The Journal of Choice Modelling 5 (2012) 157–175

Contents lists available at SciVerse ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

# Randomness in preference orderings, outcomes and attribute tastes: An application to journey time risk

Richard Batley [a,*], J. Nicolás Ibáñez [b,1]

[a] Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK
[b] Department of Transport, Technical University of Denmark, Bygningstorvet 116B, DK-2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Available online 21 May 2013

Keywords:
Discrete choice models
Probabilistic choice models
Transportation
Journey time risk
Value of reliability

ABSTRACT

Within the broad area of probabilistic modelling of individual discrete choice, we develop three strands of discussion. First, we outline a theoretical framework for the modelling of individual discrete choice under risk, distinguishing between three specific sources of randomness; in preference orderings, in outcomes, and in attribute tastes. Second, we apply this theoretical modelling framework to the domain of journey time risk (or 'reliability'), a subject which has acquired prominence in the transportation policies of many countries. Third, we apply the modelling framework empirically, based upon a Stated Preference experiment of 2395 rail travellers choosing between alternative journeys embodying different levels of journey time risk. Across the sample of travellers, we estimate a mean value of scheduled journey time of 25.62 pence/min, against a median of 18.55 pence/min. We further estimate a mean 'reliability ratio' (ratio of the value of standard deviation of journey time to the value of scheduled journey time) of 2.07, against a median of 0.85. The properties of the distribution of the reliability ratio suggest a predominant behaviour of aversion to journey time risk.



## 1. Introduction

The class of Random Utility Models (RUM) was introduced by Marschak (1960) and Block and Marschak (1960), who adopted a model form originating from the field of psychophysics (Fechner, 1859), and equipped this model with an interpretation in terms of economics. Irrespective of the disciplinary interpretation – psychophysics or economics – the model sought to formalise a behavioural phenomenon with no disciplinary allegiance, namely the phenomenon of individual discrete choice. This is where an individual decision-maker is presented with a finite and exhaustive set of $N$ alternatives $C=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$, from which he or she is invited to select their preferred alternative.

The focus of the present paper will be upon the economic interpretation of this phenomenon, which relies upon Block and Marschak's (1960) proposition of a preference ordering over discrete alternatives, and usually adopts Lancaster's (1966) representation of an alternative in terms of its attributes (including price as well as other quality features, whether observable or latent). Formalising matters, let us define 'alternative' $n$ to be a vector $\mathbf{x}_n=(x_{n1},\ldots,\ x_{nJ})$, where $x_{nj}$ is the quantity of attribute $j$, and where $x_{nj}\geq 0$ for all $j$ and $x_{nj}>0$ for at least some $j$. The axioms of 'completeness' and 'transitivity' establish a complete (weak) preference ordering[2] on $C$, which can be represented by a real-valued 'utility' function $U$.

* Corresponding author. Tel.: +44 1133431789; fax: +44 1133435334.
E-mail addresses: R.P.Batley@its.leeds.ac.uk (R. Batley), iban@transport.dtu.dk (J.N. Ibáñez).
[1] Tel.: +45 45251554; fax: +45 45936533.

[2] If, for each and every pair of alternatives within a choice set, we know whether one alternative is preferred to or indifferent to another alternative, then we have a complete (weak) preference ordering on that choice set.

1755-5345 
http://dx.doi.org/10.1016/j.jocm.2013.03.003

For any pair of alternatives within this ordering, the individual is represented as choosing the alternative $\mathbf{x}_n \in C$ that maximises his or her utility, i.e. if $\mathbf{x}_n \geq \mathbf{x}_m$ then $U(\mathbf{x}_n) \geq U(\mathbf{x}_m)$. This is what we mean by individual discrete choice.

The key contribution of Block and Marschak was to admit the possibility that an individual, repeatedly undertaking the discrete choice task described above, might not always make the same choice. This was conceptualised in terms of RUM, as follows:

$$P(\mathbf{x}_n|C) = \Pr(U(\mathbf{x}_n) > U(\mathbf{x}_m) \quad \forall m \in C, \quad m \neq n) \tag{1}$$

where $(U(\mathbf{x}_1), \ldots, U(\mathbf{x}_N))$ is a random vector $\mathbf{U}$ that is unique up to an increasing monotone transformation, and where the basic probability rules apply, that is, $P(\mathbf{x}_n|C) \geq 0$ for all $n \in C$, and $\sum_{n=1}^{N} P(\mathbf{x}_n|C) = 1$. Within RUM, utility is taken to be a random variable. On any given repetition of the choice task a complete preference ordering is established and $\mathbf{U}$ defined in the manner of (1), but this ordering may change on successive repetitions. Given the potential variability in the ordering, we now speak of the 'probability' of choosing a particular alternative $\mathbf{x}_n \in C$.

The theoretical apparatus of RUM offered simplicity and intuition, and subsequent researchers recognised its practical potential, especially McFadden (1968, but unpublished until 1975), who was responsible for the pioneering application to public policy analysis. An interesting feature of this practice, and a particular motivation for the present paper, is that different researchers – even within economics – have applied the RUM framework to the analysis of three distinct (but not necessarily mutually exclusive) sources of randomness in utility, namely:

1. Randomness in preference orderings; as motivated the original work of Block and Marschak (1960).
2. Randomness in outcomes; associated with the quantity of attribute $x_{nj}$ arising under different events.
3. Randomness in attribute tastes, associated with heterogeneity in tastes towards the $x_{nj}$ across a population of individuals.

Against this background, the primary contribution of our paper is to seek reconciliation between recent advancements in working specifications of RUM and the economic theory of individual choice underpinning the initial propositions of Block and Marschak. More specifically, the paper will comprise three strands.

First, the paper will begin by outlining a general framework for the modelling of individual discrete choice under risk. In particular, we shall seek to reconcile the three distinct sources of randomness; in preference orderings, in outcomes, and in attribute tastes; with the Neo-Classical economic theory that underpins RUM. The extant literature would seem to show a lack of clarity and/or lack of consensus in identifying and attributing distinct sources of randomness, and we shall therefore endeavour to deliver an authoritative account of such matters, drawing upon the perspectives of the most significant contributors to the literature.

Second, the paper will apply this framework to the analysis of journey time risk, a policy issue of particular pertinence to transport economists, and an interesting digression from the usual interest of economists in money risk. The paper will identify desirable properties of models of individual discrete choice under journey time risk, drawing contrast to properties embodied by models used in transport economic practice.

Third, the paper will present an empirical application to journey time risk, articulating the aforementioned three sources of randomness in RUM. Arising from this application, the paper will draw lessons for the specification of transport economic models, so as to realign practice with Neo-Classical theory.

## 2. Probabilistic models of discrete choice, and the 'three sources of randomness'

Having introduced our typology of different sources of randomness, the following section will formalise this in terms of econometric specification.

### 2.1. Randomness in preference orderings

Applying (1) under conditions of certainty, convention is to specify random utility:

$$U(\mathbf{x}_n) = V(\mathbf{x}_n) + \varepsilon_n \quad \text{for all } n \in C \tag{2}$$

where $V(\mathbf{x}_n)$ is referred to as deterministic or systematic utility (which the analyst considers 'observable'), and $\varepsilon_n$ is a random error term (which also affects $U(\mathbf{x}_n)$ but is considered by the analyst to be 'unobservable' or latent, and 'partially' independent of $\mathbf{x}_n$).

Block and Marschak (1960) are reasonably explicit in their interpretation of $\varepsilon_n$, as deriving from intra-individual variation in the preference ordering. It is interesting to contrast this perspective with McFadden's (1968, 1975) pioneering application of RUM to public policy analysis, which re-interprets $\varepsilon_n$ as deriving from inter-individual variation across a population of decision-makers. Irrespective of which perspective is adopted, we can substitute for $U(\mathbf{x}_n)$ in (1) using (2), arriving at the probability statement:

$$P(\mathbf{x}_n|C) = \Pr\{V(\mathbf{x}_n) + \varepsilon_n > V(\mathbf{x}_m) + \varepsilon_m\} \quad \text{for all } m \in C, \quad m \neq n$$

As is widely understood, different specific forms of RUM arise from different assumptions on the distribution of $\varepsilon_n$. For example, logit arises from the assuming that $\varepsilon_n$ for all $n$ are marginal Gumbel, and probit from assuming that they

instead follow a Normal distribution. It might be noted that, in applying distributional assumptions to $\varepsilon_n$, utility mutates from an ordinal construct (note Block and Marschak's reference to 'increasing monotone transformation' in (1)) to a cardinal one. Batley (2008) devotes particular attention to this point.

### 2.2. Randomness in outcomes

Thus far we have restricted attention to conditions of certainty relating actions and outcomes. Irrespective of whether or not choice is itself probabilistic, consideration of risk introduces a distinct dimension of randomness pertaining to the outcomes under different events. Even though the latter proposition is a *sine qua non* of the Journal, it will be helpful to again formalise matters, since this will assist the discussion that follows in Section 3. Retaining a discrete perspective on the choice problem, let $E=\{e_1,\ldots,e_K\}$ be a finite and exhaustive set of mutually exclusive 'events'. Furthermore, let $E$ be associated with a vector $\mathbf{w}_n=(\mathbf{x}_{n1},\ldots,\mathbf{x}_{nK};p_{n1},\ldots,p_{nK})$, which is referred to as a 'prospect', and denotes the probability $p_{nk}$ that each event $e_k\in E$ will occur, together with the attribute vector (or in Lancaster's parlance the alternative) that arises under each such event. Therefore, within this framework, $\mathbf{x}_n$ is subject to variability, where this variability could emanate from one or more of the attributes $x_{nj}\in\mathbf{x}_n$; the present paper will focus particularly upon variability in journey time.

With reference to the event probabilities $p_{nk}$, we note that some authors, notably Keynes (1921, 1936) and Knight (1921), distinguish between risk and uncertainty, describing the former as situations where probability is known to the individual decision-maker, and the latter as situations where probability is unknown. In what follows, we shall in most instances use the term risk, although this will essentially be a presentational convenience; the framework is sufficiently general to admit either risk or uncertainty.

