# An investigation of due-date assignment rules with constrained tightness



*DOI:*
10.1016/0272-6963(81)90014-0
10.1016/0095-0696(74)90008-4

*Document status and date:*
Published: 01/01/1981

*Document Version:*
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

JOURNAL OF OPERATIONS MANAGEMENT
Vol. 1, No. 3, February 1981

# An Investigation Of Due-Date Assignment Rules With Constrained Tightness

Kenneth R. Baker*
J. W. M. Bertrand**

## ABSTRACT

*We describe an experimental study of a single-machine scheduling model for a system that assigns due-dates to arriving jobs. The average tightness of the due-dates is assumed to be governed by a policy constraint, which we link analytically to the parameters of the decision rules for due-date assignment. We examine the use of different kinds of information in setting due-dates, and we investigate the relationship between the due-date assignment rule and the priority dispatching rule. On the basis of our results we identify situations in which the dispatching rule is critical to effective scheduling, others in which the due-date assignment rule is critical, and still others in which the combination of the two rules is a critical design issue.*

## INTRODUCTION

One of the most important problems in production control is scheduling jobs to meet due-dates. At one extreme, due-dates can be externally imposed parameters, represented as "given" information in the statement of a scheduling problem. This is the assumption implicit in most models that have traditionally been used to study scheduling with due-dates. At the other extreme, due-dates can be internally selected parameters, determined within the control system itself. This approach has seldom been taken in the study of scheduling with due-dates, but it is the main theme in our present research.

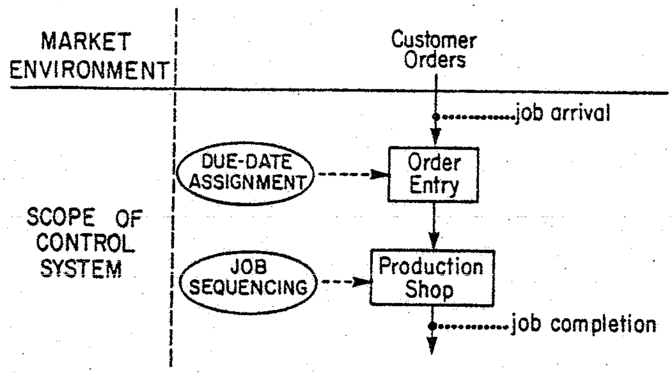
In fact, the most common practical instances probably lie somewhere between these two extremes. A useful model might be one in which due-dates are produced by a balance of production control considerations and market pressures, symbolized perhaps by negotiations between a "scheduler" and a "customer". Even in such a setting, however, the scheduler would want to have some insight into the implications of alternative due-date assignment policies, as a basis for understanding his own side of the process. Therefore, in search of this kind of insight, we examine the ideal situation in which the production control system completely determines due-dates for all jobs to be scheduled.

This type of problem has been investigated in relatively few research studies. The most relevant previous work is due to Eilon and Chowdhury [2], Elvers [4] and Weeks [7], each dealing with simulation studies of the dynamic job shop. These previous papers did not draw on analytic results in order to establish parameters for their control rules, and the emphasis of the research was on a strictly empirical approach to the problem. There remains a need to develop basic principles for the design and operation of a control system that contains due-date selection capability. Moreover, there is a need to develop some of these principles in relatively simple systems, where it is possible to focus on the due-date assignment procedure apart from the confounding influences that arise in complex systems. In this fashion, some fundamental hypotheses can be established; these help determine the issues around which subsequent investigation will be structured.

We begin our research, therefore, with the study of a relatively simple model of shop scheduling. We assume that there is a two-level hierarchy for scheduling decisions consisting of (1) due-date assignment and (2) job sequencing. (See Figure 1.) The "shop" is represented by the dynamic single-machine model, in which jobs arrive over time and in which each job consists of a single operation. Jobs are sequenced nonpreemptively according to a priority dispatching rule with the objective of meeting due-dates. The due-date of a given job is set at its arrival time, which in this case is the same as its release date. The job then enters the shop and joins the queue of jobs waiting to be processed. The jobs are selected for processing in a sequence determined by the prior-

*Amos Tuck School, Dartmouth College.
**Department of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

**FIGURE 1**
**The Structure of the Control System**




ity rule, and at each job completion it is possible to determine whether the due-date has been met.

The two levels in the hierarchy have conflicting objectives. At the level of due-date assignment, tight due-dates are preferred to loose due-dates because tight due-dates, if they can be met, tend to enforce more responsive customer service and lower in-process inventory levels. Nevertheless, at the level of job sequencing, tight due-dates are more difficult to meet than loose due-dates. Tight due-dates tend to yield more tardiness in job completions and to restrict scheduling flexibility.

In this paper we propose a specific approach to resolving this inherent conflict. We treat the aggregate tightness of job due-dates as a policy decision, made by the firm in light of such factors as the expectations of customers and the posture of competing firms. This policy decision imposes a constraint on the overall tightness of due-dates; subject to this constraint, the sequencing objective is to avoid job tardiness as much as possible. (A complementary approach, in which avoiding tardiness is treated as a constraint and in which the objective is to make due-dates as tight as possible, is described in [1].)

In order to make our formulation more precise, we introduce the following notation. For job j, let

$r_j$ = release date (arrival time)
$p_j$ = processing time
$d_j$ = due-date
$f_j$ = flow allowance (defined by $f_j = d_j - r_j$)
$C_j$ = completion time
$T_j$ = tardiness (defined by $T_j = \max\{0, C_j - d_j\}$)
n = number of jobs processed

At the level of due-date assignment, the aggregate tightness of a set of due-dates is measured by the average value of $d_j$ in a given job set. Since we assume that arrival times as well as processing times are determined exogenously, we can equivalently measure aggregate tightness by the mean flow allowance, defined to be $\bar{f}$, where

$$\bar{f} = \sum_{j=1}^{n} f_j/n$$

and where n denotes the number of jobs in the schedule. For convenience, we sometimes prefer to use instead the *allowance factor* $a = \bar{f}/\bar{p}$, which simply normalizes the mean flow allowance by dividing it by the mean processing time. Thus the firm's policy with respect to the long-run aggregate tightness of due-dates can be represented as a choice for the value of the parameter $a$.

At the level of job sequencing, performance is measured by the average tardiness, defined to be $\bar{T}$, where

$$\bar{T} = \sum_{j=1}^{n} T_j/n$$

No rewards or costs are associated with early completion of jobs.

