By
Baibing Li[1] and David A. Hensher[2]

[1] School of Business & Economics Loughborough University, Loughborough LE11 3TU, United Kingdom
[2] Institute of Transport and Logistics Studies (ITLS), The University of Sydney Business School, Sydney, Australia

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in

Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER:	Working Paper ITLS-WP-17-10
TITLE:	**Risky Weighting in Discrete Choice**

ABSTRACT: This paper presents a new approach to discrete choice analysis for risky prospects. Conventional discrete choice analysis focuses on riskless prospects and does not deal with the scenario where the alternatives that the decision-makers choose from are associated with risk. In this paper, we investigate decision-makers' risk perception and choice behaviour in choice experiments when they are facing several risky prospects. We propose a broad class of cumulative risky weighting functions, upon which a unified cumulative risky weighting function is developed. We show that this unified cumulative risky weighting function includes several existing cumulative risky weighting functions as special cases. We then develop a multivariate method for choice analysis with risky prospects to account for decision-makers' individual-specific risk perception and the impact of various factors on the value function respectively. We illustrate the developed method using an empirical study on road tolling in Australia.

AUTHORS: **Li and Hensher**

CONTACT: INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H73)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney   NSW 2006   Australia

Telephone: +612 9114 1813

E-mail: business.itlsinfo@sydney.edu.au

Internet: http://sydney.edu.au/business/itls

DATE: May 2017

1. Introduction

Discrete choice analysis is widely applied to a range of disciplines such as transport, economics, marketing, political science, etc. (see, e.g., Hensher *et al.*, 2015; Train, 2009). It investigates decision-makers' choice behaviour when they make a choice from several options. Conventional discrete choice analysis associated with a fully compensatory choice paradigm focuses on the scenario where the choice situations are assumed *riskless* for choice makers in the sense that the alternatives under investigation are completely characterised by a number of deterministic (or given) attributes (covariates).

This papers considers decision-making when decision-makers are facing several *risky* prospects, following along the research line of Prospect Theory (PT) (Kahneman and Tversky, 1979), Rank Dependent Utility model (RDU) (Quiggin, 1982), and Cumulative Prospect Theory (CPT) (Tversky and Kahneman, 1992). These theories offer an alternative behavioural paradigm to the conventional expected utility theory (EUT) (Wakker, 2010; Fehr-Duda and Epper, 2012) for the investigation of decision-making under uncertainty. They argue, with a large body of evidence in the literature, that when facing uncertainty, decision-makers will transform the given uncertainty measure and utility function, prior to decision-making, into their own decision weights and value function.

Discrete analysis and CPT are mostly investigated separately in the literature. Hensher *et al*. (2011) reviewed the contributions of CPT in transport and other literature and undertook a study to embed perceptual conditioning and risk attitude into discrete choice analysis. They investigated, in a stated choice setting, individual preferences for two designed routes in Australia with travel time subject to risk. The respondents (travellers) in the survey were asked to consider their reference alternative and two designed alternatives. Unlike the conventional discrete choice analysis where alternatives are characterised by a number of deterministic (or given) attributes, the most distinguished feature of the research in Hensher *et al*. (2011) was that the alternatives shown to the respondents during the survey were under risk and described by a probability distribution, and hence the analysis in Hensher *et al.* (2011) was based on the respondents' perceived travel time. Gao *et al.* (2010), on the other hand, developed a routing policy choice model based on CPT. They investigated a scenario where travellers know the probabilistic distributions of the link travel times, leading to a route choice problem under risk.

They also compared differences between models based on EUT and CPT. Their study demonstrates the flexibility of the CPT models to represent varying degrees of risk aversion and risk seeking depending on the outcome probabilities. In addition to Hensher *et al.* (2011) and Gao *et al.* (2010), some other studies in the transportation and traffic studies literature investigated travellers' choice behaviour with risky prospects. For example, Xu *et al.* (2011) investigated modelling traveller's route choice behaviour based on CPT. Li and Hensher (2013) investigated behavioural implications of preferences, risk attitudes and beliefs in modelling risky travel choice with travel time variability. Kemel and Paraschiv (2013) applied PT to investigate joint time and money consequences in risk and ambiguity of transport users. De Palma *et al.* (2008) provided a general overview on choice analysis for both EUT and non-EUT approaches. Hu *et al.* (2012), on the other hand, investigated a comparison of EUT and non-EUT approaches in the context of modelling travellers' risky choice.

The purpose of this paper is to generalise the approaches investigated in Hensher *et al.* (2011) and Gao *et al.* (2010). We aim to investigate the risky weighting problem in a general context so that the methodology developed in this paper can be applied to a wide range of problems of behaviour under risk, including travel behaviour and route choice. We contribute to the literature by developing a simultaneous modelling method for both decision-makers' risk perception and choice behaviour when decision-makers choose from several risky prospects. We propose a unified method for elicitation of decision-makers' risk perception. In the literature, there are several widely used risky weighting functional forms, including Goldstein and Einhorn (1987), Tversky and Kahneman (1992), Wu and Gonzalez (1996), and Prelec (1998). It is not yet clear, from a theoretical perspective, how these risky functions are related to each other and in which particular circumstance we should choose one from several candidate functions. We examine these commonly used risky weighting functions in the literature, upon which we propose a broad class of cumulative risky weighting functions (CRWFs) for the modelling of decision-makers' risky weighting behaviour. We then investigate some important theoretical characteristics of risky weighting and how these risky weighting functions are related to each other.

On the basis of the proposed CRWFs, we develop a multivariate approach to describing decision-makers' choice behaviour when choice situations are subject to risk. We consider the case of stochastic value functions in CPT, and capture the relationship between choice outcomes and the attributes of decision-makers and prospects under the framework of discrete choice analysis.

This paper is structured as follows. In Section 2, we propose a broad class of CRWFs. We then consider some important properties of risky weighting and develop a unified CRWF. In Section 3 we investigate a multivariate method for choice analysis with risky prospects. The elicitation of the risky weighting function and the value function will be discussed in Section 4. The proposed method is illustrated using a practical example in Section 5. Finally, in Section 6, we summarise the main contributions of this paper. All the proofs of theorems are presented in Appendix A. In Appendix B we discuss the maximum entropy principle.

2. Cumulative risky weighting functions

In this section, we focus on decision-makers' risky weighting and propose a general approach to the specification of the CRWF.

2.1. *Expected utility theory and cumulative prospect theory*

We briefly summarise EUT and CPT in this subsection. Consider a typical decision-making problem where there are a number of (say J) states of nature, each having a probability $p_j \geq 0$ ($j = 1, \dots, J$) with $\sum_{j=1}^{J} p_j = 1$. Each decision-maker n ($n = 1, \dots, N$) needs to choose a prospect from a choice set C_n of size M, where each prospect i is defined to be a contract (gamble):

$$g_i = \left(\{p_j\}_{j=1}^{J}, \{t_{ij}\}_{j=1}^{J} \right) \quad \text{for } i = 1, \dots, M,$$

associated with the primary outcome (payoff) variable T that takes a value of t_{ij} with probability p_j.

Let $u(t; \tilde{\mu}, \tilde{\tau})$ denote the utility function of a decision-maker with risk attitude coefficient $\tilde{\tau}$ and individual-specific parameter $\tilde{\mu}$. For decision-maker n choosing prospect i, the expected utility u_{in} is defined to be

$$u_{in} = \sum_{j=1}^{J} u\left(t_{ij}; \tilde{\mu}_{in}, \tilde{\tau}\right) p_j, \tag{1}$$

where $\tilde{\mu}_{in}$ is the individual-specific parameter of the utility function for decision-maker n choosing prospect i. The most dominant normative and descriptive model of decision making under uncertainty in the conventional decision-making theory is EUT. In EUT, a decision-maker chooses a prospect i if and only if $u_{in} \geq max_{j \in C_n}\{u_{jn}\}$, i.e., his/her expected utility attains the maximum when choosing prospect i (see, e.g., Wakker, 2010).

Allais (1953), however, considered a decision-making problem consisting of two choice situations (denoted as A and B), each with two prospects (gambles). Allais (1953) found that the choice behaviour of the same person across a series of choice situations was often inconsistent with EUT, now known as the Allais paradox.

Since the study of Allais (1953), there has been a large body of empirical evidence in the literature in relation to the Allais paradox, and considerable effort has been made to try to resolve the Allais paradox. In particular, Kahneman and Tversky (1979) developed prospect theory, which was subsequently modified to become a more general theory, known as cumulative prospect theory (CPT), in Tversky and Kahneman (1992). It is argued in CPT that, when facing uncertainty in reality, each decision-maker will transform the given uncertainty measure and his/her utility function prior to the decision making. Specifically, let $v(t; \mu, \tau)$ denote the value function with risk attitude coefficient τ and an individual-specific parameter μ, and $w_j(p_1, \dots, p_J)$ denote the risky weighting function that transforms probability p_j ($j = 1, \dots, J$) into the decision-maker's "decision weight" (probability) q_j:

$$q_j = w_j(p_1, \dots, p_J). \tag{2}$$

The expected value v_{in} after the risk perception transformation in CPT is:

$$v_{in} = \sum_{j=1}^{J} v(t_{ij}; \mu_{in}, \tau) q_j, \tag{3}$$

where μ_{in} is the individual-specific parameter of the value function for decision-maker n choosing prospect i.

In the literature, there is in general a consensus that the two components in equation (3), i.e., the value function $v(t; \mu, \tau)$ and the risky weighting function $q_j = w_j(p_1, \dots, p_J)$, must satisfy some desired behavioural properties. For example, the value function is usually assumed to be concave for gains and convex for losses; a risky weighting function $q_j = w_j(p_1, \dots, p_J)$ overweights small tail probabilities

(Tversky and Kahneman, 1992; Abdellaoui *et al.* 2010). This allows the researcher considerable flexibility in modelling. In empirical analysis, however, because there is little guidance on the choice of the value function and risky weighting function, the researcher usually has to explore a number functional forms (see, e.g., Stott, 2006; Hensher *et al.*, 2011, Balcombe & Fraser, 2015), aiming to find a suitable choice of value function $v(t; \mu, \tau)$ and risky weighting function $q_j = w_j(p_1, \dots, p_J)$ that performs best numerically. For example, Stott (2006) and Balcombe & Fraser (2015) examined 256 and 549 variants constructed from these functions respectively.

2.2. CRWFs and primary outcome transformations

We now focus on the risky weighting function in equation (2). Following the assumption in the previous subsection, each prospect is characterised by a random primary outcome variable T (e.g., travel time of a journey) with a reference point of zero and a probabilistic measure $F(t) = \Pr(T > t)$ (which is usually termed survival function in the event-to-time analysis, describing the likelihood that the primary outcome exceeds a particular level t).

In real-world decision-making problems, the primary outcome variable T can be of various data formats. In Tversky and Kahneman (1992), they confine their investigation to monetary outcomes, and hence the support of the variable is the entire real line (denoted by $\Re$). In this paper, we restrict our interest to the situation that the primary outcome variable T has the support of all nonnegative real values (denoted by $\Re^+$); this appears to line up well with the design of attribute levels in many choice experiments in transportation research. We will briefly discuss the scenario of $T \in \Re$ at the end of this section.

In many choice experiments, the continuous primary outcome variable T is often discretised and expressed as a number of outcomes, each corresponding to a state of nature with probability representing the chance that the state will occur, i.e. (t_j, p_j) with $p_j \geq 0$ (for $j = 1, \dots, J$) and $\sum_{j=1}^{J} p_j = 1$. This leads to a prospect defined to be $g = \left(\{p_j\}_{j=1}^{J}, \{t_j\}_{j=1}^{J} \right)$. Without loss of generality, it is assumed that $t_1 < t_2 < \cdots < t_J$. Therefore, p_1 and p_J are the left-tail and right-tail probabilities of the distribution

respectively. Let $P_j = \sum_{k \geq j} p_k$ denote the corresponding discrete survival function $(j = 1, \ldots, J+1)$ with $P_{J+1} = 0$.

Suppose that a decision-maker transforms the survival function $F(t) = \Pr(T > t)$ shown in a choice experiment into a new survival function $G(t)$ via a CRWF, $W(P)$:

$$G(t) = W(F(t)). \tag{4}$$

Clearly, for the given two survival functions $F(t)$ and $G(t)$, the CRWF that transforms $F(t)$ to $G(t)$ is equal to $W(P)=G[F^{-1}(P)]$. Note that the risky weighting function $q_j = w_j(p_1, \ldots, p_j)$ in equation (2) is a special case of (4) where the survival function $F(t)$ is discretised and represented by $\{P_j\}_{j=1}^{J+1}$. In this case, the risky weighting function that maps from the probabilities p_j to decision weights q_j $(j = 1, \ldots, J)$ is defined explicitly as

$$q_j := w_j(p_1, \ldots, p_J) = W(P_j) - W(P_{j+1}) = W(\sum_{l=j}^{J} p_l) - W(\sum_{l=j+1}^{J} p_l), (j = 1, \ldots, J).$$

Note that the difference between CPT and the original version of prospect theory is that it is the cumulative probability distribution function, rather than probability mass function, in CPT that is applied to the perceptual conditioning because the functional form of prospect theory violates stochastic dominance (Tversky and Kahneman, 1992; Wakker, 2010). Empirical studies also favour cumulative prospect theory over prospect theory (e.g., Fennema and Wakker, 1997).

From a probabilistic perspective, a transformation function for a probability measure implicitly derives a transformation function for the corresponding random variables. Theorem 1 below shows that a CRWF $W(P)$ is linked to a transformation function that connects the primary outcomes presented in the choice experiment to the ones that are perceived by the decision-makers.

Theorem 1. The CRWF $W(P)=G[F^{-1}(P)]$ that transforms a survival function $F(t)$ to $G(t)$ derives a transformation function $H(T) = G^{-1}[F(T)]$ such that

(i) it transforms the primary outcome variable T with survival function F to the decision-maker's perceived outcome $S = H(T)$ with survival function $G(t) = F(H^{-1}(t))$;

(ii) $W(P)$ and $H(T)$ are connected via $W(P)=F(H^{-1}[F^{-1}(P)])$.

See Appendix A for proof. This theorem indicates that $W(P)$ and $H(T)$ can be mutually determined for a given underlying survival function $F(t)$. In practice, therefore, we only need to focus on one of them if $F(t)$ is specified.

2.3. Properties of CRWFs

In this section, we propose a large class of CRWFs and investigate some important properties about decision-makers' risky weighting.

