PS918 Modelling Assignment: Risky Choice & CPT

Author: 2138473

Report Section

Introduction

Cumulative Prospect Theory (CPT), that is, a prominent model for decisions under risk, was developed from the prospect theory and was firstly introduced by Tversky and Kahneman (1992). To introduce a simple example, let's say we have a gamble which contains two options with different *payoffs* and *probabilities*; CPT models usually transform these two key elements: payoff and probabilities, and then subject the transformed values to an additional choice function to predict people's stochastic choice under uncertainty. It has been suggested that there is a general agreement on the nature of these three transformations, which involve outcome values, probabilities and choice functions (Stott, 2006). Thus, there exist various functional forms of CPT's key components.

The Current Investigation

The current enquiry investigates people's preferential choices in various scenarios, covering gains, losses, and a mixture of these. Specifically, the gambling set was designed by Rieskamp (2008) to have pairs of choices with similar expected values. The primary aim of this investigation is to re-fit the real-world preferential choice data from Rieskamp (2008) to estimate the parameters from 3 versions of CPT with different model complexity on the individual level. Each version of the CPT model will be explored and evaluated with the estimated parameters.

Methods

The data being utilised was from Rieskamp's (2008) study 2. Specifically, the data contains: 1) information of the gamble set deployed in the study; 2) participants' choices of each gamble; 3) demographic information of each participant:

The Gamble Data. The data of the gambling set contain information from a total of N = 180 gambling trials. Each gamble has two options: A and B and the outcomes of the two options were being either 1) both positive gains (N = 60 trials), 2) both negative losses (N = 60 trials) or 3) one positive gain and one negative loss (N = 60 trials) that associated with probabilities. The values of possible payoffs were drawn from 0 to 100 for gain, and -100 to 0 for losses. The probability of one outcome of an option was drawn randomly within the interval of 0 to 1 (rounded to 2 decimals).

The Demographics & Choice Data. The choice data contains the responses from 30 participants (18 females), and the average age is 24.5 years old (*SD* = 3.88). All participants' responses to the 180 gambles were recorded with no missing values. Specifically, participants' decision to choose option A was coded as "0", and option B was coded as "1".

Model Fitting. Specifically, we aim to fit three different versions of the Cumulative Prospect Theory (see Stott, 2006; Tversky & Kahneman, 1992) into the choice data. The functional forms of the models are displayed as follows:

Version 1 value function: Here, α and λ are the free parameters for the power value functions. α sometimes accounts for the trend of marginal utility of monetary gains, λ accounts for loss aversion.

$$U(x) = x^{\alpha} \quad for\ x \geq 0$$
$$U(y) = -\lambda|y|^{\alpha} \quad for\ y < 0 \tag{1}$$

Version 2 value function: Here α, β and λ are the free parameters for the power value functions.

$$U(x) = x^{\alpha} \quad for\ x \geq 0$$
$$U(y) = -\lambda|y|^{\beta} \quad for\ y < 0 \tag{2}$$

Version 3 value function: The value function of version 3 is identical to version 2, but version 3 specifically transforms the probability that is used to calculate the subjective values using a power probability function. Here γ is the free parameter:

$$w(p) = p^{\gamma} \tag{3}$$

Finally, the subjective valuation function of gamble option A with outcomes x_m > ... > x_1 ≥ 0 > y_1 > ... > y_m with corresponding probabilities p_m... p_1 and q_1... q_n and logit choice function for calculating the choice probabilities are:

$$V(A) = \sum_{i=1}^{m} w(p_i)U(x_i) + \sum_{j=1}^{n} w\left(q_j\right)U(y_j) \tag{4}$$

$$p(A) = \frac{1}{1 + e^{-\tau[V(A)-V(B)]}} \tag{5}$$

To note, here we restricted the bound values of the parameters in the CPT according to Rieskamp's (2008) study, where $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, $1 \leq \lambda \leq 10$. Furthermore, the ranges of γ, τ were set to 0 to 1, and 0 to 5, respectively

Parameters Estimation. The maximum likelihood estimation (MLE) approach has been applied extensively when modelling risky choices (e.g., Birnbaum, Patton, & Lott, 1999; Stott, 2006). MLE precisely captures the optimal parameters that maximise the log-likelihood ($ln\boldsymbol{L}$) to observe the actual data. Thus, we employed MLE to fit the models to the choice data on the individual level and estimate the optimal parameters.