Thus we examine a problem in which the sequencing objective is to minimize tardiness subject to a constraint on the aggregate tightness of the due-dates assigned. From sequencing theory we know that the sequencing portion of this problem is difficult to solve optimally, since enumerative procedures are required in general. For large problems the computational requirements of optimizing algorithms are prohbititive, and only heuristic procedures are practical. Meanwhile, an optimal procedure for due-date assignments would in general require complete information (arrival times and processing times) on all jobs to be scheduled. This type of comprehensive look-ahead information is unlikely to exist in practice. Therefore, a more practical approach is to restrict the amount of information involved in setting due-dates so that foreknowledge of future arrivals is not required.

In summary, our research is aimed at providing insights into the design and operation of a control system which assigns due-dates to jobs, using a limited information approach, and which schedules the jobs, using a heuristic sequencing procedure.

## A STUDY OF THE SIMPLE RULES

At the outset, we propose three limited-information assignment rules for the determination of job due-dates:

CON ($d_j = r_j + \gamma$) : jobs are given constant flow allowances

SLK ($d_j = r_j + p_j + \beta$) : jobs are given flow allowances that allow for equal amounts of waiting slack

TWK ($d_j = r_j + \alpha p_j$) : jobs are given flow allowances that are proportional to the total work required.

These rules utilize only job-related information (release date and processing time) together with a tightness parameter ($\alpha$, $\beta$ or $\gamma$) which relates to policy considerations. The relationships between tightness parameters and the allowance factor can be expressed as follows:

for CON $\bar{f} = \gamma$, so that $\gamma = a\bar{p}$
for SLK $\bar{f} = \bar{p} + \beta$, $\beta = (a - 1)\,\bar{p}$
for TWK $\bar{f} = \alpha\bar{p}$, $\alpha = a$

We also consider five simple dispatching rules, representative of heuristic procedures that might be used even in more complex systems:

- ERD (Earliest Release-Date): Priority is given to the job with the smallest $r_j$ (i.e. First Come First Served).
- SPT (Shortest Processing Time): Priority is given to the job with the smallest $p_j$.
- EDD (Earliest Due-Date): Priority is given to the job with the smallest $d_j$.
- EFT (Earliest Finish Time): Priority is given to the job with the smallest sum ($r_j + p_j$).
- MST (Minimum Slack Time): Priority is given to the job with the smallest difference ($d_j - p_j$).

These rules also utilize only job-related information, and they have been frequently studied in job shop and project scheduling simulations.

There are fifteen basic combinations of due-date assignment rule and priority dispatching rule. However, our experiments indicated that certain combinations are systematically less effective than others. For example, the lowest tardiness levels were always achieved by the priority sequencing rules SPT and EDD, while the other three rules were dominated. Furthermore, in combination with these two priority rules, the due-date assignment rule CON was always dominated by SLK and TWK. For this reason we emphasize the due-date assignment rules SLK and TWK and the priority sequencing rules SPT and EDD in the results which we report.

### The Simulation Model

Our observations are based on a series of simulation experiments involving the dynamic single-machine model with nonpreemptive dispatching. Each simulation required a data set describing the jobs to arrive during the run, where each job was specified by a release date ($r_j$) and a processing time ($p_j$). The data set was constructed using elementary queueing concepts. Job interarrival times were samples drawn from an exponential distribution, giving rise to a Poisson arrival process. Job processing times were samples drawn independently from another exponential distribution. Thus the model can be viewed as a simulation of the queue M/M/1 with priority sequencing.

In a companion data set, job processing times were samples drawn from a normal distribution in place of an exponential distribution. Since the exponential model is sometimes considered to have a relatively large variance, the use of the normal model provides an opportunity to investigate the effects of a smaller variance for processing times. In this case the model can be viewed as a simulation of the queue M/G/1 with priority sequencing.

For each type of processing time distribution two data sets were created, one for low utilization and another for high utilization. The utilization level was set simply by selecting the mean arrival rate (denoted $\lambda$) while in all cases the mean processing time was set at 100. The utilization levels corresponding to "low" and "high" were related to workload conditions. In each of the four cases, a standard queueing formula exists for the theoretical mean workload in the system. (See Appendix B.) Using this formula, the "low" utilization level was chosen to correspond to a mean workload of about 300, while the "high" level was chosen to correspond to a mean workload slightly above 700.

The parameters of the four main data sets are summarized in Table 1.

During the simulation, data collection was carried out for 1,000 jobs following an initial run-in period containing at least 300 completions. To calibrate the "representativeness" of the sample interval containing the 1,000 job completions, two statistics were checked: the average processing time of the completed jobs and the average workload in the system. Both of these average values

**TABLE 1**
**Summary Description of the Four Primary Data Sets**

| Data Set | I | II | III | IV |
|---|---|---|---|---|
| Processing time | | | | |
| Distribution | Exponential | Exponential | Normal | Normal |
| Mean | 100 | 100 | 100 | 100 |
| Standard Deviation | 100 | 100 | 25 | 25 |
| Utilization | 0.75 | 0.88 | 0.85 | 0.93 |
| Mean Workload | 300 | 733 | 301 | 706 |

were within 2% of their theoretical mean values during the sample interval.

### Experimental Results

The results of our initial runs are given in Table A1 of Appendix A, where the average tardiness per job is tabulated under various conditions. Some representative results are also shown graphically in Figure 2.

In our initial runs the TWK rule for due-date assignment always produced the lowest average tardiness, other things being equal. Similarly, the SLK rule always produced lower tardiness than did the CON rule. In graphical terms, if we were to plot average tardiness as a function of the allowance factor, we would find that the SLK curve lies below the CON curve in the relevant range, and that the TWK curve lies below the SLK curve. This behavior held for both utilization levels and both distribution types. These results indicate that the use of processing time information for setting due-dates (especially in the TWK form) improves the performance of the system in reducing tardiness. Furthermore, the variance of the processing time distribution had a predictable influence on these results. When the variance was small, the differences between TWK and SLK were relatively slight. (Indeed, their ability to discriminate among jobs on the basis of processing time could hardly provide much of a benefit compared to the simpler CON rule.) When the variance was larger, the relative differences among the rules became greater.