2.3.1. A class of CRWFs

From Theorem 1, in order to fully specify a CRWF, the transformation $H(T)$ and underlying survival function $F(t)$ need to be determined. We consider an important case of perceptual conditioning where the decision-makers' perceived outcomes are transformed from T to S via:

$$H: \quad S = (aT)^\lambda \qquad \text{for } \lambda > 0 \text{ and } a > 0, \tag{5}$$

where following Prelec (1998), we restrict $\lambda > 0$. Clearly transformation function (5) ensures that the support $\Re^+$ of the primary outcome variable is invariant. According to Theorem 1, equation (5) implicitly derives a broad class of CRWFs for any survival function $F(.)$ as defined below:

$$\mathcal{A} = \left\{ W(P) := F\big(\{F^{-1}(P)\}^{1/\lambda}/a\big) \middle| \text{ for any survival function } F(.) \text{ on } \Re^+, \ \lambda > 0, a > 0 \right\}. \tag{6}$$

As it will be shown latter (see Theorems 2 and 3), the parameter λ in (6) plays a central role in determining the shape of the CRWFs belonging to class $\mathcal{A}$ and is termed risk perception parameter. By definition, for any survival function $F(t)$ on $\Re^+$, a CRWF $W(P) \in \mathcal{A}$ transforms $F(t)$ to

$$G(t) = \Pr\big((aT)^\lambda > t\big) = \Pr\big(T > t^{1/\lambda}/a\big) = F(t^{1/\lambda}/a) \quad \text{for } \lambda > 0 \text{ and } a > 0. \tag{7}$$

Consequently, choosing a different survival function $F(.)$ leads to a different CRWF. Hence, there are an infinite number of CRWFs belonging to class (6); as a result, the class of risky weighting functions (6) provides flexibility to model a wide range of practical problems.

Power transformation is usually a good approximation of many monotonically increasing functions that a decision-maker may use in practice. It is widely used in many economic analyses such as the well-known Cobb–Douglas production function. The power transformation can also be justified from a statistical perspective. This transformation is known as the Box-Cox transformation in statistical analysis. A more general form for the Box-Cox transformation is $S = (T^\lambda - 1)/\lambda$. According to Box

and Cox (1964), $(T^\lambda - 1)/\lambda$ is slightly preferable for theoretical analysis because it is continuous at $\lambda = 0$. In this paper, however, we follow Prelec (1998) and restrict $\lambda > 0$; a negative value of λ transforms gains into losses, which does not make any practical sense. Mathematically, an obvious advantage of using the power transformation (5) over $S = (T^\lambda - 1)/\lambda$ is that the former ensures the support $\Re^+$ of the primary outcomes invariant but the latter may lead to negative values. Hence we use (5) throughout this paper. Technically, the parameter a in equation (5) affects the scale of the perceived outcomes. In some applications, it is taken as 1 (i.e., without a scale transformation) or $\lambda^{-1/\lambda}$. Note that in the latter case, equation (5) becomes $S = T^\lambda/\lambda$. There are some other choices for the scale parameter a in the Box-Cox transformation; see, for example, Weisberg (2005).

Next, we turn to discuss some important theoretical properties for decision-makers' risky weighting. Prelec (1998) considered four important properties that CRWFs are expected to possess, i.e., (a) regressiveness, (b) asymmetry, (c) inverse-s shape, and (d) reflectiveness. The following theorem investigates these properties for any $W(P) \in \mathcal{A}$ with $\lambda > 1$.

Theorem 2A. For any $W(P) = F(\{F^{-1}(P)\}^{1/\lambda}/a) \in \mathcal{A}$ with $a = 1$ and $\lambda > 1$, we have

(i) $Q = W(P)$ is an increasing function of P with $W(0) = 0$ and $W(1) = 1$;

(ii)   For continuous $F(t)$, there exists a fixed-point (probability) $P^* = F(1)$ satisfying $P^* = W(P^*)$ such that $W(P) \geq P$ for any $P \in (0, P^*)$ and $W(P) \leq P$ for any $P \in (P^*, 1)$. If $F(t)$ is strictly decreasing and $0 < F(1) < 1$, then the fixed-point P^* is unique on the interval $(0, 1)$;

(iii)   The left-tail decision weight $q_1 = 1 - W(1 - p_1)$ is an increasing function of λ if $p_1 < 1 - P^*$; the right-tail decision weight $q_J = W(p_J)$ is also an increasing function of λ if $p_J < P^*$. Hence the CRWF *overweights* a small left-tail decision weight and a small right-tail decision weight;

(iv)   As the risk perception parameter λ becomes sufficiently large, the left-tail and right-tail decision weights $q_1 = 1 - W(1 - p_1)$ and $q_J = W(p_J)$ approach to $1 - P^*$ and P^* respectively; all the other decision weights q_j $(j = 2, ..., J - 1)$ vanish.

The proof of Theorem 2A is given in Appendix A. Property (i) ensures that the transformed function $G(t) = W(F(t))$ is a well-defined survival function. Properties (ii)-(iii) are related to the regressiveness and inverse-s shape natures of the cumulative risky weighting curve $Q = W(P)$ ($P \in [0,1]$): $P = W(P)$ holds at the two end-points $P = 0$ and 1 respectively; the curve also intersects the diagonal line of $Q = P$ at the fixed point P^*. In addition, the curve $Q = W(P)$ overweights a small left-tail probability and a small right-tail probability. Property (iv) further shows that $G(t)$ is approximately a U-shaped distribution for a large λ with its limiting distribution of a Bernoulli distribution that takes values at the two extreme ends only, i.e. t_1 and t_J.

Barberis (2013) emphasises that risky weighting means that the decision-makers overweight the two tails of any distribution (i.e., overweight unlikely extreme outcomes). This is consistent with Theorem 2A. Note that the fixed-point P^* of the asymmetry property in Prelec (1998) is expected to lie in the range of 1/3 to 1/2; the inverse-s shape feature in Prelec (1998) also requires that the curve $Q = W(P)$ be concave within an initial interval and be convex beyond that. However, because there is little restriction imposed on the shape of survival function $F(t)$ here, these strong properties do not generally hold. A further investigation will be undertaken in the next subsection.

The following theorem discusses the case of $0 < \lambda < 1$. The proof is given in Appendix A.

Theorem 2B. For any $W(P) = F(\{F^{-1}(P)\}^{1/\lambda}/a) \in \mathcal{A}$ with $a = 1$ and $0 < \lambda < 1$, we have

(i) $Q = W(P)$ is an increasing function of P with $W(0) = 0$ and $W(1) = 1$;

(ii)   For continuous $F(t)$, there exists a fixed-point (probability) $P^* = F(1)$ satisfying $P^* = W(P^*)$ such that $W(P) \leq P$ for any $P \in (0, P^*)$ and $W(P) \geq P$ for any $P \in (P^*, 1)$. If $F(t)$ is strictly decreasing and $0 < F(1) < 1$, then the fixed-point P^* is unique on the interval $(0, 1)$;

(iii)   The left-tail decision weight $q_1 = 1 - W(1 - p_1)$ is an increasing function of λ if $p_1 < 1 - P^*$; the right-tail decision weight $q_J = W(p_J)$ is also an increasing function of λ if $p_J < P^*$. Hence the CRWF *underweights* a small left-tail decision weight and a small right-tail decision weight;

(iv)   As the risk perception parameter $\lambda > 0$ becomes sufficiently small, the decision weights degenerate to a single-point distribution such that all the decision weights q_j vanish except for one, $q_{j_0} = 1$, where j_0 is the state with $P_{j_0} \geq 1 - P^*$ and $P_{j_0+1} \leq 1 - P^*$.

Clearly, the case of $0 < \lambda < 1$ describes an unusual circumstance in practical decision-making: decision-makers underestimate the tail probabilities. This can occasionally occur in reality. For example, if the respondents believe that the tail probabilities shown in a survey experiment are substantially larger than what they have experienced before, they may make some necessary adjustment for the decision weights; see Balcombe and Fraser (2015) for some examples.

2.3.2. A unified CRWF

There are a number of existing CRWFs in the literature, some of which were formulated in an *ad hoc* manner, whereas the others were developed on the basis of a set of axioms (e.g., Prelec, 1998). In this subsection, we propose a unified CRWF that includes several commonly used weighting functions as special cases. The proposed unified CRWF helps us understand the relationships among these existing risky weighing functions, and also differentiates risky weighting in CPT from EUT. We first examine two important CRWFs in relation to the power transformation (5).

Case I (Prelec, 1998). When the underlying survival function $F(t)$ is chosen as the exponential distribution $F(t) = \exp(-t/b)$, the power transformation (5) leads to a Weibull distribution, $G(t) = \exp[-t^{1/\lambda}/(ab)]$ and the following CRWF:

$$W(P) = F(H^{-1}[F^{-1}(P)]) = \exp(-\omega[-\log(P)]^{1/\lambda}) \qquad \text{with } \omega = b^{1/\lambda-1}a^{-1}, \qquad (8)$$

where a is the scale parameter in (5) and b is the parameter of the underlying distribution $F(t) = \exp(-t/b)$. Either a or b or both can be pre-specified. For example, as mentioned earlier, a can be taken as 1 or $\lambda^{-1/\lambda}$. If the primary outcomes are scaled to have a unit variance, then b is set as 1. Alternatively, ω can be estimated in numerical analysis.

$W(P)$ in (8) is termed Prelec II function in the literature. When ω is pre-set as 1, it is termed Prelec I function. In addition, we note that when $\lambda = 1$, the CRWF in (8) becomes

$$W(P) = P^\eta.$$

with $\eta = 1/a$. Hence, a power transformation of the probability is a special case of Prelec II function. This function was also considered in De Palma *et al.* (2008), Abdellaoui *et al.* (2010), and Beaud and Willinger (2015). In particular, taking $\eta = 1$, we obtain

$$W(P) = P. \tag{9}$$

This is the case corresponding to no risky weighting, a scenario considered in EUT.

Case II (Goldstein and Einhorn, 1987). When the underlying survival function $F(t)$ is chosen as the log-logistic distribution $F(t) = 1/(1 + t/b)$, the power transformation (5) leads to a log-logistic distribution, $G(t) = 1/[1 + t^{1/\lambda}/(ab)]$ and the Goldstein-Einhorn's CRWF:

$$W(P) = F(H^{-1}[F^{-1}(P)]) = P^{1/\lambda}/[P^{1/\lambda} + \omega(1 - P)^{1/\lambda}] \quad \text{with } \omega = b^{1/\lambda - 1}a^{-1}. \tag{10}$$

Two questions arise from a practical perspective: (a) how are these existing risky weighting functions related to each other? (b) how do we choose a CRWF $W(P)$ among many candidate functions?

Clearly, if the choice is restricted within the class $\mathcal{A}$ defined in equation (6), this is equivalent to the choice of the underlying survival function $F(t)$. To address these research issues, we propose a general CRWF that is based on the following underlying survival function:

$$F(t; \kappa) = 1/[1 + t/(\kappa b)]^{\kappa} \quad \text{for } t \geq 0 \quad \text{and } \kappa > 0. \tag{11}$$

In the literature, this distribution is termed type I generalised log-logistic distribution. See, e.g., Balakrishnan (1992, Chapter 9), for a detailed discussion of the statistical properties of type I generalised logistic distribution. $F(t; \kappa) = 1/[1 + t/(\kappa b)]^{\kappa}$ is also a scaled F-distribution with the first degree of freedom of 2, the second degree of freedom of 2κ, and scale parameter b.

The underlying distribution (11) derives the following CRWF belonging to class $\mathcal{A}$:

$$W(P) = F(H^{-1}[F^{-1}(P)])$$

$$= P^{1/\lambda}/[P^{1/(\kappa\lambda)} + \omega\kappa^{1/\lambda - 1}(1 - P^{1/\kappa})^{1/\lambda}]^{\kappa} \text{ with } \omega = b^{1/\lambda - 1}a^{-1}, \tag{12}$$

that transforms $F(t; \kappa)$ to $G(t; \kappa) = 1/[1 + t^{1/\lambda}/(\kappa ab)]^{\kappa}$.

The proposed CRWF (12) includes several important CRWFs in the literature as special cases:

(i)   As $\kappa \to +\infty$ with $\lambda = 1$ and $\omega = 1$, $F(t; \kappa) = 1/[1 + t/(\kappa b)]^\kappa$ in (11) approaches the exponential distribution $F(t) = \exp(-t/b)$. Hence, the CRWF in (12) approaches the identity function in (9);

(ii)   As $\kappa \to +\infty$, $F(t; \kappa) = 1/[1 + t/(\kappa b)]^\kappa$ in (11) approaches the exponential distribution $F(t) = \exp(-t/b)$. Hence, the CRWF in (12) approaches Prelec's function in (8);

(iii) As $\kappa = 1$, $F(t; \kappa) = 1/[1 + t/(\kappa b)]^\kappa$ collapses to the log-logistic distribution $F(t) = 1/(1 + t/b)$. Hence, the CRWF in (12) becomes the Goldstein-Einhorn's function in (10);

(iv)   Let $\tilde{P} = P^\eta$. Then the CRWF in (12) can be written as $\tilde{P}^{1/(\eta\lambda)}/[\tilde{P}^{1/(\eta\kappa\lambda)} + \omega\kappa^{1/\lambda-1}(1 - \tilde{P}^{1/(\eta\kappa)})^{1/\lambda}]^\kappa$. This is related to Wu-Gonzalez's CRWF, $P^{1/\lambda}/[P^{1/\lambda} + (1 - P)^{1/\lambda}]^\kappa$ (Wu and Gonzalez, 1996), and Tversky-Kahneman's function $W(P) = P^{1/\lambda}/[P^{1/\lambda} + (1 - P)^{1/\lambda}]^\lambda$ (Tversky & Kahneman, 1992).

Case (i) corresponds to the scenario where the risky weighting function is an identity function, i.e. the case of EUT where no transformation is involved. In empirical analysis, we can statistically test if the evidence is in favour of EUT against CPT by checking if $\lambda = 1$, $\omega = 1$ and $\kappa \to +\infty$.

In addition, from cases (ii) and (iii), Prelec's function and Goldstein-Einhorn's function are two special cases of the unified CRWF (12) with parameter κ at the two extreme spectrums of its range. Again, statistical tests can be performed to decide which risky weighting function is more preferable.

We also note that case (iv) can be considered as a combination of the power transformation and (12), i.e., a scale transformation ($\lambda = 1$ and $a = 1/\eta$) with $F(t; +\infty)$, followed by a power transformation ($\lambda \neq 1$ and $a = 1$) with $F(t; \kappa)$. It is of interest to note that this is also related to the source preference model developed in Fox and Tversky (1998) for the situation where people's choice depends not only on the degree of uncertainty but also on its source, i.e. $[W(P)]^\delta$, where $\delta > 0$ is a parameter that is inversely related to the attractiveness of the source. Clearly, risky weighting with source preference can be modelled as a combination of (12) and the power transformation, i.e., a power transformation ($\lambda \neq 1$ and $a = 1$) with $F(t; \kappa)$, followed by a scale transformation ($\lambda = 1$ and $a = 1/\delta$) with $F(t; +\infty)$.

