Randomness in preferences, outcomes and tastes; an application to journey time risk

Richard Batley & J Nicolás Ibáñez
Institute for Transport Studies, University of Leeds, Leeds, UK

ABSTRACT

The Random Utility Model (RUM), as originally proposed by Block and Marschak (1960), specifies a probabilistic relation between choice and utility, within a paradigm of individual choice under certainty. According to this specification, utility is comprised of deterministic and random elements, with the latter reflecting variability in preferences over certain outcomes within and across individuals. Our paper exploits Marschak et al.’s (1963) extension to RUM, which admits the possibility of variability in the outcomes themselves, and specifies a probabilistic relation between choice and the expectation of utility across these outcomes. Emanating from Marschak et al.’s analysis is the notion of a Random Expected Utility Model (REUM), which might be seen as an analogue to RUM within a paradigm of individual choice under risk or uncertainty.

Although previous researchers have implemented various formulations of REUM, these have seemingly been rather ad hoc, with little acknowledgement of the distinction between RUM and REUM, and the implications of this distinction for model specification. A number of recent works (e.g. Batley & Daly, 2004; Michea & Polak, 2006; Polak et al., 2007) have been more instructive in this regard, and our paper begins by distilling these works and presenting a typology of methods. We build upon previous typologies by reconciling specific practical model specifications (e.g. logit, probit, random parameters logit) with different dimensions of randomness; in marginal utility (i.e. the notion of a ‘taste distribution’), in outcomes (i.e. the notion of ‘expected utility’), and in preferences (i.e. the notion of ‘random utility’).

The second stage of our paper distinguishes between different ‘currencies’ of risk. Whereas the vast literature on economic decision-making under risk or uncertainty is devoted almost entirely to variability in monetary outcomes, of more immediate concern to transport users is variability in journey time. The paper discusses various implementations of REUM in the context of journey time variability, including the so-called ‘mean-variance’, ‘scheduling’ and ‘mean lateness’ approaches. Drawing analogy with the economic literature on attitudes to risk (e.g. Pratt, 1964), we explicate the properties of the utility functions supporting these implementations. In particular, we observe that some implementations imply rather restrictive properties concerning travellers’ attitudes towards journey time risk. Moreover, we identify substantive discrepancies between the methods adopted as ‘standard’ on different transport modes.

Exploiting data from a recent Stated Preference (SP) study of rail reliability, we implement various REUM models, experimenting with different specifications of the underlying utility function (e.g. mean-variance) and different dimensions of risk (e.g. in terms of total journey time, in-vehicle journey time, or lateness). We show that Marschak et al.’s notion of REUM implies somewhat restrictive properties on error specifications, such that only a subset of model specifications can be entirely REUM-compliant. Our most general model involves a mixed logit specification, and offers the means by which we can parameterise each of the aforementioned sources of randomness; across outcomes, utilities and tastes. An important outcome from this model is insight into the distribution of attitudes to journey time risk across our population.
1. INTRODUCTION

As the International Choice Modelling conference convenes for its inaugural meeting in Harrogate, nearly 50 years will have passed since Marschak (1960) and Block & Marschak (1960) introduced the conceptual bedrock on which our community is formed. Marschak and Block observed the phenomenon that an individual, when presented with the same set of discrete alternatives on different occasions, might not always make the same choice. Their response was to propose the Random Utility Model (RUM), as a probabilistic analogue to the conventional Neo-Classical economic theory of choice. Randomness in RUM derives from the repetition of a discrete choice task; on any given repetition a complete preference ordering is established (and a utility function defined) in the manner of Neo-Classical theory, but preferences may show variability across repetitions.

The theoretical apparatus of RUM offered simplicity and intuition, and subsequent researchers recognised its practical potential. McFadden’s (1968, but unpublished until 1975) pioneering application to public policy analysis set in motion the research momentum which has supported 50 years of RUM practice, and culminates in our inaugural meeting. McFadden presented an alternative perspective on RUM, with randomness deriving not from the intra-individual variability in preferences, but rather from variability in preferences across a population of individuals. It was this shift in perspective that paved the way for the adoption of RUM as a policy analysis tool. McFadden’s analysis appealed to the focus of policymakers on markets rather than individuals, and was amenable to available data on revealed preferences across the individuals making up these markets.

Looking back on this history, half a decade hence, 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 Marschak and Block. The advancements to which we refer might be seen as arising from two observable trends within the decision-making environment. First and foremost, economic agents exhibit preferences within an increasingly uncertain world, subject to financial, political and environmental instability. Second, agents are becoming increasingly demanding in seeking goods and services bespoken to their personal attribute tastes, provoking ever more proliferation in product differentiation.

Whilst the latter trend has received considerable attention from the RUM community over the last 10 years, through the advent of mixed logit models, the former has received only limited attention. Our paper will seek to redress this balance of research effort by re-establishing risk and uncertainty as important methodological challenges for our community. In particular, we will consider an application to journey time risk, a policy issue of particular pertinence to transport economists, and an interesting digression from the more common interest of economists in money risk. More generally, our paper will seek to reconcile three distinct sources of randomness introduced above; in preferences (as described by Block and Marschak), attribute tastes (in the context of mixed logit) and outcomes (as implicit in the notions of risk and uncertainty); with the Neo-Classical economic theory that underpins RUM. Synthesising 50 years of development in RUM models, we will pose the question: ‘What is random about the Random Utility Model?’
We shall begin by introducing a typology of discrete choice models, which we categorise according to two dimensions: first, whether choice is made under certainty or uncertainty, and second, whether choice is deterministic or probabilistic.

