A comparison of due-date selection rules

Citation for published version (APA):
Baker, K. R., & Bertrand, J. W. M. (1981). A comparison of due-date selection rules. *IIE Transactions*, *13*(2), 123-131. https://doi.org/10.1080/05695558108974544

DOI:
10.1080/05695558108974544

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)

A Comparison of Due-Date Selection Rules

K. R. BAKER
The Amos Tuck School of Business Administration
Dartmouth College
Hanover, New Hampshire

J.W.M. BERTRAND
Department of Industrial Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands

Abstract: In sequencing and scheduling models it is usually assumed that due dates represent exogeneous information. In many practical settings, however, due dates can be discretionary, or at least negotiable. Relatively few studies have incorporated discretionary due dates, and even then the rules proposed for due-date selection have seldom been developed from normative, analytic results. In this research we return to very basic scheduling models in search of fundamental insights and relationships that suggest guidelines for due-date selection in more complicated situations. We exploit some fundamental results from scheduling theory involving the single-machine model in order to compare three basic strategies for due-date selection.

Due dates are usually treated as "given" information which takes the form of input to a scheduling problem; in actual practice, however, the due date can be a decision variable within the boundary of the scheduling problem. Relatively few studies have incorporated discretionary due dates (see, for example, [2], [3], and [5]), and even then the rules examined for due-date selection have not been developed from normative, analytic results. Our main purpose is to focus attention on due-date selection, which is an important and often overlooked decision process in scheduling problems involving deadlines.

Consider a model in which there are n jobs to be processed by a single machine. Job j is characterized by a ready time (r_j) and a *processing time* (p_j), both known in advance. In the *static* single-machine model, all jobs are simultaneously available, so that $r_j = 0$. In the more general *dynamic* model jobs become ready (schedulable) at different points in time. We assume that jobs can be processed in a preempt-resume mode. That is, the processing of a job can be interrupted and resumed at a later time with no need to repeat any work. Thus the total amount of time the job spends in process must equal p_j, although the interval between the job's initiation and its completion may well be longer than p_j. The *completion time* of job j, which is determined by scheduling decisions, is denoted C_j.

The due date of job j (denoted d_j) is determined by a particular selection rule. We consider three selection rules that represent sensible ways of using available information to set norms for flow allowances. (A job's flow allowance is the length of time from its ready time to its due date.) These rules are the following:

CON: jobs are given *constant* flow allowances, so that $d_j = r_j + \gamma$.

SLK: jobs are given flow allowances that reflect equal waiting times, or equal *slack*, so that $d_j = r_j + p_j + \beta$.

TWK: jobs are given flow allowances that are proportional to the *total work* they require, so that $d_j = r_j + \alpha p_j$.

Each rule contains a single parameter (α, β, or γ) still to be determined.

In this context the role of a scheduling procedure is to construct a schedule that meets some specified performance objective. In sequencing against due dates, the primary objective is usually to complete all jobs on time. When due dates are discretionary, of course, this objective can be met by allowing the due dates to be very loose. We believe, however, that in an environment where due dates

Paper was handled by the organizers of the Special Report on Production and Inventory Control, Professors William L. Berry and Vincent A. Mabert.

0569-5554/81/0600-0123/$2.00/0

can be selected, an implicit objective is to assign due dates to be as tight as possible. Tight due dates attract more customers than loose due dates in a competitive market and imply better customer service. Tighter due dates also tend to produce lower in-process inventory levels, assuming that they induce relatively more shortest-first sequencing. Tight due dates are therefore desirable in scheduling. To formalize these considerations, we define the scheduling criterion to be one of minimizing the average due date subject to the feasibility requirement that all jobs must be completed on time. Specifically, the quality of a given set of assigned due dates is measured by the average value

$$\bar{d} = \sum_{j=1}^{n} d_j/n$$

in a schedule where $C_j \leq d_j$.

In principle, an *optimal* solution to this problem can be found. For any schedule, the tightest possible set of feasible due dates is given by $d_j = C_j$. Therefore $\bar{d}$ can be minimized by a sequencing procedure that minimizes the mean completion time. For instance, in the case of a finite set of simultaneously available jobs, the optimal procedure is shortest processing time (SPT) sequencing. In this case an optimal set of due dates could be calculated by first constructing the SPT sequence and then setting $d_j = C_j$ for each job in the sequence. Nevertheless, this solution implies that the due date of each job is dependent on specific information about every job. In more realistic, complex problems it is not practical to require such an extensive information base for selecting due dates. A more practical approach is to restrict attention to such schemes as TWK, SLK, and CON, where the selection of a due date depends only on information about the job itself (r_j and p_j) and on a tightness parameter (α, β, or γ). Our purpose, then, is to study these three rules in order to determine their relative effectiveness at minimizing $\bar{d}$. In addition, the theoretically optimal value of $\bar{d}$ can be used as a benchmark in order to assess the rules' absolute effectiveness.

Analysis of the Static Model

In the static version of the problem all jobs are simultaneously available at time zero. In this case the maximum lateness is minimized by sequencing the jobs according to the Earliest Due-Date (EDD) priority rule. The EDD sequence will yield a schedule with no late jobs, if such a schedule exists. Therefore, in determining the parameters for the due date rules, it is sufficient to assume that the EDD sequence applies.

