Further Exposition of the Value of Reliability

Richard Batley
Institute for Transport Studies, University of Leeds, UK
rbatley@its.leeds.ac.uk

Paper to be presented at the 45th Congress of the European Regional Science Association, 23-27 August 2005, Vrije Universiteit Amsterdam

Abstract

The aim of this paper is to assist deeper understanding of the value of reliability, as it relates to the users of transport systems. The approach is theoretical, and follows the precedent of Noland & Small (1995) and Bates *et al.* (2001) in couching the scheduling model of Small (1982) within an objective problem of expected utility maximisation (von Neumann & Morgenstern, 1947; Savage, 1954). In contrast to these earlier works on reliability, the paper adopts a discrete representation of time; this appeals both to the context of scheduled public transport services, and implementation within Stated Preference and the Random Utility Model. The paper applies this representation to further theoretical exposition, in the following respects. The implications of Small's utility function for travellers' attitudes to unreliability are considered, finding that travellers would tend to exhibit risk aversion. Following from this observation, the paper considers the associated risk premium, i.e. the delay in arrival time that a risk-averse traveller would be willing-to-accept in exchange for eliminating unreliability in arrival time. The risk premium is then converted from time to money, thereby arriving at the 'true' value of reliability. Following from the properties of Small's function, the theoretical analysis yields two policy implications. First, a public transport operator might feasibly increase the timetabled journey time, whilst maintaining market share, provided full reliability of service is ensured. Second, some departure times carry a value of reliability, whilst others do not.

Keywords

Reliability, Valuation, Expected Utility, Risk Aversion, Risk Premium

JEL Classification

D11 Consumer Economics: Theory
D81 Criteria for Decision-Making under Risk and Uncertainty
R41 Transportation: Demand, Supply. Congestion, Safety, and Accidents

1. Introduction

Although a precise understanding has often seemed elusive, it is widely accepted that the reliability of transport systems may impact upon the choices of transport users. Research has usefully illuminated several facets of this proposition, but usually without the authority of comprehensive evidence on the *value* of reliability to transport users. That such evidence is lacking can perhaps, in turn, be attributed to the difficulty of formulating a research apparatus that carries theoretical validity, is insightful, but remains practicable. The latter aspiration is the concern of the present paper.

The review by De Jong *et al.* (2004) distinguishes between three approaches to the valuation of reliability, referred to as: I) the mean *vs.* variance approach, II) percentiles of the travel time distribution, and III) scheduling models. This paper exploits the third approach, which is founded on the hypothesis that travellers may accommodate expectations of unreliability through their trip scheduling. In the analysis of trip scheduling, Small's (1982) approach has received considerable support. Small extends the microeconomic theory of time allocation (e.g. Becker, 1965; De Serpa, 1971), accounting for scheduling constraints, through reference to Vickrey (1969), in the specification of utility and its associated constraints.

A fundamental limitation of Small's approach, however, is that individuals make choices under certainty, an assumption that is clearly unrealistic in the context of urban travel choice. The orthodox response to such challenge - at least in terms of microeconomic theory - is to reformulate the objective problem from the maximisation of utility, to one of maximising expected utility (von Neumann & Morgenstern, 1947; Savage, 1954). The latter works are exploited by Noland & Small (1995), who re-couch Small's original model of trip scheduling within an objective problem of expected utility maximisation.

Two related properties of Noland & Small's analysis should be noted. First, both the choice (i.e. departure time) and pay-off (i.e. arrival time) dimensions are specified to be continuous; this carries the attraction of permitting easy calculation of the optimal departure time. Second, interest is restricted to the morning commute of car travellers. The proposition of a continuous pay-off would appear more reasonable for car travellers than for users of public transport services, since the latter are typically constrained by fixed service intervals. Bates *et al.* (2001) develop Noland & Small (1995) further, first considering its amenability to public transport users, and then applying the analysis to derive a value of reliability from a choice between two public transport services.

The present paper adheres to the basic thesis of Noland & Small and Bates *et al*., but with the following distinctions. First, a discrete representation of time is adopted in both the departure and arrival dimensions. Not only is this more faithful to von Neumann & Morgenstern (1947) and Savage (1954), but it permits ready accommodation of public transport users. As we shall see in due course, the discrete representation would, furthermore, appear more amenable to analysis using Stated Preference (SP) and the Random Utility Model (RUM). Second, the paper employs theoretical analysis to yield a deeper understanding of the value of reliability than hitherto offered. To this end, the paper illuminates travellers' attitudes to unreliability, and articulates the notion of a *reliability premium*. The latter permits a succinct definition of the 'true' value of reliability.

2. Theory of individual choice under uncertainty

Theoretical analysis of risk and uncertainty typically involves some relation between choice and a probability distribution. The interpretation of the latter has been the source of some contention in the literature, since it is embroiled with the distinction between risk and uncertainty. Keynes (1921, 1936) and Knight (1921) are helpful in this regard, characterising *risk* as situations where probabilities are known (or knowable), and *uncertainty* as situations where probabilities may be neither knowable nor definable. Rather than distract ourselves with this debate, let us arbitrarily adopt the term *uncertainty* in what follows, without necessarily implying allegiance to the above distinction.

Despite the best efforts of experimental economists, it would seem premature to depose the orthodox paradigm of von Neumann & Morgenstern (1947) and Savage (1954). Let us then proceed with their model of expected utility maximisation, which is couched at the level of the individual. Formally:

Let E be a finite and exhaustive set of 'events':

$$E = \{e_1,...,e_K\}$$

Corresponding to E, define a 'prospect' vector:

$$\mathbf{w} = (w_1,...,w_K; p_1,...,p_K)$$

where w_k is the pay-off to the individual if event e_k occurs, and p_k is the probability (however defined) that event e_k does indeed occur. With regards

to the event probability, the necessary condition $\sum_{k=1}^{K} p_k = 1$ applies; it follows that $0 \leq p_k \leq 1$ for $k = 1,...,K$. Finally, let T be a finite and exhaustive set of N prospects, from which the individual is invited to choose his or her preferred alternative:

$T = \{\mathbf{w}_1,...,\mathbf{w}_N\}$

Having defined the relevant variables, let us state the necessary and sufficient axioms, as follows.

Completeness over prospects states that an individual is able to express weak preference between any pair of prospects. Formally:

Either $\mathbf{w}_q W \mathbf{w}_r$, or $\mathbf{w}_r W \mathbf{w}_q$, or both $\mathbf{w}_q W \mathbf{w}_r$ and $\mathbf{w}_r W \mathbf{w}_q$.

where W denotes 'weakly preferred' (i.e. indifferent to or strictly preferred to).

