PRODUCTION-INVENTORY SYSTEMS WITH UNRELIABLE MACHINES

M. BERG
University of Haifa, Haifa, Israel

M. J. M. POSNER and H. ZHAO
University of Toronto, Toronto, Ontario, Canada

(Received February 1991; revisions received February, May 1992; accepted June 1992)

A broad class of production-inventory systems is studied in which a number of producing machines are susceptible to failure following which they must be repaired to make them operative again. The machines' production can also be stopped deliberately due to stocking capacity limitations or any other relevant considerations. The interplay between the processes involved, namely, production, demand, and failure/repair or reliability, in conjunction with the shutdown policy used, determine the inventory accumulation process and possible shortages. We first obtain the stationary distribution of the inventory process for different assumptions on the random behavior of the production, demand, and reliability processes. By employing level-crossing techniques, a mathematical analysis is carried out for a "core" model, which then serves the role of the nucleus for the study of a wide range of models. We compute performance measures that characterize the operation of the production-inventory system with respect to its service-level to customers, expected inventory stocked, machines' utilization, repairmen utilization, and so on. A numerical illustration is provided which shows the effect of machine breakdowns on service and inventory levels.

We will consider a production system that consists of a number of machines, each producing the same type of item. The produced items are stored in a warehouse from which they are taken by arriving demand orders. The machines operate continuously but production will stop when they are shut down due to various considerations involving costs, maintenance, inventory capacity limit, etc. Apart from these deliberate shutdown periods there may also be machine failures which must be repaired before they become operative again.

The general purpose of this research is to study the performance of such production systems. This is done by employing different performance indices which relate to the different facets of the system, namely, the inventory accumulation process, demand fulfillment and possible shortages, machine utilization rates, production switch-on and switch-off rates, machine operability rates, etc. Such performance indices characterize the system's operation and provide input for decision making, usually in conjunction with an appropriate cost structure on various decision parameters of the system.

Obtaining concrete results requires more specificity with respect to the nature of the operation of the system. In this context, this means formulating assumptions regarding the three input processes of the problem, namely, the production process, the demand process, and the failure/repair or reliability process. These usually random processes and the interplay among them determine the behavior of the inventory accumulation process, and it is the stationary distribution of this process that we need to compute the various performance measures discussed above. Accordingly, our main task will be to derive the stationary distribution of the inventory level in terms of the three input processes and our assumptions regarding these processes.

In general, we can think of a large number of relevant assumptions, and a practical and convenient way to proceed is to start with a core model, described by a set of relatively simplified assumptions—the core assumptions. Other models are then presented in terms of their relation to the core model through the particular assumptions they alter, relax, or generalize. Looking at it from the opposite perspective, an important consideration in the construction of the core model should be its ability to serve as a starting point for other model building so that the relationship between the core model and the other ones goes both

Subject classifications: Inventory/production, uncertainty, stochastic: several machines. Reliability: machine failures.
Area of review: MANUFACTURING, OPERATIONS AND SCHEDULING.

0030-364X/94/4201-0111 $01.25

ways. This structure seems to be useful in providing insight into the problem as a whole and in clarifying the interrelations within the spectrum of model variations. Moreover, results for other models can sometimes be obtained from those of the core model once some necessary modifications have been made (for another example of such a research-structuring approach regarding a different inventory-related problem see Berg and Posner 1980).

Before we proceed to the description and analysis of the core model we will elaborate on the general scope of the study. The main thrust of this work is on the impact of the reliability dimension on the performance of the production-inventory system. At the same time, the results of the analysis here will yield, as special cases, solutions for problems concerned solely with the production and inventory facets (thereby tacitly assuming machine perfection). However, for brevity, we will consider in the brief literature review that follows only papers that feature all three facets of the problem, thereby excluding those papers that deal only with production and inventory facets. However, these other models can still be extracted as special cases from the "three-facet" analysis here. For convenience of exposition, we defer the literature review until after the listing of the assumptions of the core model.

1. THE CORE MODEL ASSUMPTIONS

We now list the modeling assumptions which were chosen to represent the core model for this study. (Remark: Throughout, the term *rate* implies per-unit of time.)

Production

P1: We have N identical machines, each producing the same type of item.
P2: Each machine can produce the items continuously and uniformly over time at a fixed production rate γ.
P3: Production on all machines is halted whenever the inventory is at level M.
P4: All operative machines are producing, simultaneously, whenever the inventory level is below M.

