POLITECNICO DI TORINO

Master of Science programme in Physics of Complex Systems

Master Degree Thesis

Horse-race between decision-making models

Quantum Decision Theory and "Classical" counterparts

Supervisor

Prof. Alessandro Pelizzola, Politecnico di Torino

Prof. Didier Sornette, ETH Zürich

Co-Supervisors

Prof. Fabrizio Dolcini, Politecnico di Torino

Dr. Tatyana Kovalenko, ETH Zürich

Candidate

Giuseppe Maria Ferro

ID Number: 235966

Academic year 2017 - 2018

Abstract

In this work, we propose another parametrization of Quantum Decision Theory (QDT), based on Rank Dependent Utility Theory (RDU). Using experimental data made of choices between pairs of lotteries, we then compare QDT with "classical" decision theories, RDU and Cumulative Prospect Theory (CPT). At aggregate level, assuming homogeneous preferences across subjects, we find that CPT-based QDT wins by far. At the individual level, we classify decision makers as RDU, CPT or QDT. Our major findings are the following: quantum factor plays a key-role in describing subjects' behavior; there is a considerable heterogeneity across subjects, so that the classic representative agent approach would be completely wrong for this sample. In light of such results, mixture models are then considered as a possible extension of the present work, in order to take into account potential heterogeneity within a subject himself.

Acknowledgements

To start with, I would like to thank my supervisor, Prof. Didier Sornette, who was always available for discussing new ideas, both from a technical and a philosophical point of view. Each time we talked I was truly inspired by his way of doing science. His great professionalism allowed me to learn a lot and I think I am going to learn even more if we will work together.

Many thanks go also to my co-supervisor, Dr. Tatyana Kovalenko, her skills guided me in overcoming the obstacles encountered during the development of this project. She helped me in reaching concrete results, saving me from wandering in unsuccessful directions.

A special thanks goes to my mother, always present for me, to whom I dedicate this thesis.

Finally, I want to thank my girlfriend Marianna, who supported me during each stressful period of this course, especially in the exam sessions. She was close to me during the achievement of this goal, always encouraging me to do better.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Descriptive Decision theory is a branch of Decision Theory which aims to understand and predict the choices of (potentially irrational) decision makers. The most famous theories are the expected utility theory (D. Bernoulli, 1738 [1]), the prospect theory (D. Kahneman and A. Tversky, 1979 [2]), the rank dependent utility theory (RDU, John Quiggin, 1982 [3]) and the cumulative prospect theory (CPT, D. Kahneman and A. Tversky, 1992 [4]). The main drawback of these theories is their deterministic approach; indeed, experimental results (F. Mosteller and P. Nogee, 1951 [5]) tend to identify decision making as a stochastic process, in principle far from a deterministic system.

Probabilistic extensions of such theories have been proposed, under the name of Random Utility Theories, firstly introduced in mathematical psychology by Duncan Luce and Anthony Marley [6]. For example, CPT has been enhanced with the probit (J.D. Hey and C. Orme, 1994 [7]) or the logit functions (E. Carbone and J.D. Hey, 1995 [8], M.H. Birnbaum and A. Chavez, 1997 [9]).
In essence, Random Utility Theory assumes that the utility of an option perceived by the decision maker is not *directly observable* from outside, and therefore it must be represented in general by a random variable. Nevertheless, it is conceptually very far from recognizing an **intrinsic** probabilistic nature of choice, assessing instead that choice is non-deterministic because of several sources of "measurement" errors (E. Cascetta, 2000 [10]).

On the other hand, Quantum Decision Theory (QDT, V.I. Yukalov and D. Sornette, 2008 [11]), by relying on the mathematics of separable Hilbert spaces, provides an intrinsic probabilistic theory able to describe entangled

decision making and non-commutativity of decisions. In this way QDT avoids all the paradoxes arising in classical decision theories such as violation of sure-thing principle [12], conjunction fallacy [13] and Ellsberg Paradox [14].
In QDT, the probability of choosing a certain prospect is the sum of two terms: the utility factor and the attraction factor. The *rational* comparison among the available alternatives is represented by the utility term; the attraction factor, instead, quantifies the *attractiveness* of a prospect, dependent on feelings, beliefs and subconscious biases. [15]. To put it differently, the attraction factor is a measure of the deviation from rationality.

Based on the work of S. Vincent et al. [16], we extend it by investigating another parametrization of QDT, based on the previously mentioned Rank-Dependent Utility Theory. The motivation underlying this study is that RDU, being a generalized expected utility theory, evaluates utilities with respect to the final wealth. CPT, on the other hand, relies on a definition of a *reference point* for discriminating between gains and losses. Such reference point is often taken as 0 for laboratory tasks, but, in real-world applications, its identification may be highly non-trivial, as highlighted by Nicholas C. Barberis in [17].
In short, an RDU-based QDT has the advantage of not depending on the (difficult) calibration of a reference point, since only changes in final wealth matter.
At the aggregate level, where homogeneous preferences are assumed across subjects, we found that QDT_{CPT} best characterizes the average behavior. On the other hand, at the individual level, our subject classification clearly shows a relevant heterogeneity of preferences in the sample, (with the majority of subjects better described by QDT_{RDU}) implying the inconsistence of assuming the existence of a universal theory valid for each subject, at least in the standard macroeconomic way of pooling data.

The data set studied comes from an experiment conducted at the Max Planck Institute for Human Development in Berlin [18]. In this experiment, decision makers had to choose the preferred risky option for several gambles. The experiment was iterated twice, so that we can fit the models to the first set of data, and then we can use data from the second time for out-of-sample prediction.
Such experiment has been already studied by R. O. Murphy and R. ten Brincke [19], where they show the usefulness of *hierarchical* maximum likelihood estimation for CPT. (subsection 3.2.2)

Chapter 2

State of the Art in Decision Theory

A typical choice faced by the decision maker in the studied experiment is the following:

$$A = (x_1^A, p_1^A; x_2^A, p_2^A) \quad \text{vs} \quad B = (x_1^B, p_1^B; x_2^B, p_2^B) \tag{2.1}$$

where $\{x_i^j\} \in R$ are the outcomes of the gambles, and $\{p_i^j\}$ are the probabilities with which they occur.
For istance, if the decision maker chooses lottery B, she will receive x_1^B with probability p_1^B or x_2^B with probability $p_2^B = 1 - p_1^B$. Since the outcome of a lottery is uncertain but the probabilities are known, decision of this kind are generically called *decisions under risk*; situations where probabilities and/or outcomes are unknown go under the name of *decisions under ambiguity or uncertainty* (Even though there seem not to be a common accepted definition of risk, ambiguity and uncertainty). We will focus only on the former type.

2.1 Expected Utility Theory(EUT)

Expected utility theory [1] is the first famous attempt to capture some decision-making mechanisms violating the expected value criterion, where an individual is supposed to choose the lottery with the highest expected value. Psychological patterns such as risk-aversion highlight that individuals do not follow such approach. Therefore, instead of merely considering the actual outcomes of a lottery, EUT individuals associate a utility to each outcome, which can be represented as a "new" choice:

$$\tilde{A} = (v(W+x_1^A), p_1^A; v(W+x_2^A), p_2^A) \quad \text{vs} \quad \tilde{B} = (v(W+x_1^B), p_1^B; v(W+x_2^B), p_2^B) \tag{2.2}$$

where v is the so called utility function and W is the initial wealth of the decision maker.
The individual will then choose the "transformed" lottery with the highest expected value, defined as:

$$E\big[\tilde{A}\big] = U(A) = p_1^A v(W + x_1^A) + p_2^A v(W + x_2^A) \tag{2.3}$$

The fact that individuals make their choices according to changes in final wealth, and not considering the gamble frame, is referred as *asset integration*: as we will see, it is the foundamental feature differentiating generalized expected utility theories on one hand and prospect theory on the other.