Transitivity over prospects imposes a consistency over cycles of weak preference. Formally:

If $\mathbf{w}_n W \mathbf{w}_q$ and $\mathbf{w}_q W \mathbf{w}_r$, then $\mathbf{w}_n W \mathbf{w}_r$

Taken together, completeness and transitivity establish a complete (weak) preference ordering of the prospects $\mathbf{w} \in T$.

Preference increasing with probability states that if the probability of a preferred pay-off within a prospect increases, while the probability of an inferior pay-off falls, then the new prospect will be preferred to the old. Formalising for the simple case $E = \{e_i, e_j\}$:

If $w_i S w_j$ and $\mathbf{w}_q = (w_i, w_j; p_{iq}, [1 - p_{iq}])$, $\mathbf{w}_r = (w_i, w_j; p_{ir}, [1 - p_{ir}])$, then $\mathbf{w}_q S \mathbf{w}_r$ iff $p_{iq} > p_{ir}$.

where S denotes 'strictly preferred'.

Continuity over prospects states that for any three prospects, it is always possible to combine the best and worst prospects in some probability mix, and arrive at a prospect that is indifferent to the middle prospect. Again formally:

If $\mathbf{w}_n W \mathbf{w}_q$ and $\mathbf{w}_q W \mathbf{w}_r$ then there exists some probability p such that $(\mathbf{w}_n, \mathbf{w}_r; p, [1-p])\, I\, \mathbf{w}_q$

where I denotes 'indifferent to'.

Strong independence states that if any pay-off within a prospect is substituted by a pay-off that is regarded as indifferent, then there will be indifference between the resulting prospect and the original one. Formally:

If $\mathbf{w} = (w_i, w_j; p_i, [1-p_i])$ and $w_i\ I\ w_k$, then $\mathbf{w}\ I\ (w_k, w_j; p_i, [1-p_i])$

Rules for combining probabilities relates to general rules for taking expectations. Suffice to say, if the preceding axioms hold, then preferences over prospects can be represented by a utility function, such that for any two prospects:

$\mathbf{w}_q = (w_{iq}, w_{jq}; p_{iq}, [1-p_{iq}])$ and $\mathbf{w}_r = (w_{ir}, w_{jr}; p_{ir}, [1-p_{ir}])$, $\mathbf{w}_q W \mathbf{w}_r$ iff:

$$Y(\mathbf{w}_q) \geq Y(\mathbf{w}_r)$$

where:

$$Y(\mathbf{w}_q) = p_{iq}U(w_{iq}) + [1-p_{iq}]U(w_{jq})$$
$$Y(\mathbf{w}_r) = p_{ir}U(w_{ir}) + [1-p_{ir}]U(w_{jr})$$

and $U(w)$ is the 'Von Neumann & Morgenstern utility index' of w.

To interpret, the individual acts so as to maximise *expected utility*. The conventional wisdom - it would appear - is that choice under uncertainty permits the mutation of utility from an ordinal metric to a cardinal one. Baumol's (1958) clarification on this is important. The above theory relies, indeed, on the proposition that $U(w)$ is cardinal. The latter is, however, derived from $Y(\mathbf{w})$ - rather than *vice versa* - where $Y(\mathbf{w})$ is an entirely ordinal construct. In other words, the dependent variable does *not* 'become cardinal' in any shape or form.

Finally, and with an eye on the potential for implementation in SP, we can translate the above presentation to RUM, exploiting the proposal of Marschak *et al.* (1963). It is important to be clear about the basis for adopting a probabilistic representation, which is as follows. Consider an individual faced with a repeated choice task under uncertainty. On any given repetition, he or she is able to order a set of prospects in terms of expected utility, but on

successive repetitions this ordering may show variability. Formally, if $(w_1,...,w_K)$ is constant across prospects, and there exists a random vector:

$$\mathbf{Y} = (Y(\mathbf{w}_1),...,Y(\mathbf{w}_N))$$

then probability can be expressed as RUM, such that:

$$P(\mathbf{w}_q|T) = \Pr\{Y(\mathbf{w}_q) \geq Y(\mathbf{w}_r)\} \text{ for all } \mathbf{w}_r \in T,\ q \neq r$$

3. The theory applied to trip scheduling

For purposes of application to trip scheduling, the above theory may be re-interpreted as follows. The pay-off is defined over the dimension of arrival time. Since arrival time is naturally an 'event', pay-offs and events become - at least for purposes of the analysis - one and the same. The prospect (or choice) set is similarly defined over the dimension of departure time. More formally:

Let A be a finite and exhaustive set of arrival times:

$$A = \{a_1,...,a_K\}$$

Let D be a finite and exhaustive set of departure times:

$$D = \{d_1,...,d_N\},$$

The latter corresponds to the choice set T:

$$T = \{\mathbf{w}_1,...,\mathbf{w}_N\}$$

wherein each prospect is defined in terms of the $a_k \in A$ for $k = 1,...,K$, together with the associated event probabilities, thus:

$$\mathbf{w}_n = (a_1,...,a_K; p_{1n},...,p_{Kn})$$
$$\sum_{k=1}^{K} p_{kn} = 1, \text{ for } n = 1,...,N$$
$$0 \leq p_{kn} \leq 1, \text{ for } k = 1,...,K \text{ and } n = 1,...,N$$

Introducing some efficiency in notation, the expected utility of any departure time $d_n \in D$ for $n = 1,...,N$ can be expressed:

$$Y_n = \sum_{k=1}^{K} p_{kn} U_{kn}$$

It remains to specify the precise form of the Von-Neumann & Morgenstern utility index U. Given its predominance in the literature of scheduling models, let us adopt Small's (1982) formulation, which is itself a development of Vickrey's (1969).