With regard to the choice of a priority rule, neither SPT nor EDD is preferable under all conditions. The results distinctly show that SPT is desirable when the allowance factor is very low, while EDD is desirable when the allowance factor is high. This effect is readily explained. The SPT rule tends to yield small mean flowtimes. This means that when a pair of waiting jobs is to be ordered, the SPT rule will place the shorter job first in order to achieve even a slight benefit in mean flowtime, even if this ordering yields some tardiness that could otherwise be prevented. This situation, where SPT yields some tardiness while EDD avoids it, tends to occur only when most jobs are completed on time, or in other words, when the allowance factor is high. Thus the EDD priority rule is desirable when $a$ is high. On the other hand, when the allowance factor is very low, most jobs are tardy under either priority rule, and improvements in mean flowtime tend to be reflected as improvements in $\overline{T}$. Thus the SPT priority rule is desirable when the allowance factor is low. The result is a crossover phenomenon: as $a$ increases, first SPT is preferable but subsequently EDD is preferable. This type of behavior is most distinct in the case of the normal distribution in Table A1.

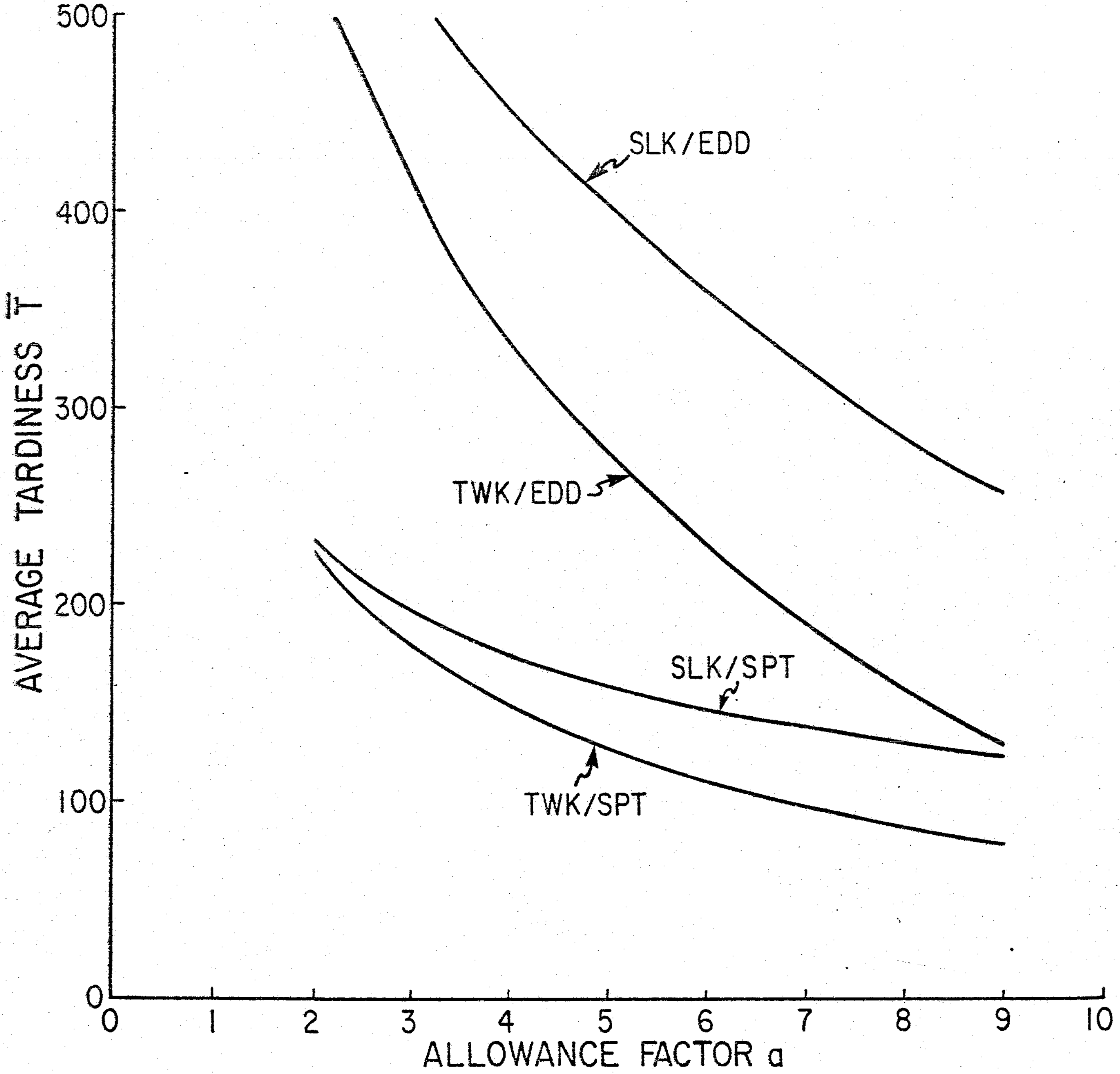
This crossover was influenced by the variance of the processing time distribution. When variance is relatively high there is a greater potential benefit in reducing mean flowtime to offset the tardiness of long jobs. Therefore, the desirability of SPT should persist over a wider range of $a$-values than when variance is relatively low. This pattern was observed in the results.

With regard to priority dispatching we conclude that the desirability of SPT or EDD depends on the tightness of the due-dates. The distinctive crossover effect has previously been recognized (in conjunction with TWK due-dates) by Elvers [4], and it can also be found in the data of Eilon and Hodgson [3], although it is not discussed explicitly. Weeks [7] did not find a crossover effect in comparing SPT and MST, but his experiments covered a much smaller range of allowance factors. A set of simulation runs beyond the $a$-values given in Table A1 confirmed that the crossover occurs for SLK and for CON as well as for TWK.

## WORKLOAD DEPENDENT ASSIGNMENT RULES

In the foregoing discussion we have dealt with due-date assignment rules that discriminate

**FIGURE 2**
**Simulation Results for Data Set II in the Initial Runs**




among jobs on the basis of their processing times. Alternatively, it might be said that these rules rely on an information base consisting of the parameters describing the individual job to which a due-date is being assigned. A more ambitious approach is to include shop status in the information base. For such a purpose, the shop workload is a natural indicator to use because it is logical that flow allowances should reflect workload conditions. In this section we examine some ways of incorporating workload information into the due-date assignment rule.