Under the CON due-date rule, $d_j = \gamma$. In other words, all jobs have the same due date and every sequence is an EDD sequence. For all jobs to be completed on time, the due dates must satisfy

$$\begin{aligned} d_j &\geq C_j \quad \text{for all jobs } j, \\ \gamma &\geq C_j, \\ \gamma &= \operatorname{Max}_j (C_j), \\ \gamma &= \sum_{j=1}^{n} p_j. \end{aligned} \tag{1}$$

Thus the requirement that the last job be on time dictates the smallest feasible value of γ, and it is clear from the conditions in (1) that the choice of a job sequence does not affect γ. The average due date generated by the CON rule is easily seen to be

$$\bar{d}_{\text{CON}} = \gamma = \sum_{j=1}^{n} p_j. \tag{2}$$

Under the SLK due-date rule, $d_j = p_j + \beta$. Jobs with different processing times will have different due dates, and the EDD sequence is equivalent to the SPT sequence. Assume that the jobs are indexed according to SPT. For all jobs to be completed on time, the due dates must satisfy

$$\begin{aligned} d_j &\geq C_j \quad \text{for all jobs } j, \\ p_j + \beta &\geq C_j, \\ \beta &= \operatorname{Max}_j (C_j - p_j), \\ \beta &= \operatorname{Max}_j \left(\sum_{i=1}^{j-1} p_i \right), \\ \beta &= \sum_{i=1}^{n-1} p_i. \end{aligned} \tag{3}$$

Thus the requirement that the last job be on time dictates the smallest feasible value of β, which is equal to the time needed to process all jobs except the longest. It is also clear from the conditions in (3) that no sequence can produce a smaller value of β than the SPT sequence. The average due date generated by the SLK rule is given by

$$\bar{d}_{\text{SLK}} = \bar{p} + \beta = \bar{p} + \sum_{j=1}^{n-1} p_j. \tag{4}$$

Under the TWK due-date rule, $d_j = \alpha p_j$. Again, the EDD sequence is equivalent to the SPT sequence, and we assume the jobs are indexed in that order. For all jobs to be completed on time, the due dates must satisfy

$$\begin{aligned} d_j &\geq C_j \quad \text{for all jobs } j, \\ \alpha p_j &\geq C_j, \\ \alpha &= \operatorname{Max}_j (C_j/p_j), \\ \alpha &= \operatorname{Max}_j \left(\sum_{i=1}^{j} p_i \Big/ p_j \right). \end{aligned} \tag{5}$$

We know that $1 \leq \alpha \leq n$, but the maximum ratio satisfying (5) can occur at any position in the sequence, depending on the specific problem data. The average due date generated by the TWK rule is given by

$$\bar{d}_{\mathrm{TWK}} = \alpha\bar{p} = \bar{p} \max_j \left(\sum_{i=1}^{j} p_i \Big/ p_j \right). \qquad (6)$$

The following example illustrates the behavior of the due-date rules:

Job j	1	2	3
p_j	1	2	16

First, consider the two-job problem consisting of the first two jobs. From the formulas in (1) - (6) we obtain

$$\gamma = 3 \qquad \beta = 1 \qquad \alpha = 1.5$$

$$\bar{d}_{\mathrm{CON}} = 3 \qquad \bar{d}_{\mathrm{SLK}} = 2.5 \qquad \bar{d}_{\mathrm{TWK}} = 2.25 \ .$$

Thus the TWK rule fares best and the CON rule worst. Next consider the three-job problem formed when job 3 is added to the job set. This time the formulas yield

$$\gamma = 19 \qquad \beta = 3 \qquad \alpha = 1.5$$

$$\bar{d}_{\mathrm{CON}} = 19 \qquad \bar{d}_{\mathrm{SLK}} = 9.3 \qquad \bar{d}_{\mathrm{TWK}} = 9.5 \ .$$

Here, the SLK rule is best and the CON rule worst. Evidently, it is possible that either TWK or SLK produces the best set of due dates in a given static problem.

Some additional insight can be gained by analyzing special cases. For this purpose we define the tightness ratio of rule R as follows:

$$T(R) = \bar{d}_R / \bar{d}_{\mathrm{OPT}} \ ,$$

where $\bar{d}_{\mathrm{OPT}}$ is the average completion time under the optimal (SPT) schedule. Below we summarize the behavior of $T(R)$ for certain very special static problems. (See Appendix 1 for detailed derivations.)

- **Indistinguishable jobs** ($p_j = p$ for $j = 1, 2, \ldots, n$). Intuitively we should expect that the three rules exhibit comparable tightness in this case because the ability of TWK and SLK to discriminate among the jobs according to processing time is of no help. In fact, it can be shown that in this case

  $$T(\mathrm{TWK}) = T(\mathrm{SLK}) = T(\mathrm{CON}) = 2n/(n+1) \ .$$

  In particular, the rules have identical tightness ratios. Furthermore, as n becomes large the ratios approach 2.

- **Dominant job** ($p_j = 1$ for $j = 1, 2, \ldots, n-1$ and $p_n = p > 2$, where $n > 2$). This case is only slightly different than the previous case, but there is a perceptible difference in the three tightness ratios. In particular as p becomes large, the following limiting ratios hold:

  $$T(\mathrm{TWK}) = n-1 \qquad T(\mathrm{SLK}) = 1 \qquad T(\mathrm{CON}) = n \ .$$

  Therefore, the tightness ratios of TWK and CON can be made arbitrarily large if sufficiently large values of n and p are chosen.