When talking about monetary outcomes, it is assumed that *more is better*, and so increasing utility functions are adopted. Moreover, the curvature of the utility function plays an important role in determining decision maker's attitude toward risk, as we will discuss in section 2.2.

The striking simplicity of expected utility is a double-edged sword: on one side, it has a wide range of applicability, being indeed the most used decision theory in microeconomics; on the other side, EUT is not able to encompass many situations such as the Allais [20] and Ellesberg [14] paradoxes. As a result, alternative theories have been proposed.

However, before listing them, it is useful to recall the main risk-aversion measures used in economics.

2.2 Risk Aversion (for EUT)

One of the most important psychological processes driving human decision-making is risk-aversion, that can be informally defined as the general tendence of preferring a situation with a more predictable payoff over a riskier, but potentially better one.
Roughly speaking, the more the utility function is concave, the more the decision maker is risk-averse (in Expected Utility, this can be easily verified by

applying Jensen's inequality). More precisely, there are two measures of risk aversion for a given utility function: **Absolute risk aversion** and **Relative risk aversion**.

2.2.1 Absolute risk aversion

Absolute risk aversion (K. Arrow and J. W. Pratt, 1965 [21]) is a measure of investors reaction to uncertainty relating to changes in their wealth, defined as:

$$A(W) = \frac{-v''(W)}{v'(W)} \tag{2.4}$$

where $v'(W)$ and $v''(W)$ are the first and second derivatives of the utility function $v(W)$ respectively and W is the current wealth level.
Table 2.1 shows several behaviors according to this measure.

Condition	Definition	Property of $A(W)$
Increasing Absolute Risk Aversion (IARA)	As wealth increases, hold fewer dollars in risky assets	$A'(W) > 0$
Constant Absolute Risk Aversion (CARA)	As wealth increases, hold same dollar amount in risky assets	$A'(W) = 0$
Decreasing Absolute Risk Aversion (DARA)	As wealth increases, hold more dollars in risky assets	$A'(W) < 0$

Table 2.1: Several behaviors according to Absolute risk aversion measure.

2.2.2 Relative risk aversion

Relative risk aversion is a measure of investors reaction to uncertainty relating to percentage changes in their wealth, defined as [22]:

$$R(W) = \frac{-Wv''(W)}{v'(W)} = WA(W) \tag{2.5}$$

Table 2.2 shows several behaviors according to this measure. CRRA is usually assumed for simplicity.

Condition	Definition	Property of $R(W)$
Increasing Relative Risk Aversion (IRRA)	Percentage invested in risky assets declines as wealth increases	$R'(W) > 0$
Constant Relative Risk Aversion (CRRA)	Percentage invested in risky assets is unchanged as wealth increases	$R'(W) = 0$
Decreasing Relative Risk Aversion (DRRA)	Percentage invested in risky assets increases as wealth increases	$R'(W) < 0$

Table 2.2: Several behaviors according to Relative risk aversion measure. CRRA is often used for simplicity.

2.3 Rank Dependent Utility Theory (RDUT)

In 1982, John Quiggin [3] suggested a generalization of the expected utility model, relaxing the assumption that the Utility functional has to be linear in the probabilities. One possible simple explanation is that individuals tend to substitute 'decision weights' for probabilities. [23]

Thanks to rank-dependence, there is the possibility of assigning two different weights to outcomes with equal probabilities, because the weight of outcome i will be not only function of the probability p_i, but of the *ranking position* of i.
Formally, given a prospect X with n outcomes, ordered from the worst to the best i.e. $x_1 < x_2 < .... < x_n$ with probabilities $p_1, ..., p_n$, the rank-dependent functional is:

$$U(X) = U(\{\vec{x};\vec{p}\}) = \sum_{i=1}^{n} v(x_i + W)h_i(\vec{p}) \tag{2.6}$$

with $h_i(\vec{p})$ being the decision weight relative to outcome x_i, defined as:

$$h_i(\vec{p}) = w(\sum_{j=i}^{n} p_j) - w(\sum_{j=i+1}^{n} p_j) = w(1 - F(x_i)) - w(1 - F(x_{i-1})) \tag{2.7}$$

where $F(x_i) = P(X <= x_i)$ is the ranking position of outcome x_i and $w(p)$ is called transformation or weighting function.
For two-valued lotteries, equation 2.6 reduces to:

$$U(X) = w(p_1^X)v(W + V_1^X) + (1 - w(p_1^X))v(W + V_2^X) \tag{2.8}$$

From equation 2.8, we can easily retrieve the EUT expression by letting $w(p) = p$.

One of the key ideas of rank-dependence is to overweight only low probability **extreme** outcomes and not low probability intermediate outcomes. A simple example, taken from [23], may help:
suppose a *pessimist* decision maker faces the lottery $(\frac{1}{3}, 10; \frac{1}{3}, 20; \frac{1}{3}, 30)$; he will pay more attention to the worst outcome, so $h_1 > \frac{1}{3}$, say $h_1 = \frac{1}{2}$. Of the remaining attention (in rank dependent theory it is assumed that $\sum_{i=1}^{n} h_i(\vec{p}) = 1$ to avoid violations of 1st order stochastic dominance [3]), being a pessimist, he will pay more attention to the second worst outcome, so $h_2 > \frac{1}{4}$, say $h_2 = \frac{1}{3}$. This implies $h_3 = \frac{1}{6}$.
If the lottery presented is now $(\frac{1}{3}, 0; \frac{1}{3}, 10; \frac{1}{3}, 20)$, the outcome 10 will receive less attention than before, being "just" an intermediate outcome, and not an extreme one.

From this simple discussion we understand that the risk-aversion measures mentioned in section 2.2 are not able to fully describe the characteristics of generalized expected utility theories.
As Quiggin [3] stresses, even supposing a linear utility function, with a convex w we can still have a risk-averse behaviour. Sticking to the linear utility function assumption for simplicity, a natural generalization of the risk aversion concept for RDU is the following: pessimist people adopt a set of decision weights that yield an expected value for a transformed risky prospect lower than the mathematical expectation.
An individual is said to be *pessimistic* $\iff w(p) \leq p \forall p$. Conversely, she is *optmistic* $\iff w(p) \geq p \forall p$.

In the general case of nonlinear utility and transformation functions, they both contribute in a nontrivial way to determine subject's behavior. Indeed, an individual with a concave utility function can be globally risk-seeking if the transformation function w is "sufficiently" optimistic.

The idea is that utility function v and transformation function w can be seen as two "forces": they "point" in opposite directions if, for example, v is concave and w optimistic. The first, having decreasing marginal utility, discourages risk-taking, but the second encourages it. The "net" behavior will be the resultant force.
One important thing to notice is that the effect of probability weighting is independent of the scale of the bet. Therefore, in the above case optimism will tend to predominate when bets are small (when all the outcomes are

near the current wealth level).

In the present report, for the probability weighting a function known as the Prelec II weighting function was chosen [24]. Having two parameters, it is very flexible: δ controls the general elevation of the curve, and γ controls its curvature.

$$w(p) = exp(-\delta(-ln(p))^{\gamma}) \quad \delta > 0 \quad \gamma > 0 \tag{2.9}$$

As for the utility functions, we chose the following:

$$v_1(x) = \begin{cases} \dfrac{x^{\alpha}-1}{\alpha}, & \alpha \neq 0 \\ ln(x), & \alpha = 0 \end{cases} \tag{2.10}$$

$$v_2(x) = \frac{1}{\beta}\Big\{1 - exp\Big[-\beta\Big(\frac{x^{\alpha}-1}{\alpha}\Big)\Big]\Big\} \tag{2.11}$$

When 2.10 (2.11) is used we call such theory $RDU_1(RDU_2)$.
The first one is CRRA (constant relative risk-aversion), while the second one is very flexible: depending on the parameter values, it can represent CARA, DARA, IRRA, CRRA and DRRA.[25]

2.4 Cumulative Prospect Theory(CPT)

Prospect theory was first introduced by D. Kahneman and A. Tversky in 1979 [2]. It presents a foundamental difference with respect to generalized expected utility theories: separations of gains and losses with respect to a reference point. The utility of a prospect does not depend anymore on the initial wealth, i.e "the carriers of value are gains and losses, not final assets" [2]. Thanks to this feature, other psychological mechanisms can be described, the most famous being *loss aversion*. The idea behind loss aversion is that the "pain" deriving from losing x€ is greater than the "joy" deriving from receiving x€. In their words, "losses loom larger than gains"[2].