Demand

D1: Demands arrive according to a jump process:

the random point process of demand epochs is homogeneous Poisson with rate λ;
the demand sizes (the jumps) are i.i.d. random variables having an exponential distribution with a mean μ^{-1}.

D2: An arriving demand which does not find all it needs in the inventory takes whatever is available there, and the remainder of its needs is lost (no backlogging).
D3: The demand process is independent of the inventory level.

Reliability

R1: The machines are identical with respect to their failure and repair processes:

the operating time of a machine is exponential with mean θ^{-1};
the repair time of a machine is exponentially distributed with mean σ^{-1}.

R2: The repair of a failed machine starts immediately after its failure; this corresponds to an "ample" repair capacity assumption; i.e., there are enough repairmen to avoid queueing of failed machines for repair (in our case N or more repairmen will ensure that).

R3: A machine can fail only during productive operation.
R4: The operation and repair times are independent within and between machines; they are also independent of the inventory level and the demand process.

2. LITERATURE REVIEW

We limit our literature review to papers that also incorporate the reliability dimension. Meyer, Rothkopf and Smith (1979) studied a different model in which the demand for the production output occurs at a constant rate D, and the production facility, when operating, produces at rate M ($>D$). The inventory, with a limited capacity X, is thus filled at rate $V = M - D$ when the machine is operating. When the inventory capacity is reached, the machine reduces its production rate to D. The failure and repair processes of the production facility are random, and while the machine is down, undergoing repair, demand is satisfied from existing inventory. When inventory is exhausted, arriving demands are not satisfied, and are considered lost (no backlogging). A similar model was also considered by Parthasarathy and Shafarali (1987). However, the failure and repair processes can be either deterministic or probabilistic.

In Shafarali's (1984) model, demand arrives according to a Poisson process, and production output is also a Poisson process while the facility is producing. The machine's operating time is exponentially distributed and its repair time is distributed arbitrarily. Production operates under an (s, S) policy.

Hsu and Tapiero (1987) analyzed a production-inventory model using queueing theory techniques. An unreliable $M/G/1$ queueing system was considered, and a maintenance process was also introduced in addition to the failure process of the production facilities. For this case, the Laplace transform of the slack capacity was derived. Note that all the above models assume, explicitly or implicitly, that only one machine is involved in the production process.

Finally, we mention some additional work which, although not directly related, still bears relevance to the study here. In a somewhat different problem setting, Mitra (1988) considered a production-consumption system characterized by a finite capacity buffer that incorporated the reliability factor into the modeling. Another application area of production theory in which a machine's unreliability plays an important part is FMS (Vinod and Solberg 1984, Vinod and Altiok 1986, Windmer and Solot 1990). In these papers, ignoring the various sources of system unreliability in a multimachine, multijob situation can result in unrealistic performance measures. The mathematical analysis for these system models can utilize concepts and methods borrowed from queueing network theory.

3. MATHEMATICAL MODEL

Let us now turn to the mathematical analysis of the core model. As mentioned above, our primary task is to find the stationary distribution of the inventory level process. The core model is indeed an extension of a basic model for the problem under consideration (Posner and Berg 1989) in which only one machine is considered. The analysis there as well as the solution method employed—the "level-crossing" technique—provide the basis for the more general study here. Following the notation in that paper, define $I(t)$, $t > 0$, as the inventory level at time t, and set $W(t) = M - I(t)$. So, $W(t)$ has the physical interpretation of the slack inventory capacity available at time t. We also let $F(\cdot)$ be the stationary cdf of the $W(\cdot)$ process, and $f(\cdot)$ its pdf (it is easy to verify that under the assumptions of the core model a stationary distribution exists, which is, furthermore, a mixture of an absolutely continuous component and point masses at 0 and M where the process is bounded from above and below). Clearly, at any given point in time there is a number $i (i = 0, 1, \ldots, N)$ of machines under repair. The evolution of the process when there are i machines under repair is described in Figure 1 which in the terminology of the level-crossing technique corresponds to "page" $i (i = 0, \ldots, N)$. There are

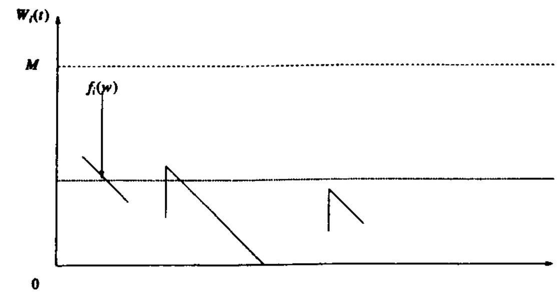

Figure 1. Sample path of page i.

