OPTIMIZING SCHEDULED TIMES, ALLOWING FOR BEHAVIOURAL RESPONSE

MALACHY CAREY*
Faculty of Business and Management, University of Ulster, Newtownabbey, U.K., BT37 0QB

(Received 30 September 1993; in revised form 12 August 1997)

Abstract—We consider transport activities for which time has to be allocated or scheduled in advance. When the schedule is implemented, the time actually taken by each activity is subject to random variation, hence can exceed the scheduled time. To reduce such over-runs or lateness, and improve reliability and costs, some extra time is usually allowed for some, or all, activities in the schedule. However, it is well known that if more time is allocated for an activity then the activity often tends to take longer. Because of this behavioural response, some or all of the benefits (in reliability, costs, etc.) of the extra time allowance are lost. To compensate for this, should even more time be allowed for each activity, or should less be allowed, and if so, how much? We consider this question here and in particular we discuss the effect of such behavioural response on expected costs and on the optimal time to allow for an activity. We find that the optimal time to allow depends, in very simple ways, on a behavioural response ratio, and on the ratio of scheduled time costs to lateness costs. The model is applicable to computing optimal times for public transport timetables, for buses, trains or airlines. It is also relevant to choosing how much time to allow for each of a set of operations in production scheduling or service scheduling.

© 1998 Elsevier Science Ltd. All rights reserved

Keywords: Scheduled transport, activity durations, behavioural response, reliability

1. INTRODUCTION

Public transport systems for buses, trains, rapid transit, or airlines are often operated to a pre-specified published schedule or timetable. Operators try to maintain punctuality of arrivals and departures at each stop, but this is often difficult since there is usually an inherent variability in the time taken to traverse each link, due to congestion, passenger boarding and alighting, maintenance, safety checks, failures, etc. One strategy to improve reliability is to allow more time in the timetable for each activity, e.g. travelling from A to B, B to C, etc. This is sometimes referred to as putting ‘slack’ or ‘recovery time’ in the schedule, to allow activities to get back on schedule if they happen to be running late. Though this improves punctuality for each activity it increases the average time taken to complete a sequence of activities, e.g. travelling from A to D. Thus, increasing the scheduled times reduces punctuality costs but increases other travel costs, and vice versa.

This trade-off is well known and discussed in the literature (e.g. Jenkins, 1976; Newell, 1977; Powell and Sheffi, 1983; Hall, 1985; Abkowitz et al., 1986; Bookbinder and Ahlin, 1990; Bookbinder and Desilets, 1992; Carey, 1992, 1994). These authors have explored the effects of variability on punctuality, wait times and travel costs. They have considered designing more reliable schedules or control strategies (e.g. bus holding strategies) in order to improve punctuality or expected travel costs.

However, inserting more time in the schedule introduces an other trade-off, which appears to have been ignored in the literature but is well known and important to transport operators and managers, and indeed to managers in other industries. It is widely observed that if more time is allowed for an activity, then the activity itself often takes longer to complete. For example, suppose bus trips from A to B are time tabled to take no more than 60 min, but in practice 10% of buses take more than 60 min, though none take more than 70 min. To eliminate lateness suppose we increase the scheduled trip time to, say, 70 min. However, we may now find that 15% of buses take more than 60 min, and 5% take more than 70 min, so that 5% of busses are still late. We will

*Author for correspondence. Tel: 01232 366352/366638; Fax: 01232 366868; e-mail: m.carey@ulst.ac.uk
refer to this phenomenon as ‘behavioural response’, since it seems due to the how operators, crews, despatchers, passengers, or ‘the system’, respond to having more time available. For example, if more time is available operators may simply take longer or start later.

Some transport managers respond to this by imposing a ‘tight’ schedule to avoid what they perceive as unnecessary time waste. Others respond by putting even more slack in the schedule, and in many transport organizations there is an on-going debate as to which approach to take. However, there does not appear to have been a formal analysis of the problem, or of whether or when to take either of these approaches, or what is the cost trade-off, or how much slack to allow when some of this will be used up by behavioural adjustment. We therefore consider these questions more formally here.

A qualification is in order here. The approach proposed in this paper provides a useful framework for exploring the problem, but is not intended as a complete solution to the problem of managing slack time or recovery time. For example, the trip time distributions which are taken as given here are not in practice unchanging physical phenomena, but can themselves be changed by actions of managers and operators. Indeed, one of the functions of good management is to try to control or manage (e.g. shorten or narrow) the distributions of activity time durations.

Consider an activity or task which has to be allocated a planned or scheduled amount of time $T$ in advance. The cost of this scheduled time is $c_s(T)$. However, the task actually takes a time $t$ to complete, where $t$ is a random variable, with probability density function $f(t)$ and distribution function $F(t)$ (Fig. 1). If the task takes longer than scheduled (i.e. $t > T$) then a cost of lateness $c_l(t - T)$ is incurred, and the expected or average cost of lateness is $E[c_l(t - T)]$. The problem considered here is that of choosing a scheduled time $T$ so as to make an optimal trade-off between the cost of scheduled time $c_s(T)$ and the expected cost of lateness. The problem is complicated by the fact that in practice the time $t$ actually taken by an activity is often affected by the time scheduled for the activity.

The costs in the model represent costs to passengers and/or vehicle operators. For passengers riding on a vehicle the trip time consists of (scheduled trip time) plus (lateness, if any) minus (earliness, if any). Passengers usually place quite different values or costs on these three types of time, and the trip cost is the sum of the three costs. Lateness has a cost to passengers as it causes them to missappointments, connections, work time, etc. It might be thought that arriving early would a benefit to passengers rather than a cost. However, that is not necessarily true. For example, passenger surveys in British Rail (in the late 1980s) found that, on average, passengers considered arriving early a cost rather than a benefit. This is presumably because on average they were then too early for their next appointment, connection, work, etc. In any case, passengers are much more concerned about lateness than earliness. The costs or benefits of earliness appear to much less than the costs of lateness.

For vehicle operators, costs can again be divided into (cost of scheduled trip time) plus (cost of lateness, if any) plus (cost of earliness, if any). The cost of scheduled trip time is the cost of running the vehicle, cost of driver time, etc. If the vehicle arrives late this can cause additional knock-on delays and costs, e.g. the vehicle or driver may be late for their next assignment, special arrangements may have to be made for passenger handling, passenger connections, etc. If the vehicle arrives earlier than scheduled this is usually not a benefit to operators, since the vehicle and driver

![Fig. 1. Illustrative pdf and probability of lateness.](image-url)
will usually be idle until their next scheduled activity. On the other hand, arriving early usually does not impose a significant cost on operators, and is usually easier to cure than lateness. For example, if a vehicle is about to arrive early it could simply slow down so as arrive on time.