In 1992 [4], by incorporating the rank-dependent weighting of RDU, they cured some theoretical issues, developing what is nowadays known as Cumulative Prospect Theory.

With this model, assuming the reference point to be zero, the utility of a binary lottery A is given by:

$$U(A) = \begin{cases} w(p_1^A)v(x_1^A) + (1 - w(p_1^A))v(x_2^A), & \text{for lotteries with only gains or only losses} \\ w(p_1^A)v(x_1^A) + w(p_2^A)v(x_2^A), & \text{for mixed lotteries} \end{cases} \tag{2.12}$$

where x_1^A and x_2^A have been ordered such that:

- $x_1^A \geq x_2^A$ if both are positive.
- $x_1^A \leq x_2^A$ if both are negative.

Within CPT, $v : R \rightarrow R_+$ is called *value* function, and it is usually convex in the domain of losses (risk-seeking) and concave in the domain of gains (risk-averse) so that it can accomodate common empirically observed behaviour [4].

We adopt the same functional forms of S. Vincent et al. [16]: for the value function we have

$$v(x) = \begin{cases} x^\alpha, & \text{if } x \geq 0 \quad \alpha > 0 \\ -\lambda(-x)^\alpha, & x < 0 \quad \lambda > 0 \end{cases} \tag{2.13}$$

As for the weighting function, Prelec II is also used (equation 2.9).

2.5 Random Utility Theories

As already said in the Introduction, the previously presented theories have no space for "deviation" from the best choice: the decision maker will always pick it, whatever the circumstances. However, this approach is too simplicistic and does not take into account the large amount of endogenous and exogenous factors affecting choice, deceiving the decision maker from the optimal move.

Random Utility Theory, firstly introduce by Duncan Luce and Anthony Marley [6], tries to deal with observed stochastic behaviour by separating the utilities of a lottery X in two terms: a deterministic component and a random component.

$$V(X) = U(X) + \epsilon_X \tag{2.14}$$

$U(X)$ is the already seen deterministic part and ϵ_X is the random part, called also *disturbance*. By making different assumptions on the distribution of the random component (precisely, on the distribution of the difference $\epsilon_A - \epsilon_B = \epsilon$), different probabilistic models are obtained, such as the linear probability model [26], the binary probit [27] and the binary logit [28]. Following the work of S. Vincent et al. [16], we adopt and present the latter, enhancing CPT and RDU with a choice function. Assuming that the disturbance difference ϵ is logistically distributed:

$$f(\epsilon) = \frac{1}{1 + e^{-\phi\epsilon}} \tag{2.15}$$

the probability p_B of picking an option B over A is then:

$$p_B = \frac{1}{1 + e^{\phi(U(A) - U(B))}} \tag{2.16}$$

$\phi >= 0$ is called sensitivity parameter, while $U(A)$ and $U(B)$ are the deterministic utilities, calculated according to one of the previously mentioned theories.

To sum up, a logit-CPT decision maker is described by five parameters: two for the value function (α, λ), two for the weighting function (γ, δ) and one for the choice function (ϕ).
On the other hand, a logit-RDU decision maker is characterized by four parameters: one or two for the utility function (α and β), two for the weighting function (γ, δ) and one for the choice function (ϕ).

2.6 Quantum Decision Theory (QDT)

Quantum Decision Theory (QDT) was introduced by V.I. Yukalov and D. Sornette in 2008 [11], whose aim was to build a general framework able to encompass all paradoxes in classical decision theory, without the need of an "ad-hoc" set-up for each fallacy.
The mathematical theory of separable Hilbert spaces is the "backbone" of QDT; we want to stress that the use of quantum theory is only a formal analogy: there is no claim at all of any quantum effect taking place within the brain.

The main idea driving QDT is indeed to exploit the most striking features of

quantum mechanics with respect to classical physics: intrinsic probabilistic nature and entanglement.
In decision-making context, the former means that a decision maker can choose different options when faced with the same task multiple times (without the intervention of memory effects). It is formally the same thing that happens in quantum physics when a measurement, repeated many times on a system in a well-known state, yiels different values. The only difference is that in physical experiments the system is "passive", while in decision-making the system, i.e the decision-maker, actively takes the decision.
The entanglement instead allows to model *interdependence* among the alternatives, in the sense that the decision maker does not evaluate each option separately and then picks the best, but they are overall "entangled". In other words, the appeal of an option will largely depend on the available alternatives.

We now briefly recall the mathematical structure of QDT, that is extensively presented in [29].

2.6.1 Mathematical structure of Quantum Decision Theory

The first object we need is the **Action ring**, i.e. a non-commutative ring formed by a set of *intended actions* (simply called events in classic probability theory)

$$\mathcal{A} = \{A_n : n = 1,2, ..., N\} \tag{2.17}$$

and two binary operations:

1. Addition (A or B or both occur): $A + B = B + A \in \mathcal{A}$
2. Multiplication (A and B occur together): $\forall A, B \in \mathcal{A} \quad AB \in \mathcal{A}$ and $A \cdot 0 = 0 \cdot A = 0 \quad \forall A \in \mathcal{A}$. In general, $AB \neq BA$, so the multiplication is non-commutative. The 0 element, the *empty action*, symbolizes an impossible action.

An action, or an event, can be realized in many different ways (e.g. becoming rich can be realized by working hard or by becoming a thief). For this reason we define the composite action A_n as:

$$A_n = \bigcup_{\mu=1}^{M_n} A_{n\mu} \quad M_n > 1 \tag{2.18}$$

where the $\{A_{n\mu}\}$ are called action modes.

We define also the *action prospect* π_n as:

$$\pi_n = \bigcap_j A_{nj} \tag{2.19}$$

where this time the $\{A_{nj}\}$ can be composite or simple actions. If the product is made of only simple action modes, each relative to a different action, the action prospect is called *elementary prospect*:

$$e_\alpha = \bigcap_n A_{nj_n} \tag{2.20}$$

Next, we put in relation each action mode A_{nj} to a quantum state $|A_{nj}\rangle$, defining their scalar product as:

$$\langle A_{nj}|A_{nk}\rangle = \delta_{jk} \tag{2.21}$$

The intuition behind orthogonality among action modes is that they are **different** ways of realizing the same intended action, therefore they are incompatible actions.
Thanks to definition 2.21, it is possible to define the Hilbert space, called *mode space*, generated by action mode states as:

$$\mathcal{M}_n = \text{Span}\{|A_{nj}\rangle : j = 1,2, ..., M_n\} \tag{2.22}$$

We can do the same mapping for elementary prospects, i.e we define $|e_\alpha\rangle$ as:

$$|e_\alpha\rangle = |A_{1j_1}...A_{Nj_N}\rangle \quad \text{and} \quad \langle e_\alpha|e_\beta\rangle = \prod_n \delta_{j_n i_n} = \delta_{\alpha\beta} \tag{2.23}$$

In the same way we defined the mode space, we define the *mind space* as the Hilbert space generated by elementary prospects:

$$\mathcal{M} = \text{Span}\{|e_\alpha\rangle\} \tag{2.24}$$

The mind space defined in 2.24 is the central "field" of our discussion, over which we can further define two types of quantum states: the *strategic state of mind* and the *prospect state*.
The former is essentially the (normalized) state that contains all the information about the mind of the decision maker:

$$|\psi_s\rangle = \sum_\alpha c_\alpha |e_\alpha\rangle \in \mathcal{M} \tag{2.25}$$

A prospect state is instead the correspondent quantum state of an action prospect defined in 2.19:

$$|\pi_n\rangle = \sum_\alpha a_\alpha |e_\alpha\rangle \in \mathcal{M} \tag{2.26}$$

Once the prospect state is defined, the corresponding *prospect operator* $\hat{P}(\pi_n)$ is given by:

$$\hat{P}(\pi_n) = |\pi_n\rangle \langle\pi_n| \tag{2.27}$$

At this point, the decision maker, described by the state of mind $|\psi_s\rangle$, will choose the prospect π_n with probability:

$$p(\pi_n) = \langle\psi_s|\hat{P}(\pi_n)|\psi_s\rangle = |\langle\pi_n|\psi_s\rangle|^2 \tag{2.28}$$

Clearly, since the decision maker must choose one option, the normalization condition holds:

$$\sum_n p(\pi_n) = 1 \tag{2.29}$$

We stress that $p(\pi_n)$ crucially depends on the particular state of mind $|\psi_s\rangle$ that contains all the characteristics of the decision maker.