In order to provide some perspective on the value of workload-dependent rules for assigning due-dates, we first describe a simple device based on predictable flowtimes. Let $W_j$ denote the workload in the system at time $r_j$ (not including the processing time of job j itself), and suppose that jobs are sequenced in arrival order, as prescribed by the ERD priority rule. In this situation $W_j$ measures the waiting time for job j, while $W_j + p_j$ represents its flowtime. Now suppose that flow allowances are set according to the rule $f_j = W_j + p_j$. It follows that each job is completed precisely at its due-date. Thus it is not difficult to utilize workload information to devise a "perfect" assignment rule in the sense that it guarantees $\overline{T} = 0$.

The drawback of this rule, of course, is that there is no means by which the tightness of the due-dates can be controlled. Therefore, this rule can only reveal an allowance factor at which tardiness is avoidable. We denote this value of the allowance factor $a^*$, and in Appendix B we derive the following formula:

$$a^* = 1 + \frac{\lambda(\overline{p}^2 + \sigma^2)}{2\overline{p}(1 - \lambda\overline{p})}$$

where $\overline{p}$ denotes the mean and $\sigma^2$ the variance of the distribution of processing times. Thus $a^*$ represents a threshhold for the aggregate tightness of due-dates beyond which job tardiness can be eliminated.

With regard to the family of assignment rules introduced earlier, workload information can certainly be taken into account. In order to preserve the characteristic that aggregate tightness is controllable, we still require $\overline{f} = a\overline{p}$.

**Proportional Workload Adjustments**

Perhaps the most direct way to incorporate workload status into the calculation of a flow allowance is to adjust the tightness parameter in proportion to the actual workload. For example, the CON rule could be modified so that:

$$f_j = \gamma W_j/W$$

where W is a scaling factor which assures that $\overline{f} = a\overline{p}$. In particular, W is the mean system workload, for which a standard formula exists. (See Appendix B.) With W defined this way, the mean value of $W_j/W$ will be 1, and we obtain $\overline{f} = \gamma = a\overline{p}$ as desired. For the SLK rule a proportional adjustment leads to the following form:

$$f_j = p_j + \beta, \text{ where } \beta = \overline{p}(a - 1)\, W_j/W$$

In this form $W_j$ again denotes the workload at $r_j$, excluding the work represented by job j itself. This definition seems appropriate because $\beta$ plays the role of a waiting time allowance. In order to satisfy the requirement $\overline{f} = a\overline{p}$, the constant W must again be defined as the mean system workload.

When this version of the SLK assignment rule is combined with ERD priority rule, the resulting tardiness performance is amenable to analysis. It is possible to show, for example, that when $a \geq a^*$ the combination SLK/ERD produces no tardiness. In addition, when $a < a^*$ the mean tardiness is given by $\overline{p}(a^* - a)$.

As it happens there is little change in performance when the EDD priority rule is substituted for the ERD rule. Thus the combination SLK/EDD produces no tardiness when $a \geq a^*$; otherwise, it is approximately true that $\overline{T} = \overline{p}(a^* - a)$. This means that SLK/EDD avoids tardiness at or above the tightness threshhold $a^*$ and produces very little tardiness in the vicinity of the threshhold.

One distinct feature of SLK/EDD involves the number of tardy jobs. When tardiness occurs under this combination, many jobs inevitably incur some tardiness, even if the mean tardiness is very low. In fact, the proportion of tardy jobs under SLK/EDD, when $a < a^*$, is approximately equal to the utilization level, commonly denoted $\rho$. Depending on the importance of secondary performance measures, this characteristic may be viewed as either desirable or undesirable.

In a similar fashion, it is possible to introduce a proportional adjustment of the tightness parameter in the TWK rule. The form of the rule becomes:

$$f_j = \alpha p_j, \text{ where } \alpha = a W_j/\overline{W}$$

For TWK, however, this form of the rule has an inescapable flaw. If the shop is empty when a job arrives then $W_j = 0$; consequently the flow allowance would be zero. It seems preferable in this case to define "workload" to include the processing time of the arriving job. This change yields the following form:

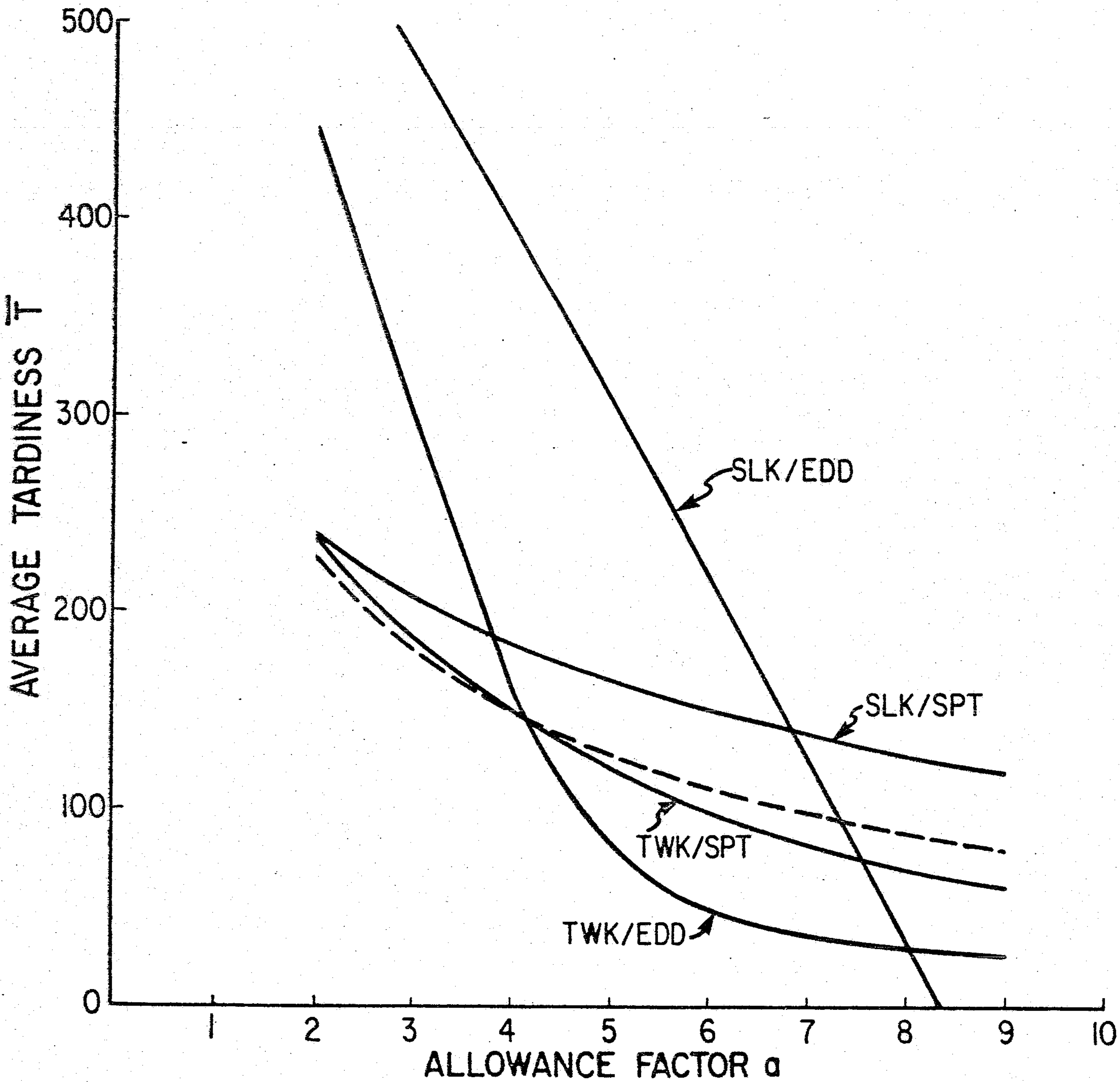
$$f_j = \alpha p_j, \text{ where } \alpha = a(W_j + p_j)/\overline{W}$$