$$U_{kn} = \alpha T_{kn} + \beta SDE_k + \gamma SDL_k + \delta L_k \qquad (1)$$

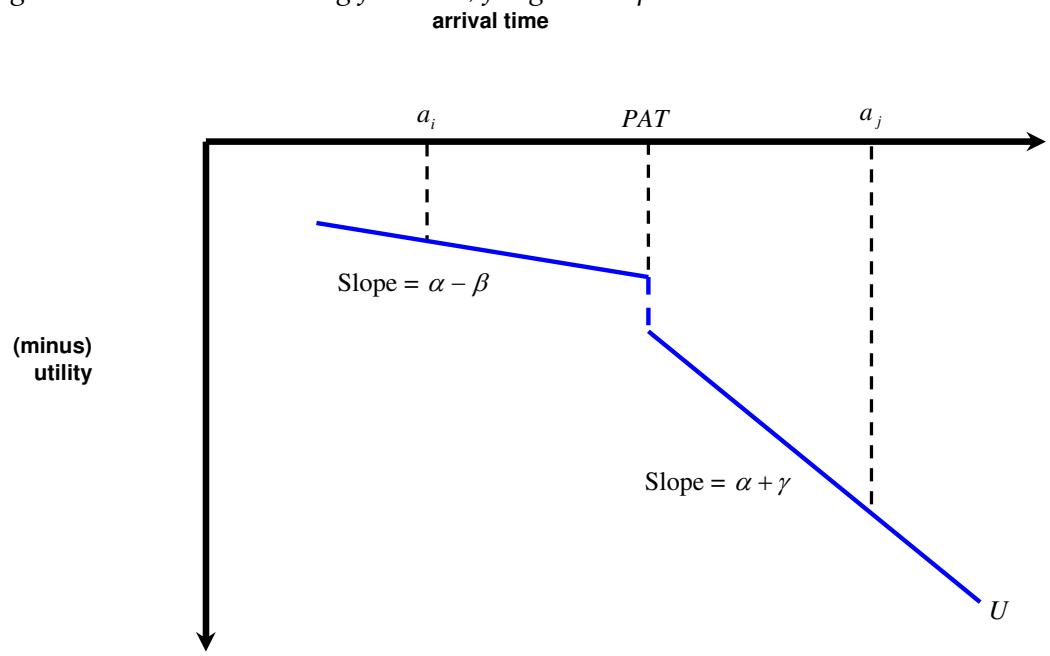
where:
T is travel time
SDE is schedule delay early
SDL is schedule delay late
L is a penalty for late arrival

The above formulation is conditioned by the preferred arrival time (PAT) of the traveller, which we take as given. On this basis, let us re-express the components of U in terms of our dimensions of interest - arrival time and departure time - for given PAT:

$$T_{kn} = a_k - d_n$$
$$SDE_k = \max\left[\left(PAT - a_k\right), 0\right]$$
$$SDL_k = \max\left[\left(a_k - PAT\right), 0\right]$$
$$L_k = 1 \text{ if } \left(a_k - PAT\right) > 0, = 0 \text{ otherwise}$$

Since all components of U are 'bads', it must be the case that $\alpha, \beta, \gamma, \delta < 0$. The empirical estimates of Small (1982), furthermore, suggest that $\gamma < \alpha < \beta$. Accepting these relations, we are able to give a schematic representation of the function U for any departure time $d \in D$. Figure 1 represents arrival time on the horizontal axis, and utility on the vertical; as U is comprised entirely of bads, we operate in the lower right quadrant.

Figure 1: Small's scheduling function, for given departure time



With reference to Figure 1, we can establish that:

At $a = a_i$: $U = \alpha(a_i - d) + \beta(PAT - a_i)$

As $(PAT - a_i) \to 0$: $U \to \alpha(PAT - d) \to \alpha(a_i - d) + \alpha(PAT - a_i)$

At $a = a_j$: $U = \alpha(a_j - d) + \gamma(a_j - PAT) + \delta$

As $(a_j - PAT) \to 0$: $U \to \alpha(PAT - d) + \delta \to \alpha(a_j - d) - \alpha(a_j - PAT) + \delta$

This now enables us to identify the respective slopes of the two portions of the function U; the upper portion must be of slope $(\alpha - \beta)$, and the lower of slope $(\alpha + \gamma)$. Given the sign and relative magnitude of the parameters, both portions must have negative slopes, and the lower portion must be steeper than the upper.

Before proceeding, three observations on the above are appropriate. First, it should be noted that T_{kn} straightforwardly accommodates the three components of travel time identified by Noland and Small (1995), namely free flow travel time, recurrent delay, and incident-related delay. Second, unlike some analyses of schedule delay, there is the facility for T_{kn} to vary by departure time d_n for $n = 1,...,N$. Third, the discrete representation of departure time is readily amenable to fixed-schedule public transport services. Furthermore, it would appear relatively straightforward - as

compared with a continuous representation - to collect data using SP, and carry out analysis using RUM.

4. Identifying the preferred departure time

Let us now turn our attention to the task of identifying the preferred departure time or, in other words, the prospect that yields maximum expected utility. To permit some expositional clarity over the general case above, we restrict attention to a binary subset of events and a binary subset of choice alternatives. Thus define:

$\tilde{A} \subset A$, $\tilde{A} = \{a_i, a_j\}$ where $a_i < a_j$ (i.e. a_j is a later arrival than a_i)

$\tilde{D} \subset D$, $\tilde{D} = \{d_q, d_r\}$ where $d_q < d_r$

$T = \{\mathbf{w}_q, \mathbf{w}_r\}$

$$\mathbf{w}_n = (a_i, a_j; p_i, p_j) = (a_i, a_j; p_i, (1 - p_i))$$

$$Y_n = \{p_{in}[\alpha T_{in} + \beta SDE_i + \gamma SDL_i + \delta L_i]\} + \{(1 - p_{in})[\alpha T_{jn} + \beta SDE_j + \gamma SDL_j + \delta L_j]\}$$

for $d_n \in \tilde{D}$, $n = q, r$

The function Y_n may be alternatively expressed:

$$Y_n = \alpha E(T_n) + \beta E(SDE_n) + \gamma E(SDL_n) + \delta E(L_n) \qquad (2)$$

Either way:

$$Y_n = \begin{Bmatrix} \{p_{in}[\alpha(a_i - d_n) + \beta \max[(PAT - a_i), 0] + \gamma \max[(a_i - PAT), 0] + \delta L_i]\} + \\ \{(1 - p_{in})[\alpha(a_j - d_n) + \beta \max[(PAT - a_j), 0] + \gamma \max[(a_j - PAT), 0] + \delta L_j]\} \end{Bmatrix}$$

Following section 2, we can determine that:

$\mathbf{w}_q \, W \, \mathbf{w}_r$ iff $Y_q - Y_r \geq 0$

$\mathbf{w}_r \, W \, \mathbf{w}_q$ iff $Y_r - Y_q \geq 0$

$\mathbf{w}_q \; I \; \mathbf{w}_r$ iff both $Y_q - Y_r \geq 0$ and $Y_r - Y_q \geq 0$