An important feature arises when the decision maker has to choose among at least one composite prospect: intuitively, the existence of several action modes for one intended action may lead to uncertainty perception.
In such situation, the probability $p(\pi_n)$ can be split into two terms: a utility factor $f(\pi_n)$ and an attraction factor $q(\pi_n)$.

$$p(\pi_n) = f(\pi_n) + q(\pi_n) \tag{2.30}$$

$$f(\pi_n) = \sum_\alpha |c_\alpha^* a_\alpha|^2 \tag{2.31}$$

$$q(\pi_n) = \sum_{\alpha\neq\beta} c_\alpha^* a_\alpha c_\beta a_\beta^* \tag{2.32}$$

The attraction factor defined 2.32 arises because of the *interference* among action modes relative to the same intended action. This interference term is responsible for non-additivity of probabilities in quantum theory.
The utility and attraction terms obey the following conditions:

- $f(\pi_n) \in [0,1]$ and $\sum f(\pi_n) = 1$ (normalization)
- $q(\pi_n) \in [-1,1]$ and $\sum q(\pi_n) = 0$ (alternation property)

2.6.2 QDT for binary choices

In our study, the decision makers face a series of binary choices, namely lottery A vs lottery B; as already reported in [16], the probabilities can be simply written as:

$$\begin{cases} p_A = f_A + q_A \\ p_B = f_B + q_B \\ q_A = -q_B \\ f_A = 1 - f_B \end{cases} \tag{2.33}$$

To allow the possible emergence of quantum interference we assume that lotteries A and B correspond to composite prospect states of the form:

$$\begin{cases} |A\rangle = a_1 |A1\rangle + a_2 |A2\rangle \\ |B\rangle = b_1 |B1\rangle + b_2 |B2\rangle \end{cases} \tag{2.34}$$

with $|A1\rangle, |A2\rangle, |B1\rangle, |B2\rangle$ are ortoghonal action modes. As suggested in [30], such decomposition reflects the uncertainty for the decision maker, arising for example from a misunderstanding of the experimental setup.
The mind space $\mathcal{M}$ in this case is given by:

$$\mathcal{M} = \text{Span}\{|A1\rangle, |A2\rangle, |B1\rangle, |B2\rangle\} \tag{2.35}$$

and the decision maker state of mind can be written as:

$$|\psi_s\rangle = c_{11} |A1\rangle + c_{12} |A2\rangle + c_{21} |B1\rangle + c_{22} |B2\rangle \tag{2.36}$$

In such binary context, as derived in [16], the attraction factor is further constrained by the following condition:

$$|q_A| = |q_B| \leq min(f_A, f_B) \tag{2.37}$$

2.6.3 QDT parametrizations

The problem now is to find suitable parametrization of the utility and attraction factor so to reach satisfatory results.
In [16], a logit-CPT function was used for the utility factor:

$$f_A = \frac{1}{1 + e^{\phi(U(B) - U(A))}} \tag{2.38}$$

where utilities $U(A)$ and $U(B)$ are calculated according to CPT criterion 2.12.
For the attraction factor they used:

$$q_A = min(f_A, f_B) tanh(a(E(A) - E(B)) \tag{2.39}$$

where expected utilities $E(A)$ and $E(B)$ are calculated utilizing a CARA utility function:

$$E(A) = p_1^A u(W + x_1^A) + p_2^A u(W + x_2^A) \tag{2.40}$$

$$u(x) = 1 - e^{-\eta x} \quad \eta \geq 0 \tag{2.41}$$

In the present work, we propose to parametrize the utility factor as a logit-RDU function:

$$f_A = \frac{1}{1 + e^{\phi(\tilde{U}(B) - \tilde{U}(A))}} \tag{2.42}$$

where utilities $\tilde{U}(A)$ and $\tilde{U}(B)$ are calculated according to RDU criterion (2.8), adopting either 2.10 or 2.11. For the attraction factor, the above mentioned parametrization is used.

When the CPT framework is adopted to build f_A we call this model *CPT-based QDT* (QDT_{CPT}); if the RDU framework is used instead, we call such model *RDU-based QDT* (QDT_{RDU}). In particular, if utility function 2.10 is used, it is referred as QDT_{RDU1}, otherwise (using 2.11) QDT_{RDU2}.

Before presenting the estimation procedure, we think it is useful to have in one page all the decision making models analyzed in practice.

2.7 Summary of models investigated

Task: two lotteries A and B, each one made of 2 possible outcomes with relative probability of occurrence.

2.7.1 logit-CPT

5 free parameters: $\alpha, \lambda, \delta, \gamma, \phi$

$$p(B \succeq A) = \frac{1}{1 + e^{\phi(U(A) - U(B))}}$$

$$U(A) = \begin{cases} w(p_1^A)v(x_1^A) + (1 - w(p_1^A))v(x_2^A), & \text{for lotteries with only gains or only losses} \\ w(p_1^A)v(x_1^A) + w(p_2^A)v(x_2^A), & \text{for mixed lotteries} \end{cases}$$

where x_1^A and x_2^A have been ordered such that:

- $x_1^A \geq x_2^A$ if both are positive.
- $x_1^A \leq x_2^A$ if both are negative.

$$v(x) = \begin{cases} x^\alpha, & \text{if } x \geq 0 \quad \alpha > 0 \\ -\lambda(-x)^\alpha, & x < 0 \quad \lambda > 0 \end{cases}$$

$$w(p) = exp(-\delta(-ln(p))^\gamma) \quad \delta > 0 \quad \gamma > 0$$

2.7.2 logit-RDU

4(5) free parameters for logit-RDU1(2).

$$p(B \succeq A) = \frac{1}{1 + e^{\phi(U(A) - U(B))}}$$

$$U(A) = w(p_1^A)v(W + x_1^A) + (1 - w(p_1^A))v(W + x_2^A)$$

where $V_1^A \geq V_2^A$ and W is the initial wealth.

$$w(p) = exp(-\delta(-ln(p))^\gamma) \quad \delta > 0 \quad \gamma > 0$$

RDU1

$$v_1(x) = \begin{cases} \dfrac{x^\alpha - 1}{\alpha}, & \alpha \neq 0 \\ ln(x), & \alpha = 0 \end{cases}$$

RDU2

$$v_2(x) = \frac{1}{\beta}\{1 - exp[-\beta(\frac{x^\alpha - 1}{\alpha})]\}$$

2.7.3 RDU1/2-based QDT and CPT-based QDT

It has 2 parameters more (η, a) than the classical theory adopted for the utility factor.

$$p(B \succeq A) = f_B + q_B$$

$$f_B = \frac{1}{1 + e^{\phi(U(A)-U(B))}}$$

where U(A) and U(B) are calculated according to one of the "classical" theories above. (Equation 2.12 or 2.8)

$$q_B = min(f_A, f_B) tanh(a(E(B) - E(A))$$

$$E(A) = p_1^A u(W + x_1^A) + p_2^A u(W + x_2^A)$$

where W is the initial wealth and the utility function used is:

$$u(x) = 1 - e^{-\eta x} \quad \eta \geq 0$$

Chapter 3

Case study and Model Calibration

3.1 Description of the studied experiment

142 subjects faced 91 binary decision tasks, where each option is a binary lottery, as shown in figure 3.1. The range of the outcomes was -100€ to 100€; at the end of each individual session one of the chosen lotteries was really played and the participants earned $\frac{1}{10}$ of the outcome. To avoid net losses, the participants were endowed with an "initial wealth" of 10€, so that the worst case would have been coming home with 0€.
Different types of lotteries have been used in order to isolate several psychological patterns (risk-aversion, loss-aversion etc.): only gains, only losses and mixed games.
The same experiment was repeated after two weeks, but randomly changing the order of questions. With "time 1" ("time 2") we will refer to the first (second) session.
It is important to stress that the random ordering of questions avoids order effect only at the aggregate level (i.e when we pool the data coming from all the participants), but not at individual level, where memory effects may play an important role in subsequent choices.