Having introduced the notion of a prospect, let us now redefine the finite choice set $\tilde{C}$ in terms of these prospects (i.e. $\tilde{C}=\{\mathbf{w}_1,\ldots,\mathbf{w}_N\}$). The seminal exposition of von Neumann and Morgenstern (1947), referred to henceforth as vN&M, introduced supplementary axioms on the above definitions, relating to preference orderings over prospects, rules for combining prospects, and rules for combining probabilities. If we accept these axioms, then we arrive at the proposition that, under risk, the individual will choose the prospect $\mathbf{w}_n\in\tilde{C}$ which maximises his or her expectation of utility across the events $e_k\in E$. Analogous to the case of deterministic choice under certainty introduced in Section 1, a complete (weak) preference ordering on $\tilde{C}$ can be represented by a real-valued 'expected utility' function $Y$, such that for any pair of prospects:

$$\text{If } \mathbf{w}_n\geq\mathbf{w}_m \quad \text{then } Y(\mathbf{w}_n)\geq Y(\mathbf{w}_m)$$

where $Y(\mathbf{w}_n)$ is the expected utility of prospect $\mathbf{w}_n$, and is itself given by

$$Y(\mathbf{w}_n)=\textstyle\sum_{k=1}^{K}p_{nk}U(\mathbf{x}_{nk}) \quad \text{for all } n\in\tilde{C} \tag{3}$$

where $U(\mathbf{x}_{nk})$ is the utility associated with event $e_k$ of prospect $\mathbf{w}_n$, and where basic probability rules apply; that is, $p_{nk}\geq 0$ for $k=1,\ldots,K$, $n=1,\ldots,N$, and $\sum_{k=1}^{K}p_{nk}=1$.

Within this analysis we shall refer to $U(\mathbf{x}_{nk})$ as 'von Neumann and Morgenstern' utility, highlighting the fact that it embodies distinct properties, not least cardinality. In what follows, our paper will, to the extent that it considers risk, adhere to the expected utility paradigm. This paradigm remains the conventional representation of choice under risk in transport economics, and will adequately support the exposition of our modelling framework. The framework would however be amenable to generalisation, should there be interest in accommodating departures from expected utility maximisation.

Combining the two sources of randomness discussed above, associated with preference orderings and outcomes, now consider an individual faced with a repeated choice task under risk. On any given repetition of the choice task, he or she orders the prospects in $\tilde{C}$ in terms of their expected utility, but on successive repetitions this ordering may show variability. Following Marschak et al. (1963), the probability of choosing a prospect $\mathbf{w}_n\in\tilde{C}$ can be expressed as RUM, such that

$$P(\mathbf{w}_n|\tilde{C})=\Pr(Y(\mathbf{w}_n)>Y(\mathbf{w}_m) \quad \forall m\in\tilde{C},\ \ m\neq n) \tag{4}$$

where $(Y(\mathbf{w}_1),\ldots,Y(\mathbf{w}_N))$ is a random vector $\mathbf{Y}$ that is unique up to an increasing monotone transformation, and where the usual rules of probability apply.

Following from the above distributional assumptions, a difficulty emerges in applying (2) and (3) to Marschak et al.'s (1963) RUM under risk (4). In principle, there is an appealing theoretical coherence between (2) and (3), in the sense that both embody some notion of cardinality. In practice however, the combination of (2) and (3) would impose restrictions on the distribution of $\varepsilon_n$. Of particular note in this regard is logit – perhaps the most commonly applied member of the RUM class – since the summation of Gumbel variates would not itself be Gumbel. Therefore, if (2) were specified as logit, then application to (4), via (3), would not necessarily maintain the logit formulation. Whilst we are aware of little if any explicit discussion in the literature, a number of researchers (e.g. Michea and Polak, 2006; Batley et al., 2007) would appear to have acknowledged such restrictions, implementing instead the following specification:

$$P(\mathbf{w}_n|\tilde{C})=\Pr(Z(\mathbf{w}_n)>Z(\mathbf{w}_m) \quad \forall m\in\tilde{C},\ \ m\neq n) \tag{5}$$

where

$$Z(\mathbf{w}_n)=Y(\mathbf{w}_n)+\varepsilon_n=\textstyle\sum_{k=1}^{K}p_{nk}V(\mathbf{x}_{nk})+\varepsilon_n$$

The re-specification (5) is far from innocuous, since it provokes the question of how one can most appropriately represent the interface between the two sources of randomness introduced thus far (in preference orderings, and in the outcomes under different events). Whilst the specification (5) is dictated by issues of tractability, this does not however mean that it necessarily offers the most accurate statement of behaviour. In passing, we note that probit would be less restrictive in this regard, since it could be specified either as (4) or as (5), and further generality still would be permitted through the use of non-parametric methods.

Interestingly, Loomes and Sugden (1995) highlight ostensibly the same issue of specification, albeit with slightly different focus and interpretation from the present paper. Loomes and Sudgen categorise (5) as a Fechner model (Fechner, 1859) under risk, as distinct from RUM under risk (4). Furthermore, they assign different behavioural interpretations to the random error $\varepsilon_n$ terms of each model, with the Fechner model (4) embodying 'processing error' and RUM (5) embodying 'inherent variability'. Although the present paper will not digress to consider the basis of Loomes and Sugden's interpretations, we close the present discussion with the conclusion that several authors have, from both theoretical and empirical perspectives, questioned the same feature of Marschak et al.'s (1963) specification. That feature is the specification of the random error term within the probabilistic model of discrete choice under risk.

### 2.3. Randomness in attribute tastes

Developing the discussion further, let us now consider the advent of random parameters models, which potentially introduce a third dimension of randomness into the analysis. Although proposed some years ago by the likes of Cardell and Dunbar (1980), such models have over the last 10 years been formalised (e.g. McFadden and Train, 2000), and applied widely. Returning to Lancaster's (1966) representation of an alternative in terms of its attributes (introduced in Section 1) we note the traditional practice of specifying deterministic (indirect) utility simply as a linear-in-parameters function of these attributes:

$$V_n = \boldsymbol{\alpha}\mathbf{x}_n \quad \text{for all } n \in C$$

where $\boldsymbol{\alpha}$ is a vector of 'taste' parameters to be estimated.

The random parameters interpretation, by contrast, admits the possibility of taste variation across the repeated choices of a given individual and/or across the choices of different individuals within a population:

$$V_{nr} = \boldsymbol{\alpha}_r\mathbf{x}_n \quad \text{for all } n \in C, \ \ r \in R \tag{6}$$

where $\boldsymbol{\alpha}_r$ is a vector of parameters relating to these variables for repetition/individual $r$.

The randomness arises from interpreting $\boldsymbol{\alpha}_r$ as the realisation of distributions representing taste variation both across individuals and across repeated choices for a given individual. Hence, if we denote these distributions by $\boldsymbol{\eta}$ and parameterise them by $(\boldsymbol{\mu}, \boldsymbol{\Omega})$, we can re-express (6) thus

$$V_{nr} = \boldsymbol{\eta}_r(\boldsymbol{\mu}, \boldsymbol{\Omega})\mathbf{x}_n$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ are vectors (or matrices) of parameters to be estimated and where $\boldsymbol{\eta}_r$ denotes the realisation of distributions $\boldsymbol{\eta}$. An interesting justification for this parameterisation is offered by Bates and Terzis (1997).

### 2.4. The three sources of randomness

At this juncture, let us summarise the preceding discussion, by distinguishing between the three distinct sources of randomness within our model framework. First, randomness in the preference ordering, whether intra-individual or inter-individual, which is represented in terms of the random error term $\varepsilon_n$. Second, randomness in outcomes under risky events, which is represented in terms of the expectation $Y(\mathbf{w}_n)$. Third, inter-individual randomness in attribute tastes, which is represented in terms of the distribution $\boldsymbol{\eta}$ of taste parameters across the repetitions/individuals $R$. Whilst this trinity identifies distinct sources of randomness within RUM, it is worth acknowledging that some practical model specifications could potentially combine one or more such sources. Indeed we shall illustrate such a specification in the empirical application of Section 4.

Before closing this section, it is appropriate to draw reference to existing papers, notably those by Batley and Daly (2004), Liu and Polak (2006) and de Palma and Picard (2005), since these make similar distinctions between sources of randomness embodied by discrete choice models. These existing papers appear to be broadly in agreement with the present paper on the first and second sources of randomness, albeit with slightly different interpretation. Both Liu and Polak and de Palma and Picard attribute the first source to 'factors unobservable to the modeller but observable to the agent', and the second source to its converse. Whilst the latter interpretation of the first source of randomness would not seem unduly controversial (and is faithful to the interpretation offered by Manski (1977)), it does not go as far as the present paper (or indeed Marschak (1960) and Block and Marschak (1960)) in attributing all randomness to a single individual agent. As regards the second source of randomness, we would not necessarily support the proposition that this is unobservable to the agent, since this would seem to restrict the analysis to the context of uncertainty (where the event probabilities would be unknown) as distinct from risk (where the event probabilities would be known). Finally, whilst implicit in de Palma and Picard, attribute tastes are not explicitly identified as a distinct source of randomness in any of the previous papers. The present paper seeks

to avoid any ambiguities by specifying three distinct dimensions of randomness within the econometric model, and attributing each such dimension to a particular feature of behaviour.

## 3. Random outcomes in the time domain

In the course of the previous two sections of the paper, we have outlined a framework for the probabilistic modelling of individual discrete choice, and identified three distinct sources of randomness which this model could feasibly embody. We shall now focus the discussion particularly upon the application of this model to the second source of randomness, namely randomness in outcomes. Whereas the usual concern of economists in this regard is variability in monetary outcomes, here we shall consider issues that arise when applying our model to the alternative domain of journey time variability.

### 3.1. Journey time variability

Journey time variability refers to the observation that, when undertaking a given journey on different occasions, a traveller may experience variability in journey time. In this context, we are concerned primarily with random (e.g. as might be caused by an unforeseen incident) as opposed to systematic (e.g. as might be reflected by a longer journey time during peak hours relative to off-peak) variability. Our interest in journey time variability is motivated by current policy challenges facing the transportation sector. These challenges are common to many countries (e.g. see HEATCO (2006) for a review of official methods of valuing and appraising journey time variability in different EU states), although the present paper will focus upon a case study from the UK. From the UK perspective, The Eddington Study (Eddington, 2006) of transport's role in the economy will have a lasting impact on the way in which transport investment is appraised. Among other issues, Eddington identified 'reliability' (i.e. reduced journey time variability) as a significant benefit typically missing from current appraisal methods.

### 3.2. From money risk to time risk

Having identified the behavioural phenomenon of interest, we can progress our analysis by drawing an analogy between the transportation planner's notion of reliability and the economist's notion of risk. That is to say, we develop an analysis of individual choice behaviour under journey time variability (or 'time risk'), as distinct from the emphasis of the economic literature on money risk. Notwithstanding this important distinction, we can straightforwardly apply the usual theoretical conventions devoted to the analysis of money risk, postulating that, in the face of journey time variability, the individual traveller will choose the travel option that maximises his or her expected utility. This is broadly the starting point for Bates et al. (2001) and Noland and Small (1995), which are notable in offering detailed account of the theory underlying the valuation of reliability. The subsequent discussion will not seek to emulate such contributions; instead we offer an intuitive overview of the theory and its correspondence with our modelling framework.