Model Comparison. The model with different parameters that generate the maximum $ln\boldsymbol{L}$ is not often a better Therefore, we employed the likelihood ratio test (LRT) and Akaike Information Criteria (AIC) for model evaluation (LRT: Felsenstein, 1981; AIC: Akaike, 1973). The LRT is mainly used to compare nested models (similar models but varies in free parameters). Let's say we have two models in the same function form but vary in the number of free parameters. LRT firstly calculate the $ln\boldsymbol{L}$ of 2 models and transforms it to a specific value chi-squared χ^2 (likelihood ratio; LHR). Here we do not intend to explain the mathematical formula in detail, but essential LTR could tell whether models are substantially different from one another chi-squaredχ^2. Similarly, we need to do some calculations involving l n bold italic cap L and the numbers of free parameters to get AIC values. Yet, the AIC value evaluates the model by the value itself; the lower the AIC value is, the better the model fits the data (Burnham & Anderson, 2002).

Parameters Recovery. We performed parameter recovery using the estimated parameters that are demonstrated above. Specifically, we first utilised the estimated parameters with the models to simulate the responses of 30 participants. Then we re-fitted the model individually with the same set of parameter bounds to generate log-likelihood and new estimated parameters.

Results & Discussion

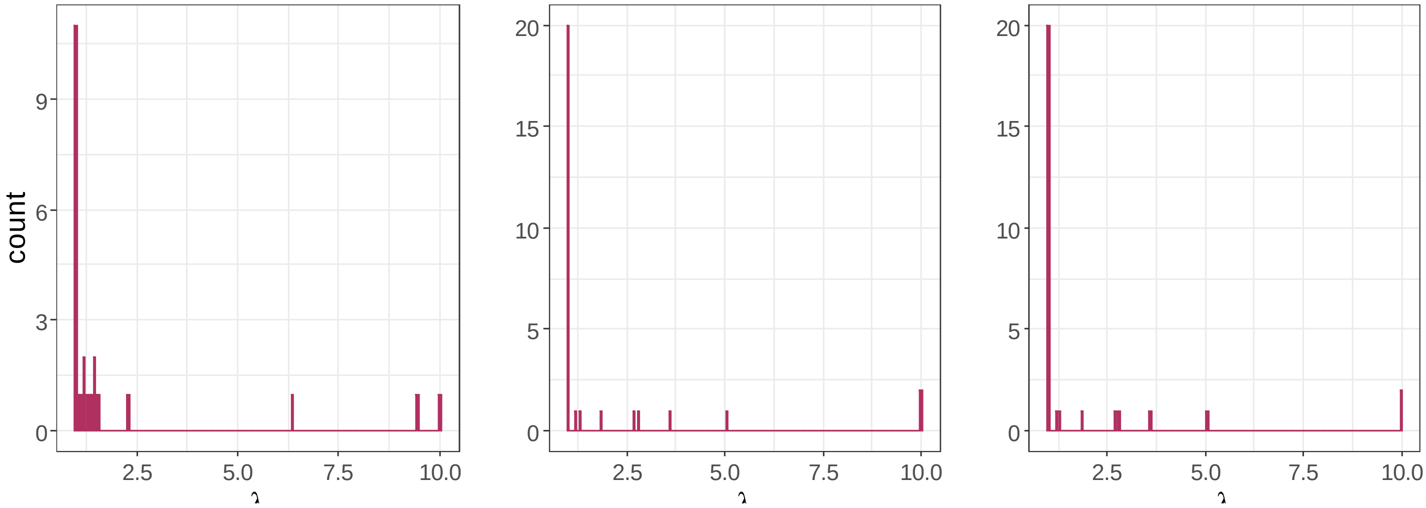

Figure 1. Distribution of parameter λ