departures to other pages as well as entrances from other pages. In this sense, W can be viewed as being partitioned into components W_i, which denotes the slack capacity while i machines are down. In steady state, the cdf and pdf of the slack capacity W_i on that page are denoted by $F_i(\cdot)$ and $f_i(\cdot)$, respectively, with the interpretation of $F_i(w)$ as the probability of the intersection of the events that the number of machines under repair is i and the inventory level does not exceed w. Note that there are probability masses at slack capacity level 0 (the inventory is full) on pages $i = 0, 1, \ldots, N-1$, and at slack capacity level M on page N (the inventory is empty and all machines are down).

Following the level-crossing techniques of Brill and Posner (1977, 1981) and Cohen (1977), we now establish the "balance equations", i.e., we equate rates of crossing of the W process into and out of appropriately chosen sets of states. The crossings are both within and between pages and the equalities are due to the existence of a stationary distribution for the process. Choosing the state interval $[0, w)$ in different pages, we have the balance equations (for $i = 0, 1, \ldots, N$)

$$\begin{aligned}(N-i)\gamma f_i(w) &+ (i+1)\sigma F_{i+1}(w) \\ &+ (N-i+1)\theta[F_{i-1}(w) - f_{i-1}^0] \\ = i\sigma F_i(w) &+ (N-i)\theta[F_i(w) - f_i^0] \\ &+ \lambda \int_0^w e^{-\mu(w-\alpha)} dF_i(\alpha), 0 \le w < M, \end{aligned} \quad (1)$$

$$(N-i)\gamma\lambda f_i(0^+) + (i+1)\sigma f_{i+1}^0 = (i\sigma + \lambda) f_i^0, \quad (2)$$

$$N\sigma f_N^M = \lambda \int_0^M e^{-\mu(M-\alpha)}\, dF_N(\alpha), \quad (3)$$

where f_i^0 is the probability mass at $W = 0$ on page i and f_N^M is the probability mass at $W = M$ on page N. (Note that by virtue of the assumptions of the core model we must have $f_N^0 = 0$.) Conventionally, $F_{-1}(\cdot) = F_{N+1}(\cdot) = 0$.

For the general equation (1), the system point theory for level crossings when multiple pages are involved gives that $(N - i)\gamma f_i(w)$ is the downcrossing rate into the interval $[0, w)$ from above on page i; $(i + 1)\sigma F_{i+1}(w)$ is the rate of entry into that interval on page i due to the repair of a failed machine from page $i + 1$; and $(N - i + 1)\theta[F_{i-1}(w) - f_{i-1}^0]$ is the rate of entry into the interval on page i due to the failure of one more operating machine on page $i - 1$. On the right-hand side of (1), $i\sigma F_i(w)$ is the rate of departure of the system point from the interval $[0, w)$ on page i due to the repair of a failed machine; $(N - i)\theta[F_i(w) - f_i^0]$ is the rate of departure from that interval on page i due to the failure of one more operating machine; and the integral term represents the effect of demands arising while $W_i = \alpha (0 \leq \alpha < w)$ with the resulting exponential jump taking the system point above w. Equation (2) is obtained by balancing rates into and out of $w = 0$ on page i, and (3), by balancing rates into and out of $w = M$ on page N. Finally, we also have the normalizing equation

$$\sum_{i=0}^{N} F_i(M) = 1. \tag{4}$$

The solution procedure for this set of equations involves the use of differential operators to transform the integral equations into differential-difference equations. The resulting set of equations are solved in the Appendix with the solution expressed in the mixed exponential form,

$$f_i(w) = \sum_{j=1}^{2N} c_j v_{ij} e^{\beta_j w}, \quad i = 0, 1, \ldots, N. \tag{5}$$

Refer to the Appendix for definitions of all the constants and parameters involved in the solution.

4. PERFORMANCE MEASURES

To characterize the performance of the production system we use performance measures which relate to all aspects of its operation. Of primary importance is the service level to customers—those that generate the demand—and possible delivery losses due to shortages. Then there is the inventory control for which information about the inventory level is needed. Information about the machines' utilization, in terms of both deliberate shutdowns and downtimes due to failures, and their switch-on and switch-off rates is useful for production control. Reliability-related measures are useful when assessing the machines' operability characteristics. Again, we find it helpful to categorize measures according to the system facet they relate to, although there may be some degree of arbitrariness in this because a certain measure can depend on more than one facet. We will now present some measures of the above-described types of potential importance, and compute them using the results derived.