The vast body of experimental evidence on money risk indicates a prevalent behaviour of risk aversion. On this basis, Fig. 1 displays the standard presentation – albeit couched in terms of model specification (5) – that risk aversion implies concavity[3] of the (deterministic) vN&M utility $V(M)$ with respect to monetary gains $M$. If, for simplicity, and still with reference to Fig. 1, we define a prospect over a pair of events ($i$ and $j$), i.e. $\mathbf{w} = (M_i, M_j;\ p_i, p_j)$, then we can derive expected utility as the chord joining the utilities at $M_i$ and $M_j$. As is well established, risk aversion results in the property that, at the expected monetary outcome $E(M)$, the expected utility of the prospect is less than the utility of the expected monetary outcome, i.e. $Y(\mathbf{w}) < V(E(M))$.

Methods for analysing reliability might be seen as analogous, but with risk manifesting in terms of journey time rather than wealth. Let us develop this analogy by re-expressing the prospect vector $\mathbf{w} = (T_i, T_j; p_i, p_j)$, where 'pay-offs' are now defined on journey time $T$. An interesting research question is whether risk aversion similarly prevails; if it does then, acknowledging that wealth is 'good' but journey time is 'bad',[4] this would basically amount to a re-orientation of Fig. 1, as shown in Fig. 2. That is to say, the individual maintains the property of risk aversion, but now in relation to journey time.

### 3.3. Mean-variance model

Deriving from work in portfolio analysis (Tobin, 1958, 1965; Markowitz, 1959), a common practical simplification is to assume that the expected utility defined in (3) can be approximated by the first and second moments of the utility distribution over the $K$ pay-offs. This offers an exact approximation under only two situations (Borch, 1969; Feldstein, 1969); either utility is quadratic, or the distribution of pay-offs is Normal. In fact, neither situation is likely to hold in practical transportation contexts, but this has failed to deter the widespread adoption of the so-called 'mean-variance' model for the valuation of reliability. Despite the terminology 'mean-variance', it has become commonplace (albeit with debatable

[3] $\partial^2 V(M)/(\partial M)^2 \leq 0$.

[4] See Bates and Whelan (2001) for a more formal diagrammatic treatment of the exchange between wealth and journey time, in the context of the value of journey time.

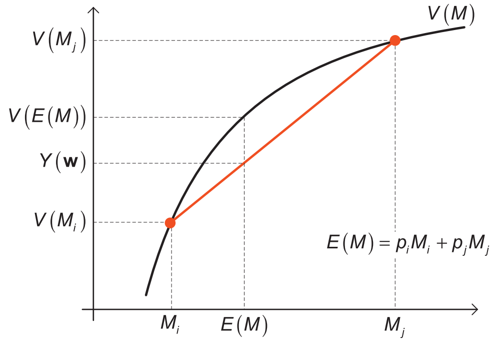


**Fig. 1.** Properties of the vN&M utility function ($V$) under aversion to money risk.

justification; see the discussion on this point in ITS Leeds et al. (2008)), for the reliability literature to measure the second moment using the standard deviation of the pay-offs rather than the variance, specifying the following approximation to expected utility[5]:

$$Y_n = \delta \overline{T}_n + \varphi \sigma_n \quad \text{for all } n \in \tilde{C} \tag{7}$$

where $\overline{T}_n$ is the mean journey time (for alternative $n$ across events $e_k \in E$), $\sigma_n$ is the standard deviation of journey time (for alternative $n$ across events $e_k \in E$) and $\delta, \varphi$ are parameters to be estimated.

A frequently cited metric in transportation policy is the so-called 'reliability ratio', defined with reference to (7) as the ratio $\varphi/\delta$, which gives the marginal rate of substitution between the expected pay-off (or in this case, when all events are taken to be equally probable, the mean) and the inherent risk (standard deviation). Journey time variability could potentially influence various choices of the traveller (e.g. departure time, mode, route, destination, transport operator), and the mean-variance model would appear sufficiently general to accommodate these various dimensions.

### 3.4. Scheduling model

Moving on from the mean-variance model, which offers a generic approximation to expected utility, let us now turn to an alternative approach devised specifically for the valuation of reliability. The basis for this approach is the so-called 'scheduling model' (Vickrey, 1969; Small, 1982), which is framed around an interest in how travellers choose their departure time when seeking to arrive at their destination at a given point in time. According to the scheduling model, the utility for a particular departure time is functional on four components; journey time ($T$), 'schedule delay early' ($SDE$), 'schedule delay late' ($SDL$), and a 'lateness' dummy variable ($D$) that is set to unity for strictly positive $SDL$. The latter three components are conditioned by the notion of a 'Preferred Arrival Time' ($PAT$[6]), as follows:

- Journeys arriving before the $PAT$ are deemed to be 'early'. In this case, the $SDE$ is derived as the difference between the $PAT$ and the actual arrival time (i.e. the number of minutes of earliness at destination), and both $SDL$ and the lateness dummy variable $D$ are null, i.e. for alternative $n$:

$$\text{If } T_n \le T(PAT) \text{ then } \begin{cases} SDE_n = T(PAT) - T_n \\ SDL_n = 0 \\ D_n = 0 \end{cases}$$

- Journeys arriving after the $PAT$ are deemed to be 'late'. In this case, the $SDL$ is derived as the difference between the actual arrival time and the $PAT$ (i.e. the number of minutes of lateness at destination), $D$ is unity, and $SDE$ is null, i.e. for alternative $n$:

$$\text{If } T_n > T(PAT) \text{ then } \begin{cases} SDE_n = 0 \\ SDL_n = T_n - T(PAT) \\ D_n = 1 \end{cases}$$

[5] Henceforth we adopt some brevity in our notation, omitting explicit reference to the attribute or pay-off vectors, and writing $U_n$ for $U(\mathbf{x}_n)$, and $Y_n$ for $Y(\mathbf{w}_n)$, etc.

[6] The travel time that, for a given departure time, delivers the traveller at his/her preferred arrival time is represented by $T(PAT)$.

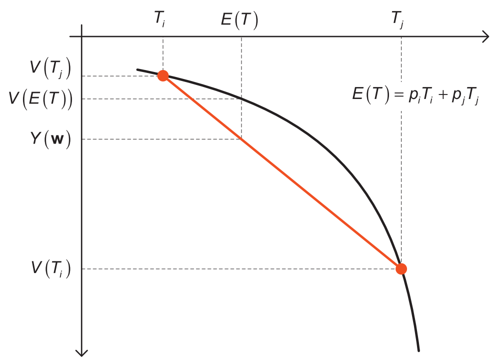


**Fig. 2.** Properties of the vN&M utility function ($V$) under aversion to time risk.

Arising from the two conditional statements above, the utility of alternative $n$ is usually specified as linearly additive in journey time, schedule delay early, schedule delay late, and the lateness dummy variable, thus

$$V_n = \phi T_n + \gamma SDE_n + \eta SDL_n + \kappa D_n \quad \text{for all } n \in \tilde{C} \tag{8}$$

where $\phi, \gamma, \eta, \kappa$ are parameters to be estimated.

An important point, often overlooked, is that Vickrey (1969) and Small (1982) restrict their interests to departure time choice under certainty. It was not until Small's later work with Noland (Noland and Small, 1995) that the scheduling model was extended to accommodate journey time variability, taking expectations of (8) over the journey time distribution (i.e. across events $e_k \in E$) thus

$$Y_n = \phi' E(T_n) + \gamma' E(SDE_n) + \eta' E(SDL_n) + \kappa' E(D_n) \quad \text{for all } n \in \tilde{C} \tag{9}$$

where $\phi'$ distinguishes from $\phi$ applying under certainty (8), and so on for the other parameters.

Having admitted variability in journey time, it is instructive to consider the functional form of (8), which might now be rationalised as the (deterministic) vN&M utility within (9). This is illustrated by Fig. 3, which shows how, for a given departure time, the vN&M utility function $V(T)$ originates at the minimum (free flow) journey time, and increases linearly with journey time, save for a discontinuity at the journey time associated with the *PAT*. Experimental evidence (e.g. Bates et al., 2001) suggests that $V(T)$ carries a negative slope throughout, but is steeper for $T > T(PAT)$ than for $T < T(PAT)$. Noting these properties, the utility function of Fig. 3 might be rationalised as a linear piecewise approximation to the ('true'[7]) continuous concave function of Fig. 2, implying the same property of aversion to journey time risk. Indeed, this notion of approximating the continuous function by means of a discontinuous function is the basic rationale behind recent work (Fosgerau and Karlström, 2009) seeking to relate the scheduling model to the mean-variance model; we will cover this point in greater detail in Section 4 below. Of course, in the special cases where the journey time 'pay-offs' are always early or always late, i.e. $T_i, T_j \le PAT$ or $T_i, T_j > PAT$, the scheduling model will exhibit localised risk neutrality; this point is discussed by Batley (2007).

### 3.5. Mean lateness model

The mean-variance and scheduling models are by far the predominant approaches to the valuation of reliability, but other approaches exist; see de Palma and Picard (2005) for a comprehensive discussion. The scope of the present paper will extend to a third and final model, since this forms the basis of the subsequent empirical analysis of Section 4. The so-called 'mean lateness' model is laid down by The Passenger Demand Forecasting Handbook or 'PDFH' (Section B4; Association of Train Operating Companies, 2005) as the standard approach for valuing rail reliability in the UK. Whilst claiming to adhere to the expected utility paradigm, the mean lateness model is distinct from the previous two models in specifying the pay-off dimension as lateness at destination ($L$) relative to scheduled journey time (*SchedT*) rather than actual journey time. Reconciling with journey time, lateness at destination would be given by the identity $L = T - SchedT$. This focus on lateness rather than journey time reflects the importance of lateness as a performance metric within the regimes governing the provision of rail infrastructure and the operation of franchised rail services in the UK.

More formally, the mean lateness model arises from an assertion that expected utility can be approximated thus

$$Y_n = \lambda GJT_n + \mu \overline{L^+}_n \quad \text{for all } n \in \tilde{C} \tag{10}$$

[7] 'True' to the extent that a plethora of empirical studies (admittedly based on money risk rather than time risk) have established the predominance of a continuous concave function.