After fitting the data of 30 subjects individually, we estimated the optimal parameters for all three models by subtracting the median values of 30 data points. Specifically, the parameters estimated for the version 1 model are: $\alpha = 0.89\ (SD = 0.37),\ \lambda = 1.14\ (SD = 2.44),\ \tau = 0.13\ (SD = 0.21)$. For version 2, $\alpha = 0.89\ (SD = 0.28),\ \lambda = 1\ (SD = 2.41)$, $\beta = 0.85\ (SD = 0.34)$, $\tau = 0.14\ (SD = 0.26)$. For version 3, $\alpha = 0.83\ (SD = 0.32),\ \lambda = 1\ (SD = 2.73),\ \beta = 0.78\ (SD = 0.31),\ \tau = 0.21\ (SD = 0.47),\ \gamma = 0.76\ (SD = 0.34)$. Similar to Rieskamp's (2008) study 2 findings, we noticed that a considerable proportion of estimated λ parameters were equal to 1, regardless of the version of the model (see Figure 1). We suspect that may be due to the gamble's outcomes are presented in a variety of scenarios, including both gains, both losses and a mixture of those. Therefore, the loss version was diminished when dealing with such noises (Rieskamp, 2008). To further inspect the parameters, we plotted the parameters pairs of the model with the largest number of free parameters (version 3) for more visual illustrations. We specifically plot the parameter pairs with identifiable trends (Figure 2). In the two subplots of Figure 2, the α and β positively correlate with γ. It might be the internal mechanism of choice bias could correlate with both gain and loss domains.

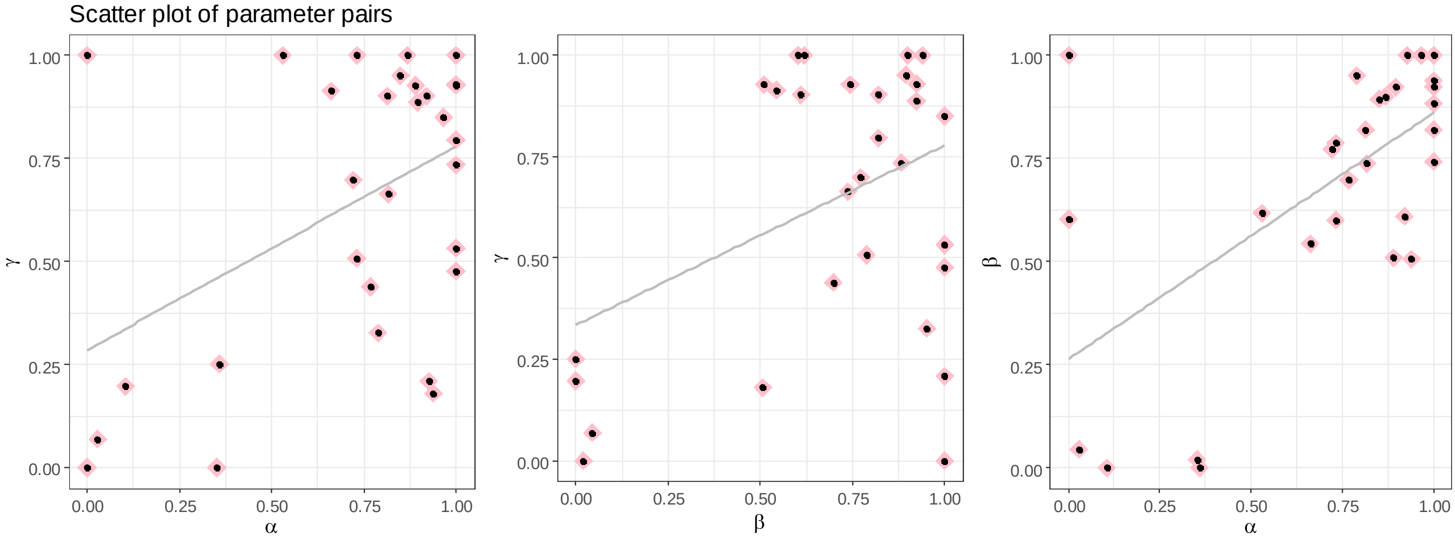

Figure 2. Scatter plots of parameter pairs with identifiable trends.