where $\overline{W}$ is a scaling constant, chosen to meet the requirement $f = a\overline{p}$. (See Appendix B.) Unfortunately, the properties of the TWK rule do not seem analytically tractable to the same extent as the properties of SLK.

**Experimental Results**

The workload-dependent rules were implemented on the same test data, and the numerical results are given in Table A2 of Appendix A. Figure 3 also displays the graph of workload-dependent rule performance for the same test data on which Figure 2 is based. On the graph, the

**FIGURE 3**
**Simulation Results for Data Set II With Workload-Dependent Rules**




broken line shows the performance of TWK/SPT, which produced the lowest mean tardiness for this data set in the earlier runs.

One interesting observation to make at the outset is that the use of workload information, in the form of a proportional adjustment of the tightness parameter, did not always yield an improvement in tardiness performance. In fact, a comparison of Tables A1 and A2 reveals that the inclusion of workload information frequently led to greater tardiness when due-dates were very tight and the SPT priority rule was used. Thus, for example, the combination TWK/SPT without workload information still produced the lowest average tardiness in the comparisons over the range $2 \leq a \leq 4$ in both high-utilization cases.

On the other hand, there were conditions under which workload information provided a substantial improvement. This effect was most pronounced in the case of SLK/EDD. This combination was always dominated in the earlier runs, when no workload information was utilized; but with the proportional adjustment SLK/EDD yielded the least amount of tardiness when due-dates were relatively loose. As noted earlier, the SLK rule produces zero tardiness when $a \geq a^*$, even in combination with the simple sequencing rule ERT.

In between the region of very tight due-dates and the region of very loose due-dates, the best combination was always TWK/EDD with the workload adjustment. We hypothesize that this combination actually harnesses some of the benefits of shortest-first sequencing without having to rely explicitly on SPT priorities, due to the fact that there is a considerable shortest-first effect imbedded in the TWK rule when used in conjunction with EDD priorities.

In summary, we can see several crossover effects at work. For small values of the allowance factor $a$, the best policy is to use the SPT priority rule with either TWK or SLK (neither rule dominates in this region, and the performance differences are quite small). No workload information need be used. As the allowance factor increases there is eventually a crossover to TWK/EDD with workload information. Ultimately, as the allowance factor approaches $a^*$, there is a crossover from TWK to SLK. At $a^*$ and above, of course, the combination SLK/EDD is perfect.

### An Alternative Use of Workload Information

It should be recognized that workload status can be incorporated in the due-date assignment rule in various ways. The proportional form introduced above is perhaps the simplest way to reflect the workload in the flow allowance, but other logical ways certainly exist. In principle the only requirement is that the form of the rule must permit control of aggregate tightness, that is: $\bar{f} = a\bar{p}$.

As an example of an alternative form, consider the following version of the TWK rule:

$$f_j = \alpha p_j \text{ where } \alpha = 2aF(W_j)$$

where F(x) denotes the cumulative distribution function for workload. The relationship $\bar{f} = a\bar{p}$ is preserved by this form because the mean value of $F(W_j)$ will always be one-half and because $W_j$ and $p_j$ are independent. Unfortunately, the function F(x) is seldom known explicitly, even though it is independent of the dispatching rule. In the case of exponential processing times, however, we know from queueing theory that F(x) itself takes the form of an exponential distribution with mean $\bar{p}/(1-\lambda\bar{p})$. Thus, for example, in the case of Data Set II we have:

$$F(x) = 1 - e^{-0.0012x} \quad (x \geq 0)$$

The virtue of such a form of TWK is that flow allowances tend to be longer for small, positive workload levels than under the proportional form, where flow allowances are often too short for small jobs. Nevertheless, our experimental results indicate that this exponential form could have either a positive or a negative influence on tardiness performance. For some $a$-values, we found that the distribution form of TWK/SPT or TWK/EDD actually produced the best tardiness performance among all the combinations tested thus far. This improvement, however, comes at the expense of more extensive information (distribution information about the workload, rather than only mean-value information) which is available in closed form only for certain special cases. Nevertheless, these results suggest that a more sophisticated information base, or at least a more detailed form of workload information, might yet provide further improvements.

## SUMMARY AND CONCLUSIONS

In order to explore the properties of a production control system containing both due-date assignment and job scheduling decisions, we formulated a simple, two-level model of such a system. At the level of due-date assignment a policy constraint dictates the average tightness of the due-dates. At the level of job scheduling a priority dispatching rule determines a processing sequence, and performance is measured by average tardiness. We assumed throughout our investigation that due-dates would remain unchanged once assigned, so that the impact of revising or renegotiating due-dates was not considered.