- **Distinct jobs** ($p_j = j$ for $j = 1, 2, \ldots, n$). This case lies at the opposite end of the spectrum, in a certain sense, from the first case. When we examine the limiting behavior of the tightness ratios as n becomes large, we find

  $$T(\mathrm{TWK}) = 1.5 \qquad T(\mathrm{SLK}) = 3 \qquad T(\mathrm{CON}) = 3 \ .$$

  Therefore the ratios lie between 1 and 3, but in this case TWK is systematically better than the other two rules.

The three special cases and the numerical example given above suggest strongly that the CON rule is the least desirable of the three rules and that the choice between TWK and SLK depends on specific conditions in the job set to be scheduled. We can indeed state this relationship formally (see Appendix 2 for proofs).

PROPERTY 1. $\bar{d}_{\mathrm{SLK}} \leq \bar{d}_{\mathrm{CON}}$ (SLK dominates CON)

PROPERTY 2. $\bar{d}_{\mathrm{TWK}} \leq \bar{d}_{\mathrm{CON}}$ (TWK dominates CON)

Thus in the static problem the CON rule is dominated by the TWK and SLK rules in the sense that they allow for tighter average due dates in a feasible schedule. Meanwhile, no dominance exists in general between TWK and SLK. In the next section the performance of the three rules is studied in a number of test problems with randomly generated processing times.

Experiments with the Static Model

In order to get an indication of how the three rules perform for a variety of job sets, several test problems were generated. In the case of the static model, a problem instance consists of a specification of n processing times. In each test problem the processing times were sampled as integers from a particular probability distribution.

In the first set of test problems, the processing times were drawn from an exponential distribution with mean 100. Problems of size n=5, 10, 20, 40, and 80 were created, and for each problem size there were 20 replications. For rule R the average due date ($\bar{d}_R$) was calculated in each test problem. Then the mean of these values was computed for the sample of 20 problems, and the frequency with which TWK or SLK yielded the best set of due dates was also recorded. A summary of these results is given in Table 1, where the mean flow allowance is expressed as a multiple of the mean processing time.

Table 1: Results for static problems with exponential samples.

Problem	Mean Flow Allowance				Frequency Best		
Size	TWK	SLK	CON	OPT	TWK	SLK	CON
$n=5$	3.09	4.23	5.23	2.26	20	0	0
$n=10$	4.22	7.97	9.95	3.16	20	0	0
$n=20$	7.32	17.02	19.51	5.60	20	0	0
$n=40$	14.76	37.74	41.04	11.31	20	0	0
$n=80$	26.27	76.57	80.69	20.60	20	0	0

Table 2: Results for static problems with truncated normal samples (mean = 100 and $n = 5$).

Standard	Mean Flow Allowance				Frequency Best		
Deviation	TWK	SLK	CON	OPT	TWK	SLK	CON
$\sigma = 0$	5.00	5.00	5.00	3.00	20	20	20
$\sigma = 100$	2.98	4.43	5.49	2.33	20	0	0
$\sigma = 200$	3.00	4.62	5.95	2.29	20	0	0
$\sigma = 400$	4.24	5.54	8.25	2.79	12	8	0
$\sigma = 800$	6.97	8.70	13.02	4.48	13	7	0

The data in Table 1 indicate that TWK is by far the most effective scheme for setting due dates among the three rules compared on the exponential samples. Similar results were encountered when processing times were drawn from uniform and normal distributions. With the exception of a small number of test problems in which $n \leq 10$, there were no problem instances encountered where SLK produced the lowest average due date. Moreover, the tightness ratio for TWK appeared to be quite insensitive to the problem size, while the ratios for SLK and CON tended to increase with larger problem sizes.

Nevertheless, as the analytic results for special cases will suggest, there are problem structures in which TWK performs less effectively. Another series of test problems was created in which processing times were sampled from a normal distribution with a relatively large variance; when ordinary sampling produced a processing time less than or equal to zero, the number was reset to 1. Thus the actual problem data represented samples from a truncated normal distribution. The relative proportion of jobs with processing time equal to one increases with the standard deviation used in the primary sampling routine. As this standard deviation is increased, the problem data begin to resemble the dominant job case considered earlier, at least in the sense that several of the jobs have the minimum processing time. The dominant job case was shown to be one in which TWK performs poorly compared to SLK; therefore, it might be expected that TWK would show less dominant performance in this second set of test problems than in the first. Table 2 shows a summary of the results for this set, with a problem size of $n=5$, where the primary sampling process involved a normal distribution with a mean equal to 100 and standard deviation as shown. Again, the mean flow allowances are shown as multiples of the mean processing time (which in this case is always larger than 100 due to truncation).

The first entry in Table 2, where $\sigma=0$, corresponds to the benchmark case of indistinguishable jobs, which was analyzed previously. When $\sigma=100$, truncation occurs for about 16% of the processing time samples, but TWK still dominated SLK and CON. As σ was increased further, the mean flow allowances increased under all of the rules. Under TWK there was considerably more variation in the tightness ratio than under SLK. This behavior accounts for the fact that SLK was frequently the better rule even though its mean flow allowance remained larger. Nevertheless, at $\sigma=800$ (where truncation occurs about 45% of the time), the TWK rule still yielded a better mean flow allowance than SLK and provided the tightest due dates in nearly two-thirds of the test problems.