As already anticipated in the introduction, we will essentially fit the models to the data relative to time 1, and then we will test their predicition power with data at time 2.

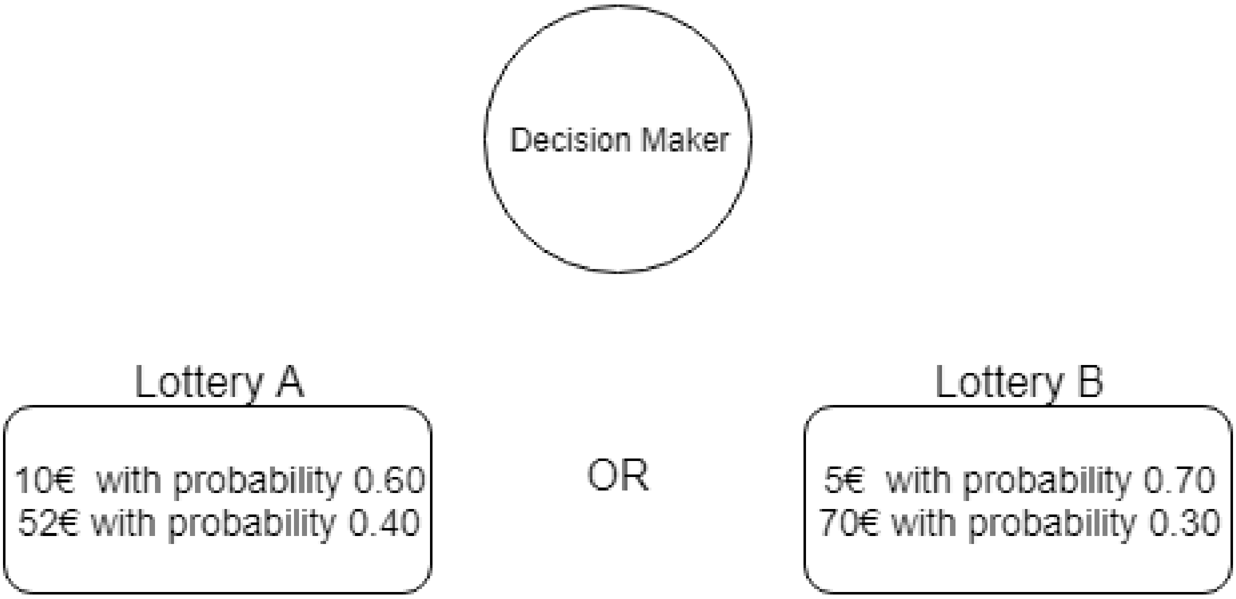


Figure 3.1: Typical decision-task of the experiment

3.2 Parameter estimation methods

3.2.1 Maximum Likelihood Estimation(MLE)

Since all the analyzed models are probabilistic, we can easily implement maximum likelihood estimations.
Adopting the same formalism of [16] for simplicity, the answers of the subject $i \in \{1...142\}$ at lottery $j \in \{1...91\}$ at time 1 are encoded in $\left(\phi_j^i\right)$, being defined as:

$$\phi_j^i = \begin{cases} 0, & \text{if subject i chooses A in the gamble j} \\ 1, & \text{if subject i chooses B in the gamble j} \end{cases} \tag{3.1}$$

We can further define a rectangular 91x142 matrix Φ, with $\Phi_{ij} = \phi_j^i$.

At the aggregate level, we gather the data coming from all the participants, i.e we aim to estimate the so called *representative agent*. Roughly speaking, it is more or less equal to assume that we have only one subject facing 91x142 decisions and we estimate the parameters such that the model best explains/predicts the data.

This can be done by maximizing the following aggregate likelihood function:

$$\begin{aligned}\mathcal{L}^{agg}(\vec{m};\Phi) &= \prod_{i=1}^{142}\prod_{j=1}^{91} p_{A_j}^{1-\phi_j^i}(\vec{m})(1-p_{A_j}(\vec{m}))^{\phi_j^i} \\ &= \prod_{i=1}^{142}\prod_{j=1}^{91} p_{A_j}^{1-\phi_j^i}(\vec{m})p_{B_j}^{\phi_j^i}(\vec{m}) \\ &= \prod_{j=1}^{91} p_{A_j}^{142-N_j}(\vec{m})p_{B_j}^{N_j}(\vec{m}) \qquad (3.2)\end{aligned}$$

where $\vec{m}$ is the parameter vector relative to the model considered (e.g. $\vec{m} = (\alpha, \delta, \gamma, \phi)$ for logit-RDU1) and $N_j = \sum_{i=1}^{142} \phi_j^i$ is the number of subjects that chose "B" in the j-th gamble.
For more computational stability we take the logarithm of expression 3.2:

$$\boldsymbol{log\mathcal{L}^{agg}(\vec{m};\Phi) = \sum_{j=1}^{91}\Big[(142-N_j)log(p_{A_j}(\vec{m})) + N_j log(p_{B_j}(\vec{m}))\Big]} \qquad (3.3)$$

$$\boldsymbol{\vec{m}_*^{agg} = \underset{\vec{m}}{\text{argmax}}\, \mathcal{L}^{agg}(\vec{m};\Phi) = \underset{\vec{m}}{\text{argmax}}\, log\mathcal{L}^{agg}(\vec{m};\Phi)} \qquad (3.4)$$

At the individual level, we can analogously maximize a likelihood function for each subject:

$$\boldsymbol{\mathcal{L}^{i}(\vec{m};\Phi) = \prod_{j=1}^{91} p_{A_j}^{1-\phi_j^i}(\vec{m})p_{B_j}^{\phi_j^i}(\vec{m})} \qquad (3.5)$$

$$\boldsymbol{\vec{m}_*^{i} = \underset{\vec{m}}{\text{argmax}}\, \mathcal{L}^{i}(\vec{m};\Phi)} \qquad (3.6)$$

Despite such approach is in principle correct, the problem of *overfitting* can arise: informally speaking, by overfitting we mean a situation when, once the model is fit to the data, it "perfectly" explains them, but it is unable to provide reliable predictions, i.e it has a low degree of generalization.
As highlighted by K.P. Burnham and D.R. Anderson [31], the less data are available, the more the possibility of overfitting is present. In other words, the noise deriving from a small sample can be misleadingly interpreted as "underlying structure".
In order to deal with this issue, several "regularization" procedures have been proposed [32]. In simple terms, a regularization is a way to constrain the fit by incorporating additional information (e.g. priors).

In our case individuals face relatively few choices (91); therefore, at the individual level overfitting should be taken into account. For CPT, such issue has already been investigated for example by J. Rieskamp et al. (2011) [33], where they showed the power of a hierarchical bayesian parameter estimation with respect to a standard MLE.

In the following subsection, we present the regularization adopted in the present work.