Within the framework of the CON, SLK and TWK structures we examined the effects of different kinds of information as a basis for due-date assignments. We found it desirable to discriminate among jobs on the basis of their processing times and also (unless due-dates are very tight) to incorporate workload status in the assignment rule.

We initially compared five common priority rules but emphasized only two of them in most of our experiments. In particular, the SPT rule is effective against tardiness when due-dates are extremely tight, while the EDD rule is effective against tardiness when due-dates are loose. These properties give rise to a crossover effect in our results, in which first SPT and then EDD is preferred as the allowance factor is relaxed. This pattern suggests that an effective approach might be to use a dynamic priority scheme which takes on SPT or EDD features depending on the tightness of the due-dates in the system at any point in time.

Our main conclusions about this type of system are as follows. When due-dates are extremely tight, few jobs can be completed on time. In this situation mean tardiness is reduced roughly as mean flowtime is reduced. This characteristic makes it very desirable to use a flowtime-oriented priority rule, such as SPT. The choice of a due-date assignment rule in this situation does not seem to be of great importance, and even an unsophisticated rule will be adequate. On the other hand, when due-dates are extremely loose, "perfect" performance can be obtained only with a specific type of due-date assignment rule. (In our framework, this is the workload-dependent form of SLK.) There is no need for a sophisticated priority rule, and even a simple rule such as ERD can be used to advantage. There is also apparently a third, intermediate situation in which it is important to select both the sequencing rule and the assignment rule in concert and in which an elaborate information base appears to be desirable. In our framework, the combination TWK/EDD, with workload-dependent due-dates, yielded especially good performance in this intermediate situation.

Certain features of our approach are worthy of note. We advocate the use of due-date assignment rules that are analytically based. In particular, the selection of tightness parameter and the structure of workload-dependent forms are linked to analytic rather than empirical considerations. We observed that workload-dependent forms of due-date assignment rules are often advantageous. Eilon and Chowdhury [2] reached a similar conclusion using queue-length as an indicator of status. The behavior of queue-length in the one-machine case is, however, dependent on the sequencing rule, unlike the amount of work in queue, which is the status indicator in our procedures. We expect that work-in-queue (for which an adequate information base would seem no more difficult to maintain than for queue-length) would be more effective in the job shop case. We limited consideration to single-parameter due-date assignment rules, unlike Eilon and Chowdhury [2] and Weeks [7]. While a two-parameter form is easily imaginable (e.g. $d_j = r_j + \alpha p_j + \beta$) and even more complex rules are also conceivable (e.g. regression forms) we believe there is good reason to limit consideration to one-parameter rules for assigning due-dates. In a comprehensive control system there will typically be other decision rules (such as order releasing or sequencing) which themselves may have parameters to be set. Since the system's decision rules must all be coordinated, we believe that it is good design philosophy to associate only one parameter with each type of decision. In such an environment it is difficult enough to coordinate the various decision rules (i.e. by coordinating their parameters) without having the added problem of selecting an efficient set of parametric values for each rule.

Our own research will continue by exploring the desirability of a dynamic priority rule for sequencing jobs and also by extending the concepts to the job shop model. The job shop setting is intrinsically more complicated in several respects. For example, the shop workload is not independent of the dispatching rule and can even be viewed as multidimensional quantity. Nevertheless, Worrall and Mert [8] have exploited workload information in a type of job shop scheduling system, and it would appear that many of our results could be adapted to the job shop environment as well.

## REFERENCES


1. Baker, K. R. and J. W. M. Bertrand. "A Comparison of Due Date Selection Rules," *AIIE Transactions* (to appear).
2. Eilon, S. and I. G. Chowdhury. "Due Dates in Job Shop Scheduling," *International Journal of Production Research*, Vol. 14, No. 2 (1976), 223–237.
3. Eilon, S. and R. M. Hodgson. "Job Shops Scheduling with Due Dates," *International Journal of Production Research*, Vol. 6, No. 1 (1967), 1–13.
4. Elvers, D. A. "Job Shop Dispatching Using Various Due-Date Setting Criteria," *Production and Inventory Management*, Vol. 14, No. 4 (December 1973), 62–69.
5. Heard, E. L. "Due-Dates and Instantaneous Load in the One-Machine Shop," *Management Science*, Vol. 23 No. 4 (December 1976), 444–450.
6. Weeks, J. K. and J. S. Fryer. "A Methodology for Assigning Minimum Cost Due-Dates," *Management Science*, Vol. 23, No. 8 (April 1977), 872–881.
7. Weeks, J. K. "A Simulation Study of Predictable Due-Dates," *Management Science*, Vol. 25, No. 4 (April 1979), 363-373.
8. Worrall, B. M. and B. Mert. "Application of Dynamic Scheduling Rules in Maintenance Planning and Scheduling," *International Journal of Production Research*, Vol. 18, No. 1 (1980), 57–74.

## APPENDIX A
## SIMULATION RESULTS

### TABLE A1
**Average Tardiness Values Observed in the Initial Simulation Runs**

| | | Exponential Distribution | | | | | | Normal Distribution | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TWK | | SLK | | CON | | TWK | | SLK | | CON | |
| Utilization Level | Allowance Factor | SPT | EDD | SPT | EDD | SPT | EDD | SPT | EDD | SPT | EDD | SPT | EDD |
| Low | 2 | 95 | 174 | 97 | 200 | 177 | 248 | 207 | 219 | 211 | 223 | 217 | 228 |
| | 3 | 65 | 109 | 71 | 157 | 89 | 201 | 175 | 161 | 184 | 168 | 189 | 172 |
| | 4 | 46 | 63 | 56 | 127 | 72 | 164 | 154 | 120 | 166 | 127 | 170 | 131 |
| | 5 | 34 | 33 | 47 | 102 | 61 | 135 | 139 | 89 | 152 | 97 | 156 | 99 |
| High | 2 | 230 | 523 | 236 | 581 | 257 | 662 | 402 | 597 | 407 | 605 | 414 | 612 |
| | 3 | 182 | 414 | 198 | 513 | 221 | 589 | 358 | 500 | 370 | 514 | 377 | 521 |
| | 4 | 150 | 333 | 175 | 455 | 196 | 524 | 329 | 409 | 346 | 429 | 351 | 436 |
| | 5 | 128 | 278 | 159 | 403 | 178 | 465 | 307 | 328 | 326 | 351 | 332 | 358 |
| | 6 | 112 | 230 | 147 | 359 | 164 | 414 | 289 | 258 | 311 | 282 | 316 | 288 |
| | 7 | 99 | 192 | 138 | 321 | 154 | 368 | 273 | 197 | 298 | 221 | 302 | 226 |
| | 8 | 88 | 158 | 130 | 287 | 145 | 327 | 250 | 147 | 287 | 169 | 291 | 173 |
| | 9 | 79 | 128 | 123 | 257 | 137 | 290 | 247 | 108 | 276 | 127 | 280 | 130 |