In view of the above, we will concentrate on costs of scheduled trip time and costs of lateness, and will omit costs or benefits of earliness. We could include these in a similar way to costs lateness, and indeed at the end of Section 2 we indicate how this can be done. However, as it is less important, simplifies the presentation, and saves space we will omit it here.

In this paper we let $c_t(T)$ denote the cost of the scheduled time $T$ for a trip, and let $c_l(l)$ denote the cost of arriving $l$ min (or h, etc.) later than scheduled. Note that $c_l(l)$ includes both a travel time costs and other costs specifically due to lateness. We could perhaps make this more explicit by instead letting $c_l(l) = c_t(t) + c_l(l)$ denote the cost of lateness, where $c_t(t)$ is a pure travel cost and $c_l(l)$ is a pure lateness cost. This means replacing $c_l(l)$ with $c_t(t) + c_l(l)$ in the various results below, but this does not appear to simplify any results. Also, the idea that the cost of lateness is the cost of pure travel time plus a pure lateness cost may be too simple. Scheduled travel time may be seen as coming out of leisure time, or at least non-work time, where as unscheduled lateness may be seen as coming out of work time. Empirical studies show that travellers put quite different values on these different types of time, work time, non-work time and leisure time. The same is true for vehicle operators. That is, the costs associated with lateness may be quite different from the costs associated with scheduled time.

We consider the basic problem in Section 2, with constant cost per unit of time. In Section 3 we let the actual time $t$ be affected by behavioural response to the scheduled time $T$, and in Section 4 we use this to extend the results from Section 2. In Section 5 we extend the results to nonlinear costs $c_t(T)$ and $c_l(l)$. In Section 6 we consider varying the extent of the behavioural response to the scheduled time $T$. In Section 7 we introduce sequences of activities—multistage trips. For example, train, bus or airline services which involve multiple scheduled stops. Section 8 is a brief summary.

### 2. COSTS, OPTIMAL TIME AND SLACK, WITH NO BEHAVIOURAL RESPONSE

Let the cost per unit of scheduled time be $c_t$ and cost per unit of lateness $l = (t - T)$ be $c_l$. Then the total cost of scheduled trip time is $c_t T$ and the expected cost of lateness, given a scheduled trip time $T$, is

$$
C_l(T) = E[(t - T)c_l|t \geq T] = \int_T^{+\infty} (t - T)c_l f(t) dt
$$

*Proposition 1 (Fig. 2):*

(a) The expected cost of lateness (1) is convex and decreasing in $T$. The expected total cost is $C(T) = C_l(T) + C_t(T)$ and is convex in $T$, which implies that any minimum of $C(T)$ is a global minimum.

![Fig. 2. Expected costs as a function of scheduled service time $T$.](image)
(b) If \( c_t > c_i \), then minimum of the expected total cost \( C(T) \) is characterised by:

\[
\text{Optimal probability of lateness } = [1 - F(T^o)] = c_t/c_i
\]

hence, optimal probability of on-time or early \( = F(T^o) = 1 - c_i/c_i \) (2a)

and, optimal scheduled time \( = T^o = F^{-1}(1 - c_i/c_i) \) (2b)

where \( T^o \) denotes the optimal value of \( T \).

(c) If \( c_i \leq c_t \), then minimum of the expected total cost \( C(T) \) is characterised by \( T^o = T_1 \).

Remark: Having obtained the optimal trip time \( T^o \), it is easy to compute the optimal slack or ‘recovery’ time which is included in the optimal scheduled time. However, how operators define the slack time depends on what they consider to be the ‘basic’ trip time, say ‘recovery’ time which is included in the optimal scheduled time. However, how operators define the latter as the absolute minimum trip time, or the 10\% percentile of the trip time distribution, or the mean or median trip time, or in some other way. In any case, the optimal slack is then \( T^o - T_b \). This remark applies to the optimal slack for all cases considered in this paper, not just the present proposition.

We can assume that \( c_t > c_i \), so that normally case (b) rather than (c) holds.

Proof:

(a) \( \frac{dC_i(T)}{dT} = -[1 - F(T)]c_i \leq 0 \) and \( d^2C_i/dT^2 = c_if(T) \geq 0 \), hence \( C_i(T) \) is convex and decreasing in \( T \). Also, \( C(t) = c_i + C_i(T) \) hence,

\[
\frac{dC(T)}{dT} = \frac{dC_i(T)}{dT} + c_i = c_i - [1 - F(T)]c_i
\]

(3)

It follows that \( d^2c(T)/dT^2 = d^2C_i(T)/dT^2 \geq 0 \) hence \( C(T) \) is convex.

(b) Since \( C(T) \) is convex its minimum occurs when \( dC(T)/dT = 0 \), or if \( dC(T)/dT \) is everywhere > 0 its minimum occurs at the minimum feasible \( T \). Setting \( dC(T)/dT = 0 \) in (3) yields (2). Since \( F(T) \) is a probability it is positive and \( \leq 1 \), hence (2) has a solution if and only if \( c_i \geq c_t \). If \( c_i < c_t \) then \( dC(T)/dT > 0 \) for all \( T \) hence the minimum of \( C(T) \) is at the minimum value of \( T \), i.e., at \( T = T_1 \).

Proposition 1(b) indicates that the optimal probability of lateness is \( [1 - F(T^o)] = (c_i/c_i) \), hence \( T^o \) is the \( \hat{c} \)th quantile of \( F(T) \), where \( \hat{c} = (1 - c_i/c_i) \). For example (see Fig. 1):

(a) If \( c_t = 2c_i \), then \( F(T^o) = 0.5 \), and the optimal scheduled time \( T^o \) is the median of the arrival distribution, i.e., 50\% of arrivals are later than scheduled.

(b) If there is no extra cost associated with lateness, i.e., \( c_t = c_i \) then \( F(T^o) = 0 \) and \( T^o = \min(t) = T_1 \), i.e., if there is no extra cost associated with lateness, set the scheduled time to the earliest possible arrival time so that all arrivals are exactly on time or late—none is early.

(c) If \( c_t \rightarrow 0 \) or \( c_i \rightarrow +\infty \) (i.e. \( c_i/c_i \rightarrow 0 \)), then \( F(T^o) \rightarrow 1 \) and \( T^o \rightarrow \max(t) \), i.e., if the cost per unit of lateness is relatively very large, set the schedule time to the latest possible arrival time so that all arrivals are exactly on time or early—none is late.