From the results in Tables 1 and 2 we conclude that TWK is empirically dominant, if not theoretically dominant, in randomly generated static problems. Compared to SLK and CON, TWK generated average due dates that are relatively less sensitive to problem size, which suggests that TWK might adapt more effectively to situations where the workload varies. In addition, on the spectrum between the case of all jobs with the same processing time and the case of many jobs with the same (minimum) processing time, TWK apparently retains its dominant performance except at the very extremes of the spectrum. This feature suggests that the TWK rule might exhibit a desirable kind of insensitivity when implemented in more complex settings.

Analysis of the Dynamic Model

In the dynamic version of the problem the jobs can have different ready times, but we assume that the jobs can be scheduled in a preempt-resume mode. In this case the maximum lateness is minimized by a preemptive version of the EDD priority rule: the job being processed should always be the job with the minimum due date among the ready jobs. This means that preemption will occur whenever an arriving job is assigned a due date earlier than that of the job in process.

Under the CON due date rule, $d_j = r_j + \gamma$, and all jobs have the same flow allowance. As a consequence, the earliest due-date order is identical to the earliest arrival order, and preemption is never necessary. For all jobs to be completed on time, the due dates must satisfy

$$d_j = r_j + \gamma \geq C_j ,$$
$$\gamma = \operatorname*{Max}_j (C_j - r_j) = \operatorname*{Max}_j (F_j) , \qquad (7)$$

where F_j denotes the flowtime of job j. In general, it is not possible to derive an explicit formula for the maximum flowtime, but it is not difficult to perform the necessary calculations. Suppose that the jobs are numbered according to their ready times, so that $r_j \leq r_{j+1}$. Then the completion times are given by the following recursive relationship:

$$C_j = \text{Max}(C_{j-1}, r_j) + p_j .$$

Furthermore, the flow times satisfy the following:

$$F_j = C_j - r_j = \text{Max}(C_{j-1}, r_j) + p_j - r_j$$
$$= \text{Max}(r_{j-1} + F_{j-1}, r_j) - r_j + p_j$$
$$= \text{Max}(F_{j-1} + r_{j-1} - r_j, 0) + p_j .$$

From this recursive relation, the values of F_j can be calculated successively and the maximum identified. The average due date is then

$$\bar{d}_{\text{CON}} = \bar{r} + \gamma = \bar{r} + \text{Max}_j(F_j) . \quad (8)$$

Under the SLK rule, $d_j = r_j + p_j + \beta$. In other words, all jobs have the same waiting time allowance. The EDD priority rule reduces to dispatching the jobs according to the minimum value of $(r_j + p_j)$, and so preemption may be necessary. For all jobs to be completed on time the due dates must satisfy

$$d_j = r_j + p_j + \beta \geqslant C_j ,$$
$$\beta = \text{Max}_j (C_j - r_j - p_j) . \quad (9)$$

No simple formula can be given for the maximum waiting time, but again direct calculations are straightforward. Since the EDD priority rule does not depend on β, an optimal schedule can be constructed by assuring that the available job with minimum $(r_j + p_j)$ is always in process. As this schedule is constructed, the waiting times can be calculated and the maximum identified. The average due date is then

$$\bar{d}_{\text{SLK}} = \bar{r} + \bar{p} + \beta = \bar{r} + \bar{p} + \text{Max}(C_j - r_j - p_j) . \quad (10)$$

Under the TWK rule, $d_j = r_j + \alpha p_j$. In this case the EDD rule cannot be implemented until α is known, in contrast to the nature of the EDD rule under CON and SLK. Therefore, it is not possible to construct an EDD schedule by proceeding from time zero with a dispatching mechanism, and a slightly more complicated scheduling procedure must be utilized. For all jobs to be completed on time the due dates must satisfy

$$d_j = r_j + \alpha p_j \geqslant C_j ,$$
$$\alpha = \text{Max}_j [(C_j - r_j)/p_j] . \quad (11)$$

Again, no simple formula can be given for the optimal parameter value in Equation (11), but a direct calculation is possible. Think of each job as having a cost function $g_j(C)$, representing the cost incurred when job j is completed at time C. In [1], an algorithm is given for minimizing the maximum cost for the single-machine problem with preemption. This algorithm can be implemented in the special case for which $g_j(C) = (C - r_j)/p_j$; the minimum value of the maximum cost is just the value of α defined by (11). Once this value has been calculated, the average due date is:

$$\bar{d}_{\text{TWK}} = \bar{r} + \alpha\bar{p} = \bar{r} + \bar{p}\,\text{Max}_j [(C_j - r_j)/p_j] . \quad (12)$$

In the dynamic model there is no dominance of the kind that occurs in Properties 1 and 2 for the static model. Consider the example below:

j	1	2	3
r_j	0	2	4
p_j	4	3	2

Under the TWK rule, job 1 completes at the end of the schedule (at time 9), so that $\alpha = 2.25$ and $\bar{d}_{\text{TWK}} = 8.75$. Under SLK, the jobs are scheduled in numerical order, so that $\beta = 3$ and $\bar{d}_{\text{SLK}} = 8$. Under CON, the jobs are scheduled in the same way; then $\gamma = 5$ and $\bar{d}_{\text{CON}} = 7$. As the example shows, the CON rule may be desirable in dynamic problems even though it is dominated in the static model.