We then compared the three versions of CPT using LHR and AIC. To perform LRT, we specifically aggregated the results individual log-likelihood of each model and obtained the average $ln\boldsymbol{L}$ values and used these values to calculate likelihood-ratios. Furthermore, unlike the LRT, we calculate the AIC values by individuals and aggregate the AICs for each model to get the average values, which could be regarded as an indicator of goodness-of-fit. Results of LRT indicate that version 3 is significantly better than version 1, $\chi^2(2) = 29.45,\ p < .001$, and version 2, $\chi^2(1) = 28.54,\ p < .001$. Moreover, results show that model version 2 did better fit than version 1 as well, $\chi^2(1) = 4.07,\ p = .044$. In alignment with the likelihood ratio test, the results of the AIC values also suggest that version 3 being the best model (average AIC = 194.66), then followed by version 2 (average AIC = 204.27), and finally version 1 (average AIC = 206.34). Also, we calculated the ΔAIC, that is, a goodness-of-fit index similar to AIC values, but the ΔAIC refers to the difference between the AIC of any model M_i and the minimal AIC in a set of models. We calculated the ΔAIC of each subject and aggregated to get the average ΔAIC between models. Notably, we found an average ΔAIC = 11.69 between version 1 and 3, which illustrate strong evidence for version 3 being better than version 1, since it has been suggested that ΔAIC > 10 indicates substantial model disadvantage (Rieskamp, 2008; Burnham et al., 2004).

Finally, the data of the re-generated parameters indicate that all 3 models' ability to