An example

Based on detailed sample surveys of passengers, the British Rail Passenger Forecasting Handbook (1986) suggests a figure of \( c_i = 2.5c_t \), where \( c_t \) is the value or disutility which passengers place on each minute of scheduled travel time and \( c_i \) is the value which they place on each minute of arrival lateness. This implies, from eqn (2), that \( F(T^o) = 0.6 \) where \( F(T) \) is the arrival time distribution, hence at an optimum 60\% of trains would arrive on time and 40\% would arrive late.

In practice, punctuality targets are usually stated not just as the percentage of trains on-time, but also as the percentage of arrivals, departures, etc., less than 5, 10, etc., min late. To illustrate computing these from Proposition 1, consider an arrival time probability distribution from Carey and Carville (1996). This is a beta distribution which is typical of arrival distributions in transportation, and in particular is typical of distributions for some busy British Rail stations. The beta distribution has a pdf \( f(x) \) defined on the interval \([0,1]\) by,
where $\alpha, \beta > 0$ and $B(\alpha, \beta)$ is the beta function, hence the name of the distribution. The $\beta$ pdf can be rescaled and translated to be defined on the interval $[T^{\min}, T^{\max}]$. The pdf (4) then becomes,

$$f(x) = \frac{(x - T^{\min})^{\alpha-1}(T^{\max} - x)^{\beta-1}}{(T^{\max} - T^{\min})^{\alpha+\beta-1} B(\alpha, \beta)}$$

(5)

We used shape parameters $\alpha = 2$, $\beta = 4$, minimum time $T^{\min}$ and maximum time $T^{\max} = T^{\min} + 20$ min.

If we set the scheduled arrival time equal to the earliest possible time $T^{\min}$ then from the above $\beta$ pdf we compute that 63.3% of trains will be more than 5 min late, 18.75% more than 10 min late, and 1.56% more than 15 min late.

On the other hand, if we assume $c_i/c_l = 2.5$ then, from Proposition 1(b), $F(T^o) = 1 - 0.4 = 0.6$, hence from the above $\beta$ distribution the optimal scheduled arrival time $T^o$ is 7.30 min after $T^{\min}$. Trains arriving $(7.30 + 5) = 12.30$ min after $T^{\min}$ are 5 min late and, from the above $\beta$ pdf, we find only 7.60% of trains will be more than 5 min late. Similarly, only 0.15% will be more than 10 min late.

The cost ratio ($c_i/c_l$) used in Proposition 1 may be different for different trains or times of day. In view of this we set out a table of results computed from Proposition 1(b) for various values of the cost ratio ($c_i/c_l$), assuming the above beta distribution of arrival times. Thus:

<table>
<thead>
<tr>
<th>Ratio ($c_i/c_l$)</th>
<th>1</th>
<th>0.8</th>
<th>0.6</th>
<th>0.4</th>
<th>0.2</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal probability of arriving on time, $F(T^o)$</td>
<td>0</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>Optimal probability of arriving late, $(1 - F(T^o))$</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>(Optimal scheduled time, $T^o$) - $T^{\min}$</td>
<td>0</td>
<td>3.37</td>
<td>5.31</td>
<td>7.30</td>
<td>9.80</td>
<td>20</td>
</tr>
<tr>
<td>% of trains $\geq$ 5 mins late (later than $T^o$)</td>
<td>63.3</td>
<td>30.56</td>
<td>16.27</td>
<td>7.60</td>
<td>1.80</td>
<td>0</td>
</tr>
<tr>
<td>% of trains $\geq$ 10 mins late (later than $T^o$)</td>
<td>18.75</td>
<td>4.43</td>
<td>1.23</td>
<td>0.15</td>
<td>0.0</td>
<td>0</td>
</tr>
</tbody>
</table>

In the above example we have ignored several factors which affect the costs $c_l$ and $c_i$, or components of $c_l$ and $c_i$, and hence may be important in estimating the optimal scheduled time, etc., from Proposition 1. For example:

(a) The quoted cost ratio $c_i/c_l = 2.5$ is based only on time costs of passengers. This is likely to reflect willingness to pay and hence long run revenues to operators. However, it ignores other costs to operators, e.g. lateness reduces utilization of resources, makes it harder to operate the service, causes delay and cancellation of other services, etc.

(b) Costs of lateness differ according to class of traveller, time of day, whether travelling to work, from work or on leisure trips, and in practice we need an appropriate weighted combination of these.

(c) In practice only a fraction of passengers alight at intermediate stops and only these passengers incur the lateness cost $c_i$.

(d) If a service is late it may miss a connecting service, which causes costs to passengers and operators.

In short, the costs or benefits $c_l$ and $c_i$ to be used in the above Proposition and formulae should include all relevant and appropriate costs and benefits, such as those in (a)–(d). We should also be clear as to whether we are optimizing private costs and benefits, to transport operators, or social costs and benefits, which would include costs to operators and passengers and perhaps others.

However, in the present paper we can take the values of $c_l$ and $c_i$ as given. The contribution of this paper is not to the estimation of $c_l$ or $c_i$. Even though that may be an interesting and important matter, to pursue it further here would distract from the contribution of this paper, which is to consider behavioural adjustment, in the following sections. For that we can take $c_l$ and $c_i$ as
given. We introduced the above example and discussion of \( c_l \) and \( c_t \) only to illustrate a context into which we will introduce behavioural adjustment.

**Lateness cost less than travel cost.** In the unlikely event that \( c_l < c_t \), \( dC(T)/dT > 0 \) for all \( T \geq 0 \), so that \( C(T) \) is minimised by setting the scheduled time \( T = 0 \). However, setting the scheduled time \( T \) less than the minimum feasible time \( T_1 \) means that the scheduled time can never actually be achieved. In practice this would lead to a loss in credibility of the scheduled time, and users would realise that the minimum scheduled time is in effect at least \( T_1 \), and \( C(T) \) would be minimised at \( T^* \geq T_1 \). [Also, setting \( T < T_1 \) is often ruled out by regulations, policy or professional standards. Indeed, regulated transport industries are often required to set \( T \) so that there is say a 0.8, or 0.9, etc, probability of \( T \) being achieved (i.e. \( F(T) \geq 0.8 \)). This lower bound on \( T \) may exceed the cost minimising \( T^* \) from eqn (2). Similar comments apply if \( c_l = c_t \), since in this case \( dC(T)/dT = -F(T) \) so that \( C(T) \) would be minimised by choosing any \( T \leq T_1 \).]