Let us then derive the quantity $Y_q - Y_r$:

$$Y_q - Y_r = \left\{ \begin{array}{l} \left\{ p_{iq} \left[\alpha T_{iq} + \beta SDE_i + \gamma SDL_i + \delta L_i \right] \right\} + \\ \left\{ \left(1 - p_{iq}\right) \left[\alpha T_{jq} + \beta SDE_j + \gamma SDL_j + \delta L_j \right] \right\} \end{array} \right\}$$

$$- \left\{ \begin{array}{l} \left\{ p_{ir} \left[\alpha T_{ir} + \beta SDE_i + \gamma SDL_i + \delta L_i \right] \right\} + \\ \left\{ \left(1 - p_{ir}\right) \left[\alpha T_{jr} + \beta SDE_j + \gamma SDL_j + \delta L_j \right] \right\} \end{array} \right\}$$

$$= \left\{ \begin{array}{l} \left\{ p_{iq} \left[\alpha \left(a_i - d_q \right) + \beta SDE_i + \gamma SDL_i + \delta L_i \right] \right\} + \\ \left\{ \left(1 - p_{iq}\right) \left[\alpha \left(a_i + \Delta a - d_q \right) + \beta \left(SDE_i - \Delta SDE \right) + \gamma \left(SDL_i + \Delta SDL \right) + \delta \left(L_i + \Delta L \right) \right] \right\} \end{array} \right\}$$

$$- \left\{ \begin{array}{l} \left\{ p_{ir} \left[\alpha \left(a_i - d_q - \Delta d \right) + \beta SDE_i + \gamma SDL_i + \delta L_i \right] \right\} + \\ \left\{ \left(1 - p_{ir}\right) \left[\alpha \left(a_i + \Delta a - d_q - \Delta d \right) + \beta \left(SDE_i - \Delta SDE \right) + \gamma \left(SDL_i + \Delta SDL \right) + \delta \left(L_i + \Delta L \right) \right] \right\} \end{array} \right\}$$

where:

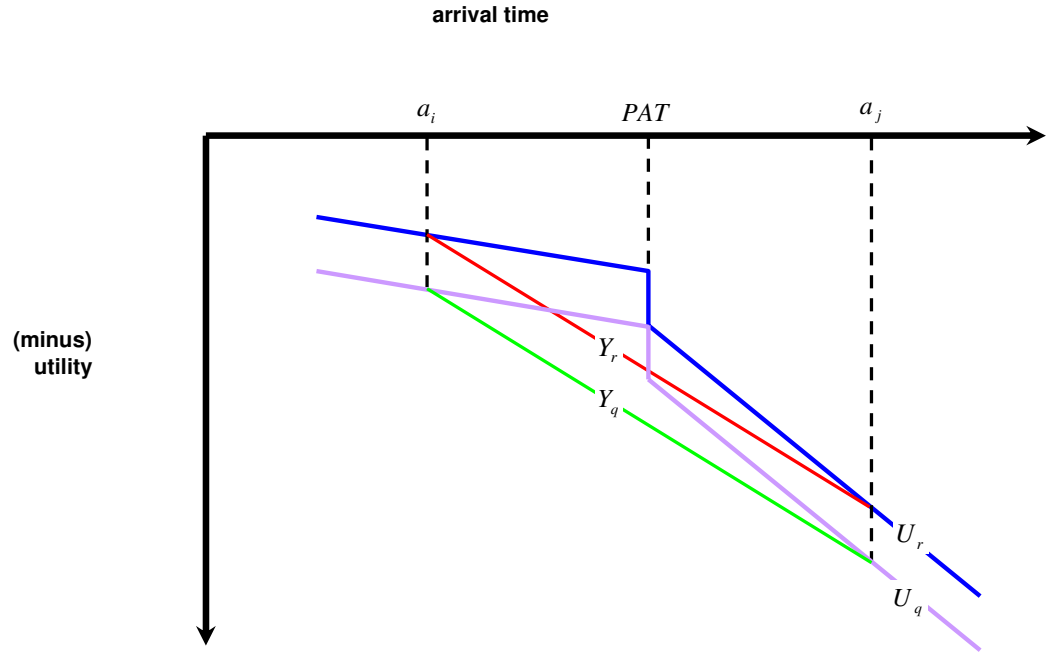
$\Delta a = \left(a_j - a_i \right) > 0$
$\Delta d = \left(d_r - d_q \right) > 0$
$\Delta SDE = \left(SDE_i - SDE_j \right) \geq 0$
$\Delta SDL = \left(SDL_j - SDL_i \right) \geq 0$
$\Delta L = 1$ if $a_i < PAT < a_j$, $= 0$ otherwise

Simplifying, we arrive at the expression:

$$\begin{array}{l} Y_q - Y_r = \alpha \left[\Delta d + \left(p_{ir} - p_{iq} \right) \Delta a \right] + \\ \qquad \beta \left[\left(p_{iq} - p_{ir} \right) \Delta SDE \right] + \\ \qquad \gamma \left[\left(p_{ir} - p_{iq} \right) \Delta SDL \right] + \\ \qquad \delta \left[\left(p_{ir} - p_{iq} \right) \Delta L \right] \end{array} \qquad (3)$$

This is shown diagrammatically in Figure 2, which features two utility functions; one pertaining to d_q and the second to d_r. It is clear from the figure that if $p_{ir} = p_{iq}$, then it is necessarily the case that $Y_r > Y_q$; this may be confirmed by reference to (3).

Figure 2: Choice between prospects with different departure times



5. Valuing reliability

Having equipped ourselves with the requisite theory, let us now proceed with our interest in the value of reliability. The concept of reliability has attracted a variety of definitions in the literature. Once again, however, we shall endeavour to steer away from such contention, by simply deferring to Bates *et al.*'s (2001) interpretation. Thus in terms of section 4, p_{iq} and p_{ir} proxy for the reliability of d_q and d_r, respectively. With reference to (3), therefore, any change in reliability will impact on expected arrival time (and hence expected travel time), expected schedule delay early, expected schedule delay late, and the expected late penalty.

We are now in a position to infer the value of reliability, using the approach followed by Bates *et al.* Before proceeding, it might be noted that the scheduling function applied empirically by Bates *et al.* (i.e. their equation (21)) shows slight divergence from the theory (as embodied by equation (2) of the present paper). Here we follow the essence of Bates *et al.*'s approach, but adhere to equation (2).

The expected utility function (2) is first supplemented with a variable representing travel cost:

$$Y_n = \alpha E(T_n) + \beta E(SDE_n) + \gamma E(SDL_n) + \delta E(L_n) + \phi C_n$$

where C is travel cost (noting that, in this case, cost is variable by departure time but not arrival time). The quantity (3) can then be re-written:

$$\begin{aligned} Y_q - Y_r = {} & \alpha\left[\Delta d + \left(p_{ir} - p_{iq}\right)\Delta a\right] + \\ & \beta\left[\left(p_{iq} - p_{ir}\right)\Delta SDE\right] + \\ & \gamma\left[\left(p_{ir} - p_{iq}\right)\Delta SDL\right] + \\ & \delta\left[\left(p_{ir} - p_{iq}\right)\Delta L\right] + \\ & \phi\Delta C \end{aligned} \qquad (4)$$

where $\Delta C = C_q - C_r$. With reference to (4), the difference in expected utility between alternatives d_q and d_r can be seen to be a function of the differences in departure time, expected arrival time, expected schedule delay early, expected schedule delay late, expected late penalty and expected cost. For a choice between prospects on the basis of maximum expected utility, moreover, we can derive the value of each component of expected utility as follows:

$$V\left(E(T)\right) = \frac{\partial Y / \partial E(T)}{\partial Y / \partial E(C)} = \frac{\alpha}{\phi}$$

$$V\left(E(SDE)\right) = \frac{\partial Y / \partial E(SDE)}{\partial Y / \partial E(C)} = \frac{\beta}{\phi}$$

$$V\left(E(SDL)\right) = \frac{\partial Y / \partial E(SDL)}{\partial Y / \partial E(C)} = \frac{\gamma}{\phi}$$

$$V\left(E(L)\right) = \frac{\partial Y / \partial E(L)}{\partial Y / \partial E(C)} = \frac{\delta}{\phi}$$

where V denotes 'value'.