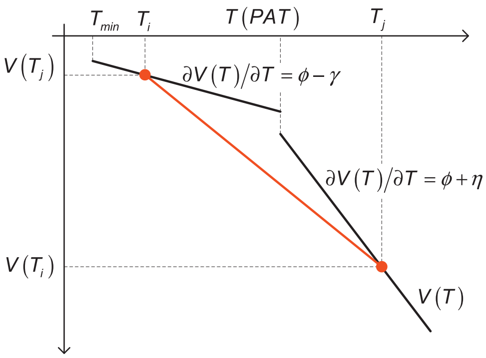


**Fig. 3.** Properties of the vN&M utility function ($V$) within the scheduling model.

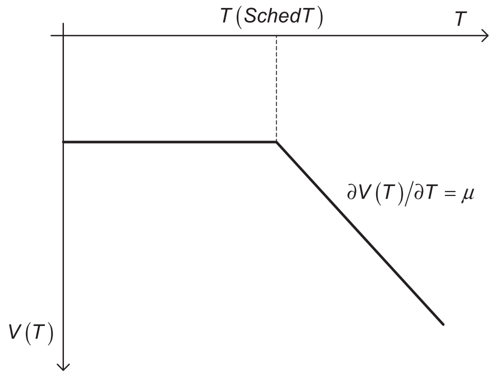


**Fig. 4.** Properties of the vN&M utility function ($V$) within the mean lateness model.

where $GJT$ is generalised journey time, $\overline{L^+}$ is the mean lateness at destination (again across events $e_k \in E$), $\lambda$, $\mu$ are parameters to be estimated, and $GJT$ is itself given by

$$GJT = SchedT + H + I$$

where $SchedT$ is total scheduled station-to-station journey time including interchange time, $H$ is a service interval penalty related to frequency of services, and $I$ is the sum of interchange penalties for the interchanges involved.

Note that, in practice, PDFH discounts the relevance of early arrivals (or 'negative lateness') at destination, such that the model considers only $L \geq 0$; hence the notation $L^+$ in (10). For simplicity, let us further assume that our journey of interest is subject to neither service interval penalties nor interchange penalties (i.e. $H = I = 0$), in which case (10) specifies $Y_n$ as a linear additive function of scheduled journey time and mean lateness at destination relative to timetable. PDFH guidance acknowledges (albeit only as a postscript) that some journeys may exhibit significant journey time variability, recommending that, in such cases, (10) be supplemented with the standard deviation of lateness. In practice, the latter specification is rarely used, and almost all analyses of rail reliability in the UK adhere strictly to (10). Indeed, such analyses focus considerable attention on the so-called 'lateness multiplier' $\mu/\lambda$, which gives the trade-off between mean lateness at destination and scheduled journey time. PDFH recommends a standard lateness multiplier of 3, but acknowledges possible variability by market segment. For example, larger multipliers are cited for long distance high speed journeys (6.1 for full fare and season; 4.2 for restricted tickets) and airport journeys (6.5 for all tickets), and a lower multiplier of 2.5 for all other services. PDFH suggests that these more disaggregated multipliers be interpreted as sensitivity analyses around the standard multiplier of 3.

As before, let us examine the properties of the vN&M utility within (10), through reference to Fig. 4. Note that, for expositional purposes, we again express this function in terms of journey time $f(T)$ rather than lateness. This facilitates easy comparison against the previous figures, and exposes its key properties, as follows. On first inspection of the vN&M utility function, it might appear that the mean lateness model is similar to the scheduling model in offering a piecewise linear approximation to Fig. 2. In contrast to the scheduling model, however, the function is referenced against the scheduled arrival time at destination rather than the $PAT$ of the traveller (although both reference points could of course coincide; we develop this point in Section 4.2.4). Empirical evidence suggests that $f(T)$ is steeper for $T > T(SchedT)$. Before drawing any inferences regarding attitudes to time risk, it is however important to amplify our earlier remark regarding the convention of PDFH to discount any earliness. The effect of this convention is to render the function undefined throughout the interval $[0, T(SchedT)]$. Moreover, since $f(T)$ is defined only for $T > T(SchedT)$, and where defined is linear in $T$, the implication follows that travellers are not risk averse to journey time, but rather risk neutral.

## 4. An empirical application to journey time risk

The purpose of this section is to demonstrate the empirical application of a probabilistic model of discrete choice under risk, particularly with a view to illustrating the conceptual interests considered in the previous two sections, namely the 'three sources of randomness', and the specification of the vN&M utility function in the context of journey time risk. To these ends, we exploit data from a recent project investigating the impact of reliability on passenger rail demand (Batley et al., 2008a, 2008b), which was conducted by the Institute for Transport Studies (ITS) at the University of Leeds for the UK Department for Transport (DfT). The primary contribution of discrete choice modelling to the aforementioned project was to deliver fresh evidence on the lateness multiplier (i.e. $\mu/\lambda$ from (10)), as well as segmentation by journey distance and purpose. The conceptual interests of the present paper did not however fall within the scope of the DfT project, and the interpretations and assertions that follow should therefore be taken as representing our own independent views and not the views of DfT.

Our source of evidence on the lateness multiplier was a self-completion mail-back questionnaire comprising two parts. The first part featured questions on the prior experiences of travellers concerning reliability, and invited retrospective reporting of their responses to changes in reliability. The second part involved a Stated Preference (SP) experiment. Whilst SP is almost *de rigueur* to the discrete choice modelling community, two factors justify our use of SP in the present context. First, SP is readily applicable to large scale field surveys of subjects (in this case, rail travellers), thereby allowing the generation of datasets sufficiently rich to support estimation of the econometric models detailed in Section 2 above. Second, SP offers a convenient vehicle for experimentation in the field as opposed to the laboratory, facilitating the recruitment of subjects who have personal experience of the policy issue of interest (in this case, rail reliability), and (perhaps more significantly) recruitment in sufficient volumes to cover segmentations relevant to policy.

Against these justifications, we should readily acknowledge a significant weakness of SP in the present context, which is the difficulty of appropriately conveying the concept of journey time risk to subjects. Indeed, the recent scoping study for the Dutch value of journey time and reliability (de Jong et al., 2007) devoted considerable effort to testing a range of survey presentations. Whilst we will not digress to consider such matters in depth, we accept that this is a challenging area, and note that our own study employed a derivative of Hollander's (2006) presentation, which might itself be seen as a simplification of the so-called 'clock face' presentation employed by Bates et al. (2001).

### 4.1. Experimental design

With reference to Fig. 5, our SP experiment offered a pair of journey options, each described in terms of fare, scheduled journey time, and the distribution of actual journey time over five repetitions of the journey. All attributes refer to a single leg of the journey. Thus, in the notation of Sections 2 and 3, $N=2$ and $K=5$. Since we did not advise subjects as to the probabilities of these five events, one might (on the face of it) characterise the prospect as uncertain rather than risky. However, for purposes of modelling (discussed subsequently in Section 4.2) we assumed that the five events were equally probable, thereby promoting ready application to the mean lateness model (Section 3.5). Whereas Hollander's (2006) presentation expressed the journey time distribution in terms of departure and arrival times *per se*, here we relate departure and arrival times to the timetable. This permits explication of any earliness or lateness, with reference to timetabling at the boarding and destination stations. Note however that, within our SP experiment, trains were never presented as departing early from the boarding station (since this occurrence would be unrealistic). Having presented a pair of services in this manner, subjects were invited to choose between the pair. Each subject was issued with five different SP experiments.

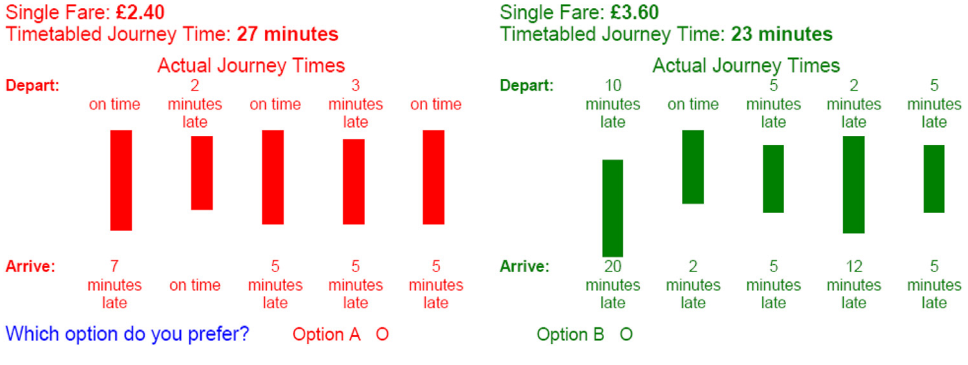


**Fig. 5.** Specimen SP experiment.

The contribution of lateness at boarding, as distinct from lateness at destination, can be explained in the context of the following identity (private correspondence with John Bates):

$$SchedT = T - L^{+} + L^{-} + B \tag{11}$$

Thus we represent scheduled journey time (*SchedT*) as the summation of in-vehicle journey time ($T$) and lateness at the boarding station ($B$), adding or abstracting any earliness ($L^{-}$) or lateness ($L^{+}$) at destination as appropriate. All elements in (11) are non-negative.

Fig. 6 illustrates identity (11) for one potential event that could appear in any of the options (A or B) in Fig. 5 and which would entail a scheduled journey time (*SchedT*) of 50 min. Assuming a scheduled departure time of 9:00, the train actually departs at 9:10, incurring lateness at boarding ($B$) of 10 min. The train then accumulates an additional 10 min of delay en route, arriving at the final destination at 10:10, and incurring lateness at destination ($L$) of 20 min.

The SP experiment was based on a 'difference' design, as follows. Having researched fare and timetabled journey time for a single trip on the origin-to-destination (O–D) of interest (see Table 1), we specified three levels of lateness in departure from the boarding station (in fixed blocks of five departure times), with each level embodying a mean and standard deviation. Different values were specified for options A and B, although in each instance the first level was characterised by the mean journey time and a standard deviation of zero (i.e. reliable). Variation in journey time around the timetabled journey time was similarly specified at three levels, again in fixed blocks of five. The sum of this variation in journey time and variation in lateness at the boarding station yielded the arrival time at destination. In essence, therefore, we constructed four variables (fare, timetabled journey time, departure time variation and journey time variation).

Once designed, we subjected the SP experiment to simulation-based testing, so as to ensure that an acceptable range of parameter ratios (e.g. pertaining to the 'lateness multiplier' $\mu/\lambda$ in Section 3.5 above) could be recovered. The experiment was further tested by means of a pilot survey, which involved the distribution of 600 questionnaires to rail travellers using Huddersfield station in the North of England on 7 December 2006. The returns from this pilot survey informed some minor refinements to the questionnaire, and gave us confidence to proceed to our field survey.

In designing the field survey, we invited the advice of key stakeholders (government, infrastructure provider, and train operating companies) from the UK rail industry on potential study locations which exhibited either, or both, of the following characteristics:

- where travellers had experience of changing levels of reliability, and
- where travellers had a choice between services offering different levels of reliability.