### TABLE A2
**Average Tardiness Values Observed in Simulation Runs with Workload-Dependent Rules for Due-Date Assignment**

| | | Exponential Distribution | | | | | | Normal Distribution | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TWK | | SLK | | CON | | TWK | | SLK | | CON | |
| Utilization Level | Allowance Factor | SPT | EDD | SPT | EDD | SPT | EDD | SPT | EDD | SPT | EDD | SPT | EDD |
| Low | 2 | 100 | 105 | 92 | 174 | 143 | 200 | 211 | 192 | 208 | 200 | 241 | 201 |
| | 3 | 67 | 56 | 64 | 88 | 118 | 100 | 165 | 88 | 175 | 101 | 208 | 101 |
| | 4 | 49 | 39 | 42 | 0 | 100 | 63 | 132 | 16 | 149 | 1 | 183 | 45 |
| | 5 | 36 | 29 | 33 | 0 | 87 | 50 | 111 | 2 | 131 | 0 | 162 | 31 |
| High | 2 | 239 | 452 | 241 | 573 | 285 | 635 | 415 | 597 | 411 | 602 | 432 | 608 |
| | 3 | 189 | 299 | 209 | 482 | 256 | 534 | 371 | 491 | 376 | 504 | 395 | 508 |
| | 4 | 152 | 163 | 185 | 397 | 231 | 433 | 339 | 385 | 351 | 405 | 369 | 408 |
| | 5 | 122 | 82 | 167 | 309 | 212 | 333 | 314 | 279 | 332 | 306 | 349 | 308 |
| | 6 | 99 | 50 | 152 | 218 | 196 | 232 | 294 | 173 | 317 | 206 | 332 | 208 |
| | 7 | 83 | 36 | 139 | 127 | 182 | 131 | 277 | 81 | 303 | 106 | 318 | 108 |
| | 8 | 71 | 29 | 127 | 33 | 170 | 64 | 262 | 23 | 291 | 6 | 306 | 29 |
| | 9 | 60 | 25 | 118 | 0 | 160 | 47 | 249 | 4 | 281 | 0 | 294 | 12 |

## APPENDIX B. WORKLOAD ANALYSIS

In our test problems the shop workload varies over time in a pattern suggested by Figure B1. At the arrival of job j, the workload is increased by $p_j$: then it declines at the rate of $-1$ until the next arrival or until the system becomes empty, whichever occurs first. The workload is independent of the priority rule as long as the machine is never permitted to be idle when there is work in the system. (This is the case under the dispatching rules we tested).

The steady-state mean workload in such a queueing system is given by the following formula:

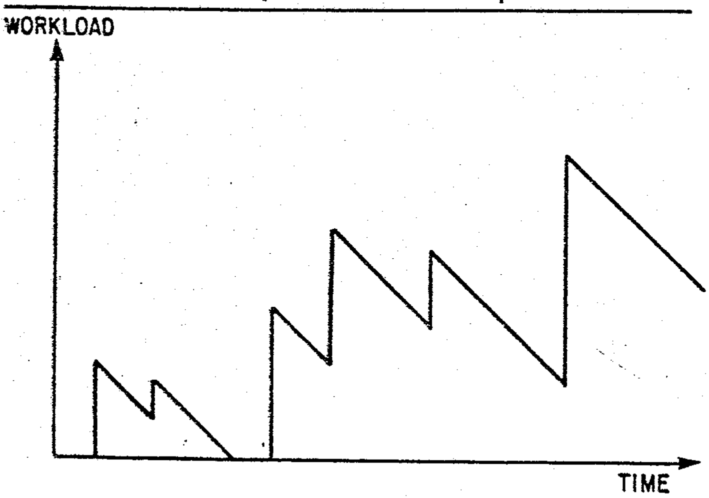
$$W = \frac{\lambda(\bar{p}^{-2} + \sigma^2)}{2(1 - \lambda\bar{p})} \qquad \text{(B1)}$$

where $\bar{p}$ denotes the mean and $\sigma$ denotes the standard deviation of the distribution of processing times.

If we observe the workload only at arrival times and include the processing time of the arriving job in the workload, then the mean workload observed will be equivalent to the mean flowtime under a first-come first-serve queue discipline. From the theory of Poisson queues, this quantity is known to be $W + \bar{p}$. The quantity $a^*$ is essentially the ratio of mean flowtime to mean processing time, by its definition. Thus

$$a^* = (W + \bar{p})/\bar{p} = 1 + W/\bar{p} \qquad \text{(B2)}$$
$$= 1 + \frac{\lambda(\bar{p}^2 + \sigma^2)}{2\bar{p}(1 - \lambda\bar{p})}$$

**FIGURE B1**
**A Sample Workload Graph**




In order for the adjusted parameters to be valid in the workload-dependent assignment rules, we require:

$$E[f_j] = a\bar{p} \qquad \text{(B3)}$$

For the SLK rule, $W_j$ is defined to exclude the processing time of the arriving job, and we obtain:

$$f_j = p_j + (a - 1)\,\bar{p}\,W_j/\overline{W}$$
$$E[f_j] = \bar{p} + (a - 1)\,\bar{p}\,E[W_j]/\overline{W}$$

It follows that (B3) is satisfied when $\overline{W} = E[W_j]$. Therefore the rule is valid when $\overline{W} = W$.

The situation is different for the TWK rule, where the workload is defined to include the processing time of the arriving job. Using the same definition of $W_j$ as the one above, we obtain:

$$f_j = a p_j(W_j + p_j)/\overline{W}$$
$$E[f_j] = a(E[p_j W_j] + E[p_j^2])/\overline{W}$$
$$= a(\bar{p}E[W_j] + \bar{p}^2 + \sigma^2)/\overline{W}$$
$$= a\bar{p}(E[W_j] + \bar{p} + \sigma^2/p)/\overline{W}$$

It follows that (B3) is satisfied when $\overline{W} = E[W_j] + \bar{p} + \sigma^2/\bar{p}$. Therefore, the TWK rule is valid when $\overline{W} = W + \bar{p} + \sigma^2/\bar{p}$.