6. Further exposition of the value of reliability

The analysis thus far basically covers the extent of Bates *et al.*'s consideration of the value of reliability. Let us now develop ideas further, with two particular interests. First, we will seek to illuminate travellers' attitudes to unreliability. Second, we will introduce the notion of a *reliability premium*, which in turn will yield the 'true' value of reliability. Before proceeding, it is fair to acknowledge that Polak (1987) considers the same two interests, albeit with different method and motivation.

6.1 Attitudes to unreliability

Further recourse to microeconomic theory provides a basis for developing our interest in attitudes to reliability, as follows. Let us calculate the utility of the expected pay-off (i.e. expected arrival time), for any departure time $d_n \in D$ for $n = 1,..., N$:

$$U(E(a_n)) = \alpha(E(a_n) - d_n) + \beta \max[(PAT - E(a_n)),0] + \gamma \max[(E(a_n) - PAT),0] + \delta L(E(a_n))$$

where
$E(a_n) = p_{in} a_i + (1 - p_{in}) a_j$
$L(E(a_n)) = 1$ if $(E(a_n) - PAT) > 0$, $= 0$ otherwise

Substituting:

$$U(E(a_n)) = \alpha(p_{in} a_i + (1 - p_{in}) a_j - d_n) + \beta \max[(PAT - p_{in} a_i - (1 - p_{in}) a_j),0] + \gamma \max[(p_{in} a_i + (1 - p_{in}) a_j - PAT),0] + \delta L(E(a_n))$$

According to theory, the subtraction $Y_n - U(E(a_n))$ yields the following inferences:

- If $Y_n = U(E(a_n))$ then the individual is *risk neutral*
- If $Y_n > U(E(a_n))$ then the individual is *risk preferred*
- If $Y_n < U(E(a_n))$ then the individual is *risk averse*

Applying this relation, three cases are of relevance:

Case 1: $a_i < a_j \leq PAT$

Here $Y_n = U(E(a_n))$, such that the individual is risk neutral.

Case 2: $PAT \leq a_i < a_j$

Again $Y_n = U(E(a_n))$, implying risk neutrality.

Case 3: $a_i < PAT < a_j$

This case is more interesting; let us consider the further dichotomy of:

Case 3.1: $E(a_n) < PAT$

This implies that $L(E(a_n))=0$, and yields the subtraction:

$$Y_n - U(E(a_n)) = (1-p_{in})[\beta(a_j - PAT)+\gamma(a_j - PAT)+\delta L_j]$$

Since $\beta,\gamma,\delta<0$ and $0\le p_{in}\le 1$ by definition, and $(a_j - PAT)>0$ by assumption, it must be the case that $Y_n < U(E(a_n))$. The individual must therefore be risk averse.

Case 3.2 $E(a_n) > PAT$

In contrast to the previous case $L(E(a_n))=1$, and the subtraction becomes:

$$Y_n - U(E(a_n)) = p_{in}[\beta(PAT - a_i)+\gamma(PAT - a_i)-\delta L]$$

where $L = L_j = L(E(a_n))$

As before $\beta,\gamma,\delta<0$ and $0\le p_{in}\le 1$, whereas this time $(PAT - a_i)>0$ by assumption. Thus unlike the previous case, it cannot be determined *a priori* which of Y_n and $U(E(a_n))$ will be the greater. Having said that, if we defer to Small's (1982) empirical estimates of β,γ,δ then it appears likely the individual will again exhibit risk aversion.

6.2 The reliability premium

Developing the ideas of section 6.1 further, let us consider the concept of a *risk premium* - again taken from microeconomic theory - to our interest in the value of reliability. As a precursor to this, we shall first introduce the concept of a *certainty equivalent*, which in the present context may be defined as follows. The certainty equivalent for a given departure time $d\in D$ is the arrival time $\tilde{a}$ that yields the same utility with certainty as the expected utility of the prospect. Two cases are of relevance, as follows.

Case I: $E(a_n) < PAT$

Let $Y_n = \alpha(\tilde{a}_n - d_n)+\beta(PAT - \tilde{a}_n)$

This enables us to identify:

$$\tilde{a}_n = \frac{Y_n + \alpha d_n - \beta PAT}{(\alpha-\beta)}$$

Case II: $E(a_n) > PAT$

Let $Y_n = \alpha(\tilde{a}_n - d_n) + \gamma(\tilde{a}_n - PAT) + \delta L(\tilde{a})$

where $L(\tilde{a}) = 1$ if $(\tilde{a}_n - PAT) > 0$, $= 0$ otherwise

Similarly, we can identify:

$$\tilde{a}_n = \frac{Y_n + \alpha d_n + \gamma PAT - \delta L(\tilde{a})}{(\alpha + \gamma)}$$

Despite the apparent clarity of the above, it should be noted that the properties of Small's (1982) utility function carry the implication that an *exact* certainty equivalent may not be empirically guaranteed. Nonetheless, let us proceed to the definition of the risk premium, as follows:

$$K_n = \tilde{a}_n - E(a_n)$$

where $K_n \geq 0$.

The risk premium K_n is thus the difference between the certainty equivalent and the expected pay-off; note that the restriction on its sign implies that it is relevant only to the case of risk aversion. Having concluded in the previous section that Small's utility function would tend to exhibit risk aversion, the concept of a risk premium would therefore seem relevant to our interests.

The risk premium may be interpreted as the individual's willingness-to-pay (in units of the pay-off) to avoid risk. Let us re-interpret this for the present context: the risk premium (or *reliability premium*, perhaps) measures, for a given departure time, the delay in arrival time that the individual would be willing-to-accept in exchange for eliminating the unreliability. This is illustrated diagrammatically in Figures 3 and 4. Figure 3 considers a late arrival, and implies that a traveller would derive equal utility from the prospect and the certain arrival time $\tilde{a}$. This yields an interesting prescriptive implication: a public transport operator, if faced with such a situation, could introduce an increased journey time and still maintain market share, provided it could ensure full reliability of service.