2.1 Deterministic choice under certainty

The following analysis considers the notion of individual ‘discrete choice’. This is where an individual decision-maker is presented with a finite and exhaustive set of \( N \) alternatives \( C = \{x_1, \ldots, x_N\} \), from which he or she is invited to select their preferred alternative. Deferring to Lancaster’s (1966) representation of an alternative in terms of its attributes, we define an ‘alternative’ to be a vector \( x = (x_1, \ldots, x_N) \), where \( x_j \) is the quantity of attribute \( j \), \( x_j \geq 0 \) for all \( j \) and \( x_j > 0 \) for at least some \( j \). With reference to Block & Marschak (1960), the axioms of completeness and transitivity establish a complete (weak) preference ordering on \( C \), which can be represented by a real-valued ‘utility’ function \( U \). For any pair of alternatives within the ordering, the individual is then represented as choosing the alternative \( x_n \in C \) that maximises his or her utility, i.e.

\[
\text{If } x_n \succeq x_m \text{ then } U(x_n) \geq U(x_m)
\]  

(1)

2.2 Probabilistic choice under certainty

Extending the analysis from deterministic to probabilistic choice, we now admit the possibility that an individual, repeatedly undertaking the deterministic choice task described above, may not always make the same decision. In particular, consider the Random Utility Model (RUM), which was defined by Block & Marschak (1960) as follows:

\[
P(x_n|C) = Pr\{U(x_n) > U(x_m)\} \text{ for all } m \in C, m \neq n
\]  

(2)

where:

\( U = (U(x_1), \ldots, U(x_N)) \) is a random vector unique up to an increasing monotone transformation, \( P(x_n|C) \geq 0 \) for all \( n \in C \), and \( \sum_{n=1}^{N} P(x_n|C) = 1 \).

Thus utility is taken to be a random variable. On any given repetition of the choice task a complete preference ordering is established and \( U \) defined in the manner of (1), but this ordering may change on successive repetitions. Given the variability in preferences, we now speak of the ‘probability’ of choosing a particular alternative \( x_n \in C \).

2.3 Deterministic choice under risk or uncertainty

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 and uncertainty introduces a second and distinct dimension of randomness pertaining to the outcomes under different events. Formalising this proposition, again in discrete terms, 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 \( w = (w_{n1}, \ldots, w_{nK}; p_{n1}, \ldots, p_{nK}) \), 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 pay-off $w_k$ to the individual should that event indeed occur. The probability rules applying to mutually exclusive events hold, i.e. $p_{nk} \geq 0$ for $k=1,...,K$, $n=1,...,N$, and $\sum_{k=1}^{K} p_{nk} = 1$. Within this framework, Keynes (1921, 1936) and Knight (1921) distinguish between risk and uncertainty, describing the former as situations where probability $p_{nk}$ is known to the individual decision-maker, and the latter as situations where $p_{nk}$ is unknown. In what follows, we shall, in most instances, use the terms risk and uncertainty interchangeably, distinguishing between the two only as necessary.

Having introduced the notion of a prospect, we now redefine the finite choice set in terms of these prospects $\tilde{C} = \{w_1,\ldots,w_N\}$. The seminal exposition of von Neumann & Morgenstern (1947) 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 or uncertainty, the individual will choose the prospect $w_n \in \tilde{C}$ which maximises his or her expectation of utility across the events $e_k \in E$. Analogous to deterministic choice under certainty, a complete (weak) preference ordering on $\tilde{C}$ can be represented by a real-valued ‘expected utility’ function $Y$. Thus for any pair of prospects:

If $w_n \succeq w_m$ then $Y(w_n) \geq Y(w_m)$

where $Y(w_n)$ is the expected utility of prospect $w_n$, and is itself given by:

$$Y(w_n) = \sum_{k=1}^{K} p_{nk} U(w_{kn}) \text{ for all } n \in \tilde{C}$$

(3)

where $U(w_{kn})$ is the von-Neumann & Morgenstern ‘sub-utility’ deriving from pay-off $w_{kn}$.

2.4 Probabilistic choice under risk or uncertainty

Marrying the two previous analyses, now consider an individual faced with a repeated choice task under uncertainty. On any given repetition of the choice task, he or she orders $\tilde{C} = \{w_1,\ldots,w_N\}$ in terms of expected utility, but on successive repetitions this ordering may show variability. Following Marschak et al. (1963), the probability of choosing a prospect $w_n \in \tilde{C}$ can be expressed as RUM, such that:

$$P(w_n|\tilde{C}) = Pr\{Y(w_n) > Y(w_m)\} \text{ for all } m \in \tilde{C}, m \neq n$$

(4)

where $Y = (Y(w_1),\ldots,Y(w_N))$ is a random vector unique up to an increasing monotone transformation, and the usual rules of probability apply.
Having introduced our typology of discrete choice models, let us now consider the implementation of probabilistic models of choice, especially under risk or uncertainty. Let us begin our discussion, however, with the more familiar context of probabilistic choice under certainty, where convention is to specify random utility:

\[ U(n) = V(n) + \varepsilon(n) \quad \text{for all } n \in C \]  

where \( V(n) \) is referred to as deterministic utility (which the analyst considers ‘observable’), and \( \varepsilon(n) \) is a random error term (which also affects \( U(n) \) but is considered by the analyst to be ‘unobservable’ or latent, and ‘partially’ independent of \( x_n \)).

Block & 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 (5) for \( U(n) \) in (2), arriving at the probability statement:

\[ P(x_n | T) = Pr\{V(n) + \varepsilon(n) > V(m) + \varepsilon(m)\} \quad \text{for all } m \in C, m \neq n \]

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 independently and identically distributed (IID) as Extreme Value Type I (or Gumbel), and probit from assuming that they follow a Normal distribution instead. It might be noted that, in applying distributional assumptions to \( \varepsilon(n) \), utility mutates from an ordinal construct (note Block & Marschak’s reference to ‘increasing monotone transformation’ in (2)) to a cardinal one. Batley (2008) devotes particular attention to this point.

Following from these distributional assumptions, a difficulty however emerges in applying (3) and (5) to Marschak et al.’s (1963) RUM under risk or uncertainty (4). With reference to logit, arguably the most popular member of the RUM class, the summation of IID Gumbel variates would not itself be Gumbel. Therefore, if (5) 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 & Polak, 2006; Batley et al., 2007) have implicitly acknowledged such restrictions, implementing instead the following specification:

\[ P(w_n | T) = Pr\{Z(w_n) > Z(w_m)\} \quad \text{for all } m \in \tilde{C}, m \neq n \]  

where:

\[ Z(w_n) = Y(w_n) + \varepsilon_n \]

\[ Y(w_n) = \sum_{k=1}^{K} p_k V(w_{kn}) \]
The re-specification (6) is far from innocuous, since it provokes the question of how one can best represent the interface between the two sources of randomness introduced thus far (in preferences, and in the outcomes under different events). Whilst the specification (6) is dictated by issues of tractability, this does not however mean that it necessarily offers the most accurate statement of behaviour.