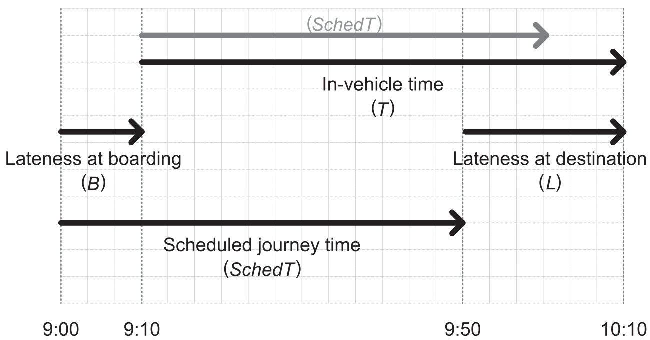


**Fig. 6.** Illustrating the time components of a rail journey.

**Table 1**
Origin–destinations of interest, by journey distance.

| Long distance | Short distance |
|---|---|
| Bristol Temple Meads–London | Brighton–London |
| Leeds–Birmingham | Edinburgh–Glasgow |
| Leeds–London | Leeds–Sheffield |
| Swindon–London | Peterborough–London |
| | Portsmouth–London |
| | Reading–London |
| | Stevenage–London |
| | Woking–London |

On this basis, it was decided that the field survey should be targeted at twelve specific O–D journeys, involving ten survey locations (the origin stations) across England and Scotland (Table 1). These twelve journeys were characterised by substantial passenger volumes, offered some variety in context (e.g. short and long distances), whilst together covering the principal operator categories (i.e. InterCity, Regional and Network South East). Having finalised this target sample, the SP experiment was customised to each O–D within the sample. The field survey was conducted in two waves; Wave 1 taking place on 7 and 8 February 2007, and Wave 2 on 28 February 2007. In all cases, we surveyed on weekdays between the hours of 6 am and 12 noon, thereby covering a range of peak and off-peak services. At each survey location (origin station), travellers were intercepted on the departing platform for the journey of interest, and invited to participate in our survey.

We distributed around 15,000 questionnaires in total, and achieved a response rate of just over 19%; this level of response is fairly typical of pen and paper surveys conducted at transport termini. Whilst we might have achieved a higher response from on-board surveys, these were not readily practicable, since our SP experiment was tailored to specific O–Ds (rather than intermediate journeys) and we were keen to achieve as many contacts as possible at the origin point. Those issues notwithstanding, Section 4.2 will show that the dataset proved sufficiently rich to yield significant estimates of all key parameters and, more generally, to illustrate the theoretical features of model specification, which is the primary focus of our paper.

With reference to Table 2, commuting traffic accounted for around half of the sample, business for around a third, and the remaining proportion of around an eighth was accounted for by leisure traffic. Two-thirds of our sample were categorised as short distance travellers, and one-third long distance. Table 2 cross-tabulates the data from Waves 1 and 2 by journey purpose and distance, reflecting the segmentations of primary interest to DfT (the sponsor of the study).

### 4.2. Empirical models

Given the five repetitions of the SP experiment for each individual, and a response rate of 19%, we were left, following cleaning, with 11,763 observations (from 2395 rail travellers) for modelling. The following discussion reports four models, all of which were implemented within the general apparatus of (5), specifying the random error term to be Gumbel (independent and identically distributed across options), and estimated by maximum likelihood.

#### 4.2.1. Model 1: randomness in preference orderings

Our starting point was to model a relatively simple but faithful representation of the SP experiment, whilst (broadly) adhering to the mean lateness model promoted by DfT. In subsequent discussion, we will generalise this specification, thereby responding to some of the weaknesses of the mean lateness model previously noted (Section 3.5). More specifically, we estimated the following approximation to expected utility:

$$Y_n = \lambda SchedT_n + \nu F_n + \pi \overline{B}_n + \mu \overline{L^+}_n \quad \text{for } n = a, b \tag{12}$$

where $n$ refers to the two options ($a$ and $b$) that individuals choose between each time (see Fig. 5), $SchedT_n$ is scheduled journey time, $F_n$ is fare for a single leg of a return journey, $\overline{B}_n$ is mean lateness at the boarding station (calculated across the five events comprising option $n$), $\overline{L^+}_n$ is mean (positive) lateness at the destination station (also calculated across the five events), and $\lambda$, $\nu$, $\pi$, $\mu$ are parameters to be estimated.

Again with reference to Section 3.5, and recalling that we have previously assumed no service interval penalties and no interchange penalties (i.e. $H = I = 0$), we note that the mean lateness model (10) is nested within (12). The distinguishing features of (12), as compared with (10), are the inclusion of fare and mean lateness at boarding. We include fare because of our interest in deriving monetary valuations of the time variables. As regards lateness at boarding, an implication of (11) is that the mean lateness model (10) already subsumes lateness at boarding within scheduled journey time, albeit under an assumption that a minute of lateness at boarding carries the same disutility as a minute of in-vehicle journey time. By introducing an explicit lateness at boarding variable to (12), we can test whether this assumption is defensible.

Drawing reference to the theoretical discussion of Section 3 more generally, an important feature of (12) is that it does not admit the possibility of early arrival at destination (i.e. $L^- = 0$). As of consequence, (12) will be undefined throughout the interval $[0, T(SchedT)]$, in effect implying a linear vN&M utility function, and a property of neutrality to journey time risk

**Table 2**
Cross-tabulation of journey purpose and distance travelled.

| Purpose | Distance | | |
|---|---|---|---|
| | Long (%) | Short (%) | Total (%) |
| Business | 20.3 | 16.1 | 36.4 |
| Commute | 12.8 | 37.9 | 50.6 |
| Leisure | 4.7 | 8.3 | 12.9 |
| Total | 37.8 | 62.2 | 100.0 |

(see Fig. 4). Moreover, reconciling with the three sources of randomness identified in Section 2, we interpret Model 1 as embodying randomness in preference orderings, but non-randomness in both outcomes and attribute tastes.

With reference to Table 3a, Model 1 demonstrates reasonable explanatory power, with an adjusted rho-squared with respect to constants of 0.24. Since all four attributes are 'bads', one would expect all coefficients to be negative, and this is indeed the case. Furthermore, all coefficients are significantly different from zero, implying that each attribute impacts materially upon choice. Of particular interest is the finding that lateness at boarding is significant and negative; this suggests that a minute of lateness at boarding is worth more than a minute of in-vehicle journey time.

Following from the discussion of Section 3, it is instructive to consider several parameter ratios deriving from Model 1, as given by Table 3b. Noting that Model 1 (and indeed all subsequent models) expresses fare in UK pounds and journey time in minutes, we can derive a value of scheduled journey time of 35.32 pence/min. Whilst this might seem rather high compared with the values of time reported in PDFH (Association of Train Operating Companies, 2005), it should be remembered that our survey was focussed upon business travellers and commuters. We will in subsequent models segment the sample of travellers, bringing deeper insight to this valuation.

Turning to the trade-off between lateness and scheduled journey time, we derive two lateness multipliers, applying to the boarding and destination stations. The lateness multiplier at destination is faithful to the PDFH definition described in Section 3; Model 1 yields a multiplier of 3.96, slightly higher than the current PDFH recommendation of 3. The lateness multiplier at boarding, at 2.41, suggests that late running at boarding incurs less disutility than lateness at destination. The latter result perhaps reflects the potential for recovery of some or all of the lateness at boarding in the course of the journey.

### *4.2.2. Model 2: randomness in preference orderings and outcomes*

Extending Model 1, we now introduce the standard deviation of journey time. Referring back to Section 3 and our distinction between the non-linear vN&M utility function of the mean-variance model and the linear vN&M utility function of the mean lateness model, the intention of Model 2 is to synthesise the two approaches. The mean lateness model (10) continues to be nested within Model 2, but we now admit the possibility of randomness in preferences through the standard deviation term, as follows:

$$Y_n = \lambda SchedT_n + \theta\sigma_n + \nu F_n + \pi\overline{B}_n + \mu\overline{L^+}_n \quad \text{for } n = a, b$$

**Table 3a**
Estimates from Models 1, 2 and 3.

| Variable name | Coefficient | | Estimate (*t*-ratios in brackets) | | |
|---|---|---|---|---|---|
| | | | **Model 1** | **Model 2** | **Model 3** |
| Fare ($F$) | Mean | $\nu$ | −0.0911 (−18.25) | −0.0892 (−18.20) | −0.3474 (−11.91) |
| | St. Dev. | $\upsilon$ | | | −0.2902 (−9.74) |
| Scheduled journey time ($SchedT$) | Mean | $\lambda$ | −0.0322 (−19.47) | −0.0369 (−20.51) | −0.0890 (−13.00) |
| | St. Dev. | $\rho$ | | | −0.1271 (−15.63) |
| Mean (positive) lateness at destination ($\overline{L^+}$) | Mean | $\mu$ | −0.1272 (−45.58) | −0.1033 (−28.12) | −0.3236 (−16.60) |
| | St. Dev. | $\xi$ | | | −0.2134 (−12.96) |
| Mean lateness at boarding ($\overline{B}$) | Mean | $\pi$ | −0.0774 (−17.45) | −0.0634 (−13.21) | −0.1109 (−8.46) |
| | St. Dev. | $\omega$ | | | −0.1822 (−8.82) |
| Standard deviation of journey time ($\sigma$) | Mean | $\theta$ | | −0.0573 (−8.96) | −0.1847 (−9.27) |
| | St. Dev. | $\tau$ | | | −0.2079 (−8.10) |
| - LL_final | | | 5682.66 | 5641.44 | 4808.44 |
| - LL_ASC | | | 7491.47 | 7491.47 | 7491.47 |
| - LL_zeros | | | 8153.49 | 8153.49 | 8153.49 |
| Rho-sq_ASC | | | 0.24 | 0.25 | 0.36 |
| Rho-sq_zeros | | | 0.30 | 0.31 | 0.41 |
| No. individuals | | | 2395 | 2395 | 2395 |
| No. observations | | | 11,763 | 11,763 | 11,763 |
| No. pseudo-random draws | | | | | 3000 |

**Table 3b**
Ratios from Models 1, 2 and 3.

| Metric | Coeff. | **Model 1** | **Model 2** | **Model 3 (Mean)** | **Model 3 (Median)** |
|---|---|---|---|---|---|
| Value of time [pence/min] | $\lambda/\nu$ | 35.32 | 41.38 | 25.62 | 18.55 |
| 'Reliability ratio' | $\theta/\lambda$ | – | 1.55 | 2.07 | 0.85 |
| Lateness multiplier_DESTINATION | $\mu/\lambda$ | 3.96 | 2.80 | 3.64 | 1.62 |
| Lateness multiplier_BOARDING | $\pi/\lambda$ | 2.41 | 1.72 | 1.25 | 0.50 |

where $\sigma_n$ is the standard deviation of in-vehicle journey time for option $n$ (calculated across the five events comprising each option $n$), and $\theta$ is a further parameter to be estimated.