Experiments with the Dynamic Model

The dynamic model is more complicated than the static model in the sense that a problem instance consists of a specification of n ready times as well as n processing times. Most computational studies of dynamic single-machine models reported in the literature rely simply on uniform distributions for generating the ready times and processing times for test problems. In this study, however, we describe two fundamentally different scenarios for the dynamic model, based on features of practical production control settings.

In a static problem the number of jobs is a reasonable measure of problem difficulty, but for dynamic problems the number of jobs is not as good a measure. Although a problem instance may contain n jobs, in the dynamic model the number of jobs competing to be the next at any point in time may always be much smaller than n. Whereas the parameter n indicates how many due dates are set and approximately how many dispatching decisions are made, it does not quite capture the difficulty of those decisions. The difficulty of a dispatching decision is related instead to the number of schedulable jobs at the time the decision must be made. Hence, we measure problem difficulty by quantifying the workload pattern underlying the dispatching decisions.

We distinguish between two opposite types of workload patterns that can be found in actual shops. The first type is a *random workload pattern*, which arises when release dates correspond to the occurrences of customer demands, which themselves occur randomly. This pattern is characteristic of production systems where the market is competitive and jobs tend to be custom orders. The second type is a *controlled workload pattern*, which arises when release dates are selected with a goal of maintaining a fairly stable workload in the processing facility. This pattern is characteristic of production systems where the market is captive or jobs tend to be internal orders for replenishment of stock. Clearly the two types are different in that the controlled pattern results from a conscious effort to insulate the workload from the randomness that characterizes external demand. In the random pattern, variations in demand intensity are transmitted directly to the workload faced by the shop.

In order to model the random workload pattern in our test problems we represented the occurrence of ready times as a Poisson process. Specifically, the interval between successive ready times was drawn from an exponential distribution with mean $1/\lambda$. We also represented processing times as samples from an exponential distribution with mean equal to 100. Thus the workload varied probabilistically in a manner that could be described by an $M/M/1$ queueing model. In particular, the mean workload, denoted $\overline{W}$, that occurs in such a system is given by the formula

$$\overline{W} = 100\lambda/(0.01 - \lambda)\,. \qquad (13)$$

In addition, the mean utilization is equal to 100λ.

In order to model the controlled workload pattern, we selected a job-releasing scheme of the following form:

Whenever the workload in the shop falls to W_0, a new job is released.

The behavior of the actual workload under this rule can be described by a "saw-tooth" graph similar to the one shown in Fig. 1. The workload drops to its minimum level W_0, at which time a new arrival is triggered. The arrival of job j increases the workload by p_j, and then the workload again falls toward its minimum level. The mean workload in this shop will be $W_0 + \bar{p}/2$. Setting this quantity equal to the mean workload given in (13) yields

$$W_0 = (150\lambda - 0.50)/(0.01 - \lambda)\,. \qquad (14)$$

Thus Equation (14) allows the parameter W_0 to be chosen so that the mean workload is the same under the controlled and random patterns. We selected the following parameter values for the test problems:

Data set	1	2	3
Utilization (random pattern)	0.80	0.85	0.90
W_0 (controlled pattern)	350	517	850
Mean number of jobs in system	4.00	5.67	9.00

Although the mean workload is the same in the problems of a given data set, we should still expect to find some differences in the behavior of average due dates in the random and controlled scenarios because system utilization under the controlled workload pattern is always 100%.

Each data set consisted of 20 problems for the random workload pattern and 20 problems for the controlled workload pattern. Each test problem contained 80 jobs. (We did not study the effect of problem size because the 80-job problem can be viewed as simply an extended version of a 20-job or 40-job problem.)

The tightness of the due dates under rule R is measured by the average flow allowance $(\bar{d}_R - \bar{r})$ in a given test problem. The mean of these values in the 20 test problems (expressed as a multiple of the mean processing time) is given in the tables. The optimal rule in the dynamic model is the Shortest Remaining Processing Time discipline [4].

The results of the test problems are summarized in Tables 3 and 4. The results in Table 3 indicate that TWK is by far the best rule for the random workload pattern: its mean flow allowance was considerably less than the mean for SLK or CON, and it produced the tightest due dates in all 60 problems. In Table 4 the observations were quite

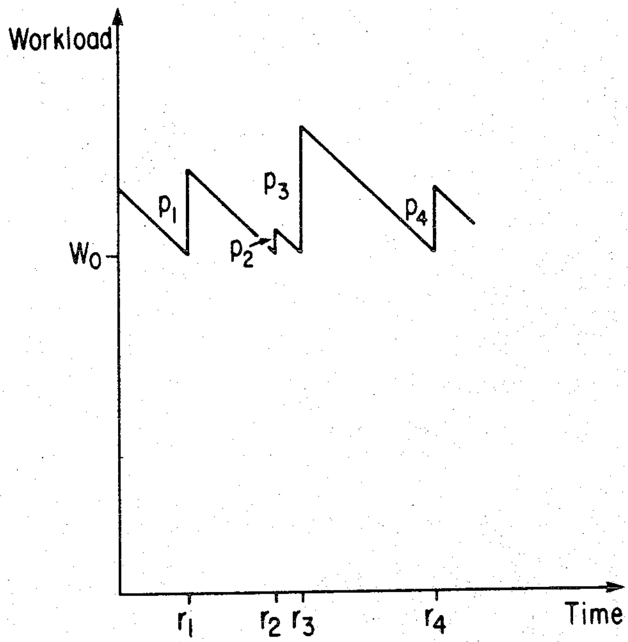

Fig. 1. An example of the controlled workload pattern.