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 & Dunbar (1980), such models have over the last 10 years been formalised (e.g. McFadden & Train, 2000), and applied widely. Returning to Lancaster’s (1966) representation of an alternative in terms of its attributes, introduced in section 2.1, we note the traditional practice of specifying deterministic utility simply as a linear-in-parameters function of these attributes:

$$V(x_n) = \alpha x_n \quad \text{for all } n \in C$$

where $\alpha$ is a vector of $J$ ‘taste’ parameters to be estimated.

The random parameters interpretation, by contrast, admits the possibility of taste variation across the population of $R$ individuals such that:

$$V(x_{nr}) = \alpha_n x_{nr}$$

where:

- $x_{nr}$ are observable attributes that characterise alternative $n$ for individual $r$
- $\alpha_j$ is a vector of parameters relating to these variables for individual $r$

If we then represent $\alpha_r$ as a distribution across the $R$ individuals, parameterised by $(\alpha, \beta)$, we can re-express (7) thus:

$$V(x_{nr}) = \alpha x_{nr} + \beta_r x_{nr}$$

where:

- $\alpha, \beta$ are vectors of parameters to be estimated

In drawing this section to a close, let us distinguish between the three distinct sources of randomness within our model. 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 or uncertain events, which is represented in terms of the expectation $Y(w_n)$. Third, inter-individual randomness in attribute tastes, which is represented in terms of the distribution of taste parameters $\alpha_r$ across the population of $R$ individuals. Whilst this trichotomy 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 5.
4. SPECIFYING VON NEUMANN & MORGENSTERN SUB-UTILITY, FOR TIME RISK AS DISTINCT FROM MONEY RISK

In the course of the previous discussion, we have defined a theoretical model of probabilistic choice under risk or uncertainty, and identified three distinct sources of randomness which this model could feasibly embody. In the present section we will develop our discussion by considering particular specification issues that arise in applying our model to the analysis of 'journey time variability'.

Our interest in journey time variability is stimulated by the current policy environment. Whereas significant investments in transport infrastructure have historically been justified on the basis of journey time benefits, scheme promoters are increasingly citing the potential benefits from improved 'reliability' (i.e. reduced journey time variability) as the principal justification. More generally, modern-day transport operators and providers are routinely assessed against performance criteria, and reliability is typically an important element of those criteria. The reliability literature is a relatively small one, especially when compared to the vast literature on money risk, but the contributions of Bates et al. (2001) and Noland & Small (1995) are notable in offering detailed account of the underlying theory. We will not set out to emulate such contributions, instead offering an intuitive overview of the theory, as follows.

Journey time variability refers to the observation that, when undertaking a given journey on different occasions, a traveller may experience variability in journey time. Although we will not concern ourselves with the possible reasons for such variability, it is worth noting that our analysis focuses particularly on unpredictable rather than predictable (e.g. discrepancy in journey time between peak and off-peak for a regular traveller) variability. Having identified the phenomenon of interest, we can progress our analysis by drawing an analogy between the transport 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 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 essentially the starting point for Bates et al. (2001) and Noland & Small (1995).

The vast body of experimental evidence on money risk indicates a prevalent behaviour of risk aversion. With reference to Figure 1, which takes notation from (6), this implies that the von Neumann & Morgenstern sub-utility \( V(w) \) is concave with respect to money. If, for simplicity, we define a prospect over a pair of events, i.e. \( w = (w_n, w_p, p, \ldots, p) \), and consider all possible values of the event probabilities from zero to unitary, then we can derive expected utility as the chord joining the utilities at \( w \) and \( w_j \). Risk aversion results in the property that the expected utility of the prospect is less than the utility of the expected monetary outcome, i.e. \( Y(w) < V(\text{E}(w)) \). Methods for analysing reliability might be seen as an analogue of the preceding discussion, but with risk manifesting in terms of journey time rather than money. Indeed, let us explicate matters by re-expressing the prospect vector \( C = (T_1, T_2, \ldots, T_K, p_1, p_2, \ldots, p_K) \), where 'pay offs' are now defined on journey time \( T \). If we assume a trade-off between journey time and money (as we typically do, see for example Bates & Whelan (2001) for an explicit treatment), then we can straightforwardly re-draw Figure 1 in terms of journey time \( T \), as shown in Figure 2. Since money is 'good' but journey time is 'bad', this basically amounts to a re-orientation of the diagram; the individual maintains the property of risk aversion, but now in relation to journey time.
4.1 Mean-Variance Model

Deriving from work in portfolio analysis (Tobin, 1958, 1965; Markowitz, 1959), a common practical simplification of the above is to assume that expected utility can be approximated by the first and second moments of the distribution of pay-offs. This offers an exact approximation under only two situations (Borch, 1969; Feldstein, 1969); either sub-utility is quadratic, or the distribution of pay-offs is Normal. Neither situation is likely to hold in practice, but this has failed to deter the widespread adoption of the so-called ‘mean-variance’ model, notably in the analysis of reliability. Despite the terminology, it is commonplace, at least in the analysis of reliability, to employ the standard deviation rather than the variance, specifying an approximation to expected utility¹:

\[ Y_n = \delta \bar{T}_n + \phi \sigma_n \quad \text{for all } n \in \tilde{C} \quad (8) \]

where:
- \( \bar{T}_n \) is the mean journey time (for alternative \( n \) across events \( e_k \in E \))
- \( \sigma_n \) is the standard deviation of journey time (again for alternative \( n \) across events \( e_k \in E \))
- \( \delta, \phi \) are parameters to be estimated

A frequently cited metric (especially by policymakers) is the so-called ‘reliability ratio’ \( \phi/\delta \), which gives the marginal rate of substitution between expected pay-off (or in this case, when all events are taken to be equally probable, the mean) and risk (standard deviation). Whilst it would seem quite reasonable to suppose that journey time variability could influence various choices of the traveller, e.g. mode, route, destination, transport operator, most interest (in the theoretical literature, at least) has been devoted to departure time choice. This reflects the proposition that an adjustment to departure time is the easiest response available to a traveller in response to journey time variability. Indeed, this is the usual context for application of the mean-variance model (8), although there would seem to be no reason in principle why the model could not be similarly applied to the other choice dimensions cited above.