With reference to Table 3a, we can see that the additional parameter is significant and negative, whilst maintaining the significance of the existing parameters and the overall fit of the model. Of particular interest, given the focus of the present study, is the so-called 'reliability ratio'. The reliability ratio, as defined here, is a slight departure from the conventional definition given in Section 3, in that the denominator is based on scheduled journey time rather than actual journey time. We will return to this point when discussing Model 4. Bates et al. (2001) note that, in the context of rail travel, previous studies have reported reliability ratios within the range 2–10, but comment that results at the lower end of this range would seem more credible. With reference to Table 3b, Model 2 yields a reliability ratio of 1.55, just outside the lower limit of that range. Whereas the correlation between the two lateness variables is relatively minor (around 0.3), Model 2 embodies a more significant correlation between lateness at boarding and the standard deviation of journey time. As can be seen from Table 3b, this causes some disturbance to the lateness multipliers, with both showing a decrease in magnitude relative to Model 1.

### *4.2.3. Model 3: randomness in preference orderings, outcomes and attribute tastes*

In terms of the 'three sources of randomness', let us now complete the picture by introducing randomness in attribute tastes, via the specification:

$$Y_{nr} = \eta_r^1(\lambda,\rho)SchedT_n + \eta_r^2(\theta,\tau)\sigma_n + \eta_r^3(\nu,\upsilon)F_n + \eta_r^4(\pi,\omega)\overline{B}_n + \eta_r^5(\mu,\xi)\overline{L^+}_n$$

for options $n=a,b$, and all individuals $r\in R$, where $\lambda,\rho,\theta,\tau,\nu,\upsilon,\pi,\omega,\mu,\xi$ are parameters to be estimated and $\eta_r^d(\mu,\Omega)$ are the values for individual $r$ arising from the distribution $d$ of tastes represented by parameters $\mu$ and $\Omega$.

That is to say, we now represent each of the parameters from Model 2 as randomly distributed across the population of $R$ individuals, but constant for all observations from a given individual $r$. Although we experimented with various distributional assumptions, the reported Model 3 specifies all parameters as Normally-distributed. This model was estimated by maximum simulated likelihood (Revelt and Train, 1998) using software developed by Nicolás Ibáñez and Richard Connors at ITS Leeds. The estimation employed 3000 draws per individual generated through Marsaglia's ziggurat algorithm (Marsaglia and Tsang, 1984).

Perusal of the $t$-ratios in Table 3a reveals that all standard deviations of the coefficients $(\rho_r,\tau_r,\upsilon_r,\omega_r,\xi_r)$ are significantly different from zero, suggesting that all of the SP design variables carry randomly-distributed coefficients. We do not here endeavour to dissect and apportion the precise sources of this randomness, but it is likely that this is a manifestation of both the repeated observations phenomenon and 'intrinsic' taste variation. Whilst specifying the random parameters as Normal typically brings some convenience in tractability, an implication is that the distribution is unbounded. It is therefore instructive to derive ratios at both the mean and the median of the estimated parameters (Table 3b), interpreting these with reference to the distribution plots in Figs. 7–10. Note that, in the case of a distribution of the ratio of two Normal variables, neither moment of the distribution is defined, although the median can always be calculated. We therefore implemented Marsaglia's (1965) formula and calculated CDFs by numerical integration. It might also be remarked that, in taking ratios, we assume independence between the two parameters involved.

With reference to Fig. 7, Model 3 yields a mean value of scheduled journey time of 25.62 pence/min, against a median of 18.55 pence/min; this indicates a slight positive skew in the PDF. The PDF shows considerable spread around these measures

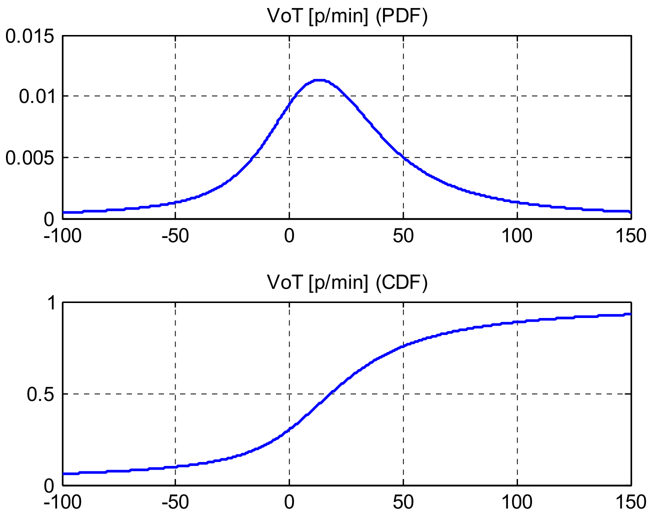


**Fig. 7.** Distribution plot of value of time from Model 3.

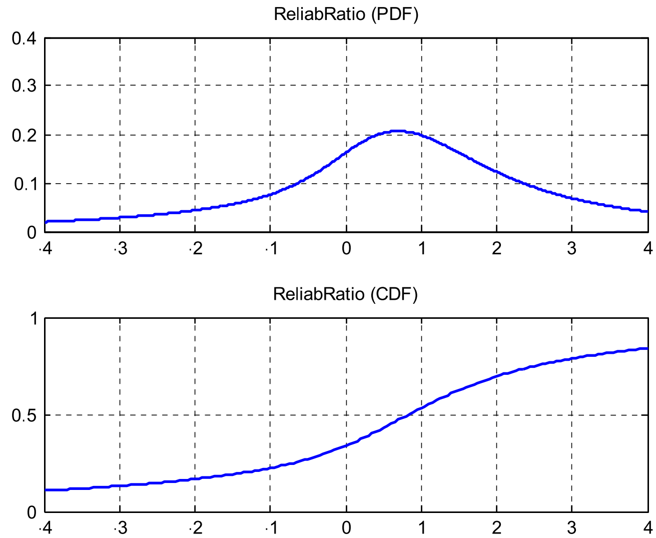


**Fig. 8.** Distribution plot of reliability ratio from Model 3.

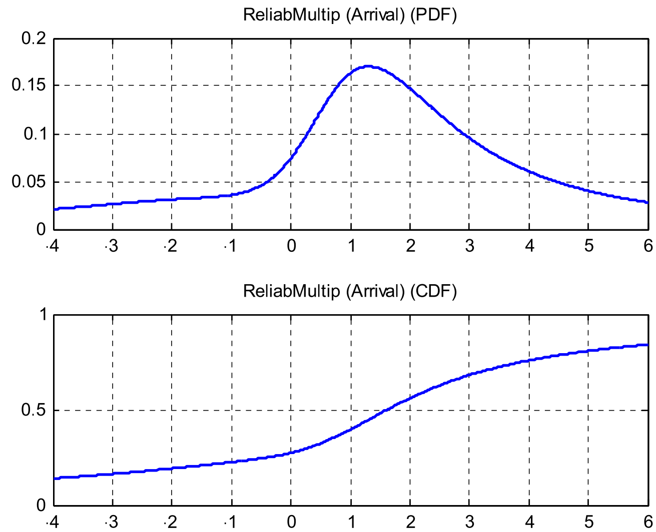


**Fig. 9.** Distribution plot of lateness multiplier at destination from Model 3.

of central tendency. The reliability metrics show a more marked (positive) skew than the value of scheduled journey time. The mean reliability ratio is estimated at 2.07, which compares with a median of 0.85 (see Fig. 8). The lateness multiplier at destination yields a mean of 3.64 and a median of 1.62 (see Fig. 9); the analogous statistics for boarding are 1.25 and 0.50 (see Fig. 10). Inspection of the associated CDF plots reveals that a significant proportion of the implied ratios carry a negative sign. Whilst many researchers reject negative valuations of time as theoretically invalid, one might note Ibáñez and Batley's (2009) rationale defending such valuations. The sign of the reliability ratio is less controversial, and may indeed yield useful interpretation. In portofolio analysis, a negative trade-off between mean and standard deviation is taken to indicate risk preference, whereas a positive trade-off indicates risk aversion. Whilst this proposition is conventionally couched in terms of money risk, it would not seem unreasonable to postulate that an analogous relationship holds for journey time risk.

#### *4.2.4. Model 4: synthesising the mean-variance and mean lateness models*

Our final model pursues a slightly different tack, by seeking to reconcile an apparent inconsistency between the model specifications officially prescribed for analyses of reliability on UK rail and road. That is to say, whereas rail represents pay-offs in terms of lateness relative to timetable at destination (and implies risk neutrality), road represents pay-offs in terms of in-vehicle journey time (and implies risk aversion). More specifically, for the valuation of reliability on road, WebTAG unit 3.5.7 (Department for Transport, 2010) prescribes the mean-variance model (7) and the headline metric of the reliability ratio, citing a standard value of the reliability ratio for car of 0.8. In what follows, we establish a simple form

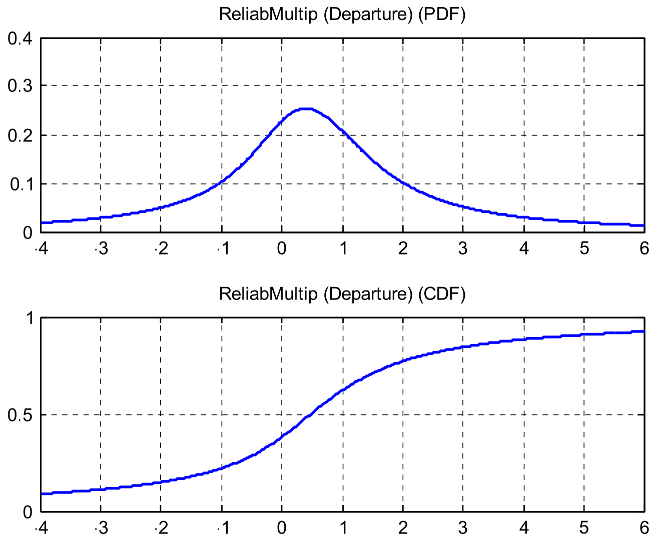


**Fig. 10.** Distribution plot of lateness multiplier at boarding from Model 3.

of correspondence between the valuations of reliability on rail and road, and in so doing harmonise the 'currencies' of risk considered by the two analyses.