Table 3: Results for dynamic problems with exponential samples and random workload.

Utilization	Mean No. of Jobs	Mean Flow Allowance				Frequency Best		
		TWK	SLK	CON	OPT	TWK	SLK	CON
0.80	4.00	4.43	9.04	10.14	1.99	20	0	0
0.85	5.67	5.63	10.37	11.39	2.41	20	0	0
0.90	9.00	6.20	11.79	12.76	2.49	20	0	0

Table 4: Results for dynamic problems with exponential samples and controlled workload.

Mean Workload (No. of Jobs)	Mean Flow Allowance				Frequency Best		
	TWK	SLK	CON	OPT	TWK	SLK	CON
4.00	5.26	4.53	8.79	2.17	3	17	0
5.67	6.51	6.23	10.09	2.69	7	13	0
9.00	8.28	9.51	13.49	3.52	17	3	0

different. In the controlled workload pattern the SLK rule emerged as the best rule in test problems with relatively light workloads. The TWK rule, however, produced due dates that were less sensitive to workload, and this characteristic led TWK to again emerge as the better rule when the workload was increased.

There are some systematic factors occurring in the case of the controlled workload pattern that help explain these results. Under CON there is no preemption, and it follows that the completion time of job j is given by $C_j = r_j + W_0 + p_j$. Therefore, $\lambda = W_0 + \text{Max}(p_j)$ and we obtain:

$$\bar{d}_{\text{CON}} = \bar{r} + W_0 + \underset{j}{\text{Max}}(p_j) .$$

Under SLK there is also no preemption (because $r_j + p_j > r_{j-1} + p_{j-1}$ in our controlled workload model), so that $C_j = r_j + W_0 + p_j$, as in the case of CON. However, for SLK this means $\beta = W_0$ and

$$\bar{d}_{\text{SLK}} = \bar{r} + \bar{p} + W_0 .$$

Thus SLK dominates CON in the controlled workload pattern.

Under TWK, however, preemption may occur. Let p^* denote the processing time of the shortest job that does not preempt other jobs while it is in the system. For this job, $C_j = r_j + W_0 + p^*$, and it follows that $\alpha p^* \geq W_0 + p^*$. Thus the parameter must satisfy

$$\alpha \geq 1 + W_0/p^* .$$

It follows that

$$\bar{d}_{\text{TWK}} \geq \bar{r} + \bar{p}(1 + W_0/p^*)$$

and

$$\bar{d}_{\text{TWK}} - \bar{d}_{\text{SLK}} = W_0(\bar{p} - p^*) / p^* . \qquad (15)$$

In other words, TWK will be dominated by SLK if $p^* < \bar{p}$, that is, if a job with processing time less than the mean has a waiting time of at least W_0. Intuitively, this sufficient condition is more likely to hold when the workload is light than when it is heavy, because there will be fewer available jobs in queue under a light workload that are candidates for preemption. Therefore, it might be anticipated that SLK tends to dominate TWK under light workloads but not under heavy workloads. This is precisely the behavior exhibited in Table 4.

In addition, the likelihood of preemption under TWK is somewhat a function of the variance of the processing times. When the variance is small there is more of a chance that $p^* < \bar{p}$; by (15) this condition means SLK will dominate TWK. (However, if the variance is sufficiently small then the rules become indistinguishable.) Because the exponential distribution might be considered a distribution with a high variance, the results in Table 4 may not be viewed as representative. In order to further explore the variance effect we repeated the experiments with processing times drawn from a normal distribution with mean 100 and standard deviation 25. [In this case, Equation (13) must be replaced with the formula corresponding to the queue $M/G/1$, and Equation (14) must be modified appropriately.] Again, we found that SLK dominated TWK, except under heavy workloads. In this set of test problems the "crossover" workload at which TWK becomes better is heavier than in the case of the exponential samples. We also noticed in the corresponding random workload problems that CON was sometimes the best rule, although TWK was still most frequently best. More significantly, perhaps, the results in Tables 5 and 6 indicate that when the variance of processing times is small there is relatively little difference among the rules TWK, SLK, and CON.

Table 5: Results for dynamic problems with normal samples and random workload.

Utilization	Mean No. of Jobs	Mean Flow Allowance TWK	SLK	CON	OPT	Frequency Best TWK	SLK	CON
0.90	4.28	7.20	7.70	7.72	3.08	16	2	2
0.95	9.59	10.06	10.70	10.75	4.24	16	2	2
0.99	52.09	10.44	10.99	11.07	4.56	20	0	0

Table 6: Results for dynamic problems with normal samples and controlled workload.