4.2 Scheduling Model

Moving on from the mean-variance model, which offers a reasonably generic approximation to expected utility, let us now turn to an alternative approximation developed specifically for the analysis of reliability. Indeed, the so-called ‘scheduling model’ (Vickrey, 1969; Small, 1982) is framed around a specific interest in how travellers choose their departure time when seeking to arrive at a given location by a given point in time. Again within a paradigm of expected utility maximisation, the sub-utility for a particular departure time is represented as functional on four components; journey time, ‘schedule delay early’ \( (SDE) \), ‘schedule delay late’ \( (SDL) \), and a ‘lateness’ dummy variable that is set to unity if \( SDL > 0 \). The latter three components are conditioned by the notion of a ‘Preferred Arrival Time’ \( (PAT) \), as follows.

- Journeys arriving at or 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 are zero, i.e.

¹ Henceforth we adopt some brevity in our notation, omitting explicit reference to the attribute or pay-off vectors, and writing \( U_n \) for \( U(x_n) \), and \( Y_n \) for \( Y(w_n) \), etc.
\[ SDE = T(PAT) - T_n \]

\[ T_n \leq T(PAT) \Rightarrow \begin{cases} 
SDE = T(PAT) - T_n \\
SDL = 0 \\
D = 0
\end{cases} \]

- 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), the lateness dummy variable is unity, and \( SDE \) is zero, i.e.

\[ T_n > T(PAT) \Rightarrow \begin{cases} 
SDE = 0 \\
SDL = T_n - T(PAT) \\
D = 1
\end{cases} \]

The four attributes of the sub-utility function are usually specified as linearly additive, thus:

\[ V_n = \varphi T_n + \gamma SDE_n + \eta SDL_n + \kappa D_n \quad \text{for all } n \in \tilde{C} \quad (9) \]

where:
- \( T_n \) is journey time
- \( SDE_n \) is schedule delay early
- \( SDL_n \) is schedule delay late
- \( D \) is unitary if \( SDL_n > 0 \), otherwise zero
- \( \varphi, \gamma, \eta, \kappa \) 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 (1995) that the scheduling model was extended to accommodate journey time variability, taking expectations of (9) over the journey time distribution thus:

\[ Y_n = \varphi' E(T_n) + \gamma' E(SDE_n) + \eta' E(SDL_n) + \kappa' E(D_n) \quad \text{for all } n \in \tilde{C} \]

where expectations are taken across events \( e_k \in E \), and \( \varphi' \) distinguishes from \( \varphi \) applying under certainty (9), and so on for the other coefficients.

Having admitted variability in journey time, it is instructive to return to (9), and consider the functional form of the sub-utility which now underpins expected utility. This is illustrated by Figure 3, which shows how, for a given departure time, the sub-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 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 sub-utility function of Figure 3 might be rationalised as a linear piecewise approximation to the continuous concave function of Figure 2, implying the same property of aversion to journey time risk.
Figure 1: Properties of the sub-utility function, under aversion to money risk

Figure 2: Properties of the sub-utility function, under aversion to time risk
Figure 3: Properties of the sub-utility function, within the scheduling model

Figure 4: Properties of the sub-utility function, within the mean lateness model
4.3 Mean Lateness Model

The mean-variance and scheduling models are by far the predominant approaches to the analysis of reliability, but other approaches exist. As background for the subsequent empirical analysis of section 5, let us introduce a third and final model, which has become standard for the analysis of reliability on rail in the UK. The Passenger Demand Forecasting Handbook or ‘PDFH’ (ATOC, 2005), the definitive reference on the modelling of UK passenger rail demand, makes two conceptual distinctions which are pertinent to our discussion. First, PDFH distinguishes between notions of ‘reliability’ and ‘punctuality’, defining the former as ‘the rate of cancellations’ and the latter as ‘the rate services exceed a given lateness standard’. Second, and applying specifically to the aforementioned definition of punctuality, PDFH introduces a further distinction between notions of ‘lateness’ and ‘delay’. Within these regimes, ‘lateness refers to the difference between the actual and publicly timetabled arrivals at destinations’ and ‘delay is used to refer to the difference between the actual and working times to pass over short route sections’. In the hope of promoting simplicity, and remaining faithful to the tone of the paper thus far, the subsequent commentary subsumes all of the aforementioned PDFH concepts within a broad definition of the term reliability, which we take to mean the propensity for a rail service to adhere to its schedule.

Whilst again adhering to the expected utility paradigm, the model prescribed by PDFH (section B4; ATOC, 2005) is distinct from the previous two models in specifying the pay-off dimension as lateness rather than journey time. This 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. Moreover, PDFH asserts the following approximation to expected utility:

\[
Y_n = \lambda SchedT_n + \mu L_n^- \quad \text{for all } n \in \tilde{C}
\]

(10)

where:

- \(SchedT\) is scheduled journey time
- \(L^-\) is the mean lateness at destination (across events \(e_k \in E\))
- \(\lambda, \mu\) are parameters to be estimated

We shall refer to (10) as the ‘mean lateness’ model, wherein expected utility is specified as a linear additive function of scheduled journey time and mean lateness at destination relative to timetable (where any earliness, or ‘negative’ lateness, is discounted). PDFH guidance acknowledges, albeit only as a postscript, that some practical cases may exhibit significant journey time variability, recommending in those cases that (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 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 sub-utility within (10), through reference to Figure 4. Note that we express this function \(f(T')\) in terms of journey time rather than lateness; this allows comparison against the previous figures, but does not obscure its key
properties, as follows. On first inspection of the sub-utility function, it would appear that the mean lateness model is similar to the scheduling model in offering a piecewise linear approximation to Figure 2. In contrast to the scheduling model, the function is referenced against the scheduled arrival time at destination rather than the preferred arrival time, although both reference points could of course coincide. 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 \([0,T(SchedT)]\). That is to say, \( f(T) \) is defined only for \( T > T(SchedT) \), implying that travellers are not risk averse to journey time, but rather risk neutral.