From a perspective of numerical computation, since the unified CRWF in (12) includes these commonly used CRWFs as special cases, the choice for risky weighting function amounts to the choice

for parameters in (12). Consequently, instead of exploring each of the commonly used CRWFs separately in empirical analysis (like it was done in Hensher *et al.*, 2011), the choice for risky weighting function can be automatically made by estimating these parameters during the data analysis stage, leading to a data-driven model-selection approach.

To better understand the proposed risky weighting function, we further explore the four important characteristics about CRWFs investigated in Prelec (1998). We show below that the unified CRWF (12) has much stronger properties in comparison with that in Theorems 2 and 3. We consider the situation where the primary outcome variable has been re-scaled to have a unit scale parameter $b = 1$ so that the following conclusions do not rely on the measurement unit of the primary outcome variable.

Theorem 3. For the unified CRWF (12) with $a = 1$ and $b = 1$, there exists a unique fixed-point (probability) $P^* = F(1; \kappa)$ satisfying $P^* = W(P^*)$ such that

(i) $0.367 < P^* < 0.5$ for $\kappa \in (1, +\infty)$;

(ii) $P^* = 0.5$ for $\kappa = 1$;

(iii) $0.5 < P^* < 1$ for $\kappa \in (0, 1)$;

(iv)   For $\lambda > 1$, $W(P)$ is concave and $W(P) > P$ for any $P \in (0, P^*)$, and $W(P)$ is convex and $W(P) < P$ for any $P \in (P^*, 1)$;

(v)   For $0 < \lambda < 1$, $W(P)$ is convex and $W(P) < P$ for any $P \in (0, P^*)$, and $W(P)$ is concave and $W(P) > P$ for any $P \in (P^*, 1)$.

The proof of Theorem 3 is given in Appendix A. The upper and lower left panels of Figure 1 illustrate Theorem 3 parts (i), (iii) and (iv), where the function $Q = W(P)$ with λ=2 is plotted against P using (12) (the real line) for $\kappa = 30$ (upper left) and for $\kappa = 0.2$ (lower left). This is compared with the straight-line $Q = P$ (the dotted line); the latter is the case where the risk perception does not involve any risky weighting, i.e. equation (9). Likewise, the upper and lower left panels of Figure 2 illustrate Theorem 3 parts (i), (iii) and (v), where the function $Q = W(P)$ with λ=0.5 is plotted against P using (12) (the real line) for $\kappa = 30$ (upper left) and for $\kappa = 0.2$ (lower left).

In addition, to demonstrate the impact of different risk perception parameter values, we consider a number of mass probabilities all equal to $p_j = 0.10$ ($j = 1, \ldots, 10$). This represents the scenario where

the primary outcome at all states of nature has an equal probability (i.e. 0.10). We then apply the risky

weighting $Q = W(P)$ in (12) with several different values of λ, i.e. 2, 4, and 10, to transform these

probabilities p_j to obtain decision weights q_j ($j = 1, \dots, 10$). The resulting decision weights are

displayed in the upper and lower right panels of Figure 1, where the original probability (0.10) and the

decision weights transformed with λ= 2, 4, and 10 are plotted. Note that the four probability values for

each of the states are clustered in the bar chart (bars from left to right at each state of nature are the

original probability of 0.1 and the associated risky weighting with λ=2, 4, and 10 respectively). Take

the lower right panel of Figure 1 as an example: the furthermost bar on the far right at each state shows

a J-shaped distribution of the decision weights which are transformed with λ=10. On the other hand,

the decision weights without risky weighting have a uniform distribution as displayed by the furthermost

bar on the far left at each state.

From Theorem 3, the asymmetry of the fixed point depends on the range of parameter κ, which in

turn has an important implication to the risk attitude of the decision-makers. Specifically, Theorem 3 (i)

shows that the fixed-point P^* is less than 0.5 for any $\kappa \in (1, +\infty)$, as displayed in the upper left graph

of Figure 1. From Theorem 2A, the left-tail decision weight $q_1 = 1 - W(1 - p_1)$ approaches to $1 -$

$P^* \in [0.5, 0.632]$ as the risk perception parameter λ becomes large. When the left-tail and right-tail

probabilities are equal, i.e. $p_1 = p_J$, the asymmetry in Theorem 3 (i) leads to a higher left-tail decision

weight q_1 than the right-tail q_J, as displayed by Figure 1 (upper right). As the risk perception parameter

λ becomes sufficiently large, the right-tail (or left-tail) decision weight approaches $P^* = F(1; 30) =$

0.374 (or $1 - P^* = 0.626$) for $\kappa = 30$. When the primary outcomes are gains, this suggests that the

perceptual conditioning tends to increase more in risk-aversion for losses than gains.

On the other hand, Theorem 3 (iii) shows that for $0 < \kappa < 1$, the fixed-point $P^* = F(1; \kappa)$ lies in

the range of $0.5 < P^* < 1$, as displayed in the lower left graph of Figure 1. As λ becomes large, the

right-tail (or left-tail) decision weight approaches $P^* = F(1; 0.2) = 0.699$ (or $1 - P^* = 0.301$) for

$\kappa = 0.2$. In other words, with $P^* > 0.5$, the decision weights are J-shaped. Hence, when the primary

outcomes are gains, the perceptual conditioning tends to increase more in risk-aversion for gains than

losses.

In addition, Theorem 3 (ii) with $\kappa = 1$ is the Goldstein-Einhorn's CRWF for which the decision-maker's risk attitude does not favour either gains or losses. Because the parameter κ reflects the degrees of asymmetry of the fixed-point P^*, as well as the skewness of the decision weights, it is termed an asymmetry parameter in this paper.

Theorem 3 (iv) shows the regressiveness and inverse-s shape natures of the unified CRWF in (12) when $\lambda > 1$. It is important to note that the fixed-point and inflection point coincide, as illustrated in the upper left and lower left graphs of Figure 1.

Finally, Theorem 3 (v), corresponding to Theorem 2B with $0 < \lambda < 1$, describes the scenario that decision-makers underweight the tail probabilities. From the upper left and lower left graphs of Figure 2, the cumulative risky weighting curve still shows the regressiveness with a unique fixed-point that coincides the inflection point. It is also clear that the curve is s-shaped (rather than inversely s-shaped). The upper right (or lower right) graph of Figure 2 shows that, as λ becomes sufficiently small, all the decision weights approaches to 0 except for one state. See also, Abdellaoui *et al.* (2010), for a discussion about the link between risky weighting function and probabilistic risk attitudes was discussed.
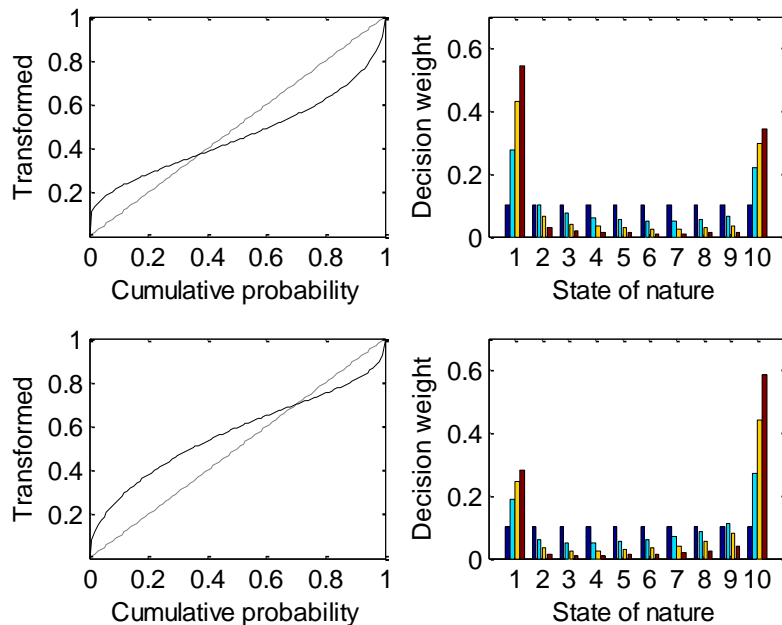

Figure 1. Overweighted tail probabilities: the left panels show CRWF $W(P)$ of the square transformation ($\lambda = 2$) in (12) when $\kappa = 30$ (upper left) and $\kappa = 0.2$ (lower left); The right panels show decision weights for the original equal mass probabilities $p_j = 0.1$ (for all states j) and using (12) with $\lambda = 2, 4, 10$ (bars from left to right at each state respectively) when $\kappa = 30$ (upper right) and $\kappa = 0.2$ (lower right).
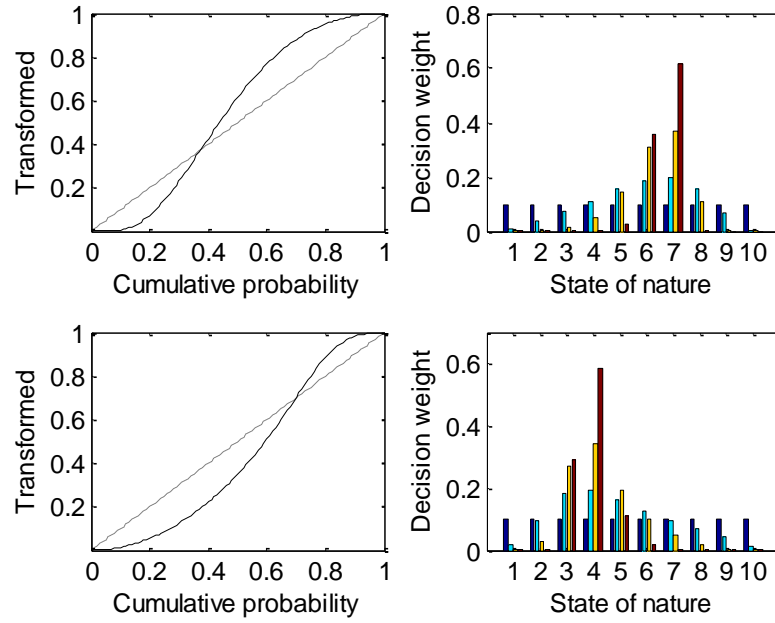
Figure 2. Underweighted tail probabilities: the left panels show CRWF $W(P)$ of the square root transformation ($\lambda = 0.5$) in (12) when $\kappa = 30$ (upper left) and $\kappa = 0.2$ (lower left); The right panels show decision weights for the original equal mass probabilities $p_j = 0.1$ (for all states j) and using (12) with λ=1/2, 1/4, 1/10 (from left to right at each state respectively) when $\kappa = 30$ (upper right) and $\kappa = 0.2$ (lower right).

In practice, most empirical evidence supports inverse-s probability weighting but s-shape weighting functions were also documented in the literature. See Prelec (1998), Balcombe and Fraser (2015) and the references therein for further discussion on inverse-s and s shapes of CRWFs.

We also note that in Theorem 3, the unique fixed-point (probability) P^* for $\kappa \in (1, +\infty)$ is bounded in the range of $P^* \in (0.367, 0.5)$. This is consistent with Prelec (1998) that indicates the fixed-point P^* lies in the range of 1/3 to 1/2. It seems that the lower bound of 0.367 may be restrictive in some applications. However, the fixed-point P^* falls in this range only for the unit scale parameter a in transformation (5). The following corollary shows that the lower bound of the fixed-point is also partially determined by the scale parameter a in transformation (5).

Corollary. For the unified CRWF (12) with $b = 1$, the unique fixed-point P^* satisfying $P^* = W(P^*)$ is given by $P^* = F(a^{\lambda/(\lambda-1)}; \kappa)$. Hence for $\lambda > 1$ and $a > 1$, we have $P^* \in (0, 0.5)$ for $\kappa \in (1, +\infty)$.

The proof of the Corollary is given in Appendix A. This shows that the scale parameter affects the location of the fixed-point P^*: by allowing the scale parameter a to vary, the fixed-point P^* can take values anywhere in the range of $(0, 0.5)$.

Remarks. (i) In many choice survey instruments, the number of states, J, of a contract $g_i = \left(\{p_j\}_{j=1}^J, \{t_{ij}\}_{j=1}^J\right)$ is usually set to be small (typically 2 to 4) to avoid overcomplicated choice situations that could confuse the respondents. The information provided in the survey instruments is hence not sufficient to elicit the entire survival function $F(t)$. We show in Appendix B that, when eliciting prior knowledge with limited information, the distribution (11) is the optimal underlying survival function that maximises a constrained entropy. In other words, the unified CRWF (12) is optimal in the sense of the maximum entropy when choosing a CRWF within class $\mathcal{A}$.

(ii) Before concluding this section, we very briefly discuss the scenario of $T \in \mathfrak{R}$. Two approaches can be used here. First, we note that the log-transformation converts any nonnegative primary outcome variable on $\mathfrak{R}^+$ to a real-valued outcome variable on the real line $\mathfrak{R}$. Hence, one approach is to use the log-transformation so that all the results obtained in this section can be applied. A more popular approach in the literature (see, e.g., Wakker, 2010) is to consider a piecewise CRWF:

$$W(P) = \begin{cases} W^+(P) & \text{if } T \geq 0 \\ W^-(P) & \text{if } T < 0 \end{cases}.$$

Clearly, for each of $W^+(P)$ and $W^-(P)$, the proposed method in this section can be used. The parameters in the unified CRWF (12) for $W^+(P)$ and $W^-(P)$ are usually set differently. As people may view gains and losses very differently (Tversky and Kahneman, 1992), the second approach is more preferable in many applications.

3.   Modelling of choice behaviour with risky prospects

In this section, we will investigate the modelling of decision-makers' choice behaviour with risky prospects on the basis of the unified CRWF (12).

We consider the stochastic form of CPT. To describe the stochastic nature of travellers' choices, we incorporate the approach of discrete choice analysis. There are three components to be considered within

this research framework, i.e. risky weighting function, value function and link function. We show in this section that the power utility function and the multiplicative discrete choice model are a natural choice for the class of risky weighting functions in (6).

3.1. *Value function*

In the previous section we focused on CRWFs. We now consider the next component in CPT, i.e. the value function. We assume that for given primary outcomes over $\Re^+$, a decision-maker's value (or utility) function of choosing a prospect is a random variable. We break down the value function into a non-stochastic systematic part $\bar{v}(t; \mu, \tau)$ termed subutility and a stochastic component ε for the given primary outcomes. The two parts are assumed to connect each other via the following multiplicative model:

$$v(t; \mu, \tau) = \bar{v}(t; \mu, \tau)\, \varepsilon. \tag{13}$$

In this subsection, we focus on the subutility $\bar{v}(t; \mu, \tau)$. The stochastic component ε will be modelled via the link function in the next subsection.