Exploiting the identity (11), we can conveniently translate an analysis based on *SchedT* and $L^+$ (i.e. faithful to the mean lateness model) into one based on $T$ (i.e. faithful to the mean-variance model) by accounting for any lateness at boarding and earliness at destination. More specifically, we approximate expected utility as follows:

$$Y_n = \lambda SchedT_n + \theta\sigma_n + \nu F_n + \pi\overline{B}_n + \mu\overline{L^+}_n + \psi\overline{L^-}_n \quad \text{for } n = a, b \tag{13}$$

where $\overline{L^-}$ is mean earliness at the destination station, and $\psi$ is an additional parameter to be estimated.

Note that we include standard deviation, so as to accommodate possible aversion to journey time risk. Substituting for *SchedT*, as per the definition (11), in (13), we then have the facility to reinterpret (13) in terms of in-vehicle journey time and the various lateness variables, as follows:

$$Y_n = \lambda T_n + \theta\sigma_n + \nu F_n + (\lambda + \pi)\overline{B}_n + (\mu - \lambda)\overline{L^+}_n + (\lambda + \psi)\overline{L^-}_n \quad \text{for } n = a, b \tag{14}$$

In seeking to synthesise the mean-variance and mean lateness models, it is appropriate to draw reference to the recent work of Fosgerau and Karlström (2009). Building upon earlier contributions by Noland and Small (1995) and Bates et al. (2001), Fosgerau and Karlström establish a formal equivalence between the mean-variance and scheduling models, which holds for any fixed standardised (absolutely continuous) travel time distribution. This basically amounts to a proposition that, within the expected utility function, the standard deviation term of the mean-variance method is suitably approximated by the expected schedule delay early and schedule delay late terms of the scheduling model. In what follows, we in effect subject this proposition to empirical testing.

If we recall the modelling assumption that the five events within each prospect are equi-probable, and further assume that the traveller's *PAT* aligns with the scheduled arrival time of the train, these simplifications together imply that: $\overline{L^-}_n = E(SDE_n)$ and $\overline{L^+}_n = E(SDL_n)$. By including both $\overline{L^-}_n$ and $\overline{L^+}_n$ together with the standard deviation term $\sigma_n$ in the same model, we can test whether the scheduling terms capture all of the inherent randomness in outcomes, as postulated by Fosgerau and Karlström, or whether there remains some residual randomness, to be captured by the standard deviation term.[8]

We implemented (13) as a multinomial logit specification, in this case segmenting by journey distance and purpose (reflecting the interests of our sponsor, DfT). Model 4 might thus be seen as embodying randomness in preference orderings and outcomes, but non-randomness in attribute tastes. With reference to Table 4a, all segmentations are significantly different from zero, with the only exception of earliness at destination for short distance leisure. In contrast to lateness, which is regarded as a 'bad', we can see that rail travellers derive positive utility from early arrival at destination. In terms of explanatory power, Model 4 improves upon Models 1 and 2, but is inferior to Model 3.

[8] As Fosgerau and Karlström themselves acknowledge, the equivalence result relies upon a continuous departure time dimension, and does not (strictly speaking) extend to discrete departure times, as would typically apply to public transport. In the present paper we overlook this restriction; one way of rationalising this would be to assume that we are dealing with a very high frequency rail service. Moreover, in seeking to align methods for valuing reliability on rail and road, further practical complications arise, not least in establishing a consistent starting point for the journey (whether that be the home or the rail station) and recognising that the notion of lateness at boarding does not readily apply to road. We will however leave such complications for future work.

**Table 4a**
Estimates from Model 4.

| Variable | | Coefficient | Estimate | $t$-ratio |
|---|---|---|---|---|
| Fare_SB | $F$ | $\nu$_SB | −0.1330 | −6.60 |
| Fare_SC | | $\nu$_SC | −0.2119 | −10.98 |
| Fare_SO | | $\nu$_SO | −0.1101 | −4.47 |
| Fare_LB | | $\nu$_LB | −0.0767 | −10.73 |
| Fare_LC | | $\nu$_LC | −0.0731 | −8.87 |
| Fare_LO | | $\nu$_LO | −0.0860 | −5.84 |
| Scheduled Journey Time_SB | *SchedT* | $\lambda$_SB | −0.0515 | −8.68 |
| Scheduled Journey Time_SC | | $\lambda$_SC | −0.0493 | −11.80 |
| Scheduled Journey Time_SO | | $\lambda$_SO | −0.0240 | −3.25 |
| Scheduled Journey Time_LB | | $\lambda$_LB | −0.0433 | −15.19 |
| Scheduled Journey Time_LC | | $\lambda$_LC | −0.0294 | −7.83 |
| Scheduled Journey Time_LO | | $\lambda$_LO | −0.0296 | −5.55 |
| Mean Lateness at Destination_SB | $\overline{L^+}$ | $\mu$_SB | −0.1382 | −15.11 |
| Mean Lateness at Destination_SC | | $\mu$_SC | −0.1535 | −23.12 |
| Mean Lateness at Destination_SO | | $\mu$_SO | −0.1245 | −10.68 |
| Mean Lateness at Destination_LB | | $\mu$_LB | −0.0770 | −15.97 |
| Mean Lateness at Destination_LC | | $\mu$_LC | −0.0589 | −11.01 |
| Mean Lateness at Destination_LO | | $\mu$_LO | −0.0524 | −6.61 |
| Mean Earliness at Destination_SB | $\overline{L^-}$ | $\psi$_SB | 0.2930 | 2.09 |
| Mean Earliness at Destination_SC | | $\psi$_SC | 0.4325 | 4.33 |
| Mean Earliness at Destination_SO | | $\psi$_SO | 0.1968 | 1.12 |
| Mean Earliness at Destination_LB | | $\psi$_LB | 0.2303 | 3.59 |
| Mean Earliness at Destination_LC | | $\psi$_LC | 0.2663 | 3.43 |
| Mean Earliness at Destination_LO | | $\psi$_LO | 0.2922 | 2.51 |
| Mean Lateness at Boarding_SB | $\overline{B}$ | $\pi$_SB | −0.0674 | −4.56 |
| Mean Lateness at Boarding_SC | | $\pi$_SC | −0.0874 | −8.11 |
| Mean Lateness at Boarding_SO | | $\pi$_SO | −0.0608 | −3.09 |
| Mean Lateness at Boarding_LB | | $\pi$_LB | −0.0437 | −6.43 |
| Mean Lateness at Boarding_LC | | $\pi$_LC | −0.0377 | −3.85 |
| Mean Lateness at Boarding_LO | | $\pi$_LO | −0.0381 | −2.70 |
| Standard Deviation of Journey Time | $\sigma$ | $\theta$ | −0.0644 | −10.32 |
| -LL_final | | | 5449.89 | |
| -LL_ASC | | | 7491.50 | |
| -LL_zeros | | | 8153.50 | |
| Rho-sq_ASC | | | 0.27 | |
| Rho-sq_zeros | | | 0.33 | |
| No. individuals | | | 2395 | |
| No. observations | | | 11,763 | |

N.B.: S—short distance, L—long distance; B—business purpose, C—commuting purpose, and O—other purpose.

Focussing particularly on the lateness at destination variable, it is useful to make some observations regarding differences across segments. First, in the case of short distance journeys, the marginal disutility of (positive) lateness at destination for commuting is significantly greater than for both business and leisure. Second, in the case of long distance, the marginal disutility of (positive) lateness at destination for business is significantly greater than for leisure. Third, across all journey purposes, the marginal disutility of (positive) lateness at destination is significantly greater for short distance relative to long; that is to say, the marginal disutility of a minute of lateness diminishes with scheduled journey time.

Turning to the implied ratios, we consider, for reasons of brevity, simply the reliability ratios and lateness multipliers at destination (Table 4b). The first column presents the lateness multipliers at destination deriving from (13). If we then re-state the model as (14), we can derive reliability ratios faithful to the conventional definition, in terms of in-vehicle time rather than in terms of scheduled journey time; these are given in the second column. However, in re-stating the model in terms of in-vehicle time rather than scheduled journey time, this provokes a re-calibration of the lateness multipliers, as given in the third column. In this way, we illuminate the nature of the relationship between the reliability metrics employed on UK road and rail, demonstrating the implications of anchoring a model on scheduled journey time as distinct from in-vehicle journey time, and vice versa.

Last but not least, we return to the earlier discussion concerning Fosgerau and Karlström (2009). Whilst a degree of covariance between estimates was evident, this did not undermine our ability (with the exception of mean earliness for one segment) to derive significant estimates for *both* the scheduling terms and the standard deviation. We therefore conclude that, in the case of Model 4, the underlying vN&M utility function is highly non-linear, and not fully specified by the linear piecewise function of the scheduling model (i.e. the scheduling terms do not adequately proxy for the standard deviation). This finding does not dispute the validity of Fosgerau and Karlström *per se*, but does expose the fact that their result is a theoretical one which, depending on the properties of the vN&M utility function, may or may not be defended empirically.

**Table 4b**
Reliability metrics from Model 4.

| Segment | Reliability ratio ($\theta/\lambda$) (based on in-vehicle time) | Lateness multiplier_DESTINATION ($\mu/\lambda$) (based on scheduled journey time) | Lateness multiplier_DESTINATION ($(\mu-\lambda)/\lambda$) (based on in-vehicle time) |
|---|---|---|---|
| SB | 1.25 | 2.68 | 1.68 |
| SC | 1.31 | 3.12 | 2.12 |
| SO | 2.69 | 5.19 | 4.19 |
| LB | 1.49 | 1.78 | 0.78 |
| LC | 2.19 | 2.00 | 1.00 |
| LO | 2.18 | 1.77 | 0.77 |

N.B.: S—short distance, L—long distance; B—business purpose, C—commuting purpose, and O—other purpose.

## 5. Summary and conclusions

Nearly 50 years have elapsed since Marschak (1960) and Block and Marschak (1960) proposed the concept of the Random Utility Model (RUM), as a probabilistic analogue to the Neo-Classical economic theory of individual choice. McFadden's (1968, but unpublished until 1975) pioneering application to public policy analysis demonstrated the practical potential of RUM, paving the way for the plethora of applications which have followed. Looking back on this history, half a decade later, our paper sought reconciliation between recent advancements in working specifications of RUM and the economic theory of individual choice underpinning the initial propositions of Marschak and Block. The paper developed three principal strands of discussion.

First, we outlined a general framework for the modelling of individual discrete choice under risk, reconciling three distinct sources of randomness in RUM; in preference orderings (as described by Block and Marschak (1960)), in outcomes (as implicit in the notion of risk), and in attribute tastes (in the context of mixed logit); with the Neo-Classical economic theory that underpins RUM. A particular objective of this discussion was to articulate precisely what is meant by a model of probabilistic choice under risk.