5. AN EMPIRICAL APPLICATION TO JOURNEY TIME RISK

5.1 Background

The purpose of this section is to demonstrate the empirical application of a probabilistic model of discrete choice under uncertainty, particularly with a view to illustrating the conceptual interests considered in the latter two sections, namely the 'three sources of randomness', and the specification of the sub-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 the Department.

The primary feature of the study design was a self-completion mailback questionnaire comprising two sections, as follows. The first section featured questions on the prior experiences of travellers concerning reliability, and invited retrospective reporting of their responses to changes in reliability. The second section involved a Stated Preference (SP) experiment. As previous researchers have discovered (e.g. Bates et al., 2001), reliability is not the easiest concept to present in the context of a SP experiment. Our experiment applied an extension to Hollander’s (2006) presentation, which might itself be seen as a simplification of the so-called ‘clock face’ presentation employed by Bates et al.

With reference to Figure 5, we presented a pair of travel options for a given journey, each described in terms of fare for a single leg, scheduled journey time, and the distribution of journey time over 5 repetitions of the journey. Thus, in the notation of sections 3 and 4, \( N = 2 \) and \( K = 5 \). Since we did not offer advice to respondents on the probabilities of these 5 events (i.e. the \( p_k \) ’s were unknown to respondents), one might characterise these as uncertain, as distinct from risky, prospects. As we shall see subsequently, for purposes of analysis however, we assumed the 5 events to be equi-probable, thereby promoting ready application to the mean lateness model. Whereas Hollander’s 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. Having presented a pair of services in this manner, respondents were invited to choose between the pair.

13
The SP experiment was based on a ‘difference’ design, as follows. Having researched cost and timetabled journey time for a single trip on the origin-to-destination (O-D) of interest, we specified 3 levels of lateness in departure from the boarding station (in fixed blocks of 5 departure times), with each level embodying a mean and standard deviation. Different values were specified for options A and B, although in each instance level 1 was characterised by the mean journey time and a standard deviation of 0 (i.e. reliable). Variation in journey time around the timetabled journey time was similarly specified at 3 levels, again in fixed blocks of 5. 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 4 variables (cost, timetabled journey time, departure time variation and journey time variation). Following design, the SP experiment was tested by means of simulation, assuring ourselves that the design would be capable of recovering parameter ratios within acceptable ranges, and a pilot survey.

Having satisfied ourselves as to the robustness of the SP design, we proceeded to the field survey. This survey was guided by an industry consultation, which sought to identify potential study areas which exhibited either, or both of, the following characteristics:

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

On this basis, it was decided that the field survey should be targeted at 12 specific O-D journeys, involving 10 survey locations (Table 1). These 12 journeys were characterised by substantial passenger volumes, offered some variety in context, 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 7th and 8th February 2007, and Wave 2 on 28th February 2007, in all cases between the hours of 6am and 12noon.

### Table 1: Origin-destinations of interest, by journey distance

<table>
<thead>
<tr>
<th>LONG DISTANCE</th>
<th>SHORT DISTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bristol Temple Meads</td>
<td>Brighton-London</td>
</tr>
<tr>
<td>Leeds-Birmingham</td>
<td>Edinburgh-Glasgow</td>
</tr>
<tr>
<td>Leeds-London</td>
<td>Leeds-Sheffield</td>
</tr>
<tr>
<td>Swindon-London</td>
<td>Peterborough-London</td>
</tr>
<tr>
<td></td>
<td>Portsmouth-London</td>
</tr>
<tr>
<td></td>
<td>Reading-London</td>
</tr>
<tr>
<td></td>
<td>Stevenage-London</td>
</tr>
<tr>
<td></td>
<td>Woking-London</td>
</tr>
</tbody>
</table>
From a distribution of around 15,000 questionnaires, we achieved a response rate of just over 19% - a typical level of response for a self-completion mailback questionnaire. 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 primary interests of the DfT study.

Table 2: Cross-tabulation of journey purpose and distance

<table>
<thead>
<tr>
<th>PURPOSE</th>
<th>DISTANCE</th>
<th>Long</th>
<th>Short</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td></td>
<td>20.3%</td>
<td>16.1%</td>
<td>36.4%</td>
</tr>
<tr>
<td>Commute</td>
<td></td>
<td>12.8%</td>
<td>37.9%</td>
<td>50.6%</td>
</tr>
<tr>
<td>Leisure</td>
<td></td>
<td>4.7%</td>
<td>8.3%</td>
<td>12.9%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>37.8%</td>
<td>62.2%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

5.2 Empirical model

The following discussion reports four models, all of which were implemented within the apparatus of (6), specifying the random error term to be IID Gumbel, and estimated by maximum likelihood using CIMLogit (Connors & Ibáñez, 2009).

Model 1: Randomness in preferences

Our initial model was reasonably faithful to the mean lateness model (10), applying the following approximation to expected utility:

$$Y_n = \lambda SchedT_n + v C_n + \pi \bar{B}_n + \mu \bar{L'}_n \quad \text{for } n = a, b$$  \hspace{1cm} (11)

where:

$SchedT$ is scheduled journey time

$C$ is cost for a single leg of a return journey

$\bar{B}$ is mean lateness at the boarding station (across events $k = 1, \ldots, 5$)

$\bar{L'}$ is mean (positive) lateness at the destination station (again across events $k = 1, \ldots, 5$)

$\lambda, v, \pi, \mu$ are parameters to be estimated

Reconciling against the discussions of sections 3 and 4, it might be seen that this model, whilst referred to as an expected utility specification, embodies randomness in preferences but non-randomness in both outcomes and attribute tastes. Equation (11) introduces two additional variables beyond (10). The inclusion of cost is conventional, and permits the derivation of valuations for the time variables. Mean lateness at boarding is less conventional, and deserves more explanation. 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):

$$SchedT = T - L + B$$  \hspace{1cm} (12)

where $T$ is in-vehicle journey time. That is to say, scheduled journey time is the sum of in-vehicle journey time and lateness at boarding, net of any lateness at destination. Figure 6 illustrates this identity for a journey that does incur lateness at destination (i.e. $L > 0$). Consider a train which is scheduled to depart at 9:05, with a scheduled journey time of
$T=140$ minutes. The train actually departs at 9:20, incurring lateness at boarding of $B=15$. The train makes up 5 minutes of this late-running en route, arriving at the final destination at 11:30, and incurring lateness at destination of $L=10$.