Figure 3: The reliability premium of a late arrival, for given departure time

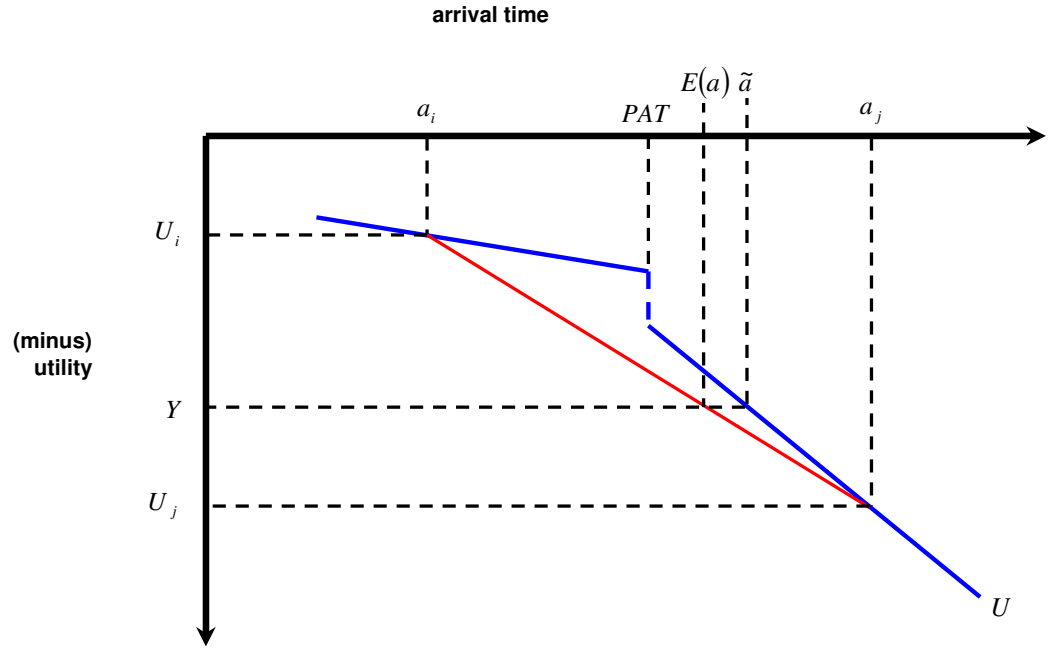
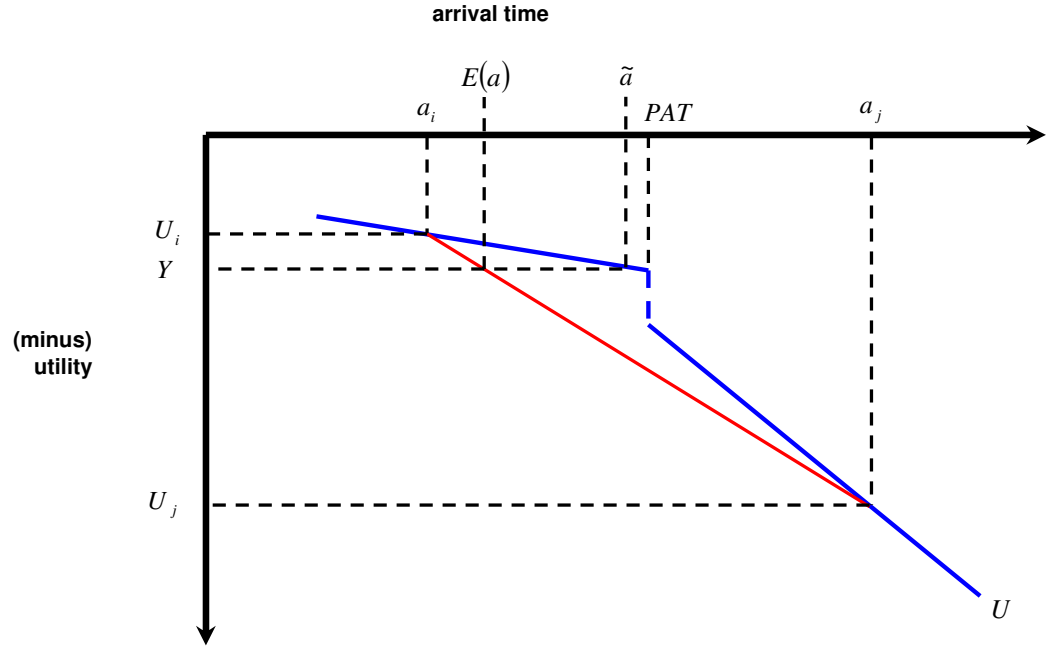


Figure 4 applies analogously to the case of early arrival, with similar implication; in this case, the traveller would be indifferent between the prospect and a certain arrival time $(\tilde{a})$ just early of the PAT .

Figure 4: The reliability premium of an early arrival, for given departure time



It remains to derive the value of the reliability premium K_n, which is elicited by means of a conversion from time to cost:

$$\begin{aligned} V(K_n) = {} & V(E(T)) * (\tilde{a}_n - E(a_n)) + \\ & V(E(SDE)) * [\max((PAT - \tilde{a}_n), 0) - \max((PAT - E(a_n)), 0)] + \\ & V(E(SDL)) * [\max((\tilde{a}_n - PAT), 0) - \max((E(a_n) - PAT), 0)] + \\ & V(E(L)) * [L(\tilde{a}) - L(E(a_n))] \end{aligned}$$

Moreover the above calculation constitutes, arguably, the 'true' value of reliability; for given departure, it measures the willingness-to-pay of the traveller to eliminate the unreliability of arrival time.

7. Worked example

Let us now illustrate the above theory by means of a worked example. With reference to Table I, consider a one-way commute with a departure time profile of 7:00am to 8:15am, in increments of 5 minutes. Arrival times are similarly defined in increments of 5 minutes, and reveal a minimum journey time of 30 minutes. This could be representative of a high-frequency scheduled public transport service; alternatively, it could be a discrete approximation to a car-based journey. Note that all times in the table are quantified in minutes after midnight. The body of the table displays the event probabilities by departure and arrival times. Note also that we consider a more general (and perhaps realistic) set of arrival times than the binary set considered in the preceding theoretical analysis. As a consequence, the clear conclusions of section 6.1 can no longer be relied upon, and we defer instead to the empirical findings, as follows.