**Costs of earliness.** In the discussion above we could introduce expected costs of earliness \( (T - t) \) analogous to the expected costs of lateness \( (t - T) \). Thus let \( c_e \) denote the cost per unit of earliness \( (T - t) \), so that the total cost of earliness is \( c_e(T - t) \) and the expected cost of earliness is

\[
C_e(T) = E[c_e(T - t)] = \int_0^T c_e(T - t)f(t)dt = \int_{T_1}^T c_e(T - t)f(t)dt
\]

Optimality conditions are derived by setting \( C'(T) = 0 = C'_e(T) + C'_l(T) + C'_c(T) \). However, we do not pursue this further for reasons already set out towards the end of the previous section.

### 3. BEHAVIOURAL RESPONSE

In the above section we let the scheduled time \( T \) exceed the minimum possible time \( T_1 \). The difference \( (T - T_1) \) is often referred to as the ‘slack’ or ‘time buffer’ or ‘performance allowance’ in a schedule or timetable. Increasing this allowance increases the probability that the activity will be completed by its scheduled time. If the activity is one of a sequence of activities, each of which has to be completed before the next one can begin, then \( (T - T_1) \) is sometimes referred to as ‘recovery time’. This is because it allows time to ‘recover’ from lateness and get back on (or nearer) schedule before beginning the next activity or task.

However, it is well known in transportation that allowing more slack time or recovery time can result in the activities all taking longer. For example, suppose a train has a scheduled dwell time of, say, 2 min at a station, but is often not ready to depart until up to 5 min of dwell time, i.e. 3 min late. In view of this, suppose the scheduled dwell time is changed to 5 min. It might seem this would ensure the train is always ready to depart on time. However, it may also make the various people involved with the train more relaxed about the task. If the driver, dispatcher, engineers, and passengers all think they have 5 min, the train may not be ready as quickly as when they thought they had only 2 min. In practice it is often found that the train continues to depart late, though not late as often as when only 3 min were allowed. Adding 3 min slack has reduced the problem of lateness, but note eliminated it since some of the slack has been used up by the response of the operators, etc.

More generally, it often happens that if more time \( T \) is allowed then operators, users, etc., take advantage of this by taking, on average, more time for the task. There are various reasons for this, some technical, some behavioural. Perceptions as to how much time an activity should take or when to start on sub-tasks within it may depend on how much time is allowed or how much time remains. Ignoring this phenomenon could make the optimal scheduled time derived in Section 2 seriously misleading.

To formalize this phenomenon, suppose that allowing more time \( T \) simply shifts the pdf of \( t \) to the right, without otherwise changing the shape of the pdf (Fig. 3). The position of the pdf can be defined by \( T_1 = \min(t) \), but \( T_1 \) now depends on the scheduled time \( T \). Let \( T'_1 \) denote the smallest possible value of \( T_1 \), which occurs if we allow no slack in the scheduled time \( T \). If we set the scheduled time \( T \) so that \( T > T'_1 \) then we can assume that a fraction \( \theta \) of this slack \( (T - T'_1) \) is used up or ‘wasted’ by operators, etc., so that the minimum trip time increases by an amount \( \theta(T - T'_1) \). In other words, the minimum trip time becomes
The probability density of \( t \) is now \( f(t - \theta(T - T_{m}^n)) \) rather than \( f(t) \).

If the scheduled time is set lower than the minimum feasible time \( T_{m}^m \), the ‘slack’ \( (T - T_{m}^n) \) is negative, so there is no slack for operators, etc., to use up or ‘waste’. In that case we assume there is no behavioural response \( \theta \), as there is no slack to respond to, hence \( T_1 = T_{m}^m \).

Instead of the behavioural response beginning when \( T \) exceeds \( T_{m}^m \), as assumed above, it may begin only when \( T \) exceeds say the 5th or 10th percentile of \( f(t) \), or perhaps only when it exceeds say the mean, median or mode of \( f(t) \). It is a simple matter to adapt the above model to reflect this, by simply redefining \( T_{m}^m \) as the 5th or 10th percentile, etc., of \( f(t) \). This slightly alters the results in Sections 2–6 below, but does not significantly alter the nature of the results. A further advantage of redefining \( T_{m}^m \) in this way is that it may make it much easier to obtain accurate estimates of \( T_{m}^m \) than if \( T_{m}^m \) is defined as \( \min(t) \). This is especially so if \( f(t) \) has a relatively long left hand tail. When collecting data, relatively few observations will be found in the tail, so that it will be difficult to establish the absolute minimum of \( t \). On the other hand, estimating \( \min(t) \) may be no more difficult than estimating a percentile of \( t \) if \( f(t) \) is, say, a negative exponential or gamma distribution.

To illustrate extreme cases in the discussion below it is often convenient to assume that the pdfs \( f(t) \) start and end at finite points on the \( x \)-axis, i.e. at \( T_1 = \min(t) \) and \( T_2 = \max(t) \). This will be true for any empirically observed distribution of activity times, and reflects the fact that in practice activities take at least a non negative amount of time and cannot take an infinite amount of time—activities will eventually be aborted or terminated if they take ‘too’ long. This finite beginning and end is also true for say beta distributions, or truncated forms of any other theoretical distributions. If as an approximation we use say a normal or negative exponential distribution, in which the maximum value of \( t \) is infinite, this does not affect the analysis, since the end point of the pdf does not enter our analysis, and is shown as \( T_2 \) in Figs 1 and 3 only for illustrative purposes.

There are, of course, other ways to model behavioural response, in particular by allowing the shape of the pdf to change, as well as shifting it to the right. However, the model considered here has the advantage of being perhaps the simplest and being intuitively appealing. There is also some evidence to support it for the case of scheduled public transport.

4. COSTS, OPTIMAL TIME AND SLACK, WITH BEHAVIOURAL RESPONSE

Now consider the expected costs, and the optimal scheduled time \( T \), in the presence of the above behavioural response, eqn (6). We can assume that the scheduled time is not earlier than the minimum feasible time, i.e. \( T \geq T_{m}^m \). (We consider \( T \leq T_{m}^m \) in the final paragraph of this section).

The cost of scheduled trip time \( T \) is still \( C_t = c_1 T \), but the expected cost of lateness is now

\[
C_l(T) = E[c_l(t - T \mid t \geq T)] = \int_t^{+\infty} (t - T)c_lf(t - \theta(T - T_{m}^n))dt
\]  

(7*)

To keep the derivative of this with respect to \( T \) simple, it is useful to first rewrite this as follows. Let \( \tilde{t} = t - \theta(T - T_{m}^n) \). Using this to substitute for \( t \) in eqn (7*) reduces \( (t - T)c_lf(t - \theta(T - T_{m}^n)) \)
It follows that $dC_{\hat{t}}$ extends Proposition 1(a) and Proposition 3 extends Proposition 1(b) and (c).