For the non-stochastic subutility function, Tversky and Kahneman (1992) considered the constant relative risk aversion (CRRA) utility function:

$$\bar{v}(t; \mu, \tau) = \mu t^{1-\tau}/(1 - \tau) \tag{14}$$

where τ is risk attitude coefficient and μ is an individual specific parameter. The CRRA utility is also considered in many other studies, e.g., Prelec (1998), Gao *et al*. (2010), Hensher *et al*. (2011), and Beaud and Willinger (2015).

As shown in Theorem 1, perceptual conditioning is a process that the decision-makers transform the primary outcome variable T into their perceived outcome $S = H(T)$. Under this transformation, the subutility $\bar{v}(S; \mu, \tau)$ becomes $\bar{v}(H(T); \mu, \tau)$:

$$\bar{v}(H(T); \mu, \tau) = \mu[H(T)]^{(1-\tau)}/(1 - \tau). \tag{15}$$

Hence, the consistency in functional form of the value (utility) function is retained via the power transformation if the decision-maker incorporates the CRRA utility form. This is summarised below.

Theorem 4. Suppose that the subutility function $\bar{v}(.\,;\mu,\tau)$ for a decision-maker is of the CRRA form as given in (14). Then the power transformation in (5) ensures that the subutility function $\bar{v}(H(.)\,;\mu,\tau)$ is also of the CRRA form.

To characterise each decision-maker's value function, we assume that the parameter μ in equation (14) is linked to a vector of covariates termed attributes $\mathbf{x}_{in} = [x_{in1}, \ldots, x_{inl}]^T$, denoted as μ_{in}:

$$\mu_{in} = \exp(\boldsymbol{\beta}^T \mathbf{x}_{in}), \tag{16}$$

where $\boldsymbol{\beta} = [\beta_{1,\ldots,}\beta_l]^T$ is a vector of parameters.

Now we consider the expected value function that integrates the two components of CPT, i.e. the risky weighting and value functions. By definition, the expected value function with respect to the primary outcome variable is given by $E = -\int v(t;\mu,\tau)dG(t)$. In the case of a finite number of states of nature, it reduces to (3), i.e.

$$E = \sum_{j=1}^{J} v(t_{ij};\mu_{in},\tau)q_j(\lambda)$$

for each prospect i, where $q_j(\lambda) = W(P_j;\lambda) - W(P_{j+1};\lambda)$ and $W(P;\lambda) = P^{1/\lambda}/[P^{1/(v\lambda)} + \omega v^{1/\lambda-1}(1 - P^{1/v})^{1/\lambda}]^v$ is given by (12). Consequently, from equation (13), the value function of decision-maker n for choosing prospect i is

$$v_{in} = \bar{v}(\mathbf{x}_{in};\boldsymbol{\beta},\lambda,\tau)\varepsilon_{in} = \varepsilon_{in}\exp(\boldsymbol{\beta}^T \mathbf{x}_{in})(1-\tau)^{-1}\sum_{j=1}^{J} t_{ij}^{(1-\tau)} q_j(\lambda), \tag{17}$$

with the subutility $\bar{v}(\mathbf{x}_{in};\boldsymbol{\beta},\lambda,\tau) = \exp(\boldsymbol{\beta}^T \mathbf{x}_{in})(1-\tau)^{-1}\sum_{j=1}^{J} t_{ij}^{(1-\tau)} q_j(\lambda)$.

3.2. *Link function and choice probability*

Now we turn to consider the stochastic component ε_{in} in the multiplicative model (17).

We consider the scenario where a decision-maker n is facing a decision-making problem with several *risky* prospects. Rather than assume that the prospects under investigation are riskless in conventional discrete choice analysis (see, e.g., Hensher *et al.* 2015; Train 2009), we investigate a more general scenario where each prospect is risky and characterised by a contract (gamble) $g_i = \left(\{p_j\}_{j=1}^{J}, \{t_{ij}\}_{j=1}^{J}\right)$. Following CPT, we take into consideration risky weighting of the decision-maker so that his/her decision-making is based on $v_{in} = \bar{v}(\mathbf{x}_{in};\boldsymbol{\beta},\lambda,\tau)\varepsilon_{in}$ in (17). A prospect i is chosen if $v_{in} \geq max_{k \in C_n}\{v_{kn}\}$. The probability that this event occurs,

$$P_{in} = \Pr\{v_{in} \geq max_{k \in C_n} v_{kn}\} \ , \tag{18}$$

is used to characterise the decision-making problem (see, e.g., Ben-Akiva and Lerman, 1985).

To derive an explicit form of choice probability (18), we need to specify a statistical model for the stochastic component ε_{in} of the value function in equation (17). Specifically, for the multiplicative model (17), we follow Fosgerau and Bierlaire (2009) and Li (2011), and assume that the independent error terms ε_{in} have a Weibull distribution with $\alpha > 0$ the scale parameter of $\log(\varepsilon_{in})$. When $\tau > 1$, the choice probability (18) is given by

$$P_{in} = P_{in}(\boldsymbol{\beta}, \lambda, \tau, \kappa | \mathbf{x}_n)$$

$$= \Pr\{v_{in} \geq max_{k \in C_n} v_{kn}\} = \Pr\{-\log(-v_{in}) \geq max_{k \in C_n} -\log(-v_{kn})\}$$

$$= \frac{\exp\left(-\alpha\{\boldsymbol{\beta}^T \mathbf{x}_{in} + |1-\tau| + \log[\sum_{j=1}^{J} t_{ij}^{(1-\tau)} q_j(\lambda)]\}\right)}{\sum_{k \in C_n} \exp(-\alpha\{\boldsymbol{\beta}^T \mathbf{x}_{kn} + |1-\tau| + \log[\sum_{j=1}^{J} t_{kj}^{(1-\tau)} q_j(\lambda)]\})} \ ,$$

where $\mathbf{x}_n$ is a collection of all $\mathbf{x}_{kn}$ for $k \in C_n$. We note that the common term $|1 - \tau|$ is cancelled out in the above expression. Hence, we obtain

$$P_{in}(\boldsymbol{\beta}, \lambda, \tau, \kappa | \mathbf{x}_n) = \frac{\exp\left(-\alpha\{\boldsymbol{\beta}^T \mathbf{x}_{in} + \log[\sum_{j=1}^{J} t_{ij}^{(1-\tau)} q_j(\lambda)]\}\right)}{\sum_{k \in C_n} \exp(-\alpha\{\boldsymbol{\beta}^T \mathbf{x}_{kn} + \log[\sum_{j=1}^{J} t_{kj}^{(1-\tau)} q_j(\lambda)]\})} \ .$$

Similar to the above analysis, we can obtain the following choice probability for $\tau < 1$:

$$P_{in}(\boldsymbol{\beta}, \lambda, \tau, \kappa | \mathbf{x}_n) = \Pr\{v_{in} \geq max_{k \in C_n} v_{kn}\} = \Pr\{\log(v_{in}) \geq max_{k \in C_n} \log(v_{kn})\}$$

$$= \frac{\exp\left(\alpha\{\boldsymbol{\beta}^T \mathbf{x}_{in} + \log[\sum_{j=1}^{J} t_{ij}^{(1-\tau)} q_j(\lambda)]\}\right)}{\sum_{k \in C_n} \exp(\alpha\{\boldsymbol{\beta}^T \mathbf{x}_{kn} + \log[\sum_{j=1}^{J} t_{kj}^{(1-\tau)} q_j(\lambda)]\})} \ .$$

Let $\rho = 1 - \tau$. The above two expressions can be written in a unified form:

$$P_{in}(\boldsymbol{\beta}, \lambda, \rho, \kappa | \mathbf{x}_n) = \frac{\exp\left(\alpha \, sign(\rho)\{\boldsymbol{\beta}^T \mathbf{x}_{in} + \log[\sum_{j=1}^{J} t_{ij}^{\rho} q_j(\lambda)]\}\right)}{\sum_{k \in C_n} \exp(\alpha \, sign(\rho)\{\boldsymbol{\beta}^T \mathbf{x}_{kn} + \log[\sum_{j=1}^{J} t_{kj}^{\rho} q_j(\lambda)]\})} \ ,$$

with $\alpha > 0$. $sign(\rho)$ is the sign function of ρ.

Finally, if the primary outcomes are losses rather than gains, a prospect i is chosen if $-v_{in} \geq max_{k \in C_n} \{-v_{kn}\}$. We can then obtain the following result by some simple manipulations:

$$P_{in}(\boldsymbol{\beta}, \lambda, \rho, \kappa | \mathbf{x}_n) = \frac{\exp\left(-\alpha \, sign(\rho)\{\boldsymbol{\beta}^T \mathbf{x}_{in} + \log[\sum_{j=1}^{J} t_{ij}^{\rho} q_j(\lambda)]\}\right)}{\sum_{k \in C_n} \exp(-\alpha \, sign(\rho)\{\boldsymbol{\beta}^T \mathbf{x}_{kn} + \log[\sum_{j=1}^{J} t_{kj}^{\rho} q_j(\lambda)]\})} \ .$$

We offer the following remarks before concluding this section. First, although the CRRA model has been shown to be a natural choice in Section 3.1, this is not a restriction on the model specification; the value function and risky weighting function can be chosen separately.

Secondly, the multiplicative model in (13) essentially leads to a weibit link funciton (Fosgerau and Bierlaire, 2009; Li, 2011); the weibit link is transformed into the logit link function via the log-transformation. Li (2011) proposed a distribution family consisting of an infinite number of link functions and including both logit and weibit links as special cases. We will discuss some potential extensions with this distribution family in Section 6.

4. Elicitation of risky weighting and value functions

In this section, we discuss the elicitation of risky weighting and value functions via statistical inference.

4.1. *Individual-specific risk perception*

So far, we have implicitly assumed that all decision-makers have a common risk perception parameter λ. In many applications, this assumption may not be realistic (DePalma *et al*., 2008; Glöckner and Pachur, 2012). We now relax this assumption. In doing so, the elicitation of the risky weighting function will be more accurate.

Specifically, rather than assume the same transformation function $H(.)$ applying to all the decision-makers, we incorporate a more realistic assumption that each decision-maker n has his/her own risk perception parameter λ_n, and hence the transformation is denoted as $S = H(T; \lambda_n) = (a_n T)^{\lambda_n}$ in equation (5) with the individual-specific risk perception parameter $\lambda_n > 0$ being written explicitly. For simplicity, here we consider the case that a_n is constant across all the individuals, i.e. $a_n = a$; otherwise, a_n or its log-transformation $\log(a_n)$ need to be included into the parameter vector $\boldsymbol{\theta}_n$ to be defined in equation (19) below. The corresponding CRWF is denoted as $W(P; \lambda_n)$. The decision weights in (2) are denoted as $q_j(\lambda_n) = w_j(p_1, \dots, p_J; \lambda_n)$ $(j = 1, \dots, J)$.

Let $\varphi_n = \log(\lambda_n)$. This transforms a positive parameter λ_n to a real-valued parameter φ_n. The following statistical model is specified for parameter φ_n for individual-specific risk perception:

$$\varphi_n \sim p(\varphi_n; \psi_0, \sigma^2),$$

where ψ_0 is the location parameter and σ is the scale parameter characterising the centre of parameter φ_n across the whole population as well as its variability.

In practice, distribution $p(\varphi_n; \psi_0, \sigma^2)$ can be specified as a normal distribution; this is equivalent to stipulating a log-normal distribution for λ_n. However, we point out that the methodology developed in this paper does not impose such a restriction; any reasonable distribution can be used. In addition, the risk perception parameter can potentially be linked to a number of covariates. See Section 6 for further discussion.

In many applications, the coefficients in vector $\boldsymbol{\beta}$ can be treated as random variables. This approach is particularly useful for the scenario where each decision-maker faces more than one choice situation (which is typical in many choice experiments). In this case, the error term ε_{in} and hence value function $v_{in} = \bar{v}(\mathbf{x}_{in}, ; \boldsymbol{\beta}, \lambda)\varepsilon_{in}$ in equation (17) are not independent of each other. We now follow McFadden and Train (2000) and use a random-coefficient discrete choice model to deal with this problem.

In the random-coefficient choice model, each decision maker n has a different coefficient vector, denoted as $\boldsymbol{\beta}_n = [\beta_{1n,\dots,}\beta_{ln}]^T$. $\boldsymbol{\beta}_n$ is assumed to follow a statistical distribution, $\boldsymbol{\beta}_n \sim p(\boldsymbol{\beta}_n; \tilde{\tilde{\boldsymbol{\xi}}})$, with an unknown parameter vector $\tilde{\tilde{\boldsymbol{\xi}}}$. A commonly used specification for the distribution of $\boldsymbol{\beta}_n$ is a normal (or lognormal) distribution. In this case, $\tilde{\tilde{\boldsymbol{\xi}}}$ includes both mean vector $\tilde{\boldsymbol{\psi}} = [\psi_1, \dots, \psi_K]^T$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$. In the case of normal distributions for $\boldsymbol{\beta}_n$, parameters $\psi_1, \dots, \psi_K$ represent the population means of $\boldsymbol{\beta}_n$ which are usually of primary interest in practice.

Unless it is known *a priori* that λ_n and $\boldsymbol{\beta}_n$ are independent of each other, we can in general assume a joint distribution for them:

$$\boldsymbol{\theta}_n \sim p(\boldsymbol{\theta}_n; \boldsymbol{\psi}, \boldsymbol{\Sigma}), \tag{19}$$

where $\boldsymbol{\theta}_n = [\varphi_n, \boldsymbol{\beta}_n^T]^T$ and $\boldsymbol{\psi} = [\psi_0, \tilde{\boldsymbol{\psi}}^T]^T$. The covariance matrix $\boldsymbol{\Sigma}$ consists of σ^2, $\tilde{\boldsymbol{\Sigma}}$, and the covariances of φ_n and $\boldsymbol{\beta}_n$.