**Figure 6: Dissecting the time components of a rail journey**

![Diagram showing time components of a rail journey](image)

Lateness at boarding might thus be seen as a component of $SchedT$ within the mean lateness model (9), but under the assumption that a minute of lateness at boarding carries the same disutility as a minute of in-vehicle journey time. By adding an explicit lateness at boarding variable, we can test whether this assumption is defensible. In practice, trains were never presented as arriving early at the boarding station within our SP experiment, such that $B \geq 0$. Arrivals at the destination station could, by contrast, either be early or late, such that $L$ could potentially be negative or positive. For purposes of modelling however, and in accordance with rail industry convention, negative lateness was reset to zero.

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 at 1%, 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 4, 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 cost in £ and journey time in minutes, we can derive a value of time of 35.32p/min. Whilst this might seem rather high compared with the values of time reported in PDFH (ATOC, 2005), it should be remembered that our survey targeted 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 4; 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.
Table 3a: Estimates from Models 1, 2 and 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model 1</td>
</tr>
<tr>
<td>COST</td>
<td>Mean</td>
<td>$\nu$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-18.25)</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>$\nu_j$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-9.74)</td>
</tr>
<tr>
<td>SCHED JOURNEY TIME</td>
<td>Mean</td>
<td>$\lambda$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-19.47)</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>$\rho_j$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-15.63)</td>
</tr>
<tr>
<td>MEAN (POSITIVE) LATENESS AT DESTINATION</td>
<td>Mean</td>
<td>$\mu$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-45.58)</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>$\varepsilon_j$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-12.96)</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING</td>
<td>Mean</td>
<td>$\pi$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-17.45)</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>$\omega_j$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8.82)</td>
</tr>
<tr>
<td>ST.DEV. JOURNEY TIME</td>
<td>Mean</td>
<td>$\theta$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8.96)</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>$\tau_j$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8.10)</td>
</tr>
<tr>
<td>-LL_final</td>
<td></td>
<td>5682.66</td>
</tr>
<tr>
<td>-LL_ASC</td>
<td></td>
<td>7491.47</td>
</tr>
<tr>
<td>-LL_zeros</td>
<td></td>
<td>8153.49</td>
</tr>
<tr>
<td>Rho-sq_ASC</td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td>Rho_sq_zeros</td>
<td></td>
<td>0.30</td>
</tr>
<tr>
<td>No. individuals</td>
<td></td>
<td>2395</td>
</tr>
<tr>
<td>No. observations</td>
<td></td>
<td>11763</td>
</tr>
<tr>
<td>No. pseudo random draws</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3b: Ratios from Models 1, 2 and 3

<table>
<thead>
<tr>
<th>Metric</th>
<th>Model 1</th>
<th>Model 2 (Mean)</th>
<th>Model 3 (Median)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of time</td>
<td>$\lambda/\nu$</td>
<td>35.32p/min</td>
<td>41.38p/min</td>
</tr>
<tr>
<td>'Reliability ratio'</td>
<td>$\theta/\lambda$</td>
<td>1.55</td>
<td>2.07</td>
</tr>
<tr>
<td>Lateness multiplier_DESTINATION</td>
<td>$\mu/\lambda$</td>
<td>3.96</td>
<td>2.80</td>
</tr>
<tr>
<td>Lateness multiplier_BOARDING</td>
<td>$\pi/\lambda$</td>
<td>2.41</td>
<td>1.72</td>
</tr>
</tbody>
</table>

Model 2: Randomness in preferences and outcomes

$$Y_n = \lambda SchedT_n + \theta \sigma_n + \nu C_n + \pi \bar{B}_n + \mu L^+_n$$ for $n = a, b$

where:

$\sigma$ is the standard deviation of in-vehicle journey time, and $\theta$ is a parameter to be estimated.

Extending Model 1, we now introduce randomness in outcomes through the standard deviation of journey time. 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 4, 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.

Model 3: Randomness in preferences, outcomes and attribute tastes

In terms of the ‘three sources of randomness’, we now complete the picture by introducing randomness in attribute tastes, via the specification:

\[
Y_{nj} = \lambda SchedT_{nj} + \rho_j SchedT_{nj} + \theta_n\sigma_{nj} + \tau_j\sigma_{nj} + \nu C_{nj} + \omega_j B_{nj} + \pi B_{nj} + \xi_j L_{nj}^+ + \mu L_{nj}^+
\]

for \( n = a, b \)

where we estimate \((\lambda, \rho)\) as distributed across the \( R \) individuals, and so on for the other parameters.

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 \( j \). Although we experimented with various distributional assumptions, the reported Model 3 specifies all parameters as Normally-distributed. This model was again estimated using CIMLogit, by way of 3000 draws per individual, and exploiting Marsaglia’s ziggurat algorithm (Marsaglia and Tsang, 1984).

Perusal of Table 3a reveals that all standard deviations of the coefficients are significantly different from zero at 1%, 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 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 Figures 5 to 8. 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.
Figure 5: Distribution plot of value of time from Model 3

Figure 6: Distribution plot of reliability ratio from Model 3
Figure 7: Distribution plot of lateness multiplier at destination from Model 3

Figure 8: Distribution plot of lateness multiplier at boarding from Model 3
With reference to Figure 5, Model 3 yields a mean value of time of 25.62p/min, against a median of 18.55p/min; this indicates a slight positive skew in the PDF. The PDF shows considerable spread around these measures of central tendency. The reliability metrics show a more marked (positive) skew than the value of time. The mean reliability ratio is estimated at 2.07, which compares with a median of 0.85. The lateness multiplier at destination yields a mean of 3.64 and a median of 1.62; the analogous statistics for boarding are 1.25 and 0.50. 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 & Batley’s (2009) rationale defending such valuations. The sign of the reliability ratio is less controversial, and may indeed yield useful interpretation. In portfolio 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 unduly controversial to postulate that an analogous relationship holds for journey time risk.