First, let N_m be the stationary number of operating (i.e., nonfailed *and* producing) machines. Since $F_i(M)$ is the probability of the process being in page i (i.e., i failed machines), it follows that $F_i(M) - f_i^0$ is the probability of $N - i$ operating machines. Hence, the expected number of machines operating is given by

$$E(N_m) = \sum_{i=0}^{N-1} (N - i)[F_i(M) - f_i^0] = \sum_{i=0}^{N-1} \sum_{j=1}^{2N} (N - 1) c_j v_{ij} \frac{e^{\beta_j M} - 1}{\beta_j}. \tag{6}$$

p1: The effective (total) production rate (taking into account all nonproduction periods due to shutdown and downtimes).

Clearly, $p1 = \gamma E(N_m)$.

p2: Machines' utilization rate

$$p2 = \frac{E[N_m]}{N},$$

by definition.

d1: The fraction of demand satisfied.

$$d1 = \frac{p1}{\lambda/\mu}, \tag{7}$$

because the demand arrival rate times the mean demand size represents the total demand rate λ/μ, and **p1** is the actual total consumption rate (or satisfied demand). The finite inventory capacity, within the assumption set of the model, guarantees that *all* produced items are consumed.

d2: The loss rate of demands (due to shortages):

$$d2 = \frac{\lambda}{\mu} - p1 = \frac{\lambda}{\mu}(1 - d1).$$

d3: The rate of not-fully-satisfied customers:

$$d3 = \lambda \int_0^M \Pr(\text{demand} > M - w \mid W = w) \cdot d\Pr(W < w) = \lambda \left[\sum_{i=0}^{N} \sum_{j=1}^{2N} c_j v_{ij} \frac{e^{\beta_j M} - e^{-\mu M}}{\mu + \beta_j} + e^{-\mu M} \sum_{i=0}^{N-1} f_0^i + f_N^M \right], \tag{8}$$

because a demand is not fully satisfied if the available inventory upon its arrival is below its size.

i1: The mean inventory size: Since $I(t) = M - W(t)$ for all $t \geq 0$,

$$i1 = M - E(W), \tag{9}$$

where

$$E(W) = \int_0^M wf(w)\,dw + Mf_N^M$$
$$= \sum_{i=0}^{N}\sum_{j=1}^{2N} \frac{c_j v_{ij}}{\beta_j}\left[\left(M - \frac{1}{\beta_j}\right)e^{\beta_j M} + \frac{1}{\beta_j}\right] + Mf_N^M.$$

r1: Machines' failure rate: $r1 = \theta E(N_m)$, since, by assumption **R3**, only an operating machine can fail.

r2: Mean number of failed machines:

$$r2 = \sum_{i=1}^{N} iF_i(M)$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{2N} ic_j v_{ij}\frac{e^{\beta_j M} - 1}{\beta_j}.$$

p3: Switch-on rate of machines:

$$p3 = \lambda \sum_{i=0}^{N-1} (N - i)f_i^0 + \sigma \sum_{i=1}^{N} i[F_i(M) - f_i^0]$$
$$= \sum_{i=0}^{N-1} [\lambda N + (\sigma - \lambda)i]f_i^0 + N\sigma f_N^M$$
$$+ \sigma \sum_{i=1}^{N}\sum_{j=1}^{2N} ic_j v_{ij}\frac{e^{\beta_j M} - 1}{\beta_j},$$

because following the completion of a repair (the rate of repair is σ times the number of failed machines), a machine is switched on provided that the inventory is not at full capacity at that moment. Also, by assumption **P4**, all nonfailed machines are switched on—terminating a deliberate shutdown period—when a demand arrives to a full inventory.

As an illustration of the above procedure we have carried out some numerical evaluations for the special case of a 2-machine system, i.e., $N = 2$. Specifically, we have calculated **d1** and **i1** for various parameter values, setting arbitrarily $\gamma = 1$, $\lambda = 0.3$, $\mu = 0.5$, and $M = 10$. In line with the main thrust of this work, we approached this from the reliability angle, namely, we investigated how the failure and repair factors affect these focal performance measures. The results are depicted in Figures 2 and 3, and they clearly illuminate the meaningful impact of the machines' imperfection on the performance of production-inventory systems.