The expected total cost is

$$t_{\hat{t}} = \min.$$  \hfill (7)

Fig. 4. Cost functions for low or high values of $t_{\hat{t}}$ to (a). Since $t_{\hat{t}}(T - T_{m}^{m})$, in response to inserting the slack $(T - T_{m}^{m})$ in the schedule. The optimality condition (8) is illustrated in Fig. 5, and can be interpreted in much the same way as already discussed for eqn (2) in Section 2.

(b) If $\theta > \tilde{c}$ then $T_{o} = T_{1} = T_{1}^{m}$, i.e. the optimal scheduled time is set to the earliest feasible time, so that all arrivals are late, and there is no slack in the schedule.

(c) If $\theta = \tilde{c}$ then both results (b) and eqn (8) hold.

Proof:

(a) If $\theta < \tilde{c}$ then the optimal scheduled time is $T_{o} > T_{1} > T_{1}^{m}$ and $T_{o}$ can be computed from,

$$[1 - F(T_{o} - \theta(T - T_{m}^{m}))] = c_{l}/c_{l}(1 - \theta)$$

or

$$[1 - F^{v}(T_{o})] = c_{l}/c_{l}(1 - \theta)$$

or

$$T_{o} = F^{v-1}[1 - c_{l}/c_{l}(1 - \theta)]$$

(b) If $\theta > \tilde{c}$ then $T_{o} = T_{1} = T_{1}^{m}$, i.e. the optimal scheduled time is set to the earliest feasible time, so that all arrivals are late, and there is no slack in the schedule.

(c) If $\theta = \tilde{c}$ then both results (b) and eqn (8) hold.

Proof:

(a) See Fig. 4(a). Since $C(T)$ is convex it has a minimum if $dC(T)/dT = 0$. From (7), $dC(T)/dT = 0$ implies,

$$C = C_{t} + C_{l}$$

Costs

$T_{1}$ $T$

$C_{t}$ $C_{l}$

$C_{l}$ $C_{t}$

If $c_{l} > c_{t}$
If $c_{l} = c_{t}$
If $c_{l} < c_{t}$

Fig. 4. Cost functions for low or high values of $\theta$, relative to $\tilde{c} = (c_{l} - c_{l})/c_{l}$. (a) Low $\theta$ (i.e. $\theta < \tilde{c}$), implies $T_{o} > T_{1} > T_{1}^{m}$; (b) high $\theta$ (i.e. $\theta > \tilde{c}$), implies $T_{o} = T_{1} = T_{1}^{m}$.
which immediately gives eqn (8) in the proposition. However, we must check whether or when this has a solution. The term \([1 - F(T - \theta(T - T_1^m))]\) is a probability, hence is positive and \(\leq 1\). It follows that eqn (8'), and hence eqn (8), has a solution only if \(c_i/(1 - \theta)c_l < 1\). To see that eqn (8') always has a solution if \(c_i/(1 - \theta)c_l < 1\), note that by varying \(T\) from \(T_1^m\) to \(+\infty\) we can always find a \(T\) to make \(F(T - \theta(T - T_1^m))\) take any desired value between 0 and 1. [If \(T = T_1\) then \(F(T - \theta(T - T_1^m))\) reduces to \(F(T_1^m)\) which by definition = 0. And as \(T\) goes to \(+\infty\), \(F(T - \theta(T - T_1^m))\) goes to \(F(+\infty) = 1\).] In summary, the optimality condition (8) has a solution if and only if \(c_i/(1 - \theta)c_l > 1\), but rearranging this gives \(\theta < \hat{\theta}\) where \(\hat{\theta} = 1 - c_i/c_l\), as in (a) of the proposition.

(b) See Fig. 4(b). In eqn (7') the \([1 - F(T - \theta(T - T_1^m))]\) term is a probability hence is always positive and \(\leq 1\). It follows that \(dC(T)/dT > 0\) if \(c_i > (1 - \theta)c_l\). This holds for all values of \(T\). But \(C(T)\) is convex, hence if \(dC(T)/dT > 0\) for all \(T\) then \(C(T)\) has a unique global minimum at the minimum feasible value of \(T\), which is defined as \(T_1^m\). In summary, \(C(T)\) has a minimum at \(T = T_1^m\) if \(c_i > (1 - \theta)c_l\). Rearranging the latter gives \(\theta < \hat{\theta}\) where \(\hat{\theta} = 1 - c_i/c_l\), as stated in (b) of the proposition.

(c) It can easily be shown that if \(\theta = \hat{\theta}\) then both results (a) and (b) hold.

Equation (8) of the proposition has several interesting and intuitively very reasonable implications. For example, if the cost per unit of lateness \((c_l)\) is very high relative to the cost per unit of scheduled time \((c_i)\), then intuitively we expect the optimal scheduled time \(T\) to be chosen to ensure that the probability of being late is very low \((= 0)\). That is, we expect a large scheduled time and large slack. To see that this does occur in eqn (8), note that if \(c_i > c_l(c_i/c_l) \to 0\) hence, from eqn (8), \([1 - F^\theta(T^o)] \to 0\) and \(F^\theta(T^o) \to 1\). That is, the optimal probability of being on time tends to 1. Or equivalently, the optimal probability of lateness (the area cut off in the tail of the pdf in Fig. 5) tends to zero.

An example
To further illustrate this proposition, we continue the example from Section 2. We use the ratio \((c_i/c_l) = 2.5\) quoted there. This gives \(\hat{\theta} = 1 - c_i/c_l = 0.6\), which implies that it is optimal to put no slack in the schedule if \(\theta \geq 0.6\), i.e. if 0.6 or more of the slack would be ‘wasted’ as described earlier. Note that putting no slack in the schedule is much more likely to occur here than in Section 2. There it was implicitly assumed that \(\theta = 1\) so that it was optimal to put no slack in the schedule only if \(c_i \leq c_l\).