Applying Small's (1982) estimates of α, β, γ and δ, and assuming that $PAT = 525$ (i.e. 8:45am), Figure 5 displays the expected travel time, expected SDE, expected SDL, and expected late penalty of each departure time in Table I, plotted against expected arrival time. Figure 6 plots the expected utility and utility of the expected arrival time, for each departure; it can be seen that departure $d = 465$ yields the highest expected utility. Comparing these two lines, expected utility and the utility of the expected arrival are equal in all but three cases; these cases pertain to arrivals surrounding the PAT. More specifically:

At both $d = 465$ and $d = 470$, $Y < U(E(a))$, implying risk aversion.

At $d = 475$, $Y > U(E(a))$, implying risk preference.

Figure 5: Expected travel time, SDE, SDL and late penalty vs. expected arrival time

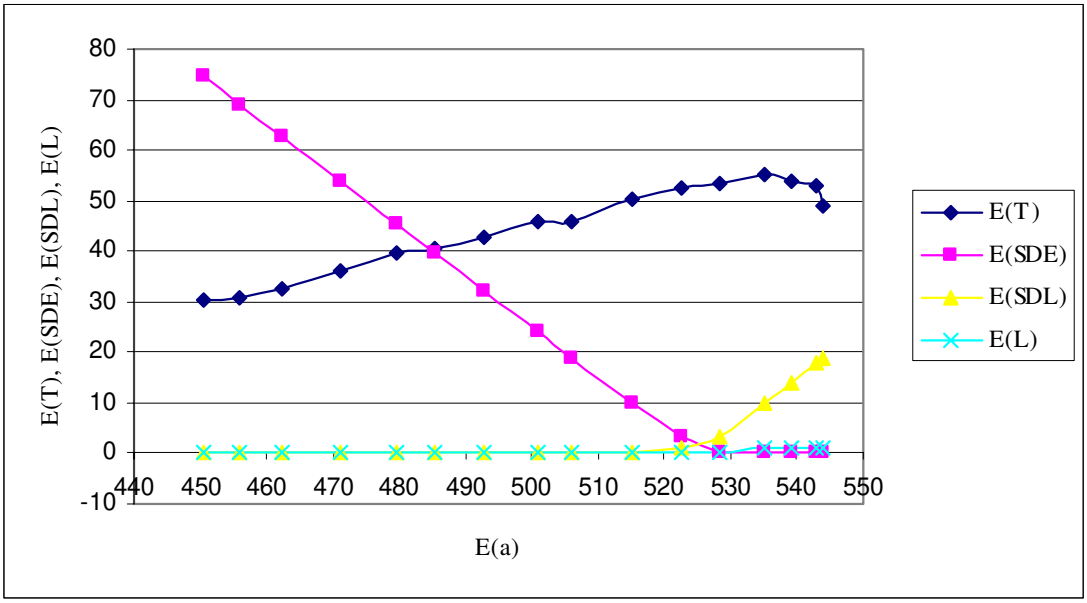
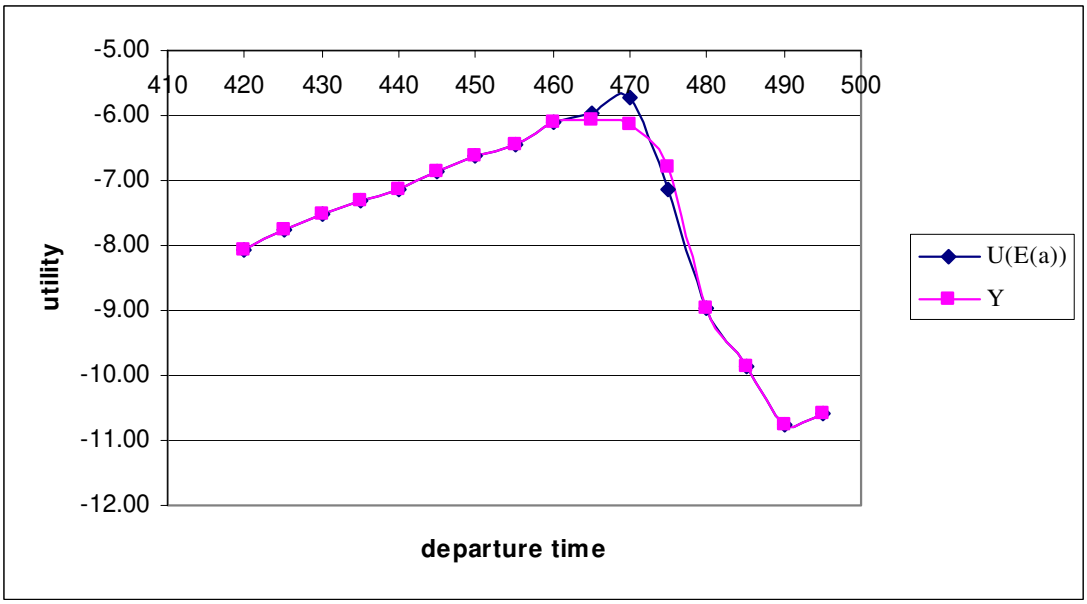
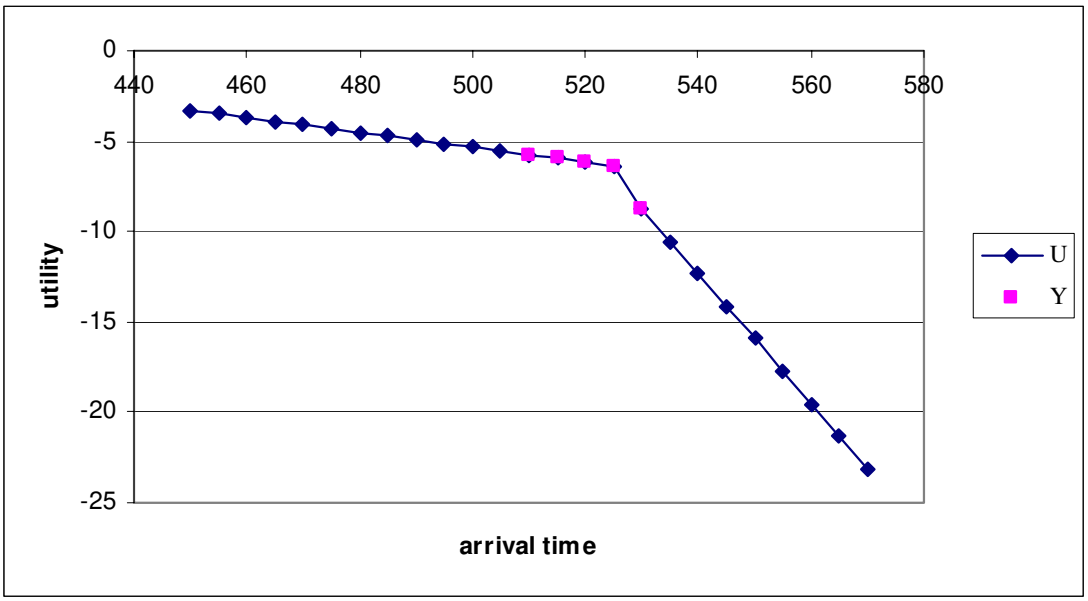


Figure 6: Expected utility and utility of expected arrival time, by departure time



Finally, let us consider an example of the reliability premium, taking the particular case of $d = 465$ (since this departure is characterised by risk aversion). The empirical utility function for this departure is shown in Figure 7, and follows the characteristic shape of the theoretical utility functions of Figures 1 to 4. The empirical expected utility function, in contrast, cannot be shown in the manner of the theoretical examples; this is because we have expanded the set of arrival times beyond the binary. Suffice to say, the arrival time window for $d = 465$ extends from $a = 510$ to $a = 530$; hence the points labelled Y.