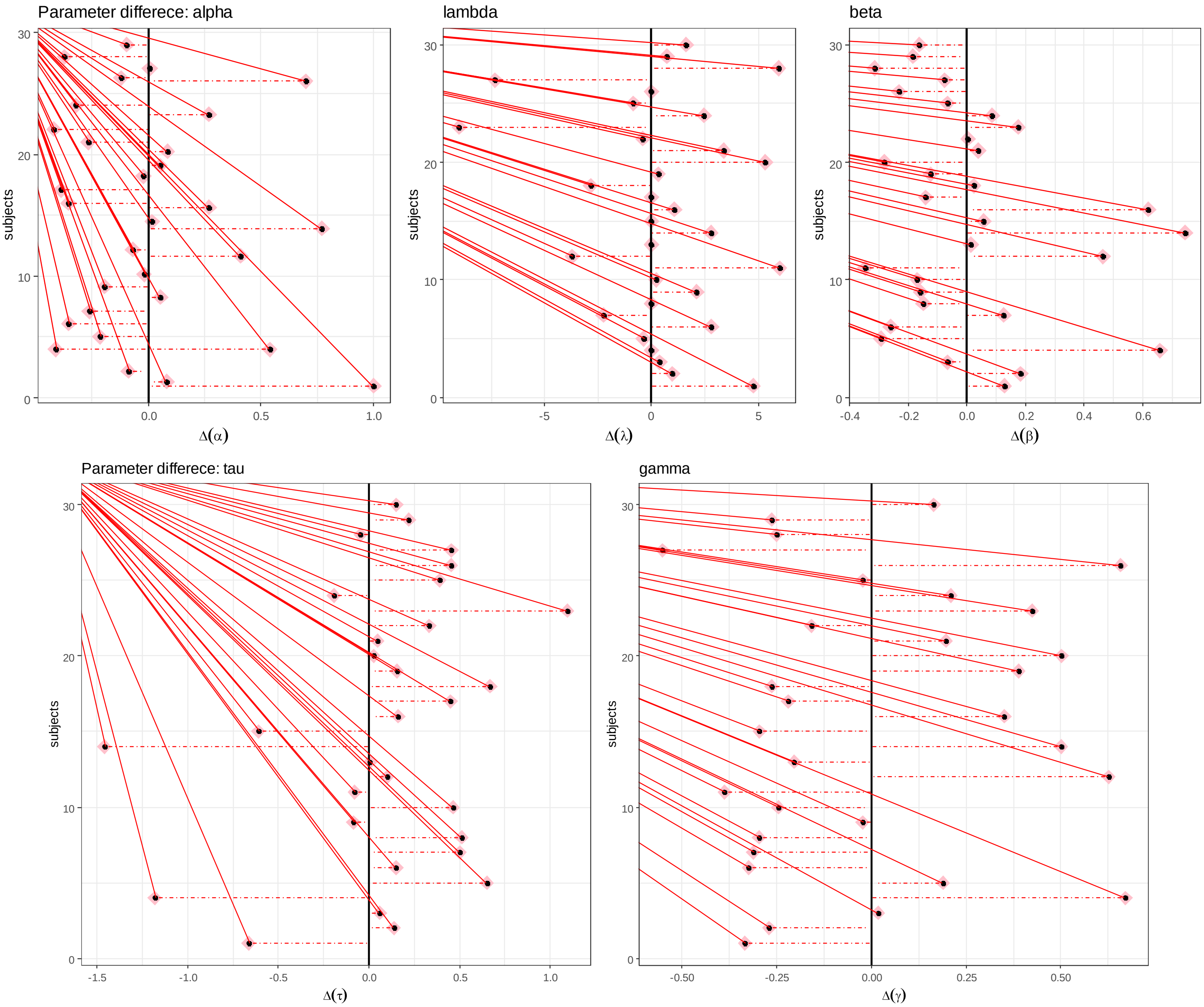

Figure 2. Point plot of the parameter differences: initial vs recovered; each point corresponds to one unique simulated subject.

recover its parameters is relatively poor. For example, the data from parsimonious version (version 1) model recovery shows that 10 simulated subjects' estimated parameters are not empirically identifiable. Furthermore, we visualised the differences between the initially estimated and recovered parameters per subject. As shown in figure 3, we can see that most of the differences between parameters does not equal to 0 and mostly larger than 0.1, except for λ. However, the high recovered proportion of the λ parameter is that λ was easily estimated to 1, and this might be due to the diminished loss aversion as discussed above. The second reason of the poor recoverability might be our model restricted the parameters τ, γ in a less sensible range, or the function forms in the models are being relatively simple and such that the models are not computationally flexible enough to deal with the noisy choice context. Thus, it might be of future interest to fit the current choice data with CPT models by using broader range of functional formulas (see Stott, 2006).

Reference:

Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., *International Symposium on Information Theory*, 267-281.

Birnbaum, M., Patton, J., & Lott, M. (1999). Evidence against Rank-Dependent Utility Theories: Tests of Cumulative Independence, Interval Independence, Stochastic Dominance, and Transitivity. *Organizational Behavior And Human Decision Processes*, *77*(1), 44-83. https://doi.org/10.1006/obhd.1998.2816

Burnham, Kenneth & Anderson, David. (2004). Model Selection and Multimodel Inference. A Practical Information-theoretic Approach. 10.1007/978-0-387-22456-5_5.

Felsenstein J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, *17*(6), 368–376. https://doi.org/10.1007/BF01734359

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34* (6), 1446–1465. doi: 10.1037/a0013646

Stott, H. P. (2006). Cumulative prospect theory’s functional menagerie. *Journal of Risk and Uncertainty, 32* (2), 101–130. doi: 10.1007/s11166-006-8289-6

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5* (4), 297–323. doi: 10.1007/BF00122574