Model 4: Synthesising the mean-variance and mean lateness models

Our final model pursues a slightly different tack, developing the issues discussed in section 4 concerning the form of the von Neumann & Morgenstern sub-utility function. In particular, we endeavour to reconcile an apparent inconsistency, in terms of their risk properties, between the prescribed model specifications for the analysis of reliability on UK road and rail. In so doing, we also seek to reconcile the ‘currencies’ of risk considered by the two analyses; whereas road represents risk in terms of in-vehicle journey time, rail represents risk in terms of lateness relative to timetable at the destination station.

On road, reliability is conventionally measured in terms of in-vehicle journey time for a given O-D journey; more specifically, a driver is represented as choosing his/her time of departure from the origin, subject to variability in the journey time experienced in reaching the destination. In these terms, modelling and forecasting of reliability is developed using the mean-variance model (8), yielding the reliability ratio as the headline metric. WebTAG unit 3.5.7 cites ‘standard values’ for the reliability ratio of ‘about 0.8 for private passenger travel’, drawing comparison to analogous values for public transport of 0.6 to 1.5. Whilst WebTAG is less than explicit, it may well be the case that this comparison is not entirely like-for-like, since the reliability ratio for public transport is often formulated in terms of lateness rather than in terms of in-vehicle journey time.

In seeking to align methods for analysing reliability on rail and road, an important consideration is to establish a consistent starting point for the journey. On road, the starting point is reasonably unambiguous; the journey begins as and when the driver starts up his/her vehicle and departs the origin. On rail, the mean lateness model focuses on scheduled journey time, implying that the starting point for the journey is the timetabled departure time at the boarding station, irrespective of whether the journey actually commences on time. Developing these interests, let us extend the identity previously introduced in (12) to distinguish between earliness and lateness at the destination station:

\[ SchedT = T - L^+ + L^- + B \quad \text{where} \quad L^+, L^- \geq 0 \quad (13) \]

Thus we represent scheduled journey time 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. It might be remarked that, in the case of road, the notion of lateness at boarding does not readily apply, and we can therefore take \( B = 0 \). Reconciling with Figure 4, note that, by admitting the possibility of earliness, we now arrive
at a function which does embody a linear piecewise approximation to risk aversion in journey time.

Exploiting the identity (13), we can 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:

$$Y_n = \lambda SchedT_n + \theta \sigma_n + \nu C_n + \pi \overline{B}_n + \mu \overline{L^+}_n + \psi \overline{L^-}_n \quad \text{for } n = a, b$$  \hspace{1cm} (14)

where:

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

It might be reminded that our SP experiment never offered the possibility of early departure from the boarding station, such that $B \geq 0$. Substituting for $SchedT$, as per the definition (13), in (14), we can re-state our model in terms of in-vehicle journey time and the various boarding variables, as follows:

$$Y_n = \lambda T_n + \theta \sigma_n + \nu C_n + (\lambda + \pi) \overline{B}_n + (\mu - \lambda) \overline{L^+}_n + (\lambda + \psi) \overline{L^-}_n \quad \text{for } n = a, b$$  \hspace{1cm} (15)

Appealing to the interests of policy-makers, we implement this as a logit specification, but with segmentation by journey distance and purpose. Model 4 might thus be seen as embodying randomness in attributes tastes (as well as randomness in preferences and outcomes), but with a coarser resolution of the taste distribution through segmentation than the explicit random parameters specification of Model 3.