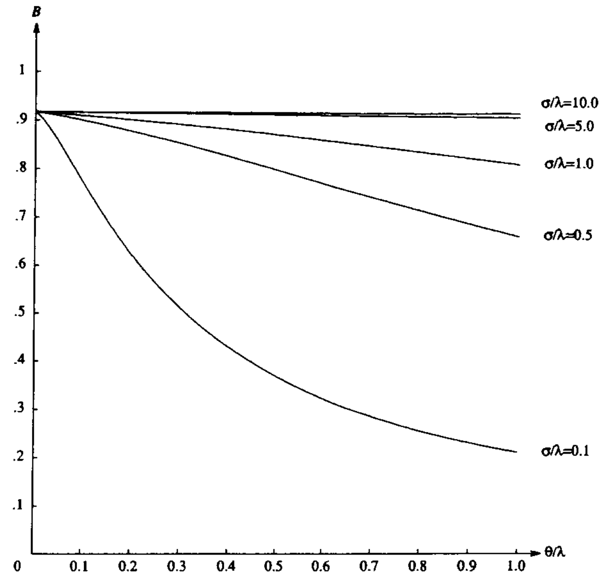

Figure 2. The effects of machine breakdowns on the service level.

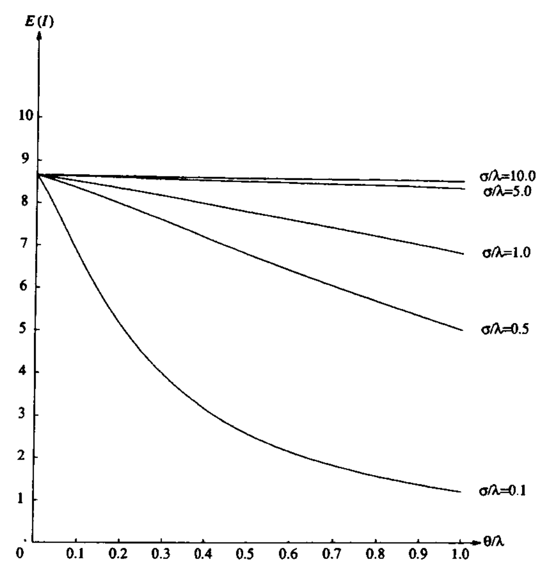

Figure 3. The effects of machine breakdowns on the average inventory level.

5. OTHER MODELS

As discussed earlier, the core model serves as a starting point for modeling a variety of production-inventory systems. To demonstrate that we now look into some model variations.

5.1. Queueing for Repair

It may not be feasible or cost-wise to employ enough repairmen to guarantee that no queueing for repair will occur. We thus relax, or change, assumption **R2**. Suppose that we now only have K ($<N$) repairmen. The repair capacity of the system is expressed as

$$\xi(i) = \begin{cases} i\sigma, & \text{if } i < K \\ K\sigma, & \text{if } K \leq i \leq N, \end{cases}$$

where i is the number of simultaneously failed machines. The balance equations of the core model are still valid once modifications accounting for the change in the system's repair capacity are introduced. The new set of equations equivalent to (1) and (2) is (for $i = 0, 1, \ldots, N$):

$$(N - i)\gamma f_i(w) + \xi(i + 1)F_{i+1}(w) + (N - i + 1)\theta[F_{i-1}(w) - f_{i-1}^0] = \xi(i)F_i(w) + (N - i)\theta[F_i(w) - f_i^0] + \lambda \int_0^w e^{-\mu(w-\alpha)}\, dF_i(\alpha),$$

$$(N - i)\gamma f_i(0^+) + \xi(i + 1)f_{i+1}^0 = [\xi(i) + \lambda]f_i^0,$$

with all else unchanged.

5.2. Demands Backlogging

Suppose that unsatisfied demand is backlogged rather than lost; the assumption altered is then **D2**. In this case, the I process is no longer bounded from above, and subsequently, the W process is unbounded from below, taking values in the $[0, \infty]$ range (the physical interpretation of W as the slack capacity is now only partially relevant, but this has no effect on the mathematics). To ensure the existence of a stationary distribution for the W process, and likewise for the I process, we need to impose the requirement that the total demand rate should be less than the effective production rate of the system, i.e., $\rho = (\lambda/\mu)/p1 < 1$, where p1 is as in the core model but with M replaced by ∞. As for the balance equations, they still hold except that the range of values of w is now the entire R^+, and (3) becomes redundant (the probability mass at $W = M$ in the Nth page vanishes as the sample-space point $W = M$ when all machines are down is no longer of special character). Remark: A model that combines the modifications of both models in subsections 6.1 and 6.2 is clearly solvable by simply superimposing the modifications of the individual models. This indicates the generality and effectiveness of the solution technique employed.