We follow Train (2009) and extend the developed model to the case where the coefficient vector $\boldsymbol{\beta}_n$ is a set of random variables. When the primary outcomes are gains, the choice probabilities are:

$$P_{in}(\boldsymbol{\beta}_n, \lambda_n, \rho, \kappa|\mathbf{x}_n) = \frac{\exp\left(\alpha sign(\rho)\{\boldsymbol{\beta}_n^T \mathbf{x}_{in} + \log[\sum_{j=1}^J t_{ij}{}^\rho q_j(\lambda_n)]\}\right)}{\sum_{k \in C_n} \exp(\alpha\, sign(\rho)\{\boldsymbol{\beta}_n^T \mathbf{x}_{kn} + \log[\sum_{j=1}^J t_{kj}{}^\rho q_j(\lambda_n)]\})} \qquad (20)$$

with $\alpha > 0$. The results for the case that the primary outcomes are losses can be worked out similarly:

$$P_{in}(\boldsymbol{\beta}_n, \lambda_n, \rho, \kappa|\mathbf{x}_n) = \frac{\exp\left(-\alpha sign(\rho)(\{\boldsymbol{\beta}_n^T \mathbf{x}_{in} + \log[\sum_{j=1}^J t_{ij}{}^\rho q_j(\lambda_n)]\})\right)}{\sum_{k \in C_n} \exp(-\alpha\, sign(\rho)\{\{\boldsymbol{\beta}_n^T \mathbf{x}_{kn} + \log[\sum_{j=1}^J t_{kj}{}^\rho q_j(\lambda_n)]\}\})} \;. \qquad (21)$$

Equation (20) (or (21)) is a multivariate model for choice analysis with risky prospects. It includes a generalised linear equation (16) that links the value function to a number of attributes in **x**; it also characterises risk perception by the individual-specific risk perception parameter λ_n.

4.2. Statistical inference

To elicit the risky weighting function and the other components in CPT, a commonly used approach is the maximum likelihood method (see, e.g., Wakker, 2010). Recently, Toubia, *et al.* (2013) have considered Bayesian parameter estimation for CPT. Li (2011), on the other hand, have considered Bayesian multinomial choice modelling.

In this sub-section, we briefly outline a simple Bayesian approach to statistical inference for the developed model. As it will be seen, a Bayesian approach is particularly convenient when dealing with individual-specific random coefficients within the context of choice experiments.

We consider a general situation where each decision-maker n in a choice experiment faces K different choice situations. Let $\mathbf{y}_n = [i_1, \dots, i_K]^T$ denote the corresponding sequence of the choice alternatives in the K choice situations.

It is well known that a logit model is identifiable only up to a scale parameter (see, e.g., Hensher *et al.*, 2015; Train, 2009; Fosgerau and Bierlaire, 2009; Li, 2011). Hence, we set the scale parameter α to be unity in this paper.

Let $\boldsymbol{\phi}$ denote the collection of all fixed effect parameters, including ρ and κ, and $\varphi = \log(\lambda)$ (if a common risk perception is assumed), plus the coefficients of any fixed effects in the attribute vector **x**. Let $\boldsymbol{\theta}_n$ denote the collection of all random coefficients, including $\varphi_n = \log(\lambda_n)$ (if assuming an individual-specific risk perception parameter), and the random coefficients in $\boldsymbol{\beta}_n$.

For each decision-maker n, the contribution to the likelihood is given by:

$$L_n(\boldsymbol{\theta}_n, \boldsymbol{\phi}| \mathbf{y}_n, \mathbf{x}_n) = \prod_{k=1}^K P_{i_k n}(\boldsymbol{\theta}_n, \boldsymbol{\phi}|\mathbf{x}_n),$$

with $P_{i_k n}(\boldsymbol{\theta}_n, \boldsymbol{\phi}|\mathbf{x}_n)$ is given by (20) or (21).

The above likelihood depends on the unknown vector $\boldsymbol{\theta}_n$ which is treated as a latent vector in the Bayesian analysis and specified by (19). Hence, the joint distribution for decision-maker n is given by $L_n(\boldsymbol{\theta}_n, \boldsymbol{\phi}| \mathbf{y}_n, \mathbf{x}_n)p(\boldsymbol{\theta}_n; \boldsymbol{\psi}, \boldsymbol{\Sigma})$. Taking into account all the decision-makers, we can write out the likelihood function as $\prod_{n=1}^N L_n(\boldsymbol{\theta}_n, \boldsymbol{\phi}| \mathbf{y}_n, \mathbf{x}_n)p(\boldsymbol{\theta}_n; \boldsymbol{\psi}, \boldsymbol{\Sigma})$.

To fully specify the problem, we also need the prior distributions of $\boldsymbol{\phi}$, $\boldsymbol{\psi}$ and $\boldsymbol{\Sigma}$, denoted as $\pi(\boldsymbol{\phi})$ and $\pi(\boldsymbol{\psi}, \boldsymbol{\Sigma})$ respectively. When no external information is available in the analysis, non-informative priors are used for $\boldsymbol{\psi}$ and $\boldsymbol{\Sigma}$. Following Train (2009) and Gelman *et al*. (2009), the prior for $\boldsymbol{\psi}$ can be chosen as $\pi(\boldsymbol{\psi}|\boldsymbol{\Sigma}) \propto 1$ and the prior for $\boldsymbol{\Sigma}$ can be chosen to be inverse Wishart: $\pi(\boldsymbol{\Sigma}) = IW(K, \mathbf{I})$, where K is the dimension of $\boldsymbol{\theta}_n$ and $\mathbf{I}$ is the identity matrix of size K.

Then we apply Bayes' rule to obtain the posterior distribution

$$p(\boldsymbol{\theta}_n, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\Sigma}| \mathbf{y}_n, \mathbf{x}_n, \forall n) \propto \pi(\boldsymbol{\phi})\pi(\boldsymbol{\psi}, \boldsymbol{\Sigma}) \prod_{n=1}^N L_n(\boldsymbol{\theta}_n, \boldsymbol{\phi}| \mathbf{y}_n, \mathbf{x}_n)p(\boldsymbol{\theta}_n; \boldsymbol{\psi}, \boldsymbol{\Sigma}). \qquad (22)$$

The above posterior distribution is not analytically tractable, and hence it needs to be evaluated numerically. The Markov chain Monte Carlo (MCMC) method with data augmentation can be used. We outline this iterative method as follows.

Let $\boldsymbol{\psi}^{(k)}$, $\boldsymbol{\Sigma}^{(k)}$, $\boldsymbol{\phi}^{(k)}$, $\boldsymbol{\theta}_n^{(k)}$ denote the draws of $\boldsymbol{\psi}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}_n$ at the kth iteration of the MCMC. We set the initial guess $\boldsymbol{\phi}^{(0)}$, $\boldsymbol{\psi}^{(0)}$ and $\boldsymbol{\Sigma}^{(0)}$ for $k = 0$. Then for each k=1, 2, …, we draw a sample alternately: (a) simulate vectors $\boldsymbol{\theta}_n^{(k)}$ from (22) for given $\boldsymbol{\phi}^{(k-1)}$ $\boldsymbol{\psi}^{(k-1)}$ and $\boldsymbol{\Sigma}^{(k-1)}$; and (b) simulate parameters $\boldsymbol{\phi}^{(k)}$, $\boldsymbol{\psi}^{(k)}$ and $\boldsymbol{\Sigma}^{(k)}$ from (22) for given $\boldsymbol{\theta}_n^{(k)}$. This continues until convergence.

The simulation in step (a) can be based on the Metropolis-Hasting algorithm (see, e.g., Gelman *et al*., 2009). The implementation for the simulation in step (b) depends on the assumption about $p(\boldsymbol{\theta}_n; \boldsymbol{\psi}, \boldsymbol{\Sigma})$ in equation (19) and specification of the priors $\pi(\boldsymbol{\phi})$ and $\pi(\boldsymbol{\psi}, \boldsymbol{\Sigma})$. When the location parameters are assumed to follow a normal distribution and the scale parameters are assumed to follow an inverse Wishart (or inverse gamma) distribution, Gibbs sampler can be embedded into the algorithm; see Train (2009) for details. Following Gelman *et al*. (2009), we use the deviance information criterion (DIC) for model comparison and selection in the Bayesian analysis.

5. Application

In this section, we consider an empirical application on road tolling to illustrate the developed method. Road tolling has attracted a growing number of choice model applications in the last two decades. See the general overviews and discussion of Saleh (2005), Palma *et al*. (2006), and the references therein; see, in particular, Li and Hensher (2010) for an overview on road tolling in Australia. The application considered here is based on a stated preference survey undertaken in Australia. In the literature, choice analysis with risky prospects has been investigated with both stated preference data (e.g. Hess *et al*., 2008; Razo and Gao, 2013) and revealed data (e.g. Hu *et al*., 2012).

5.1. Data

The data used in the analysis of this section were drawn from a large-scale study undertaken in 2008 in Brisbane, Australia, in the context of toll vs. free roads, which utilised a stated choice experiment involving two alternatives (i.e., route A and route B) pivoted around the knowledge base of travellers (i.e., the current trip). The original large-scale study was designed and implemented to capture a large number of travel circumstances and to determine how each individual traveller trades-off different levels of travel times and trip time variability with various levels of proposed tolls and vehicle running costs, in the context of tolled and non-tolled roads. 752 respondents were included into the original analysis. Each respondent considered 16 choice situations, hence resulting in 12,032 choice situations in total.

The primary outcome variable in this analysis was travel time; it was assumed to be subject to a great deal of uncertainty, depending on the traffic conditions. In the survey the respondents were advised that departure time remained unchanged and the prospects were described with three travel scenarios (states of nature), i.e., 'arriving earlier than expected', 'arriving at the time expected', and 'arriving later than expected'; each was associated with a corresponding probability of occurrence. Table 1 gives an example of the payoffs table and Figure 3 is an illustrative choice scenario.

Table 1. Example of payoffs (travel time in hours) in one choice situation of a traveller with the corresponding probability in parentheses

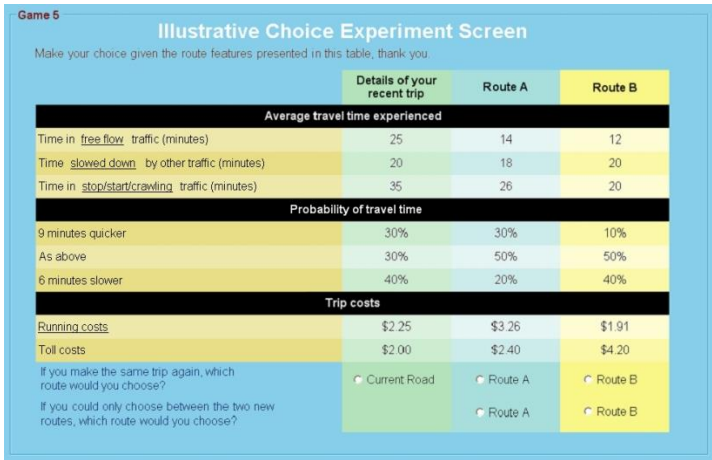| State of nature	arriving earlier than expected	arriving at the time expected	arriving later than expected
Prospect 1	0.28 (0.2)	32 (0.5)	42 (0.3)
Prospect 2	0.37 (0.4)	0.40 (0.5)	0.50 (0.1)
…………			

Figure 3 Illustrative Choice Scenario

There were in total 734 travellers included into the analysis below after the initial data pre-screening. Table 2 gives descriptive statistics of the sample on the payoffs (travel times in hours) and the corresponding probabilities. We also included other attributes in the analysis, namely running costs (x_1) and toll costs (x_2), as well as the socioeconomic profile of the respondents: (a) age (x_3) (years); (b) gender (x_4) (male =1 and female=0); (c) hours of work in a typical week (x_5); (d) annual personal income before tax under \$10,000 (x_6) (binary indicator); (e) annual personal income before tax over \$120,000 (x_7) (binary indicator). Hence, we have $\mathbf{x}=[x_1, x_2, x_3, x_4, x_5, x_6, x_7]^T$. The summary statistics of these attributes are displayed in Table 3.

<div align="center">

Table 2. Travel times and probabilities of occurrence

</div>

	Mean	Std. Dev.	Min	Max
Early arrival time (hours)	0.631	0.324	0.117	4.583
Probability of arriving earlier than expected	0.25	0.11	0.10	0.40
Arrival on time (hours)	0.717	0.358	0.167	4.750
Probability of arriving at the time expected	0.50	0.15	0.20	0.80
Late arrival time (hours)	0.891	0.453	0.183	6.067
Probability of arriving later than expected	0.25	0.11	0.10	0.40

<div align="center">

Table 3. Descriptive statistics of the attributes

</div>

	Running costs (x_1)	Toll costs (x_2)	Age (x_3)	Gender (x_4)	Work hours per week (x_5)	Income under $10,000 ($x_6$)	Income over $120,000 ($x_7$)
Mean	$4.51	$2.10	42.70	0.57	30.75	0.09	0.07
Std. Dev.	$7.48	$1.37	14.96	0.50	18.56	0.29	0.25

In the following analysis, variables x_3 (age) and x_5 (work hours per week) were divided by 100 so that all the socio-economic factors have a similar magnitude to each other.

5.2. Model and prior specification

5.2.1. Models

In the following analysis, the value function was chosen as the CRRA utility function (14), $\bar{v}(t; \mu, \tau) = \mu t^{1-\tau}/(1-\tau)$. The generalised linear model for the value function (16), $\mu_{in} = \exp(\boldsymbol{\beta}_n{}^T \mathbf{x}_{in})$, was incorporated, where $\mathbf{x}_{in}$ is a vector of factors that potentially influenced the value function. In the analysis, $\mathbf{x}_{in}$ initially included all seven attributes in Table 3 and hence $\boldsymbol{\beta}_n = [\beta_{1n}, \dots, \beta_{7n}]^T$. The unified CRWF in equation (12) was used, $W(P) = P^{1/\lambda}/[P^{1/(\kappa\lambda)} + \omega\kappa^{1/\lambda-1}(1 - P^{1/\kappa})^{1/\lambda}]^\kappa$. The choice probability model was chosen as equation (21):

$$P_{in}(\boldsymbol{\beta}_n, \lambda, \rho, \kappa | \mathbf{x}_n) = \frac{\exp\left(-sign(\rho)(\{\boldsymbol{\beta}_n^T \mathbf{x}_{in} + \log[\sum_{j=1}^J t_{ij}{}^\rho q_j(\lambda_n)]\})\right)}{\sum_{k \in C_n} \exp\left(-sign(\rho)\{\{\boldsymbol{\beta}_n^T \mathbf{x}_{kn} + \log[\sum_{j=1}^J t_{kj}{}^\rho q_j(\lambda_n)]\}\right)} \tag{23}$$

where t_{ij} are the primary outcomes (travel times). Note that as mentioned in the previous sections, we have $\rho = 1 - \tau$. In addition, the parameter α in (21) is scaled to unity. Importantly, there is no beta coefficient for travel time. This is because the multiplicative model is used here, equivalent to the case that each element of $\boldsymbol{\beta}_n$ is divided by this coefficient. In addition, we set $a = 1$ and set b to be the average travel time across all respondents and alternatives so that the analysis was independent of the

measurement unit of the primary outcome variable. This also allows us to compare with the existing risky weighting functions, such as Prelec's and Goldstein-Einhorn functions, more easily.