Average workload (No. of Jobs)	Mean Flow Allowance TWK	SLK	CON	OPT	Frequency Best TWK	SLK	CON
4.28	6.65	5.31	5.90	4.02	0	20	0
9.59	12.35	10.61	11.18	7.71	0	20	0
52.09	48.53	53.10	53.64	30.31	20	0	0

Summary and Conclusions

A common assumption in scheduling research is that due dates are given. In many production environments, however, it is appropriate to treat due dates as decision variables which are determined within the production control system. In this paper we have utilized a basic scheduling model in order to gain some insight into the problem of selecting due dates. Specifically, the problem was formulated in terms of making due dates as tight as possible, subject to the constraint that all jobs complete on time, in the context of the single machine model with dynamic job arrivals.

We investigated three basic rules for setting due dates: CON (in which flow allowances are constant), SLK (in which waiting time allowances are constant), and TWK (in which flow allowances are proportional to processing times). Of these rules, only CON fails to discriminate among the jobs on the basis of their processing times. This property might intuitively suggest that the CON rule is not as effective as the other rules. In fact, we demonstrated analytically that the CON rule is dominated in the static model and also in the dynamic model with controlled workload. Even in the remaining scenarios (i.e., the dynamic model with random workload) there were few problem instances where CON was desirable. We conclude from this pattern of results that processing time information is relevant in due-date selection, and that a rule for determining the flow allowance of a job should be based (at least in part) upon the job's length.

We discovered that the TWK rule produced tight due dates most of the time but very loose due dates some of the time. As a consequence, TWK frequently exhibited better average performance than SLK, although in unfavorable circumstances TWK could be considerably worse. From this pattern we conclude that the proportional strategy for setting flow allowances under TWK (as opposed to the additive strategy under SLK) may produce relatively good average performance, but it is also susceptible to poor performance under "worst case" conditions. In particular, an unfavorable job mix for the TWK rule appears to be several equivalently short jobs together with a small number of very long jobs. In this situation, the relatively long waiting time allowance that is given to one of the short jobs must also be given (in the same proportion) to each of the long jobs. The pattern of results for the TWK rule (i.e., good "mean" performance but poor "maximum" performance) is analogous to the flow-time performance of the SPT sequencing rule in job shops. This analogy further suggests that the pure TWK rule might have to be modified in practice, in order to provide some measure of protection against the undesirable performance that might occasionally result under the pure rule.

We examined two workload scenarios and found that the comparison between the TWK rule and SLK rule was difference under random workloads than under controlled workloads. We conclude from this result that in complex production control systems it might be desirable to develop a strategy for due-date selection that depends on the strategy for order releasing, since the latter will affect workload behavior.

From the observations given above, certain topics arise as relevant for future research. One topic is the development of a suitable modification of the pure TWK rule to provide the kind of protection discussed above. Another topic is the investigation of the relationship between due-date selection and workload control. Thirdly, the same kind of problem can be studied in the context of a more complicated model. In this connection, we note that the single-machine model is convenient because a zero-tardiness schedule can be found readily. In more complex models, this optimizing module might well have to be replaced by a heuristic scheduling routine; and it might also be convenient to treat the tightness of the due dates as a constraint instead of a criterion.

Appendix 1

- **Indistinguishable jobs** ($p_j = p$ for $j = 1, 2, \ldots, n$).
Under OPT, the due dates (completion times) will be p, $2p, 3p, \ldots, np$,

$$\bar{d}_{\text{OPT}} = \left(p \sum_{k=1}^{n} k\right) \Big/ n = (n+1)p/2:$$

Under each of the rules CON, SLK, and TWK, the due date of each job will be np. Thus

$$\bar{d}_R = np\,.$$

Therefore the tightness ratio $\bar{d}_R/\bar{d}_{\text{OPT}} = 2n/(n+1)$.

- **Dominant job** ($p_j = 1$ for $j = 1, 2, \ldots, n-1$ and $p_n = p > 2$, where $n > 2$).
Under OPT the due dates (completion times) are as follows:

$$d_j = j \quad \text{for } j = 1, 2, \ldots, n-1 \text{ and } d_n = n-1+p\,.$$

Therefore,

$$\bar{d}_{\text{OPT}} = \left[\sum_{k=1}^{n} k + (n-1+p)\right] \Big/ n = [n(n-1)/2 + (n-1+p)]/n$$

$$= (n^2 + n + 2p - 2)/2n\,.$$

Under CON, we obtain $\bar{d}_{\text{CON}} = d_j = n-1+p$. As a consequence,

$$\bar{d}_{\text{CON}}/\bar{d}_{\text{OPT}} = 2n(n-1+p)/(n^2+n+2p-2).$$

As $p \to \infty$ this ratio approaches $2np/2p$. Hence

$$\lim_{p\to\infty} (\bar{d}_{\text{CON}}/\bar{d}_{\text{OPT}}) = n\,.$$

Under SLK, we obtain $\beta = n-1$. Since $\bar{p} = (n-1+p)/n$, it follows that

$$\bar{d}_{\text{SLK}} = (n-1) + (n-1+p)/n = (n^2-1+p)/n\,.$$

As a consequence,

$$\bar{d}_{\text{SLK}}/\bar{d}_{\text{OPT}} = 2(n^2-1+p)/(n^2+n+2p-2)\,.$$

For $p \to \infty$ this ratio approaches $2p/2p$. Hence

$$\lim_{p\to\infty} (\bar{d}_{\text{SLK}}/\bar{d}_{\text{OPT}}) = 1\,.$$

Under TWK, we obtain $\alpha = n-1$. Thus

$$\bar{d}_{\text{TWK}} = (n-1)(n-1+p)/n = (n^2-2n+np+1-p)/n\,.$$

As a consequence,

$$\bar{d}_{TWK}/\bar{d}_{OPT} = 2(n^2 - 2n + np + 1 - p) / (n^2 + n + 2p - 2) .$$