5.3. More General Distributions

In the core model the distributions of the basic random variables involved are all exponential. The solution procedure, however, is valid even for the far more general family of phase-type distributions (e.g., Neuts 1981), only the number of balance equations (within the set of equations (1)–(3)) will increase, and with it, the computational complexity. To demonstrate this, suppose that the interarrival times of demands follow an Erlang distribution with shape parameter k (i.e., a "k-phase" Erlang distribution). Then, for the whole system, W is divided into $N + 1$ pages, with each page associated with a different number of machines working. Correspondingly, the distributions of inventory level, $f(\cdot)$ and $F(\cdot)$, are partitioned into $f_0(\cdot), f_1(\cdot), \ldots, f_N(\cdot)$ and $F_0(\cdot), F_1(\cdot), \ldots, F_N(\cdot)$, respectively. In an Erlang arrival system, we must keep track of how many phases have passed since the last arrival. Therefore, to fully define the state of the system, it is insufficient to know only on which pages the system is operating; rather, it is also essential to know the phase of the arriving customer. Hence, $f_i(\cdot)$ and $F_i(\cdot)$ should be further partitioned into $f_{i1}(\cdot), f_{i2}(\cdot), \ldots, f_{ik}(\cdot)$, and $F_{i1}(\cdot), F_{i2}(\cdot), \ldots, F_{ik}(\cdot)$, respectively. Now, for all states $s = (i, j)$ corresponding to i failed machines and the arrival process being in phase j, we can construct balance equations in the same spirit as before and the solution procedure is then applied.

Furthermore, we may even go beyond the phase-type family of distributions and keep the basic set of balance equations valid (once some appropriate modifications are made). In the core model, if we let the demand-size distribution be arbitrary, say $G(\cdot)$, then the set of equations (1)–(3) still holds, but with $e^{-\mu y}$ replaced by $1 - G(y)$. The existence of the equations implies that an exact solution, using numerical methods, is still possible (subject to computational limitations), but obtaining an analytic solution is a different technical challenge. This could be part of further studies into the various extensions of the core model.

6. CONCLUSIONS

A general framework has been developed in this paper for studying production-inventory systems. The main thrust of the work is the incorporation of machines' imperfection into the analysis, but the approach is general enough to contain various production-inventory models as special cases. The derivation of the stationary distribution of the inventory level is the focus of the mathematical analysis. Using the level-crossing technique, or indeed a multidimensional version of it, we obtained a set of equations whose solution yields the desired distribution. Once the distribution of the inventory level is obtained we are able to compute various performance measures which characterize the different facets of the system's operation relating to inventory, production, demand, and reliability. The results also can be used as input for decision making on various model parameters, usually within the framework of a larger model which includes the different cost factors involved.

APPENDIX

To solve the system of equations represented by (1)–(4), we first introduce the differential operator $\langle D\rangle \equiv d/dw$ and apply $\langle D+\mu\rangle$ to (1) twice to obtain

$$\langle a_{i1}D^2 + a_{i2}D + a_{i3}\rangle f_i(w) + \langle b_{i1}D + b_{i2}\rangle f_{i+1}(w) + \langle c_{i1}D + c_{i2}\rangle f_{i-1}(w) = 0, \quad i = 0, 1, \ldots, N, \qquad \text{(A.1)}$$

where

$a_{i1} = (N-i)\gamma$, $a_{i2} = \mu a_{i1} - i\sigma - (N-i)\theta - \lambda$,

$a_{i3} = -i\mu\sigma - (N-i)\mu\theta$

$b_{i1} = (i+1)\sigma$, $b_{i2} = \mu b_{i1}$, $c_{i1} = (N-i+1)\theta$,

$c_{i2} = \mu c_i$, $i = 0, 1, \ldots, N$.