Two scenarios were considered: (a) the CRWF with a risk perception parameter λ common to all individuals; and (b) the individual-specific risky weighting functions where each individual traveller has his/her own value of risk perception parameter λ_n in order to reflect the individual effect on perceptual conditioning. Following Train (2009), we also consider running cost and total cost to be random effects in Model (b). Hence, two models were investigated:

- Model (a): The perception parameter λ is a fixed-effect parameter and x_1 - x_7 are all fixed effects; Hence, $\boldsymbol{\beta}_n$ for $n = 1, \dots, 734$ are all equal, denoted as $\boldsymbol{\beta} = [\beta_1, \dots, \beta_7]^T$;

- Model (b): The risk perception parameter λ_n is an individual-specific random parameter; the coefficients of variables x_1 and x_2 are also individual-specific random variables (denoted as β_{1n} and β_{2n}). The other five attributes were fixed effects. Hence $\boldsymbol{\beta}_n = [\beta_{1n}, \beta_{2n}, \beta_3, \dots, \beta_7]^T$.

5.2.2. *Prior specification*

Next, we specify the priors of the parameters in the models. The prior of coefficient ρ was set as a non-informative prior $\pi(\rho) \propto 1$.

As mentioned early, when the asymmetry parameter κ in (12) is sufficiently large, it approaches Prelec's function (8). In the subsequent analysis, the range of κ for both models was restricted to be (0, 100), and the prior of the asymmetry parameter κ was chosen to be a uniform distribution on (0, 100). Clearly, any estimate of κ close to the upper bound of 100 is in favour of Prelec's weighting function.

For Model (a), we choose the prior of $\boldsymbol{\beta}$ to be non-informative, i.e. $\pi(\boldsymbol{\beta}) \propto 1$. From a practical perspective, only a few transformations in the Box-Cox transformation are usually of interest, such as square transformation and cubic transformation (see, e.g., Weisberg, 2005). In the analysis we restricted the range of λ to be less than or equal to 5 in transformation (5). In addition, from Theorem 2A, we further restricted $\lambda \geq 1$ in the analysis. Hence, the prior of $\varphi = \log(\lambda)$ was set as a uniform distribution on $[0, \log(5)]$. This range is usually wide enough to include all the values of the perception parameter of practical importance.

For Model (b), the risk perception parameter λ_n was treated as a random coefficient. We specified the distribution of λ_n as log-normal $LN(\psi_0, \sigma^2)$ restricted to interval $[1, 5]$; hence, the prior distribution of $\varphi_n = \log(\lambda_n)$ was assumed to be a normal distribution $N(\psi_0, \sigma^2)$ truncated to the interval $[0, \log(5)]$. For Model (b), the coefficients of x_1 and x_2 were also treated as individual-specific random variables and were denoted as β_{1n} and β_{2n} respectively. For each respondent n, we specified the distribution of $[\beta_{1n}, \beta_{2n}]^T$ as a normal distribution:

$$[\beta_{1n}, \beta_{2n}]^T \sim N([\psi_1, \psi_2]^T, \widetilde{\Sigma}), \tag{24}$$

where $\widetilde{\Sigma}$ is a 2×2 covariance matrix of the normal distribution. Let $\psi = [\psi_0, \psi_1, \psi_2]^T$ and Σ be the covariance of φ_n and $[\beta_{1n}, \beta_{2n}]^T$. Hence, ψ_1 and ψ_2 are the population means of β_{1n} and β_{2n}. The prior of ψ and Σ, $\pi(\psi, \Sigma) = \pi(\psi|\Sigma)\pi(\Sigma)$, was taken as non-informative prior: $\pi(\psi|\Sigma) \propto 1$ and $\Sigma \sim IW(3, \mathbf{I})$ respectively, where IW is an inverse-Wishart distribution. $\mathbf{I}$ is a 3×3 identity matrix. In addition, we chose a non-informative prior for β_j, $\pi(\beta_j) \propto 1$ (j=3,…,7). See Train (2009) for further discussion on prior specification.

The total number of iterations of the MCMC was chosen as 10000, among which the first 5000 iterations were treated as the burn-in period and the draws in this period were discarded. The results reported below were based on the rest of the 5000 draws. The posterior means were used as the estimates of the parameters.

5.3. Results

Table 4 displays the main empirical results. We first consider Model (a). It can be seen from Table 4 that the coefficients of all the socio-economic factors, β_j ($j = 3, …, 7$), are statistically insignificant at the 5% level. Hence, there is no evidence suggesting that these factors influenced the travellers' choice. However, the coefficients associated with the running costs x_1 and toll costs x_2 are both significant at the 5% level: $\hat{\beta}_1 = 0.591$ and $\hat{\beta}_2 = 0.667$. Hence there is evidence that the running costs and toll costs affected the value function of the respondents and hence their choice probabilities. Turning to Model (b), we can see that it outperformed Model (a) in terms of the likelihood and DIC. Like Model (a), the coefficients of the socio-economic factors in Model (b), β_j ($j = 3, …, 7$), are insignificant at the 5% level.

For Model (b), the coefficients of running costs x_1 and toll costs x_2 for each traveller n, β_{1n} and β_{2n} ($n = 1, \ldots, 734$), are random variables. From equation (24), ψ_1 and ψ_2 are the population means of the random coefficients β_{1n} and β_{2n} across all the individual travellers $n = 1, \ldots, 734$, representing the average effects of the two travel-cost variable (running costs and toll costs). From Table 4, we can see that the population means associated with the running costs x_1 and toll costs x_2 are significant at the 5% level: $\hat{\psi}_1 = 0.733$ and $\hat{\psi}_2 = 0.771$. In addition, the estimate of the population mean of the logarithm risk perception parameter is also significant at the 5% level, $\hat{\psi}_0 = 0.856$.

Table 4. Estimates of parameters with 95% Bayesian credible intervals

	coefficient	Model (a)	Model (b)	Model (c)
mean log-risk parameter	ψ_0		0.856	0.638
			(0.683, 0.999)	(0.370, 0.949)
log-risk parameter	φ	0.762		
		(0.244, 1.376)		
mean running cost effect	ψ_1		0.733	0.730
			(0.651, 0.822)	(0.629, 0.834)
mean toll cost effect	ψ_2		0.771	0.759
			(0.700, 0.849)	(0.680, 0.846)
running cost effect	β_1	0.591		
		(0.538, 0.637)		
toll cost effect	β_2	0.667		
		(0.620, 0.702)		
Age effect	β_3	-0.050	-0.168	-0.224
		(-0.304, 0.234)	(-0.378, 0.064)	(-0.339, -0.105)
Gender effect	β_4	-0.037	0.001	
		(-0.226, 0.181)	(-0.093, 0.094)	
work hour per week	β_5	-0.081	-0.068	
		(-0.221, 0.118)	(-0.300, 0.152)	
income under \$10,000	β_6	-0.151	-0.057	
		(-0.397, 0.062)	(-0.250, 0.119)	
income over \$120,000	β_7	0.017	-0.073	
		(-0.074, 0.109)	(-0.289, 0.131)	
asymmetry parameter	κ	7.371	7.376	9.328
		(3.776, 10.827)	(4.302,10.112)	(3.206, 15.232)
risk attitude	τ	0.943	0.945	0.948
		(0.821, 0.998)	(0.839, 0.999)	(0.843,0.997)
Log-likelihood		-6628.0	-5625.3	-5645.7
DIC		13228.1	12375.4	12368.9

Note: ψ_j is the mean of random coefficient β_{jn} (j=1,2) and ψ_0 is the mean of random coefficient $\varphi_n = \log(\lambda_n)$ averaged across all individual travellers n (n=1,…, 734)

Next, we removed insignificant socio-economic factors, one at a time, from Model (b), resulting in a refined model, Model (c). Variable x_3 (age) became significant after having removed the other socio-economic factors, suggesting young and senior travellers have different value functions. The lower left graph in Figure 4 displays the posterior distribution of the coefficient β_3 of variable x_3 (age).

The estimated population means of the two travel-cost related random coefficients, i.e., running costs and toll costs, are $\hat{\psi}_1 = 0.730$, $\hat{\psi}_2 = 0.759$ for Model (c), respectively; both are significant at the 5% level. See the upper left and upper right graphs of Figure 4 for the posterior distributions of ψ_1 and ψ_2. Noting the negative sign in front of the exponent in equation (23), the positive $\hat{\psi}_1$ and $\hat{\psi}_2$ indicate that higher levels of the running costs and toll costs corresponds to lower probability of choosing that particular route.

The estimated covariance matrix of $\boldsymbol{\psi} = [\psi_0, \psi_1, \psi_2]^T$ is

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.123 & -0.004 & 0.008 \\ -0.004 & 0.308 & 0.248 \\ 0.008 & 0.248 & 0.403 \end{bmatrix}.$$

It can be seen that the population means of the two travel-cost related random coefficients, ψ_1 and ψ_2, are closely correlated each other with correlation coefficient of 0.704, whereas the relationship between ψ_0 and (ψ_1, ψ_2) is much weaker: the correlation coefficients are -0.022 and 0.034 respectively; both of them are insignificant at the 5% level. The estimated τ in Model (c) is 0.948 with a 95% credible interval of (0.843,0.997).
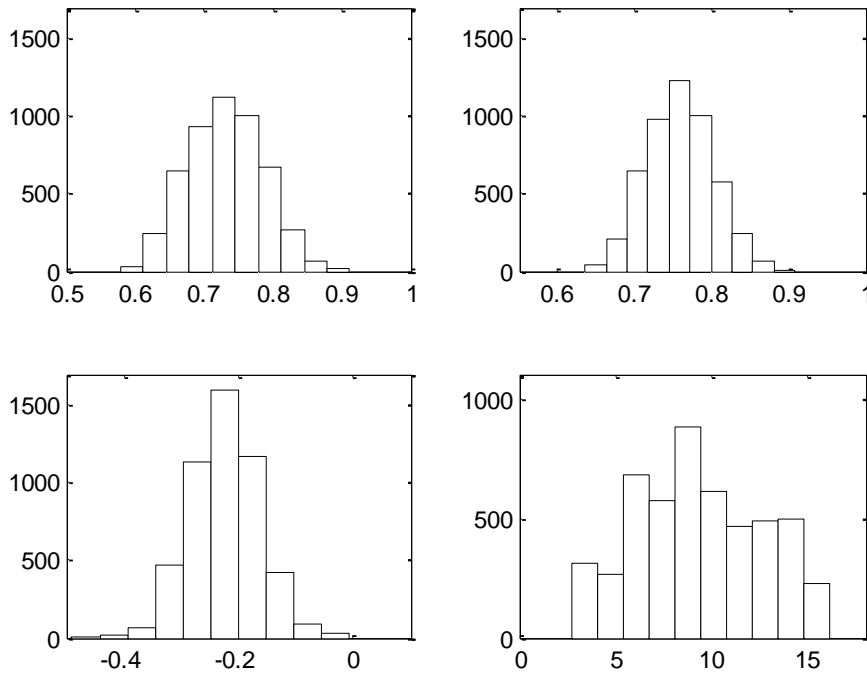
Figure 4. Posterior distributions of ψ_1 (upper left), ψ_2 (upper right), β_3 (lower left) and κ (lower right).

We now turn to consider the elicited risky weighting function for Model (c). The estimated asymmetry parameter is $\hat{\kappa} = 9.328$ with 95% Bayesian credible interval of (3.206, 15.232); see the lower right graph of Figure 4 for the posterior distribution of κ. This suggests that for this particular case, the elicited CRWF is neither Prelec's (with $\kappa = +\infty$) nor Goldstein-Einhorn's (with $\kappa = 1$) function. This has demonstrated an advantage of using the unified CRWF (12): rather than use the approaches in the existing studies, e.g., Stott (2005), Hensher *et al.* (2011) and Rasouli and Timmermans (2014), where a number of CRWFs were explored one by one, here we use a data-driven approach to choosing a CRWF.

The average risk perception parameter $\hat{\lambda} = \exp(\hat{\psi}_0) = 1.893$ is significantly different from 1, with a 95% credible interval of (exp(0.370), exp(0.949))=(1.448, 2.583). This piece of evidence about the risk perception parameter λ, in conjunction with $\hat{\kappa} = 9.328$ with 95% Bayesian credible interval of (3.206, 15.232), suggests that the risky weighting function is not the identity function $W(P) = P$ in (9). Hence, the survey data provides evidence of risky weighting: the respondents indeed transformed the

given probabilities into their own decision weights when making their choices. Hence, the evidence is in favour of CPT against EUT for this particular study.

Figure 5 displays the elicited CRWF when ψ_0 takes the values of the posterior mean $\hat{\psi}_0 = 0.638$, and the two end-points of the 95% Bayesian credible interval respectively. It is evident from Figure 5 that the two CRWFs corresponding to the end-points of the 95% Bayesian credible interval do not overlap with the diagonal straight-line; the latter is the identity function $W(P) = P$.

The asymmetry of the fixed-point is evident from Figure 5: $P^* = 0.387$. This has great implications to risky weighting: when the survival probability $P_j = \sum_{k \geq j} p_k$ is large (close to 1), the transformed survival probability $Q_j = \sum_{k \geq j} q_k$ will be substantially lower than the original survival probability P_j. As a result, the left-tail decision weight $q_1 = 1 - W(1 - p_1)$ is substantially larger than the original left-tail probability p_1.