3.2.2 Hierarchical Maximum Likelihood(HML)

The hierarchical maximum likelihood method can be seen as a "trade-off" between two extrema: on one side the representative agent approach, which describes *each* individual in the same way, and on the other the individual level approach, which characterizes each decision maker separately. Informally speaking, the idea is that subjects are not *too different* from each other, i.e their parameters come from a common density distribution.
Farrell and Ludwig (2008) [34] introduced this procedure because in psychology it is common to have samples of small size. In this way, the individual likelihood is *regularized* by this group level prior distribution.

R. ten Brincke and R.O. Murphy [19] applied this method to the presented experiment for CPT; we present the same method for RDU1.

The procedure is made of two steps:
We *assume* that the distributions of α, γ and δ are lognormal. The distribution of the sensitivity ϕ is not evaluated, since it in principle depends on the other parameters, and so ϕ is fixed with its aggregate level estimate. Each distribution is defined by location (μ) and scale (σ) parameters:

$$\alpha \sim LogN(\mu_\alpha, \sigma_\alpha) \quad \gamma \sim logN(\mu_\gamma, \sigma_\gamma) \quad \delta \sim logN(\mu_\delta, \sigma_\delta) \tag{3.7}$$

$$P_\alpha = \{\mu_\alpha, \sigma_\alpha\}; P_\delta = \{\mu_\delta, \sigma_\delta\}; P_\gamma = \{\mu_\gamma, \sigma_\gamma\} \tag{3.8}$$

We remind that a random variable is lognormally distributed if its logarithm is normally distributed.
The first step consists in estimating the $\{\mu\}$ and $\{\sigma\}$, that amounts to compute the prior distribution from the aggregate level point of view.

This is done in the following way:

$$\{P_\alpha^*, P_\delta^*, P_\gamma^*\} = \underset{P_\alpha, P_\delta, P_\gamma}{\text{argmax}} \prod_{i=1}^{142} \iiint \Big[\prod_{j=1}^{91} p_{A_{j,i}}^{1-\phi_j^i} p_{B_{j,i}}^{\phi_j^i} \Big] ln(\alpha|P_\alpha) ln(\gamma|P_\gamma) ln(\delta|P_\delta) d\alpha d_\gamma d_\delta \tag{3.9}$$

The idea behind expression 3.9 is to weight each individual likelihood function by the same *global* prior distribution, so that each parameter set $\vec{m} = (\alpha, \delta, \gamma, \phi)$ contributes relevantly to the integral not only if it maximizes the likelihood, but also if it is likely to be generated by the group level distribution.
In [19], they approximated the integral by a Monte Carlo method generating 2,500 uniform random values for each dimension. We decided to evaluate the integral by *importance sampling*, i.e we generate values directly from the lognormal distributions inside the integral. We preferred the latter because the drawback of uniform sampling is that "if points are chosen evenly in volume, we rarely consider the points close to the *peak* which give the dominant contribution" [35].
Being the computational process very intense we have implemented it only for RDU1.

Once the prior distribution has been computed, the real optimization can be carried out by maximing the regularized individual likelihood as follows:

$$\boldsymbol{\mathcal{L}_r^i(\vec{m}; \Phi) = ln(\alpha|P_\alpha^*) ln(\gamma|P_\gamma^*) ln(\delta|P_\delta^*) \prod_{j=1}^{91} p_{A_j}^{1-\phi_j^i}(\vec{m}) p_{B_j}^{\phi_j^i}(\vec{m})} \tag{3.10}$$

$$\boldsymbol{\vec{m}_*^i = \underset{\vec{m}}{\text{argmax}}\, \mathcal{L}_r^i(\vec{m}; \Phi)} \tag{3.11}$$

3.3 Model comparison Criteria

In this section, we present the statistical tests and criteria used to select and compare the models fit to data.
Before doing that, it is important to recall the possible relations among different models because, depending on the situation, different tests have to be adopted.

3.3.1 Possible relations among models

As reported by C. Gourieroux and A. Monfort (1994) [36], we consider T observations of pairs of variables (y_t, z_t), $t = 1....T$ where in our case the $\{y_t\}$ are the binary variables representing the choices of subjects, while the $\{z_t\}$ are actually constants describing the lotteries.
We aim to discover the true (unknown) conditional distribution generating the observations, which is denoted as $h^0(y|z)$.
Two rival conditional models $\mathbf{F}_\theta = \{f(y|z;\vec{\theta}), \vec{\theta} \in \vec{\Theta} \subset R^p\}$ and $\mathbf{G}_\gamma = \{g(y|z;\vec{\gamma}), \vec{\gamma} \in \vec{\Gamma} \subset R^q\}$ can be *nested*, *partially non-nested* (overlapping) or *globally non-nested*.

$\mathbf{G}_\gamma$ is nested in $\mathbf{F}_\theta \iff I_G(\vec{\gamma}, z) = 0$ for all possible values of $\vec{\gamma}$ and z, where $I_G(\vec{\gamma}, z)$ is defined as the minimal Kullback–Leibler divergence between $g(y|z;\vec{\gamma})$ and $\mathbf{F}_\theta$:

$$I_G(\vec{\gamma}, z) = \inf_{\vec{\theta} \in \vec{\Theta}} I_{GF}(\vec{\gamma}, \vec{\theta}; z) \tag{3.12}$$

$$I_{GF}(\vec{\gamma}, \vec{\theta}; z) = log\frac{g(0|z;\vec{\gamma})}{f(0|z;\vec{\theta})}g(0|z;\vec{\gamma}) + log\frac{g(1|z;\vec{\gamma})}{f(1|z;\vec{\theta})}g(1|z;\vec{\gamma}) \tag{3.13}$$

In simpler terms, the nested condition means that any distribution $g(y|z;\vec{\gamma})$ can be obtained with $\mathbf{F}_\theta$ for some value of the parameter vector $\vec{\theta}$ (the same definitions hold for $\mathbf{F}_\theta$ nested in $\mathbf{G}_\gamma$).

$\mathbf{F}_\theta$ and $\mathbf{G}_\gamma$ are globally non-nested $\iff I_G(\vec{\gamma}, z) \neq 0$ AND $I_F(\vec{\theta}, z) \neq 0$ for all possible values of $\vec{\gamma}$, $\vec{\theta}$ and z. If this is not the case, they are said to be partially non-nested.

3.3.2 Information Criteria

One first way to select the best model describing the data is by adopting the so-called *information criteria*. Within decision theory, the most used are the Akaike Information Criterion (AIC) [37] and Bayesian information criterion (BIC), also called Schwarz information criterion [38].
For Akaike, the best model is the one with the smallest AIC:

$$AIC = -2logL + 2p \tag{3.14}$$

where L is the likelihood of the considered model and p is the number of parameters to estimate.

The Schwarz criterion proposes instead to select the best model according to the smallest BIC:

$$BIC = -2logL + plog(n) \tag{3.15}$$

where n is the sample size.

In case the competing models have the same number of the parameters, the one with the biggest likelihood is selected.
The main drawback of these methods is that they don't admit non-discrimination as a possible outcome of the comparison, forcing us somehow to choose a winner. Therefore, an alternative way of selecting models is presented.

3.3.3 Nested Hypothesis Testing

The simplest situation is when the one of the competing models is nested in the other. In such case, according to Wilks' Theorem [39], we first define:

$$\Lambda = \frac{L_{small}}{L_{big}} \tag{3.16}$$

where L_{small} is the likelihood of the nested model (null hypothesis) and L_{big} is the likelihood of the nesting model. If the null hypothesis is true, when the sample size approaches ∞, we have:

$$D = -2log(\Lambda) \sim \chi^2(k) \tag{3.17}$$

where χ^2 is a chi-squared distribution with k degrees of freedom, k being the difference between the number of parameters of the "big" model and the "small" one.