Second, the paper applied this theoretical modelling framework to the context of journey time risk (or 'reliability'), a policy issue of particular pertinence to transport economists. We identified desirable properties of models emanating from the framework, drawing contrast to the properties embodied by models used in practical studies of reliability. This discussion yielded a number of observations. From an econometric perspective, we noted that the interface between randomness in preference orderings and randomness in outcomes implied restrictions upon the distributional form of the random error term within RUM; this issue has been alluded to and/or accommodated by various authors, but has never been explicitly addressed. From a theoretical perspective, we noted the allegiance of the reliability literature to von Neumann and Morgenstern's (1947) paradigm of expected utility maximisation, and devoted particular attention to the properties of the vN&M utility function underpinning expected utility. We distinguished between three alternative formulations of that function; the so-called mean-variance, scheduling and mean lateness models. The first two formulations dominate the academic literature on the economics of reliability, whilst the third formulation is the official approach of the UK rail industry. We noted that whereas the mean-variance and scheduling models embody the property of risk aversion in journey time, as would seem theoretically desirable, the mean lateness model embodies risk neutrality.

The third strand of the paper presented an empirical application of the modelling framework to journey time risk, articulating the aforementioned three sources of randomness in RUM, and synthesising the mean-variance and mean lateness representations of the vN&M utility function. Our data was taken from a recent Stated Preference (SP) experiment, which collected data from 2395 rail travellers choosing between a pair of services on the basis of fare, scheduled journey time and journey time variability. We developed our modelling framework incrementally, each time adding one of the three sources of randomness. The most general model involved a random parameters specification, and offered a means by which we could parameterise randomness across preference orderings, outcomes, and attribute tastes. By combining these different sources of randomness, we were able to derive some interesting insights, notably the distribution of attitudes to journey time risk across our sample of rail travellers. More specifically, we estimated a mean value of scheduled journey time of 25.62 pence/min, against a median of 18.55 pence/min. We further estimated a mean reliability ratio of 2.07, against a median of 0.85. Since the distribution of the reliability ratio exhibited a marked (positive) skew, we were able to infer a predominant behaviour of aversion to journey time risk.

Our final model was something of a digression, illuminating an apparent inconsistency in the methods conventionally applied to the valuation of reliability on UK road and rail. In particular, we exposed the implications of specifying a model on scheduled journey time (namely the mean lateness model) as distinct from in-vehicle journey time (the mean-variance model), and vice versa. We also drew parallels with the recent work of Fosgerau and Karlström (2009), which presented a formal equivalence between the mean-variance and scheduling models. This notion of equivalence amounts to a proposition that, within the expected utility function, the standard deviation term of the mean-variance model is approximated by the expected schedule delay early and late terms of the scheduling model. Although a degree of covariance was evident, our own empirical application yielded significant estimates of *both* the scheduling terms and the standard deviation within the

same model. We conclude that, in our application at least, the underlying vN&M utility function is highly non-linear, and not fully specified by the linear piecewise function of the scheduling model (i.e. the scheduling terms do not adequately proxy for the standard deviation). This finding does not call into question the validity of Fosgerau and Karlström's work, but does expose the fact that their notion of equivalence is a theoretical one; depending on the properties of the vN&M utility function, equivalence may or may not hold empirically.

## Acknowledgements

We are grateful for the assistance of the UK Department for Transport (DfT) in allowing us to analyse data collected as part of a recent project investigating the impact of reliability on passenger rail demand (for more details see Batley et al., 2008a, 2008b). That said, the analysis reported in this paper falls outside the scope of the aforementioned project, and the interpretations and assertions offered should therefore be taken as representing our own independent views and not the views of DfT.

## References

Association of Train Operating Companies, 2005. Passenger Demand Forecasting Handbook. Association of Train Operating Companies, London.
Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. Transportation Research Part E 37, 191–229.
Bates, J., Terzis, G., 1997. Stated preference and the 'ecological fallacy'. In: Proceedings of the European Transport Conference 1997, Cambridge, UK. URL: ⟨http://abstracts.aetransport.org/paper/index/id/665/confid/3⟩.
Bates, J., Whelan, G., 2001. Size and Sign of Time Savings. Working Paper 561. Institute for Transport Studies, University of Leeds.
Batley, R., 2007. Marginal valuations of travel time and scheduling, and the reliability premium. Transportation Research Part E 43 (4), 387–408.
Batley, R., 2008. On ordinal utility, cardinal utility and random utility. Theory & Decision 64 (1), 36–63.
Batley, R., Daly, A., 2004. Departure time choice under uncertainty: a model of random utility maximisation for wagers. In: Proceedings of the European Transport Conference 2004, Strasbourg, France. URL: ⟨http://abstracts.aetransport.org/paper/index/id/1897/confid/10⟩.
Batley, R., Ibáñez, J.N., Wardman, M., Shires, J., Whelan, G., 2007. A discrete choice study to assess the impact of reliability on passenger rail demand. In: Proceedings of the European Transport Conference 2007, Noordwijkerhout, The Netherlands. URL: ⟨http://abstracts.aetransport.org/paper/index/id/2745/confid/13⟩.
Batley, R., Wardman, M., Ibáñez, J.N., Shires, J., Whelan, G., 2008a. Revealed Preference Study to Assess Impact of Reliability on Passenger Rail Demand. Deliverable 3: Report of Discrete Choice Analysis, Version 3.1. 24 October 2008.
Batley, R., Dargay, J., Wardman, M., 2008b. Revealed Preference Study to Assess Impact of Reliability on Passenger Rail Demand. Deliverable 4: Report of Econometric Analysis, Version 3.0. 31 August 2008.
Block, H.D., Marschak, J., 1960. Random orderings and stochastic theories of responses. In: Marschak, J. (Ed.), Economic Information, Decision and Prediction: Selected Essays, vol. 1. , D. Reidel, Dordrecht, pp. 172–217.
Borch, K., 1969. A note on uncertainty and indifference curves. Review of Economic Studies 36, 1–4.
Cardell, N.S., Dunbar, F.C., 1980. Measuring the societal impact of automobile downsizing. Transportation Research Part A 14 (5–6), 432–434.
Department for Transport, 2010. WebTAG Unit 3.5.7: Reliability. URL: ⟨http://www.dft.gov.uk/webtag/⟩.
Eddington, R., 2006. The Eddington Transport Study: Main Report. HM Treasury, London.
Fechner, G.T., 1859. In: Elemente der PsychophysikBreitkopf and Härtel, Leipzig.
Feldstein, M.S., 1969. Mean-variance analysis in the theory of liquidity preference and portfolio selection. Review of Economic Studies 36, 5–12.
Fosgerau, M., Karlström, A., 2009. The value of reliability. Transportation Research Part B 44 (1), 38–49.
HEATCO, 2006. Proposal for Harmonised Guidelines. HEATCO Deliverable 5 (2nd revision). IER, Stuttgart. URL: ⟨http://heatco.ier.uni-stuttgart.de/⟩.
Hollander, Y., 2006. Direct versus indirect models for the effects of unreliability. Transportation Research Part A 40 (9), 699–711.
Ibáñez, J.N., Batley, R.P., 2009. On the Duality Relations for Discrete-Continuous Demands. Working Paper. Institute for Transport Studies, University of Leeds.
ITS Leeds, Imperial College, John Bates Services, 2008. Multimodal Travel Time Variability. Final Report to the Department for Transport.
de Jong, G., Tseng, Y., Kouwenhoven, M., Verhoef, E., Bates, J., 2007. The Value of Travel Time and Travel Time Reliability. Survey design. Final Report for the Netherlands Ministry of Transport, Public Works and Water Management.
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.H., 1921. Risk, Uncertainty and Profit. Houghton Mifflin, Boston.
Lancaster, K.J., 1966. A new approach to consumer theory. Journal of Political Economy 74 (2), 132–157.
Liu, X., Polak, J.W., 2006. From random utility to random expected utility: theory and application to departure time choice. In: Proceedings of the European Transport Conference 2006, Strasbourg, France. URL: ⟨http://abstracts.aetransport.org/paper/index/id/2528/confid/12⟩.
Loomes, G., Sugden, R., 1995. Incorporating a stochastic element into decision theories. European Economic Review 39 (3–4), 641–648.
McFadden, D., 1968. The Revealed Preferences of a Government Bureaucracy. Department of Economics, University of California, Berkeley, California, unpublished.
McFadden, D., 1975. The revealed preferences of a government bureaucracy: theory. The Bell Journal of Economics and Management Science 6 (2), 401–416.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15 (5), 447–470.
Manski, C., 1977. The structure of random utility models. Theory and Decision 8, 229–254.
Markowitz, H.M., 1959. Portfolio Selection. Wiley, New York.
Marsaglia, G., 1965. Ratios of normal variables and ratios of sums of uniform variables. Journal of the American Statistical Association 60 (309), 193–204.
Marsaglia, G., Tsang, W.W., 1984. A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. SIAM Journal on Scientific and Statistical Computing 5, 349–359.
Marschak, J., 1960. Binary choice constraints and random utility indicators. In: Marschak, J. (Ed.), Economic Information, Decision and Prediction: Selected Essays, vol. 1. , D. Reidel, Dordrecht, pp. 218–239.
Marschak, J., Becker, G.M., DeGroot, M.H., 1963. Stochastic models of choice behavior. In: Marschak, J. (Ed.), Economic Information, Decision and Prediction: Selected Essays, vol. 1. , D. Reidel, Dordrecht, pp. 254–279.
Michea, A., Polak, J.W., 2006. Modelling risky choice behaviour: evaluating alternatives to expected utility theory. In: Proceedings of the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.
Neumann, J. von, Morgenstern, O., 1947. Theory of Games and Economic Behavior, Second Edition Princeton University Press, Princeton.
Noland, R., Small, K., 1995. Journey 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.

de Palma, A., Picard, N., 2005. Route choice decision under travel time uncertainty. Transportation Research Part A 39, 295–324.
Revelt, D., Train, K., 1998. Mixed logit with repeated choices of appliance efficiency levels. Review of Economics and Statistics 80 (4), 647–657.
Small, K., 1982. The scheduling of consumer activities: work trips. American Economic Review 72 (3), 467–479.
Tobin, J., 1958. Liquidity preference as behaviour towards risk. Review of Economic Studies 25, 65–86.
Tobin, J., 1965. The theory of portfolio selection. In: Hahn, F.H., Brechling, F.P.R. (Eds.), The Theory of Interest Rates, MacMillan for the IEA, New York, pp. 3–51.
Vickrey, W.S., 1969. Congestion theory and transport investment. American Economic Review 59 (2), 251–260.