To gain further insight into the elicited risky perceptual weighting, we consider a concrete example with the mean probabilities of 'early arrival', 'arrival on time', and 'late arrival' events in Table 1, i.e. $p_1 = 0.25$, $p_2 = 0.50$, and $p_3 = 0.25$ respectively. With the average risk perception parameter $\hat{\lambda} = \exp(\hat{\psi}_0) = 1.893$, the obtained CRWF transforms these three probabilities to $q_1 = 0.44$, $q_2 = 0.29$ and $q_3 = 0.27$ respectively. Hence, the risky weighting in this analysis substantially overestimated the left-tail probability (from $p_1 = 0.25$ to $q_1 = 0.44$), whereas it kept the right-tail probability almost the unchanged: $p_3 = 0.25$ compared to $q_3 = 0.27$. This seems to indicate that the respondents tended to be more optimistic about the required travel time than what was shown in the survey questionnaire. This is a phenomenon well documented in the literature (Tversky and Kahneman, 1992; Prelec, 1998).
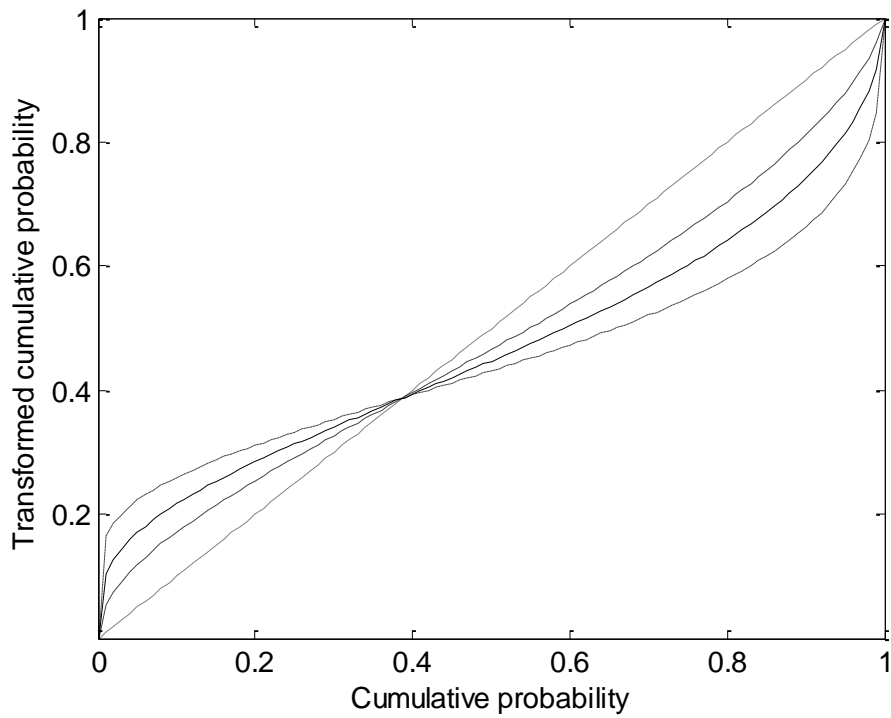
Figure 5. Elicited unified CRWF in (12) scaled to $b = 1$ with risk perception parameter λ equal to the posterior mean exp(0.638) (——), and equal to the two end-points of the 95% Bayesian credible interval, i.e. exp(0.370) (– –), and exp(0.949) (– · – ·) respectively.
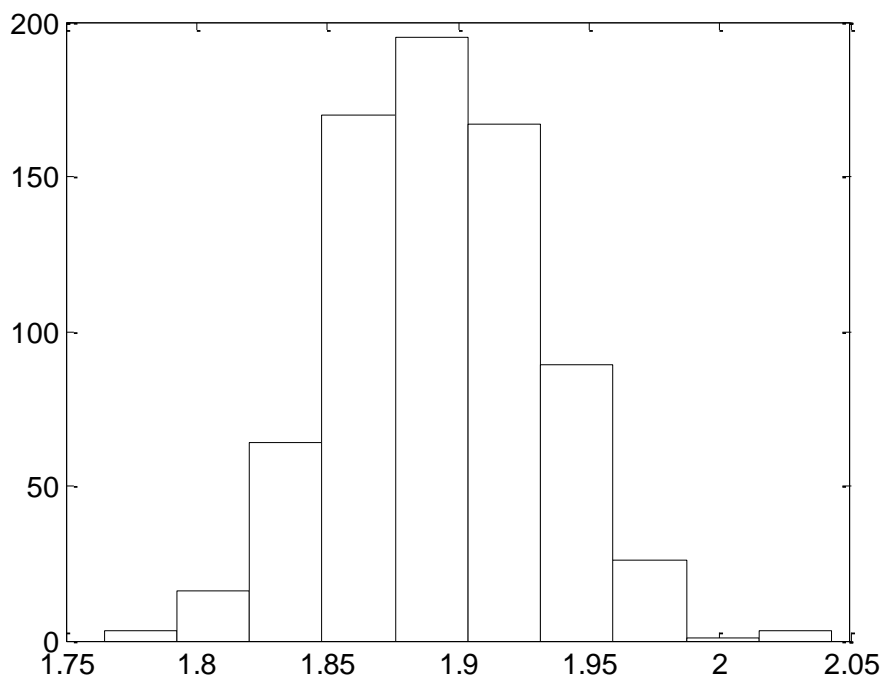

Figure 6. Histogram of the estimated values of the individual-specific risk perception parameter λ_n

across the 734 respondents.

To show the variability of the risk perception parameter, Figure 6 displays the histogram of the estimated λ_n $(n = 1, ..., 734)$ across the individual travellers. It can be seen that most values of the individual-specific risk perception parameter λ_n lie within the interval [1.85, 1.95], suggesting a transformation close to the square transformation. In addition, it is reinforced by Figure 6 that a restriction of λ_n on the interval of (0,5] is a reasonable choice; the likelihood that λ_n is larger than 5 is negligible.

In summary, the case study in this section provides empirical evidence of risky weighting for travellers' route choices as the empirical results are in favour of CPT relative to EUT. In the existing literature, there are some comparisons between CPT and EUT based models (e.g., Gao et al. 2010). However, these studies do not address this issue statistically. Secondly, the case study shows that neither the Prelec's risky weighting function nor Goldstein-Einhorn's function is a suitable choice to model risky weighting for the problem considered here; in this case, the developed unified CRWF outperforms these specific forms. Again, there are some existing studies that compare the performances of various risky weighting functions (e.g., Hensher et al. 2011). However, without a proper statistical model, these studies cannot address the issue formally and statistically. The results also suggest that if the researcher only selects risky weighting functions from a few existing forms, the results could potentially be biased. Finally, the study also provides evidence that risk perception varies from traveller to traveller, supporting the argument in De Palma et al. (2008). Hence, assuming a uniform risk perception parameter, like it was done in the most existing literature, could be problematic.

6.   Concluding remarks

In this paper, we have investigated the modelling of decision-makers' risky weighting and choice behaviour when they are facing several risky prospects. We have proposed a broad class of CRWFs and investigated some important properties of the CRWFs within the class. On the basis of this, we have developed a unified CRWF that includes several important CRWFs in the literature as special cases. We have also shown that in practice, the choice for CRWFs within this class comes down to the selection of the asymmetry parameter κ. In addition, the asymmetry parameter κ, together with the individual risk

perception parameter λ_n, can be used to test about CPT and provides empirical evidence about risky weighting. Hence, the CRWFs can be used to differentiate CPT from EUT in empirical studies.

We have developed a multivariate method for discrete choice analysis with risky prospects to account for decision-makers' individual-specific risk perception and the impact of various factors on decision making. Individual-specific risk perception makes much more practical sense than a common risk perception parameter because in reality people's risk perception does vary from person to person. The multivariate method also provides an approach to a better understanding about how decision-makers form their value function.

The research in this paper can be extended to accommodate more complicated problems. We outline a couple potential future research directions below. First, it is well documented in the literature that individuals' risk perception could be affected by some common psychological factors and social factors (see, e.g., Fox and Tversky 1998; Weber and Hsee, 1998; Sjöberg, 2000; Braisby and Gellatly, 2012; among many others). An important extension of the research in this paper is to link the individual risk perception parameter λ_n to a number of psychological and social covariates, i.e. $\lambda_{in} = \exp(\boldsymbol{\gamma}^T \mathbf{z}_{in})$, where $\mathbf{z}_{in}$ is a vector of covariates and $\boldsymbol{\gamma}$ is the vector of the corresponding coefficients. The modelling of risk perception parameter will give us an opportunity to investigate the impact of various psychological and social factors on risky weighting.

Second, we have focused on the multiplicative model, equation (13), where the link function is weibit and the underlying distribution of the stochastic part is a Weibull distribution. Li (2011) developed a large class of link functions where the underlying distribution does not have to be a Weibull distribution. Importantly, this model class includes both the additive model and multiplicative model as special cases. Clearly, extensions can be made to substantially relax the assumption of the multiplicative model used in this paper.

Finally, similar to the multinomial logit model, the scale parameter α in models (22) and (23) is assumed to be the same for all decision-makers and all alternatives throughout this paper. In some applications, this may not be realistic. The issue of scale heterogeneity across individuals has been

investigated in the literature (see, e.g., Greene and Hensher, 2010). Following these existing studies, the method developed in this paper can be extended accordingly.

Appendix A. Proofs of the theorems

Proof of Theorem 1. The result in (i) follows by noting that $G(t) = \Pr(S > t) = \Pr(H(T) > t) = \Pr(T > H^{-1}(t)) = F(H^{-1}(t))$ and hence $H(T) = G^{-1}[F(T)]$. Substituting $G(t) = F(H^{-1}(t))$ into $W(P)=G[F^{-1}(P)]$ yields $W(P)=F(H^{-1}[F^{-1}(P)])$. This completes the proof.

Proof of Theorem 2A. (i) The results of $W(0) = 0$ and $W(1) = 1$ are immediate from the fact that $F(0) = 1$ and $F(+\infty) = 0$. In addition, since $F(t)$ is a decreasing function, we have $F^{-1}(P_1) \geq F^{-1}(P_2)$ for any $P_1 < P_2$. We further have $\{F^{-1}(P_1)\}^{1/\lambda} \geq \{F^{-1}(P_2)\}^{1/\lambda}$. This implies that $F(\{F^{-1}(P_1)\}^{1/\lambda}) \leq F(\{F^{-1}(P_2)\}^{1/\lambda})$, i.e. $W(P_1) \leq W(P_2)$. Hence, $W(P)$ is an increasing function.

(ii) Because survival function $F(t)$ is a decreasing function, we obtain

$$W(P) \leq P \quad \text{if and only if} \quad \{F^{-1}(P)\}^{1/\lambda} \geq F^{-1}(P).$$

Hence, we have either $F^{-1}(P) = 0$ or $F^{-1}(P) \leq 1$ since $\lambda > 1$ and $F^{-1}(P) \geq 0$ (note that the domain of $F(t)$ is $\Re^+$). From $F^{-1}(P) \leq 1$ we obtain $P \geq F(1) = P^*$. Similarly, we can show that $W(P) \geq P$ if $P \leq P^*$. Due to the continuity of $W(P)$, there exists a fixed-point $P^* = F(1)$ satisfying $P^* = W(P^*)$. Furthermore, if $F(t)$ is a strictly decreasing function, then we have $\{F^{-1}(P)\}^{1/\lambda} = F^{-1}(P)$ from $P^* = W(P^*)$. For $\lambda > 1$, the equation $x^{1/\lambda} - x = 0$ has a unique real root of 1 on interval $(0, +\infty)$. Hence, $P^* = F(1)$ is the unique fixed-point if $0 < F(1) < 1$.

(iii) For any $1 \leq \lambda_1 \leq \lambda_2$, we have $\{F^{-1}(P)\}^{1/\lambda_1} \geq \{F^{-1}(P)\}^{1/\lambda_2}$ if $F^{-1}(P) \geq 1$. Noting that $F(t)$ is a decreasing function, we obtain that $F(\{F^{-1}(P)\}^{1/\lambda_1}) \leq F(\{F^{-1}(P)\}^{1/\lambda_2})$ if $F^{-1}(P) \geq 1$, i.e., $P \leq F(1) = P^*$. Hence, for all $\lambda \geq 1$, $W(P)$ is an increasing function of λ if $P \leq P^*$. The final conclusion in (iii) is immediate from the fact that $q_1 = p_1$ and $q_J = p_J$ when $\lambda = 1$.

(iv) It follows immediately by noting $\{F^{-1}(P)\}^{1/\lambda}$ approaches to 1 for any $P \in (0, 1)$ as λ becomes sufficiently large. This completes the proof.

Proof of Theorem 2B. The proof of (i) is the same as Theorem 2A (i).

(ii) Clearly $W(P) \leq P$ if and only if $\{F^{-1}(P)\}^{1/\lambda} \geq F^{-1}(P)$. Hence, we have either $F^{-1}(P) = 0$ or

$F^{-1}(P) \geq 1$ since $\lambda < 1$. This implies $P \leq F(1) = P^*$. The remaining part is the similar to the

counterpart of Theorem 2A(ii).

(iii) The proof that $W(P)$ is an increasing function of λ for all $0 < \lambda \leq 1$ is the same as that of Theorem

2A (iii). The conclusion in (iii) is immediate as $q_1 = p_1$ and $q_J = p_J$ when $\lambda = 1$.

(iv) As λ becomes sufficiently small, $\{F^{-1}(P)\}^{1/\lambda}$ approaches to 0 (or $+\infty$) if $F^{-1}(P) < 1$ (or

$F^{-1}(P) > 1$). Hence $W(P) = F(\{F^{-1}(P)\}^{1/\lambda})$ approaches to 1 (or 0) if $F^{-1}(P) < 1$ (or $F^{-1}(P) > 1$).

This indicates that decision weights $q_j = W(P_j) - W(P_{j+1}) = 0$ if $P_{j+1} > P^*$ or $P_j < P^*$. Hence,

there exists an index j_0 such that $W(P_{j_0}) = 1$ and $(P_{j_0+1}) = 0$, i.e. $q_{j_0} = 1$.

Proof of Theorem 3.

(i) Since $F(t; \kappa)$ is a strictly decreasing function of t, the fixed-point $P^* = F(1; \kappa) = 1/[1 + 1/\kappa]^\kappa$

given in Theorems 2A and 2B is unique. In addition, because $1/[1 + 1/\kappa]^\kappa$ is monotonic in κ, we have

$e^{-1} \leq F(1; \kappa) \leq 0.5$ for any $\kappa \in (1, +\infty)$. The proof of (iii) is similar and proof of (ii) is trivial.

(iv) It is easy to verify that

$$\partial^2 W(P)/\partial P^2 = \lambda^{-2}\kappa^{-1}\{1 + \kappa^{-1}[\kappa(P^{-\frac{1}{\kappa}} - 1)]^{\frac{1}{\lambda}}\}^{-\kappa-2}[\kappa(P^{-\frac{1}{\kappa}} - 1)]^{\frac{1}{\lambda}-2}P^{-\frac{1}{\kappa}-2}J(P),$$

where

$$J(P) = [\kappa(P^{-\frac{1}{\kappa}} - 1)]^{\frac{1}{\lambda}}\left\{(\kappa + \lambda)P^{-\frac{1}{\kappa}} - \lambda(\kappa + 1)(P^{-\frac{1}{\kappa}} - 1)\right\} + \kappa(\lambda - 1)P^{-\frac{1}{\kappa}} - \lambda\kappa(\kappa + 1)(P^{-\frac{1}{\kappa}} - 1).$$

It is straightforward to verify that $J(P^*) = 0$ and hence $\partial^2 W(P)/\partial P^2 = 0$ at P^*. Therefore, P^* is an

inflection point of $W(P)$ for $P \in [0, 1]$. In addition, we have $J(1) = \kappa(\lambda - 1) > 0$ and $J(0) \rightarrow -\infty$.