From eqn (8), the optimal probability of arriving on time is \(F^\theta(T^o) = 1 - c_i/c_l(1 - \theta)\). Again assume \((c_i/c_l) = 2.5\) hence \((c_i/c_l) = 0.4\). Then, \(F^\theta(T^o) = 1 - 0.4/(1 - \theta)\), i.e. \(F^\theta(T^o)\) is a downward sloping function of \(\theta\), between \(0 \leq \theta \leq 0.6\). For example:

<table>
<thead>
<tr>
<th>Fraction of slack time ‘wasted’, (\theta)</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>(\geq 0.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal probability of arriving on time, (F^\theta(T^o))</td>
<td>0.6</td>
<td>0.55</td>
<td>0.5</td>
<td>0.429</td>
<td>0.33</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

So far in this section we have assumed that the scheduled time \(T\) is at least the minimum feasible time, i.e. \(T \geq T_1^m\). Now consider the setting \(T\) less than the minimum feasible time \(T_1^m\). In this case
the ‘slack’ in the schedule \((T - T_m^\dagger)\) is negative, so there is no slack for operators, etc., to respond to or waste. There is no behavioural response \(\theta\), hence \(T_1 = T_m^\dagger\), and the cost model reduces to that in Section 2 for \(T \leq T_1\). There is no point in setting the scheduled time \(T\) less than the minimum feasible time \(T_m^\dagger\). To do so would only increase expected lateness, and the costs of lateness, without bringing any benefits.

5. NONLINEAR COSTS OF SCHEDULED TIME AND LATENESS

In Sections 2 and 4 we assumed that the cost \(c_i\) per unit of scheduled trip time \(T\) and cost \(c_l\) per unit of lateness is constant, independent of \(T\). However, the cost per min of lateness \(l = (t - T)\), may increase with lateness, and the cost per min of scheduled time \(T\) may increase with the amount of scheduled time. Thus \(c_i T\) and \(c_l l\) become \(c_i(T)\) and \(c_l(l)\). We can assume that these costs \(c_i(T)\) become \(c_l(l)\) are convex and increasing. This ensures that the expected total costs in Sections 2 and 4 remain convex, so that a minimum of \(C(T)\) is still a global minimum.

Replacing \(c_i T\) with \(c_i(T)\) in Section 2 changes the optimality condition (2.3) from \(T^o = F^{-1}(1 - c_i/c_l)\) to
\[
T^o = F^{-1}(1 - c_i'(T^o/c_l)) \tag{9}
\]
where \(c_i'(T^o) = dc_i(t)/dt\) evaluated at \(T^o\). Similarly, introducing \(c_i(T)\) in Section 4 changes the optimality condition (8) from \(T^o = F^{*^{-1}}[1 - c_i/c_l(1 - \theta)]\) to
\[
T^o = F^{*^{-1}}[1 - c_i'(T^o)/c_l(1 - \theta)] \tag{10}
\]
Using \(c_i'(T^o)\) rather than \(c_i\) makes it slightly more difficult to compute \(T^o\).

Replacing \(c_l\) with \(c_l(l)\) in Sections 2 and 4 complicates the optimality conditions more than does replacing \(c_i T\) with \(c_i(T)\). Recall that \(C(T) = C_i(T) + C_l(T)\) and the general optimality condition is \(dC(T)/dT = 0\) hence \(dC(T)/dT = dC_i(T)/dT + dC_l(T)/dT = 0\). When \(C_l(T)\) is from eqn (1) in Section 2 the optimality condition becomes,
\[
\begin{align*}
  c_i'(T^o) &= \int_{T^o}^{+\infty} c_i'(t - T^o) f(t) dt \\
  &= \int_{T^o - \theta(T^o - T_1^\dagger)}^{+\infty} -(1 - \theta)c_i'(t - (T^o - \theta(T^o - T_1^\dagger))) f(t) dt
\end{align*}
\tag{11}
\]
instead of (2a)–(2b). Similarly, when \(C_l(T)\) is given by eqn (7) in Section 4 the optimality condition becomes
\[
\begin{align*}
  c_l'(T^o) &= \int_{T^o - \theta(T^o - T_1^\dagger)}^{+\infty} -(1 - \theta)c_l'(t - (T^o - \theta(T^o - T_1^\dagger))) f(t) dt \\
  &= \int_{T^o - \theta(T^o - T_1^\dagger)}^{+\infty} -(1 - \theta)\frac{dC_l(t)}{dT} f(t) dt
\end{align*}
\tag{12}
\]
instead of eqn (8). These optimality conditions (11)–(12) of course reduce again to those in Sections 2 and 4 if we set \(c_i(T) = c_i T\) and \(c_l(l) = c_l l\).

The simplest nonlinear function is perhaps a two-step piecewise linear cost function, as follows. If a vehicle is late by a relatively small amount (say, \(l \leq l_1\)), passengers and operators may consider the cost of lateness the same as the cost of ordinary scheduled travel time (i.e. \(c_l = c_l\) if \(l < l_1\)). That is, they may not perceive any additional cost associated with lateness. For example, this used to be the case for British Rail for many trains late by less than a few minutes. But if a vehicle is late by more than \(l_1 \geq l_1\), the cost per minute of lateness may be greater, \(c_l > c_l\). More generally, a piecewise linear cost function may have any number of steps or segments; e.g. let the cost per min of lateness [the gradient of \(c_i(l)\)] be \(c_l = c_1\) when \(0 \leq l \leq l_1\), \(c_l = c_2\) when \(l_1 \leq l \leq l_2\), ..., \(c_l = c_i\) when \(l_{i-1} \leq l \leq l_i\), ... We can show that in this case the optimality condition (11) reduces to
\[
c_l = \sum_{i=1}^{n} [F(T^o + l_i) - F(T^o + l_{i-1})] c_i \tag{13}
\]
That is, the optimal scheduled time \(T\) is the \(T^o\) which satisfies eqn (13). \([F(T + l_i) - F(T + l_{i-1})]\) is the probability that lateness will be in the time interval \((T + l_i)\) to \((T + l_{i-1})\), and the cost per min
in this time interval is $c_i$, hence the right hand side of eqn (13) is the expected cost per min of lateness. Thus, eqn (13) states that at the optimum scheduled time, the travel cost per min ($c_i$) exactly equals the expected lateness cost per min.

Introducing behavioural response $\theta$ into the above discussion we can show that the optimality condition (13) becomes

$$c_1 = \sum_{i=1}^{n} (F(T^o_\theta + \lfloor \frac{\theta}{\theta} \rfloor) - F(T^o_\theta + \lfloor \frac{\theta}{\theta} \rfloor + 1) c_i$$

(14)

where $T_\theta = (T - \theta(T - T^m))$. That is, if there is behavioural response $\theta$, and a piecewise linear cost of lateness, then the optimal scheduled time $T$ is the $T^o$ which satisfies eqn (14).

The above piecewise linear optimality conditions (13) and (14) are approximations to the general nonlinear optimality conditions (11) and (12). As the number of ‘pieces’ becomes large, the piecewise summation equations tend to the general nonlinear integral equations. A computer programme can easily be written to compute the optimal $T^o$ from the above equations, given any empirical distribution of lateness $F(\cdot)$ and any piecewise linear cost function.