An additional useful relation is found by summing all equations in (1) and applying $\langle D+\mu\rangle$ once to give

$$\lambda f_N(w) = \sum_{i=0}^{N-1} [\langle D+\mu\rangle(N-i)\gamma - \lambda] f_i(w). \qquad \text{(A.2)}$$

Since (A.1) is of the order two we can introduce subsidiary variables $g_i(w)$ and write

$$g_i(w) = \frac{df_i(w)}{dw}, \quad i = 0, 1, \ldots, N-2. \qquad \text{(A.3)}$$

These will have the effect of converting the first $N-2$ equations of (A.1) into first-order relations. For $i = N-1$ we can use (A.2) directly as the first-order relation

$$\frac{df_{N-1}(w)}{dw} + \sum_{i=0}^{N}\left[\mu(N-i) - \frac{\lambda}{\gamma}\right] f_i(w) + \sum_{i=0}^{N-2}(N-i)g_i(w) = 0. \qquad \text{(A.4)}$$

Furthermore, for $i = N$ in (A.1), we can use (A.4) to obtain

$$\frac{df_N(w)}{dw} - \frac{c_{N1}}{a_{N2}} \cdot \sum_{i=0}^{N-2}\left[\mu(N-i) - \frac{\lambda}{\gamma}\right] f_i(w) + \left[\frac{c_{N2}}{a_{N2}} - \frac{c_{N1}}{a_{N2}} \cdot \left(\mu - \frac{\lambda}{\gamma}\right)\right] f_{N-1}(w) + \left[\frac{a_{N3}}{a_{N2}} + \frac{c_{N1}}{a_{N2}} \cdot \frac{\lambda}{\gamma}\right] f_N(w) - \frac{c_{N1}}{a_{N2}} \cdot \sum_{i=0}^{N-2}(N-i)g_i(w) = 0. \qquad \text{(A.5)}$$

Finally, substituting (A.3) into (A.1) and rearranging yields

$$\frac{dg_i(w)}{dw} + \frac{c_{i2}}{a_{i1}} f_{i-1}(w) + \frac{a_{i3}}{a_{i1}} f_i(w) + \frac{b_{i2}}{a_{i1}} f_{i+1}(w) + \frac{c_{i1}}{a_{i1}} g_{i-1}(w) + \frac{a_{i2}}{a_{i1}} g_i(w) + \frac{b_{i1}}{a_{i1}} g_{i+1}(w) = 0, \quad i = 0, 1, \ldots, N-2. \qquad \text{(A.6)}$$

Thus, (A.3)–(A.6) constitute a system of first-order differential equations which can be solved readily using methods of matrix differential equations (Spiegel 1981). Overall, these $2N$ differential equations include the $N-1$ unknown functions $g_i(w)$, $i = 0, 1, \ldots, N-2$, and the $N+1$ functions, $f_i(w)$, $i = 0, 1, \ldots, N$. Expressed in matrix form, the $2N$ equations can be written as

$$\frac{du}{dw} + Au = 0, \qquad \text{(A.7)}$$

where $u = [f_0(w), f_1(w), \ldots, f_N(w), g_0(w), \ldots, g_{N-2}(w)]^T$, and the form of A is given by

$$A = \begin{bmatrix} \mathbf{0} & R_1 \\ R_2 & R_3 \end{bmatrix}.$$

Here, $\mathbf{0}$ is an $(N-1) \times (N+1)$ null matrix, $R_1 = -I_{N-1}$, an $(N-1) \times (N-1)$ identity matrix, R_2 is an $(N+1) \times (N+1)$ matrix, and R_3 is an $(N+1) \times (N-1)$ matrix. The latter two matrices are basically tridiagonal, with their elements corresponding to the various coefficients in (A.4)–(A.6).

Now, the homogeneous matrix differential equation (A.7) has an overall general solution

$$u = \sum_{j=1}^{2N} c_j v_j e^{\beta_j w},$$

where $-\beta_1, -\beta_2, \ldots, -\beta_{2N}$ are the eigenvalues of A, $v_j = [v_{0j}, v_{1j}, \ldots, v_{(2N-1)j}]^T$ is the eigenvector corresponding to $-\beta_j$, and the c_j are arbitrary constants.

In terms of our explicit concern, we can extract the solutions required:

$$f_i(w) = \sum_{j=1}^{2N} c_j v_{ij} e^{\beta_j w}, \quad i = 0, 1, \ldots, N. \qquad \text{(A.8)}$$

Thus, we now have the desired functional forms, but these include $3N+1$ unknown constants; namely the $2N$ c_j's, f_i^0 $(i = 0, 1, \ldots, N-1)$, and f_N^M. A sufficient set of linearly independent relations to find these can be found as follows.