Finally, by some algebra it can be shown that $J(P)$ is a strictly increasing function of P. Therefore P^* is

the unique root of $J(P)$ on $[0, 1]$, and hence it is the unique inflection point of $W(P)$ on $[0, 1]$.

(v) The proof is similar to that of (iv).

Proof of Corollary.

By definition, the fixed-point P^* satisfies $W(P^*) = P^*$, i.e. $\{F^{-1}(P^*; \kappa)\}^{\frac{1}{\lambda}}/a = F^{-1}(P^*; \kappa)$. Hence, we

obtain $F^{-1}(P^*; \kappa) = a^{\lambda/(\lambda-1)}$ or equivalently $P^* = F(a^{\lambda/(\lambda-1)}; \kappa)$. From the proof of Theorem 3,

$1/[1 + 1/\kappa]^{\kappa}$ is monotonic in κ. Hence, we obtain we have $F\left(a^{\lambda/(\lambda-1)}; +\infty\right) \leq P^* \leq F\left(a^{\lambda/(\lambda-1)}; 1\right)$

for any $\kappa \in (1, +\infty)$. When $\lambda > 1$ and $a > 1$, we have $a^{\lambda/(\lambda-1)} > 1$. Hence $F\left(a^{\lambda/(\lambda-1)}; +\infty\right)$

approaches 0 as $a^{\lambda/(\lambda-1)}$ becomes large. In addition, $F\left(a^{\lambda/(\lambda-1)}; 1\right)$ approaches $F(1; 1) = 0.5$ as

$a^{\lambda/(\lambda-1)}$ approaches 1. Hence, we obtain $0 \leq P^* \leq 0.5$.

Appendix B. The maximum-entropy principle

The maximum entropy principle is widely incorporated in the literature for statistical inference with

limited prior knowledge; see Golan *et al*. (1996) for a general review, and Park and Bera (2009) for an

example in econometric modelling. In Bayesian analysis, it is also used to work out prior probability

measures (see, e.g.,, Gelman *et al*., 2009). Here we consider the problem of choosing an underlying

survival function $F(t)$ with very limited information in a choice experiment.

We wish to find a probability density function $f(t|\eta)$ for the primary outcome variable $T \in \mathfrak{R}^+$

with the mean parameter η. When the mean η is fully specified, it is well known that the maximum

entropy distribution is the exponential distribution $f(t|\eta) = \eta^{-1} \exp(-t/\eta)$ (see, e.g., Golan *et al*.,

1996; Park and Bera, 2009). Now we consider the problem from a Bayesian perspective and assume that

the mean of the underlying distribution $f(t|\eta)$ is not completely specified; rather, it is subject to random

variation and it follows a prior distribution $\pi(\eta)$ which is also unknown but some information about η

is available and is expressed as the moment conditions. Consequently, the objective is to find the joint

distribution $g(t, \eta) = f(t|\eta)\pi(\eta)$ from the maximum entropy principle. The following lemma shows

that under some conditions, we can determine a joint distribution $g^*(t, \eta) = f(t|\eta)\pi(\eta)$ from the

maximum entropy principle, upon which the marginal probability density function of $T \in \mathfrak{R}^+$ can be

obtained: $f^*(t) = \int f(t|\eta)\pi(\eta)d\eta$, and hence the underlying survival function is chosen as $F^*(t) =$

$\int_t^{+\infty} f^*(x)dx$.

Lemma. Consider the entropy $I = -\iint g(t, \eta) \log[g(t, \eta)] \, dt d\eta$ for a joint continuous probability

density function $g(t, \eta) = f(t|\eta)\pi(\eta)$, where $\pi(\eta)$ is the marginal density function of η and $f(t|\eta)$ is

the density function of $T \in \mathfrak{R}^+$ conditional on η. Then for any $g(t, \eta)$ on $\mathfrak{R}^+ \times \mathfrak{R}^+$, the solution that

$$\text{maximises} \quad I = -\iint g(t,\eta) \log[g(t,\eta)]\, dt d\eta \tag{A1}$$

$$\text{subject to:} \ \iint g(t,\eta) dt d\eta = 1 \tag{A2}$$

$$\int t f(t|\eta) dt = \eta \ < +\infty \tag{A3}$$

$$\int \log(\eta^{-1})\, \pi(\eta) d\eta = r_1 < +\infty \tag{A4}$$

$$\int \eta^{-1} \pi(\eta) d\eta = r_2 < +\infty \tag{A5}$$

is given by $g^*(t,\eta) = c_0 \eta^{-\omega_1} \exp[-(t+\omega_2)/\eta]$ with $c_0 > 0$ the normalisation constant, and $\omega_i > 0$ ($i = 1,2$) are two constant values. The marginal survival distribution of $T \in \mathfrak{R}^+$ is $F^*(t) = \int_t^{+\infty} [\int_0^{+\infty} g(x,\eta)\, d\eta] dx = 1/[1 + t/\omega_2]^{\omega_1}$.

Proof . We form the Lagrangian function as

$$L = -\iint g(t,\eta) \log[g(t,\eta)]\, dt d\eta + \omega_0 \left[\iint g(t,\eta) dt d\eta - 1 \right]$$

$$+ \omega_1 [\iint \log(\eta^{-1})\, g(t,\eta) dt d\eta - r_1] + \omega_2 [r_2 - \iint \eta^{-1} g(t,\eta) dt d\eta] + [\iint \omega_3(\eta)(\eta - t) g(t,\eta) dt d\eta],$$

where ω_j ($j = 0,1,2,3$) are the Lagrange multipliers. By using the calculus of variations, the optimal condition is the Euler equation given by

$$-\log[g(t,\eta)] - 1 + \omega_0 + \omega_1 \log(\eta^{-1}) - \omega_2 \eta^{-1} + \omega_3(\eta)(\eta - t) = 0.$$

Solving the above equation together with the constraints (A2)-(A5), we obtain

$$g(t,\eta) = \exp(\omega_0 - 1) \eta^{-\omega_1} \exp\{-\omega_2 \eta^{-1} + \omega_3(\eta)(\eta - t)\}. \tag{A6}$$

Integrating out t, we obtain the marginal density function of η:

$$\pi(\eta) = \int g(t,\eta) dt = \exp(\omega_0 - 1) \eta^{-\omega_1} \exp\{-\omega_2 \eta^{-1} + \omega_3(\eta)\eta\}/\omega_3(\eta). \tag{A7}$$

Hence, the distribution of T conditional on η is $f(t|\eta) = g(t,\eta)/\pi(\eta) = \omega_3(\eta) \exp\{-\omega_3(\eta)t\}$. From (A3), we conclude that $\omega_3(\eta) = 1/\eta$, and hence

$$f(t|\eta) = \eta^{-1} \exp\{-t/\eta\}. \tag{A8}$$

Consequently, $\pi(\eta)$ in (A7) becomes $\pi(\eta) = \exp(\omega_0) \eta^{-\omega_1} \exp\{-\omega_2 \eta^{-1}\}$. To ensure the convergence of the integrals in (A4) and (A5), both ω_1 and ω_2 must be positive. Therefore, $g(t,\eta)$ in (A6) reduces to $g^*(t,\eta) = c_0 \eta^{-\omega_1} \exp\{-(\omega_2 + t)\eta^{-1}\}$, where $c_0 = \exp(\omega_0)$ is constant. Finally, integrating out η, we obtain the marginal density function of $T \in \mathfrak{R}^+$:

$$f^*(t) = \int g^*(t,\eta)d\eta = \int c_0 \eta^{-\omega_1} \exp\{-(\omega_2 + t)\eta^{-1}\}d\eta = \omega_2^{-1}(1 + t/\omega_2)^{-\omega_1 - 1}.$$

Hence, the underlying survival function is $F^*(t) = (1 + t/\omega_2)^{-\omega_1}$. This completes the proof.

We note that constraint (A2) in the lemma is simply the normalisation condition. Constraint (A3) ensures that the mean of $f(t|\eta)$ exists and is equal to η. The mean η here is not assumed to be fully specified; rather, it follows an unknown distribution $\pi(\eta)$. The information about η is expressed by the two moment conditions, i.e. (A4) and (A5). Constraints (A4) and (A5) are commonly used in the literature to derive (inverse) gamma distributions via the maximum entropy principle (Kagan, *et al.*, 1973; Park and Bera, 2009).

The result in the lemma shows that when eliciting prior knowledge with limited information, the distribution (11) is the optimal underlying survival function in the sense of the maximum entropy.

References

Abdellaoui, M., l'Haridon, O., Zank, H., 2010. Separating curvature and elevation: A parametric probability weighting function. *Journal of Risk and Uncertainty* 41 (1), 39-65.

Allais, M., 1953. Le comportement de l'homme rationnel devant le risque. *Econometrica* 21 (4), 503-546.

Balakrishnan, N., 1992. *Handbook of the Logistic Distribution*. Marcel Dekker, New York.

Balcombe, K., Fraser, I., 2015. Parametric preference functionals under risk in the gain domain: A Bayesian analysis. *Journal of Risk and Uncertainty* 50 (2): 161-187.

Barberis, N.C., 2013. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives* 27 (1), 171-196.

Beaud, M, Willinger, M., 2015. Are people risk vulnerable? *Management Science* 61 (3), 624-636.

Ben-Akiva, M.E., Lerman, S.R., 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press, London.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society* B 26 (2), 211–252.

Braisby, N., Gellatly, A. 2012. *Cognitive Psychology*, 2nd ed. Oxford University Press, Oxford.

De Palma, A., Ben-Akiva, M., Brownstone, D., Holt, C., Magnac, T., McFadden, D., Moffatt, P., Picard, N., Train, K., Wakker, P. and Walker, J., 2008. Risk, uncertainty and discrete choice models. *Marketing Letters* 19(3-4), 269-285.

Fehr-Duda, H., Epper, T., 2012. Probability and risk: Foundations and economic implications of probability-dependent risk preferences. *Annu. Rev. Econ.* 4 (1), 567-593.

Fennema, H., Wakker, P., 1997. Original and cumulative prospect theory: A discussion of empirical differences. *Journal of Behavioral Decision Making* 10 (1), 53-64.

Fosgerau, M., Bierlaire, M., 2009. Discrete choice models with multiplicative error terms. *Transportation Research* Part B 43 (5), 494–505.

Fox, C.R., Tversky, A., 1998. A belief-based account of decision under uncertainty. *Management Science* 44 (7), 879–895.

Gao, S., Frejinger, E. and Ben-Akiva, M., 2010. Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation research part C: emerging technologies* 18(5), 727-740.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin D., 2009. *Bayesian Data Analysis,* 3rd ed. Chapman & Hall/CRC, London.

Glöckner, A., Pachur, T., 2012. Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition* 123 (1), 21-32.

Golan, A, Judge, G., Miller, D., 1996. *Maximum Entropy Econometrics Robust Estimation with Limited Data*. Wiley, New York.

Goldstein, W.M., Einhorn, H.J. 1987. Expression theory and the preference reversal phenomena. *Psychological Review* 94 (2), 236-254.

Greene, W.H., Hensher, D.A., 2010. Does scale heterogeneity across individuals matter? A comparative assessment of logit models. *Transportation* 37 (3), 413-428.

Hensher, D.A., Greene, W.H., Li, Z., 2011. Embedding risk attitude and decisions weights in non-linear logit to accommodate time variability in the value of expected travel Time savings. *Transportation Research Part B* 45 (7), 954-972.

Hensher, D.A., Rose, J., Greene, W., 2015. *Applied Choice Analysis*. Second Edition, Cambridge University Press, Cambridge.

Hu, G., Sivakumar, A., Polak, J.W., 2012. Modelling travellers' risky choice in a revealed preference context: a comparison of EUT and non-EUT approaches. *Transportation* 39(4), 825-841.

Kagan, A.M., Linik, Y.V., Rao, C.R., 1973. *Characterization Problems in Mathematical Statistics*. Wiley, New York.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263-292.

Kemel, E., Paraschiv, C., 2013. Prospect theory for joint time and money consequences in risk and ambiguity. *Transportation Research Part B: Methodological* 56, 81-95.

Li, B., 2011. The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis. *Transportation Research Part B* 45 (3), 461-473.

Li, Z., Hensher, D.A., 2010. Toll roads in Australia: an overview of characteristics and accuracy of demand forecasts. *Transport Reviews* 30 (5), 541-569.

Li, Z., Hensher, D.A., 2013. Behavioural implications of preferences, risk attitudes and beliefs in modelling risky travel choice with travel time variability. *Transportation* 40(3), 505-523.

McFadden, D., Train, K., 2000. Mixed MNL models of discrete response. *Journal of Applied Econometrics* 15 (5), 447-470.

Palma, A. D., Lindsey, R., Proost, S. 2006. Research challenges in modelling urban road pricing: An overview. *Transport Policy* 13 (2), 97-105.

Park, S.Y., Bera, A.K., 2009. Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics* 150 (2), 219-230.

Prelec, D., 1998. The probability weighting function. *Econometrica* 66 (3), 497-527.

Quiggin, J., 1982. A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3 (4), 323–343.

Rasouli, S., Timmermans, H., 2014. Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research. *Travel Behav. Soc.* 1, 79–90.

Saleh, W., 2005. Road user charging: Theory and practice. *Transport Policy* 12 (5), 373-376.

Sjöberg, L., 2000. Factors in risk perception. *Risk Analysis* 20 (1), 1-11.

Stott, H.P., 2006. Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty* 32 (2), 101-130.

Toubia, O., Johnson, E, Evgeniou, T., Delquié, P., 2013. Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science* 59 (3), 613-640.

Train, K., 2009. *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press, New York.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representations of uncertainty. *Journal of Risk and Uncertainty* 5 (4), 297-323.

Wakker, P.P., 2010. *Prospect Theory: for Risk and Ambiguity*. Cambridge University Press, Cambridge.

Weber, E.U., Hsee, C., 1998. Cross-cultural differences in risk perception, but cross-cultural similarities in attitudes towards perceived risk. *Management Science* 44 (9), 1205-1217.

Weisberg, S., 2005. *Applied Linear Regression*, 3rd ed. Wiley, New York.

Wu, G., Gonzalez, R., 1996. Curvature of the probability weighting function. *Management Science* 42 (12), 1676-1690.

Xu, H., Zhou, J., Xu, W., 2011. A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. *Transportation Research Part C: Emerging Technologies* 19(2), 218-228.