It is intuitive that, other things being equal, increasing the cost of lateness $c_i$ increases the expected cost of lateness, especially for low values of $T$. This in turn increases the optimal scheduled time $T^o$ which will be needed to minimise expected total cost.

6. VARYING THE BEHAVIOURAL RESPONSE FACTOR, $\theta$

From eqn (7),

$$\frac{dC_i}{d\theta} = (T - T^m) c_i |1 - F(T - \theta(T - T^m))| \geq 0 \quad \text{and} \quad \frac{d^2 C(T)}{dT^2} = \frac{d^2 C_i(T)}{d\theta^2} = -(T - T^m) c_i f(T - \theta(T - T^m)) \leq 0,$$

hence $C_i(T)$ and $C(T)$ are increasing and concave in $\theta$. That is, they increase at a decreasing (or non-increasing) rate with $\theta$. Consider two special cases, $\theta = 0$ and $\theta = 1$.

If $\theta = 0$ the model reduces to that in Section 2.

If $\theta = 1$, consider two cases, $T \geq T^m$ and $T \leq T^m$ separately. First, if $T \geq T^m$ and $\theta = 1$ then, from eqn (6), $T = T_1$ hence $dC_i(T)/dT = 0$ and $dC(T)/dT = c_i > 0$, so that $C(T)$ has a minimum at $T = T^m$. On the other hand, if $T \leq T^m$ then, from the last paragraph of Section 4, $dC(T)/dT$ reduces to $(c_i - c_i)$. This is normally negative $(c_i > c_i)$ in which case $C(T)$ again has a minimum at $T = T^m$. Combining the two cases ($T \leq T^m$ or $T \geq T^m$) gives $T^o = T^m$ when $\theta = 1$.

Above we considered the effect of varying the parameter $\theta$, with $\theta$ independent of $T$. However, in practice the behavioural response $\theta$ to slack $(T - T^m)$ in the schedule often depends on the amount of slack. If there is little or no slack in the schedule (i.e. if there is a very tight schedule) then any slack which is inserted is more likely to get used up or absorbed by changes in the behaviour of operators, etc., than if there is a lot of slack. This suggests that the behavioural response $\theta$ decreases as $T$ increases, and we can assume that $\theta$ decreases at a decreasing rate with $T$, so that $\theta = \theta(\text{slack}) = (T - T^m)$ is convex and decreasing in $T$.

Replacing $\theta$ with the function $\theta = \theta(T - T^m)$ in Section 4 changes $dC(T)/dT$ from eqn (7) to $c_i - (1 - \theta'(T - T^m)) c_i [1 - F(T - \theta(T - T^m))]$, hence changes the optimality condition (8) from $|1 - F(T^o)| = c_i / (1 - \theta)$ to $|1 - F(T^o)| = c_i / (1 - \theta(T^o - T^m))$, where $F(T^o)$ is still defined as before. The former is obviously a special case of the latter, since the derivative $\theta'(T^o - T^m)$ reduces to $\theta$ when $\theta'(\cdot)$ is linear. Using $\theta'(T^o - T^m)$ rather than $\theta$ makes it slightly more difficult to compute $T^o$. Note that $\theta'(s)$ is likely to become small as the slack $s$ increases, in which case the above expression for $T^o$ reduces to that in Section 2.

Finally, we consider how the optimal scheduled time $T^o$ is affected by varying $\theta$. At an optimum we have, from eqn (7'),

$$0 = c_i - (1 - \theta) c_i [1 - F(T^o - \theta(T^o - T^m))] = g(T^o, \theta)$$

hence $dg = (\partial g/\partial T^o) dT^o + (\partial g/\partial \theta) d\theta$. But $dg = 0$ hence $dT^o/\partial \theta = -(\partial g/\partial \theta)/(\partial g/\partial T^o)$.

$$\partial g/\partial T^o = + (1 - \theta)^2 c_i f(T^o - \theta(T^o - T^m)) \geq 0$$

and

$$\partial g/\partial \theta = + c_i [1 - F(T^o - \theta(T^o - T^m))] - (1 - \theta)(T^o - T^m) c_i f(T^o - \theta(T^o - T^m))$$
\( \partial g / \partial \theta \) can be positive or negative, hence \( dT^o / d\theta \) can be positive or negative, depending on the shape of the pdf \( f(\cdot) \). If \( \theta \geq 1 \) then \( \partial g / \partial \theta \geq 0 \) hence \( dT^o / d\theta \leq 0 \).

7. INTRODUCING SEQUENCES OF ACTIVITIES

For simplicity, we have so far considered a single activity, for example a scheduled nonstop bus, train or airline service from A to B. The discussion of course also applies to a set of activities which are independent of each other. However, in practice the same bus, train or plane, or their crews, may continue on to undertake a sequence of scheduled services. Since these are scheduled in sequence, any lateness in completing one of these may affect the start time of the next activity, and so on.

The analysis in this paper can easily be extended to such sequences of activities. However, in this case we generally find numerical rather than analytical solutions. This is not because of difficulties caused by the behavioural adjustment phenomenon introduced in the present paper. Instead, it is due to the fact that existing discussions of sequences of activities (e.g. multi-stage transport services) have found only numerical rather than analytical solutions, even without introducing any behavioural adjustment.

For example, Carey (1992, 1994) and Powell and Sheffi (1983), etc., consider a route consisting of several links. They introduce random variation in link traversal times and in wait times at stops, and consider the effects of these on the distributions of arrival and departure times at scheduled stops along the route. They do not consider behavioural adjustment but even in this case they do not obtain any analytical solution. Instead, they derive sets of recursive equations which can be used to compute the arrival and departure time distributions at all stops from the known distributions of trip times on previous links and wait times at previous stops.

We can easily extend this approach to introduce behavioural response. To do this we simply shift the pdf of each activity to the right by a behavioural adjustment term \( \theta(T - T^o_i) \), as in Sections 3 and 4 above. Thus if \( f_i(t) \) is the pdf of the \( i \)th activity, this becomes \( f_i(t - \theta(T_i - T^o_i)) \).