First, substitute (A.8) into original equations (1) and compare coefficients of common exponential terms. A comparison of coefficients of $e^{-\mu w}$ in each equation gives the $N + 1$ relations

$$f_i^0 - \sum_{j=1}^{2N} \frac{c_j v_{ij}}{\beta_j + \mu} = 0, \quad i = 0, 1, \ldots, N. \qquad \text{(A.9)}$$

A comparison of the constant terms also provides $N - 1$ linearly independent relations

$$(i + 1)\sigma\left[f_{i+1}^0 - \sum_{j=1}^{2N} \frac{c_j v_{(i+1)j}}{\beta_j}\right] - (N - i + 1)\theta \sum_{j=1}^{2N} \frac{c_j v_{(i-1)j}}{\beta_j} = i\sigma f_i^0 - (i\sigma + (N - i)\theta) \sum_{j=1}^{2N} \frac{c_j v_{ij}}{\beta_j},$$
$$i = 0, 1, \ldots, N - 2. \qquad \text{(A.10)}$$

The remaining $N + 1$ relations are obtained by considering the probability mass flow rates between adjacent pages. The rate from page i into page $i + 1$ is the rate of failure of one more machine when i are already down, and the rate from page $i + 1$ into page i is the overall repair rate while $i + 1$ machines are down. Equating these two rates yields the $N - 1$ independent relations

$$(N - i)\theta[F_i(M) - f_i^0] = (i + 1)\sigma F_{i+1}(M),$$
$$i = 1, \ldots, N - 1. \qquad \text{(A.11)}$$

The final two relations are simply obtained by putting the general solutions (A.8) into (3) and (4).

REFERENCES

BERG, M., AND M. J. M. POSNER. 1990. Customer Delays in M/G/∞ Repair Systems With Spares. *Opns. Res.* **38,** 344–348.

BRILL, P. H., AND M. J. M. POSNER. 1977. Level Crossings in Point Processes Applied to Queues: Single-Server Case. *Opns. Res.* **25,** 662–674.

BRILL, P. H., AND M. J. M. POSNER. 1981. The System Point Method in Exponential Queues: A Level Crossing Approach. *Math. Opns. Res.* **6,** 31–49.

COHEN, J. W. 1977. On Up and Down Crossings. *J. Appl. Prob.* **14,** 405–410.

HSU, L., AND C. S. TAPIERO. 1987. Maintenance of an Unreliable M/G/1 Queueing Job Shop. *Queue. Syst.* **2,** 333–350.

MEYER, R. R., M. H. ROTHKOPF AND S. A. SMITH. 1979. Reliability and Inventory in a Production-Storage System. *Mgmt. Sci.* **25,** 799–807.

MITRA, D. 1988. Stochastic Theory of a Fluid Model of Multiple Failure-Susceptible Producers and Consumers Coupled by a Buffer. *Adv. Appl. Prob.* **26,** 646–676.

NEUTS, M. 1981. *Matrix-Geometric Solutions in Stochastic Models.* Johns Hopkins University Press, Baltimore, Maryland.

PARTHASARATHY, P. R., AND M. SHAFARALI. 1987. On the Unutilized Capacity of a Production-Storage System. *Opns. Res. Letts.* **6,** 233–238.

POSNER, M. J. M., AND M. BERG. 1989. Analysis of a Production-Inventory System With Unreliable Production Facility. *Opns. Res. Letts.* **8,** 339–345.

SHAFARALI, M. 1984. On a Continuous Review Production-Inventory Problem. *Opns. Res. Lettrs.* **3,** 199–204.

SPIEGEL, M. R. 1981. *Applied Differential Equations.* Prentice-Hall, Englewood Cliffs, New Jersey.

VINOD, B., AND J. J. SOLBERG. 1984. Performance Models for Unreliable Flexible Manufacturing Systems. *Omega* **12,** 299–308.

VINOD, B., AND T. ALTIOK. 1986. Approximating Unreliable Queueing Networks Under the Assumption of Exponentiality. *J. Opnl. Res. Soc.* **37,** 309–316.

WIDMER, M. AND P. SOLOT. 1990. Do Not Forget the Breakdowns and the Maintenance Operations in FMS Design Problems I. *Int. J. Prod. Res.* **28,** 421–430.