Figure 7: Utility and expected utility functions for $d = 465$



For $d = 465$, we can calculate that:

$\tilde{a} = 517.90$
$E(a) = 515.25$

Thus a certain arrival time 2.65 minutes later than the expected arrival time would yield the same utility as the expected utility of the prospect. This reliability premium carries a value:

$$V(K) = [V(E(T)) - V(E(SDE))] \times 2.65$$

In the absence of comprehensive evidence, let us adopt Bates *et al.*'s (2001) valuation of $E(SDE)$ - relating to long distance rail services - which was 56.0 pence/minute, and approximate $E(T) = E(SDE) \times 1.6308$, where $\alpha/\beta = 1.6308$ from Small (1982). On this basis $V(K)$ is calculated to be 93.61 pence/minute. Contrast this with the case of $d = 460$, where $Y = U(E(a))$ and $V(K)$ is therefore zero. This yields another important implication for the likes of public transport operators. Specifically, some departure times carry a value of reliability (i.e. Case 3 of section 6.1), whereas others do not (Cases 1 and 2). Any investment to improve the reliability of the latter will - at least in terms of the reliability premium - yield zero benefit.

8. Summary and conclusion

The aim of this paper was to assist deeper understanding of the value of reliability, as it relates to the users of transport systems. The approach was theoretical, and involved couching the scheduling model of Small (1982) within an objective problem of expected utility maximisation (von Neumann & Morgenstern, 1947; Savage, 1954). Although similar analysis has been undertaken previously by Noland & Small (1995) and Bates *et al.* (2001), the present paper differed in the following respects. First, the paper adopted a discrete representation of time. Not only does this appeal to the context of scheduled public transport services, but it would seem more amenable to implementation in SP and RUM. A second, and more substantive, distinction was the scope of the theoretical exposition, which offered some extensions beyond the extant reliability literature, as follows.

Drawing analogy with the theoretical literature on attitudes to risk, the paper considered the implications of Small's (1982) utility function for travellers' attitudes to unreliability. It was found that, given Small's function, the departure time choices of travellers would tend to imply risk aversion. Following from this observation, the paper introduced the notion of a *reliability premium*. The reliability premium measures, for given departure time, the delay in arrival time that a risk-averse traveller would be willing-to-accept in exchange for eliminating unreliability in arrival time. The paper reported the policy implication that a public transport operator could feasibly increase the timetabled journey time, but still maintain market share, provided it could ensure full reliability of service. Finally, the reliability premium was converted to a monetary measure, thereby arriving at the 'true' value of reliability. Accounting once again for the properties of Small's function, it was noted that expected utility would, for some departure times, deviate from the utility of the expected arrival time, whilst for other departures times it would not. This yielded another policy implication; that some departure times carry a value of reliability, whilst others do not.

References

Bates, J., Polak, J., Jones, P. and Cook, A. (2001) The valuation of reliability for personal travel. *Transportation Research E*, 37, pp191-229.

Baumol, W. (1958) The cardinal utility which is ordinal. *The Economic Journal*, 68 (272), pp665-672.

Becker, G.S. (1965) A theory of the allocation of time. *The Economic Journal*, 75 (299), pp493-517.

De Jong, G., Kroes, E., Plasmeijer, R., Sanders, P. and Warffemius, P. (2004) The value of reliability. Paper presented at the Applied Methods stream, European Transport Conference, Strasbourg, 4-6 October 2004.

De Serpa, A.C. (1971) A theory of the economics of time. *The Economic Journal*, 81 (324), pp828-846.

Keynes, J.M. (1921) *A Treatise on Probability*. Macmillan, London.

Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*. Macmillan, London.

Knight, F. (1921) *Risk, Uncertainty and Profit*. Houghton Mifflin, Boston.

Marschak, J., Becker, G.M. and DeGroot, M.H. (1963) Stochastic models of choice behavior. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1). D. Reidel, Dordrecht.

Noland, R. and Small, K. (1995) Travel time uncertainty, departure time choice and the cost of the morning commute. Working paper UCI-ITS-WP-95-1, Institute of Transportation Studies, University of California, Irvine.

Polak, J. (1987) Travel time variability and departure time choice. Transport Studies Group Discussion Paper No. 15, Polytechnic of Central London.

Small, K. (1982) The scheduling of consumer activities: work trips. *American Economic Review*, 72 (3), pp467-479.

Savage, L.J. (1954) *The Foundations of Statistics*. John Wiley, New York.

Vickrey, W.S. (1969) Congestion Theory and Transport Investment, *American Economic Review*, 59 (2) 251-260.

von Neumann, J. and Morgenstern, O. (1947) *Theory of Games and Economic Behavior* (Second Edition). Princeton University Press, Princeton.

Table I: Pay-off matrix

		a 450.00	a 455.00	a 460.00	a 465.00	a 470.00	a 475.00	a 480.00	a 485.00	a 490.00	a 495.00	a 500.00	a 505.00	a 510.00	a 515.00	a 520.00	a 525.00	a 530.00	a 535.00	a 540.00	a 545.00	a 550.00	a 555.00
d	420.00	0.90	0.10																				
d	425.00		0.85	0.10	0.05																		
d	430.00			0.60	0.30	0.10																	
d	435.00				0.20	0.50	0.20	0.10															
d	440.00						0.30	0.50	0.15	0.05													
d	445.00							0.20	0.60	0.10	0.10												
d	450.00								0.05	0.50	0.30	0.15											
d	455.00										0.20	0.50	0.20	0.10									
d	460.00											0.10	0.70	0.10	0.10								
d	465.00													0.30	0.50	0.10	0.05	0.05					
d	470.00														0.05	0.60	0.20	0.10	0.05				
d	475.00																0.60	0.20	0.10	0.10			
d	480.00																	0.30	0.50	0.10	0.10		
d	485.00																		0.50	0.30	0.10	0.10	
d	490.00																			0.60	0.20	0.20	
d	495.00																			0.50	0.30	0.10	0.10