With reference to Table 4a, all segmentations are, with the exception of earliness at destination for short distance leisure, significant at 5%, and most are indeed significant at 1%. 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. 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 (at 5%) and leisure (at 1%). Second, in the case of long distance, the marginal disutility of (positive) lateness at destination for business is significantly greater than for leisure (at 1%). 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, derived directly from (14). If we then re-state the model as (15), 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.
Table 4a: Estimates from Model 4*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Estimate</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>COST_SB</td>
<td>ν_SB</td>
<td>-0.1330</td>
<td>-6.60</td>
</tr>
<tr>
<td>COST_SC</td>
<td>ν_SC</td>
<td>-0.2119</td>
<td>-10.98</td>
</tr>
<tr>
<td>COST_SO</td>
<td>ν_SO</td>
<td>-0.1101</td>
<td>-4.47</td>
</tr>
<tr>
<td>COST_LB</td>
<td>ν_LB</td>
<td>-0.0767</td>
<td>-10.73</td>
</tr>
<tr>
<td>COST_LO</td>
<td>ν_LO</td>
<td>-0.0731</td>
<td>-8.87</td>
</tr>
<tr>
<td>SCHED JOURNEY TIME_SB</td>
<td>λ_SB</td>
<td>-0.0515</td>
<td>-8.68</td>
</tr>
<tr>
<td>SCHED JOURNEY TIME_SC</td>
<td>λ_SC</td>
<td>-0.0493</td>
<td>-11.80</td>
</tr>
<tr>
<td>SCHED JOURNEY TIME_SO</td>
<td>λ_SO</td>
<td>-0.0240</td>
<td>-3.25</td>
</tr>
<tr>
<td>SCHED JOURNEY TIME_LB</td>
<td>λ_LB</td>
<td>-0.0433</td>
<td>-15.19</td>
</tr>
<tr>
<td>SCHED JOURNEY TIME_LO</td>
<td>λ_LO</td>
<td>-0.0294</td>
<td>-7.83</td>
</tr>
<tr>
<td>MEAN LATENESS AT DESTINATION_SB</td>
<td>μ_SB</td>
<td>-0.1382</td>
<td>-15.11</td>
</tr>
<tr>
<td>MEAN LATENESS AT DESTINATION_SC</td>
<td>μ_SC</td>
<td>-0.1535</td>
<td>-23.12</td>
</tr>
<tr>
<td>MEAN LATENESS AT DESTINATION_SO</td>
<td>μ_SO</td>
<td>-0.1245</td>
<td>-10.68</td>
</tr>
<tr>
<td>MEAN LATENESS AT DESTINATION_LB</td>
<td>μ_LB</td>
<td>-0.0770</td>
<td>-15.97</td>
</tr>
<tr>
<td>MEAN LATENESS AT DESTINATION_LC</td>
<td>μ_LC</td>
<td>-0.0589</td>
<td>-11.01</td>
</tr>
<tr>
<td>MEAN LATENESS AT DESTINATION_LO</td>
<td>μ_LO</td>
<td>-0.0524</td>
<td>-6.61</td>
</tr>
<tr>
<td>MEAN EARLINESS AT DESTINATION_SB</td>
<td>ψ_SB</td>
<td>0.2930</td>
<td>2.09</td>
</tr>
<tr>
<td>MEAN EARLINESS AT DESTINATION_SC</td>
<td>ψ_SC</td>
<td>0.4325</td>
<td>4.33</td>
</tr>
<tr>
<td>MEAN EARLINESS AT DESTINATION_SO</td>
<td>ψ_SO</td>
<td>0.1968</td>
<td>1.12</td>
</tr>
<tr>
<td>MEAN EARLINESS AT DESTINATION_LB</td>
<td>ψ_LB</td>
<td>0.2303</td>
<td>3.59</td>
</tr>
<tr>
<td>MEAN EARLINESS AT DESTINATION_LC</td>
<td>ψ_LO</td>
<td>0.2663</td>
<td>3.43</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING_SB</td>
<td>π_SB</td>
<td>-0.0674</td>
<td>-4.56</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING_SC</td>
<td>π_SC</td>
<td>-0.0874</td>
<td>-8.11</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING_SO</td>
<td>π_SO</td>
<td>-0.0608</td>
<td>-3.09</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING_LB</td>
<td>π_LB</td>
<td>-0.0437</td>
<td>-6.43</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING_LC</td>
<td>π_LC</td>
<td>-0.0377</td>
<td>-3.85</td>
</tr>
<tr>
<td>MEAN LATENESS AT BOARDING_LO</td>
<td>π_LO</td>
<td>-0.0381</td>
<td>-2.70</td>
</tr>
<tr>
<td>ST.DEV. JOURNEY TIME</td>
<td>θ</td>
<td>-0.0644</td>
<td>-10.32</td>
</tr>
</tbody>
</table>

-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: 11763
Table 4b: Reliability metrics from Model 4

<table>
<thead>
<tr>
<th>Segment</th>
<th>Lateness multiplier_DESTINATION (based on scheduled journey time; $\mu/\hat{\lambda}$)</th>
<th>Reliability ratio (based on in-vehicle time; $\theta/\hat{\lambda}$)</th>
<th>'Lateness multiplier'_DESTINATION (based on in-vehicle time; $[\mu - \hat{\lambda}]/\hat{\lambda}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB</td>
<td>2.68</td>
<td>1.25</td>
<td>1.68</td>
</tr>
<tr>
<td>SC</td>
<td>3.12</td>
<td>1.31</td>
<td>2.12</td>
</tr>
<tr>
<td>SO</td>
<td>5.19</td>
<td>2.69</td>
<td>4.19</td>
</tr>
<tr>
<td>LB</td>
<td>1.78</td>
<td>1.49</td>
<td>0.78</td>
</tr>
<tr>
<td>LC</td>
<td>2.00</td>
<td>2.19</td>
<td>1.00</td>
</tr>
<tr>
<td>LO</td>
<td>1.77</td>
<td>2.18</td>
<td>0.77</td>
</tr>
</tbody>
</table>

6. SUMMARY AND CONCLUSION

Nearly 50 years have passed since Marschak (1960) and Block & 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 hence, 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.

In defining these working specifications, we observed two trends in the decision-making environment, towards greater uncertainty in the outcomes from choice, and towards greater bespeaking of products to individual attribute tastes. Within this environment, we identified three distinct sources of randomness in RUM; in preferences (as described by Block and Marschak); in outcomes (as associated with risk and uncertainty); and in attribute tastes (in the context of random parameters logit). We articulated this trichotomy in the course of two discussions; first, by presenting a typology of theoretical models for discrete choice, categorising according to whether choice is certainty or risky/uncertainty, and whether choice is deterministic or probabilistic; and second, by considering the implementation of an empirical model for probabilistic choice under risk/uncertainty.

Having arrived at a general theoretical statement of RUM under risk/uncertainty, we considered the application of this model to a specific interest in journey time risk, or ‘journey time variability’. Noting the allegiance of the journey time variability literature to von Neumann-Morgenstern’s (1947) paradigm of expected utility maximisation, we devoted particular attention to the formulation of the ‘sub-utility’ function within expected utility. We distinguished between three such formulations; the so-called mean-variance, scheduling and mean lateness models. We noted that whereas the first two models embody the property of risk aversion in journey time (as one would typically expect), the latter model embodies risk neutrality.

Synthesising the mean-variance and mean lateness models, we demonstrated an empirical application of RUM under risk/uncertainty, using data from a recent Stated Preference (SP) experiment. The experiment invited rail passengers to choose between a pair of services on the basis of fare, scheduled journey time and journey time variability. We developed the model incrementally, each time introducing an additional source of randomness in accordance with the aforementioned trichotomy. The most general model involved a random parameters specification, and offered a means by which we could parameterise randomness across preferences, outcomes, and attribute tastes. Furthermore, an important outcome from this model was an insight into the distribution of attitudes to journey time risk across our population.
Our final model was something of a digression, illuminating an apparent inconsistency in the methods conventionally applied to the analysis of journey time variability on UK road and rail. In particular, we sought to illuminate the implications of specifying a model on scheduled journey time (as with the mean lateness model) as distinct from in-vehicle journey time (as with the mean-variance model), and vice versa.

REFERENCES