In order to quantify the idea of statistical significance of evidence, a p-value is computed. The p-value is defined as the probability, under the null hypothesis, of obtaining a result equal to or more extreme than what was actually observed. Informally, the smaller the p-value, the stronger is the evidence against the null hypothesis (nested model).

Recalling the form of the CDF of the $\chi^2(k)$:

$$F(x;k) = P(\chi^2(k) \leq x) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})} \tag{3.18}$$

where $\gamma(s,t)$ is the lower incomplete gamma function and $\Gamma(k)$ the ordinary gamma function, the p-value corresponds to:

$$p - value = P(\chi^2(k) \geq D) = 1 - F(D;k) = 1 - \frac{\gamma(\frac{k}{2}, \frac{D}{2})}{\Gamma(\frac{k}{2})} \tag{3.19}$$

3.3.4 Non-nested Hypothesis Testing

A more complicated and controversial case is with non-nested hypotheses. The first attempts to deal with non-nested case were proposed by Cox (1961) [40] and Atkinson (1969) [41].
As White (1982) points out [42], the classical properties of the maximum likelihood estimator relies on a fundamental assumption, often implicitly stated: *correct specification*. In simple terms, a model is correctly specified if the true distribution is known to be "contained" in it.
In many situation, ours included, this is not known a-priori. What happens if we do not assume correct specification for any model?

Relaxing this assumption, White provided "simple conditions under which the maximum likelihood estimator is a strongly consistent estimator for the parameter vector which minimizes the Kullback-Leibler Information Criterion (KLIC)" between the true distribution $h^0(y|z)$ and the conditional model analyzed $\mathbf{F}_\theta = \{f(y|z;\vec{\theta}), \vec{\theta} \in \vec{\Theta} \subset R^p\}$.

This being said, it is natural then to adopt statistical tests based on minimization of KLIC, where there are no assumption of correct specification. One such test is the Vuong Test [43].

Another problem, often overlooked, is the possibility of overlapping between models: for example, G. Harrison and J. Swarthout (2016) [44] state that RDU and CPT are non-nested models, seeming to neglect such possibility. Indeed, looking at the definitions in subsection 3.3.1, the global non-nested condition is fairly restrictive. On the other hand, $\mathbf{F}_\theta$ and $\mathbf{G}_\gamma$ can overlap even if $I_G(\vec{\gamma}, z) = 0 \quad \text{OR} \quad I_F(\vec{\theta}, z) = 0$ for only one combination of $\vec{\gamma}$, $\vec{\theta}$ and z.
Therefore, we implement the most general procedure proposed by Vuong, in which we neither assume correct specification nor global non-nesting.

Despite the precedent remark, we decided to follow "in spirit" [44] adopting not only the Vuong Test, but also the Clarke Test [45]. It is a good idea to use both because they compensate each other in a way: the Clarke Test has a slightly higher probability of choosing the wrong model, while the Vuong Test has a slightly higher probability of choosing neither model (See [46] for a formal comparison of the trade-offs between these errors).

A detailed description of the tests used is provided in Appendix .1.

Chapter 4

Results

The implementation of logit-RDU vs RDU-based QDT has been made using the code for logit-CPT vs CPT-based QDT developed by S.Vincent et al [16] with suitable modifications. In the following, we will refer to these models as RDU, QDT_{RDU}, CPT, QDT_{CPT}, respectively. We will use the shorthand RDU_{type} models when we generically refer to RDU and QDT_{RDU} and similarly CPT_{type} models for CPT and QDT_{CPT}.
Regarding the proper parametrization of our models, see section 2.7. For the RDU_{type} ones we recall again the adoption of two different utility functions:

- $v(x) = \frac{x^\alpha - 1}{\alpha} \rightarrow RDU_1, QDT_{RDU1}$
- $v(x) = \frac{1}{\beta}\{1 - exp[-\beta(\frac{x^\alpha - 1}{\alpha})]\} \rightarrow RDU_2, QDT_{RDU2}$

4.1 Aggregate level

At the aggregate level, we first look at the mean squared error (MSE) of the estimated probability of choosing gambles B with the empirical probability at time 1 (FIT MSE) and at time 2 (Predicted MSE). The results are in Table 4.1, where we report also the corresponding logLikelihood and the values of the estimated parameters. If we refer only to these estimations, the main conclusions are the following:

- no general improvement adopting RDU_{type} models instead of CPT_{type} ones; QDT_{CPT} performs significantly better than the others.
- QDT_{RDU} estimation is better than RDU one, especially in RDU_2 vs QDT_{RDU2}

From the conceptual point of view, the following comments are important:

- QDT_{RDU1} has $a = -0.0839$, slightly <0. Descriptively speaking, it may be a signal of irrational behavior.
- The major flexibility of RDU_2 and QDT_{RDU2} results in the presence of an inflection point in their utility function.

The second point deserves a deeper analysis. The presence of an inflection point tells us that there is a domain region where the utility function is risk-seeking ($v''(x) > 0$) and another region where it is risk-averse ($v''(x) < 0$). It is interesting to see where this inflection point is, in order to compare it with the reference point taken in CPT_{type} models to discriminate between gains and losses. Figure 4.1 shows $v''(x)$ of RDU_2 (subplot 4.1a) and $v''(x)$ of QDT_{RDU2} (subplot 4.1b)

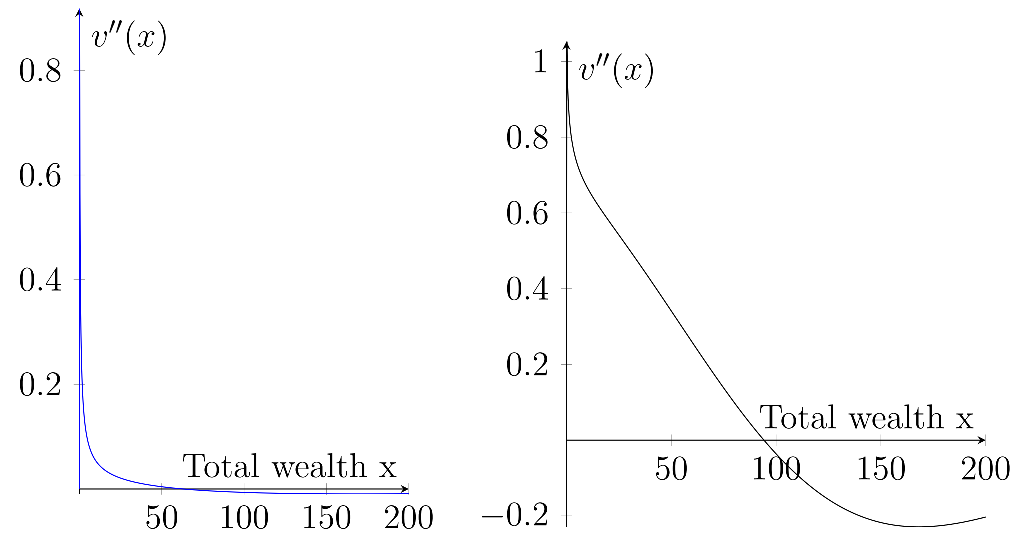
For RDU_2, the inflection point is at $x \simeq 62$; recalling that $W = 100$ is the initial wealth provided to the participants, it would imply a risk-seeking behavior only when losses are (in modulus) bigger than $62 - 100 = -38$.

For QDT_{RDU2} instead, it is at $x \simeq 92$, that would imply a risk-seeking behavior only when losses are (in modulus) bigger than $92 - 100 = -8$. This is very close to 0, the "reference point" taken in CPT_{type} models. Such result may be informally regarded as a confirmation that, at the aggregate level, CPT_{type} models best characterize the *average* behavior.

In order to better understand our general results, we evaluated the same quantities dividing the gambles by type: only gains, only losses and mixed. (Table 4.2). Again, QDT_{CPT} outperforms the others. Moreover we have that:

- QDT_{RDU2} performs significantly better than RDU_2 in each type of gamble.
- QDT_{RDU1} performs slightly better than RDU_1, except for only gain region, where they are almost the same.
- RDU, in particular RDU_1, performs better than CPT when gambles with only losses are concerned.

