) = 0\]

where the matrix A corresponds to the flow balance equations.
Appendix B: Limiting Behavior Costs as the Number of Containers is Increased in a Two-stage System

As the number of containers at either stage is increased, certain costs asymptotically reach limiting values as illustrated in Figures 7 and 8. These limiting values can be calculated a priori, and serve as bounds. First consider the holding costs at Stage 2, as 1, the number of Stage 1 containers is increased. Writing the flow balance between the set of nodes with indices j and (j+1):

\[ \sigma \sum_{i=0}^{\infty} p(i,j) = \mu \sum_{i=0}^{\infty} p(i,j+1) \]

or

\[ \sigma p_{..j} = \mu p_{..j+1} - \mu p_{..j+1} \]

where

\[ p_{..j} = \sum_{i=0}^{\infty} p(i,j) \]

As I goes to \( \infty \), the probabilities \( p_{..j+1} \) go to zero, giving

\[ \sigma p_{..j} = \mu p_{..j+1} \]

for the limiting case. These equations, together with \( \sum_{j} p_{..j} = 1 \), can be solved to give

\[ p_{..0} = \frac{1}{1 + \sum_{j=0}^{\infty} (\sigma/\mu)^j} \]

\[ p_{..j} = p_{..j}(\sigma/\mu)^j, \quad j=1,2,\ldots,J \]
Thus, the inventory cost at Stage 2, as \( J \) goes to \( \infty \) is given by
\[
\sum_{j=0}^{J} p(j, j|q_{2j})
\]

Similarly, considering the flow balance between the set of nodes with index \( i \) and those with index \( i+1 \), as \( J \) goes to \( \infty \), we have
\[
\mu p(i, .) = \lambda p(i+1, .) \quad i=0,1,\ldots,(I-1)
\]

Which gives
\[
p(0, .) = 1/(\sum_{i=0}^{I} (\mu/\lambda)^i)
\]

\[
p(i, .) = p(0, .)(\mu/\lambda)^i \quad i=1,2,\ldots,I
\]

The holding cost for Stage 1 inventory in the limit, is given by
\[
\sum_{i=0}^{I} p(i, .)(i|q_{1i}) \quad \text{and the shortage cost at Stage 1 by} \quad (b p(0, .)).
\]
References