Similarly, suppose that \( f^a_i(t) \) is the pdf of the completion time of the \( i \)th activity (e.g. the departure time from a stop) and \( f^s_i(t) \) is the pdf of the time taken by the next activity (e.g. the trip time on the next link). Then in the absence of behavioural adjustment the pdf of arrival times at the next stop is given by the convolution

\[
f^a_{i+1}(t) = \int_0^t f^a_i(\tau)f^s_i(t - \tau)d\tau
\]

—see Carey (1994). If we introduce behavioural adjustment in the link trip time this simply becomes,

\[
f^a_{i+1}(t) = \int_0^t f^a_i(\tau)f^s_i(t - \tau - \theta(T_i - T^o_i))d\tau
\]

Carey (1994) also discusses how to compute iteratively scheduled arrival and departure times so as to optimize a timetable for sequence of several trains serving a sequence of several stops. This is applied in Carey and Seckington (1992) to finding optimal timetables for a simple passenger railway system. Behavioural adjustment can be introduced into this iterative optimization approach by making only minor adjustments. Basically, as above, replace pdfs \( f_i(t) \) with \( f_i(t - \theta(T_i - T^o_i)) \) throughout. It would also be of interest to introduce the behavioural adjustment process of the present paper into the other papers listed in the introduction. It would presumably be done in a similar way to that suggested here.

8. SUMMARY AND CONCLUDING REMARKS

Here we summarise the optimality conditions under different assumptions. From these optimality conditions the optimal scheduled time \( T^o \) can be computed. We need only consider the case when there is behavioural response, since the case without behavioural response is then a special case obtained by simply setting the response parameter \( \theta \) to zero.
1. Suppose the cost per unit of scheduled time is a constant and per unit of lateness is a constant $c_t$, and per unit of lateness is a constant $c_l$. Then:

(a) If $\theta < \tilde{c}$ (see Section 4) then

$$[1 - F^\theta(T^o)] = 1 - \tilde{c}$$

or

$$T^o = F^{\theta-1}[\tilde{c}]$$

where $\tilde{c} = 1 - c_l/c_t(1 - \theta)$ and $F^\theta(T) = F(T^o - \theta(T^o - T^m))$, i.e. $F(T)$ shifted to the right amount $\theta(T - T^m)$.

(b) If $\theta > \tilde{c}$ then $T^o = T^m$.

(c) If $\theta = \tilde{c}$ then both results (a) and (b) hold.

Note that the optimal solution, the optimal $T$, can easily be computed in (a) or (b). Also, (b) is a remarkable result. It implies that only the absolute minimum feasible time should be scheduled if more than a fraction $\tilde{c}$ of the additional time would be ‘wasted’. This implies that the activity (trip, etc.) will usually be completed late.

2. If the cost of scheduled time $c_t(T)$ is nonlinear but the cost of lateness $c_l(l)$ is linear, then the optimality conditions are as in eqn (10) in Section 6. However, it is very unlikely that the time cost will be nonlinear and the lateness cost linear. The reason is, that the lateness cost includes a cost of time as well as lateness. Hence it is much more realistic to consider nonlinear costs of lateness. In that case the optimal scheduled time $T^o$ is that which satisfies

$$c'_i(T^o) = \int_{T^o-\theta(T^o-T^m)}^{+\infty} c'_i(t - (T^o - \theta(T^o - T^m))) f(t) dt$$

(12)

If the lateness cost function $c_l(l)$ is also piecewise linear, with break points at $l_1, \ldots, l_n$, then $T^o$ is computed from

$$c_t = \sum_{i=1}^n [F(T^o + [l_i] - F(T^o + [l_{i-1}]) c_i$$

(14)

where $T^o = (T - \theta(T - T^m))$. This states that the travel cost per min ($c_t$) exactly equals the expected lateness cost per min.

3. If we set the behavioural response $\theta$ to zero, the effects on the above results are obvious. Thus eqn (8) reduces to,

$$[1 - F(T^o)] = c_t/c_l$$

or

$$T^o = F^{-1}(1 - c_t/c_l)$$

(2)

as in Section 2, eqn (12) reduces to,

$$c'_i(T^o) = \int_{T^o}^{+\infty} c'_i(t - T^o) f(t) dt$$

(11)

as in Section 4, and eqn (14) reduces to

$$c_t = \sum_{i=1}^n [F(T^o + [l_i] - F(T^o + [l_{i-1}]) c_i$$

(13)

Acknowledgements—This research was supported by U.K. Engineering and Physical Science Research Council (EPSRC) grants GR/H/50432, GR/H48033 and GR/K/75798, and by cooperation with British Rail. The author would like to thank the EPSRC and British Rail for their support. The author would also like to thank an associate editor and two anonymous referees for their helpful comments and suggestions.
REFERENCES


APPENDIX

Notation

The following notation is defined and used in the paper and is repeated here for ease of reference.

\( t \) A random variable denoting the amount of time actually taken for an activity, e.g. a trip from A to B.

\( f(t) \) and \( F(t) \) Probability density function and distribution function respectively for \( t \).

\( T \) Amount of time scheduled for an activity.

\( T^o \) The optimal value of \( T \), to minimise the cost of an activity (e.g. travel costs plus lateness costs).

\( c_iT \) or \( c_i(T) \) Cost of lateness \( \mathcal{I} = (t - T) \), \( t > T \). Expected or average cost of lateness is \( \mathcal{C}_\mathcal{I} = C_i(T) = E[(t - T)c_i|t \geq T] \).

\( \bar{\mathcal{C}} = (1 - c_i/c_l) = (c_l/c_l - c_i/c_l) \), or \( (1 - c_i/c_l(1 - \theta)) \). (We find that \( T^o \) is given by \( F(T^o) = \bar{\mathcal{C}} \) or \( F^\theta(T^o) = \bar{\mathcal{C}} \)).

\( \mathcal{C}(T) = C_i(T) + C_l(T) \) Total cost of an activity: (trip time cost) + (expected lateness cost).

\( T_1 = \min(t) \) Minimum time required for an activity. Hence \( f(t) = F(t) = 0 \) for all \( t < T_1 \).

\( T_1^m \) The smallest possible value of \( T_1 \). This occurs if we allow no slack in the scheduled time \( T \).

\( \theta \) Fraction of the slack time \( (T - T_1^m) \) which is used up by operators, etc., so that the minimum trip time increases by an amount \( \theta(T - T_1^m) \), hence becomes \( T_1 = T_1^m + \theta(T - T_1^m) \), for \( T \geq T_1^m \). The probability density of \( t \) is now \( f(t - \theta(T - T_1^m)) \) rather than \( f(t) \).

\( F^\theta(T) = F(T^o - \theta(T^o - T_1^m)) \) i.e. \( F(T) \) shifted to the right by the amount \( \theta(T - T_1^m) \), in response to inserting the slack \( (T - T_1^m) \) in the schedule. The inverse of \( F^\theta(T) = x \) is written \( T = F^{\theta^{-1}}(x) \).