For $p \to \infty$, this ratio approaches $2(n-1)p/2p$. Hence

$$\lim_{p \to \infty} (\bar{d}_{TWK}/\bar{d}_{OPT}) = n - 1 .$$

- **Distinct jobs** ($p_j = j$ for $j = 1, 2, \ldots, n$)

Under OPT, the due dates (completion times) will be given by

$$d_j = 1 + 2 + \ldots + j = j(j+1)/2 .$$

Therefore

$$\bar{d}_{OPT} = \left[\sum_{k=1}^{n} k(k+1)/2 \right] \Big/ n = (n+1)(n+2)/6 .$$

Under CON, we obtain $\bar{d}_{CON} = \gamma = n(n+1)/2$. As a consequence,

$$\bar{d}_{CON}/\bar{d}_{OPT} = 6n(n+1)/2(n+1)(n+2) = 3n/(n+2) ,$$

$$\lim_{n \to \infty} (\bar{d}_{CON}/\bar{d}_{OPT}) = 3 .$$

Under SLK, we obtain $\beta = 1 + 2 + \ldots + (n-1) = n(n-1)/2$. Since $\bar{p} = (n+1)/2$, it follows that

$$\bar{d}_{SLK} = [n(n-1) + (n+1)]/2 = (n^2 + 1)/2 .$$

As a consequence,

$$\bar{d}_{SLK}/\bar{d}_{OPT} = 6(n^2+1)/2(n+1)(n+2)$$

$$= 3(n^2+1)/(n+1)(n+2) ,$$

$$\lim_{n \to \infty} (\bar{d}_{SLK}/\bar{d}_{OPT}) = 3 .$$

Under TWK, we obtain $\alpha = C_n/p_n = n(n+1)/2n = (n+1)/2$. Thus

$$\bar{d}_{TWK} = [(n+1)/2]^2 = (n+1)^2/4 .$$

As a consequence,

$$\bar{d}_{TWK}/\bar{d}_{OPT} = 6(n+1)^2/4(n+1)(n+2) = 1.5(n+1)/(n+2),$$

$$\lim_{n \to \infty} (\bar{d}_{TWK}/\bar{d}_{OPT}) = 1.5 .$$

Appendix 2

- **Proof of Property 1:** $\bar{p} + \sum_{j=1}^{n-1} p_j \leqslant \sum_{j=1}^{n} p_j$.

The inequality is obviously valid if and only if

$$\bar{p} \leqslant p_n .$$

Since SPT job ordering applies, p_n is the longest processing time, and this inequality will always hold. □

- **Proof of Property 2:** $\bar{p} \operatorname{Max}_j \left(\sum_{i=1}^{j} p_i/p_j \right) \leqslant \sum_{j=1}^{n} p_j$.

Since SPT applies, $\Sigma_{i=1}^{j} p_i \leqslant j p_j$. Thus the property holds if

$$\bar{p} \operatorname{Max}_j (j p_j/p_j) \leqslant \sum_{j=1}^{n} p_j ,$$

$$\bar{p}(n) \leqslant \sum_{j=1}^{n} p_j .$$

This is clearly an equality. □

References

[1] Baker, K. R., Lawler, E. L., Lenstra, J. K., and Rinnooy Kan, A.H.G., "Preemptive Scheduling of a Single Machine to Minimize Maximum Cost Subject to Release Dates and Precedence Constraints," Report BW 128, Mathemisch Centrum, Amsterdam (August (1980).

[2] Eilon, S., and Chowdhury, I. G., "Due Dates in Job Shop Scheduling," *International Journal of Production Research* **14**, 2, 223-237 (1976).

[3] Elvers, D. A., "Job Shop Dispatching Using Various Due-Date Setting Criteria," *Production and Inventory Management* **14**, 4, 62-69 (December 1973).

[4] Horn, W., "Some Simple Scheduling Algorithms," *Naval Research Logistics Quarterly* **21**, 4, 177-185 (March 1974).

[5] Weeks, J. K., "A Simulation Study of Predictable Due-Dates," *Management Science* **25**, 4, 363-373 (April 1979).

Kenneth R. Baker is Professor of Operations Management at the Amos Tuck School of Business Administration at Dartmouth College. Professor Baker is the author of the textbook *Introduction to Sequencing and Scheduling* and has written several articles on various aspects of scheduling. During the academic year 1979-80 he was a visiting faculty member at Eindhoven University of Technology, in the Netherlands.

Will Bertrand is a faculty member in the Department of Industrial Engineering at Eindhoven University of Technology in the Netherlands. He has served as a member of the department's staff since receiving his Engineer degree in 1970 and has worked during this time on various aspects of production and inventory control. Over the last four years he has led a design project involving the creation and implementation of a production control system for a semiconductor manufacturer. At present he is completing the requirements for his doctoral degree, writing a thesis based on his production control research.