## APPENDIX C. Properties of the SLK Rule

In the workload-dependent form of the SLK rule, the due-date for job j is calculated as follows:

$$d_j = r_j + p_j + \bar{p}(a - 1)W_j = r_j + p_j + \theta W_j \qquad \text{(C1)}$$

When the coefficient $\theta$ is equal to 1, the due-date is equal to the completion time that would be realized for job j under ERD priority sequencing. In other words, the combination SLK/ERD will yield no tardiness at all, if and only if $\theta \geq 1$.

The allowance factors corresponding to this range can be determined as follows. From (C1), the average flow allowance is seen to be $\bar{f} = \bar{p} + \theta W$, where W denotes the mean workload. Note that

$$a = \bar{f}/\bar{p} = 1 + \theta W/\bar{p}$$

and recall from (B2) that $a^* = 1 + W/\bar{p}$. Therefore the range $\theta \geq 1$ is identical to the range $a \geq a^*$.

Now consider the case $\theta < 1$. In this case all jobs will be tardy except those jobs that arrive to an empty system. From the theory of Poisson queues we know that the steady-state proportion of such jobs is equal to $\rho$, the system utilization. Furthermore, the tardiness of job j is simply $(1-\theta)\, W_j$ under the rule in (Cl). Thus:

$$T_j = (1-\theta)\, W_j = W_j[1 - \bar{p}\,(a-1)/W] = W_j\,([W - \bar{p}\,(a-1)]\,/W)$$

Since this is a proportional relationship between $T_j$ and $W_j$, it is straightforward to write $\bar{T}$ in terms of W, the average value of $W_j$. We obtain:

$$\bar{T} = W([W - p\,(a-1)]/W) = W - \bar{p}\,(a-1)$$

From equation (B2) we substitute $W = \bar{p}(a^* - 1)$. Hence:

$$\bar{T} = \bar{p}(a^* - 1) - \bar{p}\,(a-1) = \bar{p}\,(a^* - a)$$

In summary, the combination SLK/ERD yields the following tardiness performance:

$$\begin{aligned} \bar{T} &= \bar{p}(a^* - a) && \text{for } a < a^*\ (\theta < 1) \\ &= 0 && \text{for } a \geqslant a\ \ (\theta \geqslant 1) \end{aligned} \qquad \text{(C2)}$$

We next consider the effect of using the EDD priority rule in place of ERD. We know that the tardiness performance of EDD will be given by (C2) as long as EDD yields the same job sequence as ERD. Therefore, we investigate the conditions under which the rules yield different sequences. Suppose that jobs j and k are two jobs that arrive consecutively, with $r_j < r_k$. Assuming that job j has not been initiated at time $r_k$, the relationship between workloads occurring at arrival times is the following:

$$W_k = W_j + p_j - (r_k - r_j) \qquad \text{(C3)}$$

(If job j has been initiated at time $r_k$, then, of course, the jobs are processed in arrival order.) In order for EDD and ERD to yield a different sequence for the pair j and k we must have $d_k < d_j$; that is,

$$r_k + p_k + \theta W_k < r_j + p_j + \theta W_j$$

From (C3) we obtain

$$\begin{aligned} r_k + p_k + \theta(W_j + p_j - r_k + r_j) &< r_j + p_j + \theta W_j \\ p_k &< (\theta - 1)\,(r_k - r_j - p_j) \end{aligned} \qquad \text{(C4)}$$

It follows directly that when $\theta = 1$ ($a = a^*$) there will be no difference between EDD and ERD.

For the case $\theta > 1$ we have already seen that ERD yields no tardiness. Tardiness would be incurred under EDD only if it induced a different sequence for jobs j and k, the condition for which is given by (C4), and additionally if job j is tardy as a result of its delay. This condition is simply $C_j > d_j$, where $C_j$ is the time at which job k would have completed had there been no change in sequence. We thus require:

$$\begin{aligned} r_k + p_k + W_k &> d_j \\ r_j + p_j + p_k + W_j &> r_j + p_j + \theta W_j \\ p_k &> (\theta - 1)\, W_j \end{aligned} \qquad \text{(C5)}$$

In order for (C4) and (C5) to both be satisfied, it would be necessary to have:

$$\begin{aligned} W_j &< r_k - r_j - p_j \\ r_j + p_j + W_j &< r_k \end{aligned}$$

However, this condition dictates that job j would be completed (and the system would be idle) by the release date of job k. Therefore, no interchange could occur. This argument demonstrates that there can be no tardiness under EDD when $\theta > 1$. (Although we have considered only the case in which job j is delayed one position in sequence, the argument can be extended by letting k denote the last job past which job j is delayed and by replacing $p_k$ by the length of its delay due to the change in sequence.)

Now consider the case $\theta < 1$. To the extent that EDD yields the same job sequence as ERD, we know that (C2) will describe the behavior of mean tardiness. Again we know that (C4) gives a sufficient condition for EDD to differ from ERD, for the pair of adjacent jobs j and k. This inequality can equivalently be written:

$$r_k - r_j < p_j - p_k/(1-\theta) \qquad \text{(C6)}$$

Obviously, this condition depends on the processing times of the pair of jobs involved as well as on the length of the interval between their arrival times. The condition is quite likely to fail when $\theta$ is close to 1, and it is certain to fail whenever $p_j < p_k/\,(1-\theta)$. To get some idea of how likely this latter (sufficient) condition is to occur, consider the case in which $p_j$ and $p_k$ are samples from a normal distribution with mean 100 and standard deviation 25, and suppose $\theta = 0.5$. It is straightforward to calculate that

$$P[p_j < p_k/\,(1-\theta)] = 1 - \phi\,(1.79) = 0.963$$

Thus, with probability of at least 0.96, it is impossible for jobs j and k to arrive sufficiently close in time to cause the EDD rule to differ from ERD order. In addition, the complementary probability of 0.04 includes many event combinations in which the right-hand side of (C6) is positive but in which the condition still fails. For this reason, it is logical to expect SLK/EDD to produce tardiness performance closely resembling that of SLK/ERD.