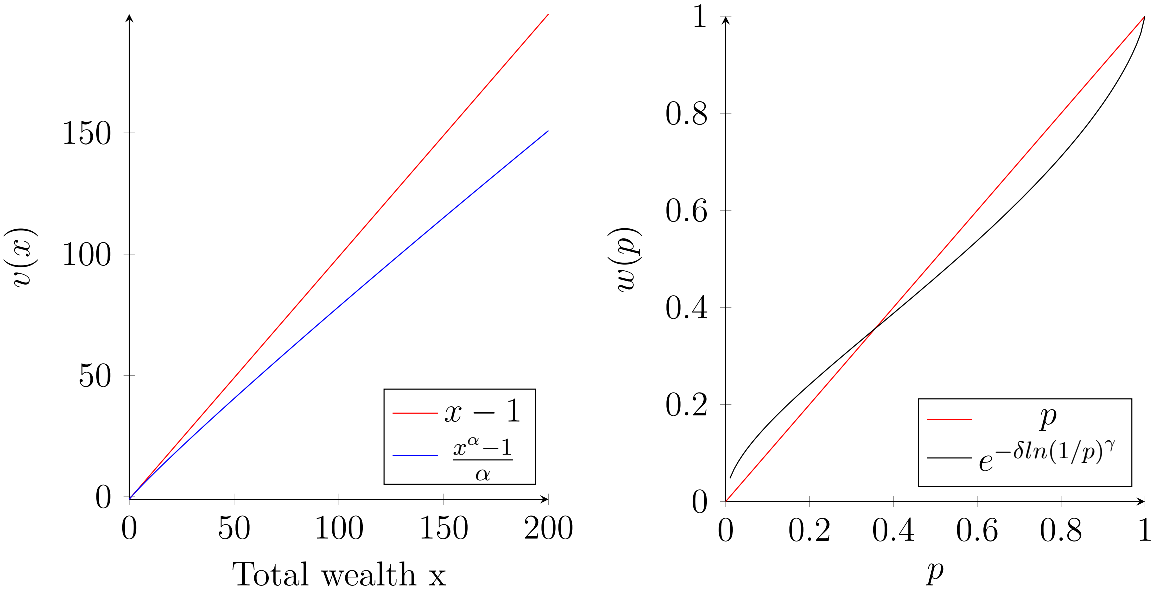
The latter point may indicate that the asset integration (participants *earn less* instead of perceiving losses) does play a role especially with this type of gambles.



(a) Second derivative of RDU_2 utility function

(b) Second derivative of QDT_{RDU2} utility function

Figure 4.1: Second derivatives of RDU_2 and QDT_{RDU2} utility functions. Their change of sign shows a change of behavior from risk-seeking to risk-averse.



(a)

(b)

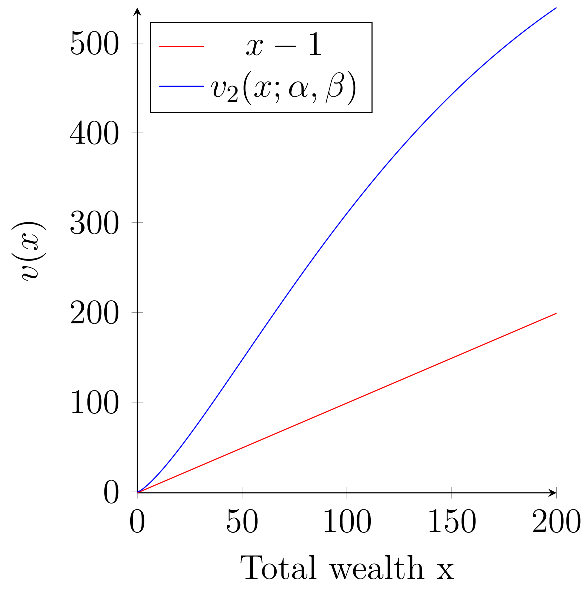
Figure 4.2: RDU_1.(a) utility function with $\alpha = 0.9357$; (b) weighting function with $\delta = 1.0090$, $\gamma = 0.7233$.

Estimation	RDU_1	QDT_{RDU1}	RDU_2	QDT_{RDU2}	CPT	QDT_{CPT}
Fit MSE	1.0920	1.0165	1.0025	0.7584	0.7294	0.5244
Predicted MSE	0.9978	0.9508	0.9430	0.7577	0.7594	0.5886
logLikelihood	-7541.4	-7518.2	-7511.5	-7442.0	-7431.5	-7371.9
α	0.9357	0.7423	1.3049	1.8888	0.7309	0.6892
β	-	-	0.0014	1.7809e-04	-	-
λ	-	-	-	-	1.1129	1.0156
δ	1.0090	0.9894	0.9524	0.9318	0.8771	0.8862
γ	0.7233	0.7572	0.7257	0.7138	0.6514	0.6280
ϕ	0.1066	0.2697	0.0339	0.0030	0.2954	0.3702
η	-	0.0408	-	0.1014	-	0.0523
a	-	-0.0839	-	2.1271	-	1.4735

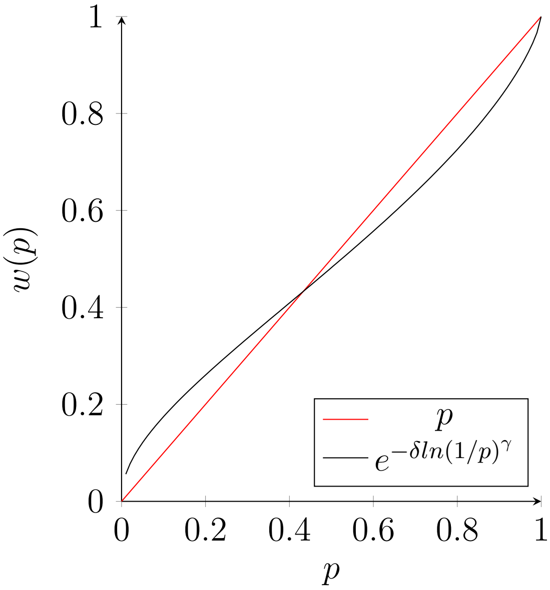
Table 4.1: Estimation Results. We report the Mean Square Error of the estimated probability of choosing gamble B both with the empirical probability at time 1 (Fit MSE) and with the empirical probability at time 2 (Predicted MSE). α is the exponent of the power function (utility/value function), β is the additional parameter for the utility function of RDU_2 and QDT_{RDU2}, δ and γ are the parameters of the Prelec transformation function, λ is the loss aversion parameter, ϕ is the sensitivity of the choice function, η and a are the parameters for the attraction factor.

FIT/PRED MSE	RDU_1	QDT_{RDU1}	RDU_2	QDT_{RDU2}	CPT	QDT_{CPT}
Fit MSE	1.0920	1.0165	1.0025	0.7584	0.7294	0.5244
Predicted MSE	0.9978	0.9508	0.9430	0.7577	0.7594	0.5886
ONLY GAINS Fit MSE	0.4403	0.4477	0.3878	0.3287	0.2418	0.2085
ONLY GAINS Predicted MSE	0.4676	0.4737	0.4241	0.3819	0.2947	0.2785
MIXED Fit MSE	0.4386	0.3803	0.3925	0.2735	0.2662	0.1707
MIXED Predicted MSE	0.3371	0.3014	0.2813	0.2262	0.2058	0.1840
ONLY LOSSES Fit MSE	0.2131	0.1885	0.2223	0.1561	0.2214	0.1452
ONLY LOSSES Predicted MSE	0.1931	0.1757	0.2374	0.1496	0.2589	0.1261

Table 4.2: Detailed MSE. QDT_{CPT} outperform the others. QDT_{RDU} models perform better than RDU ones. RDU_1 performs significantly better than CPT when gambles with only losses are concerned.

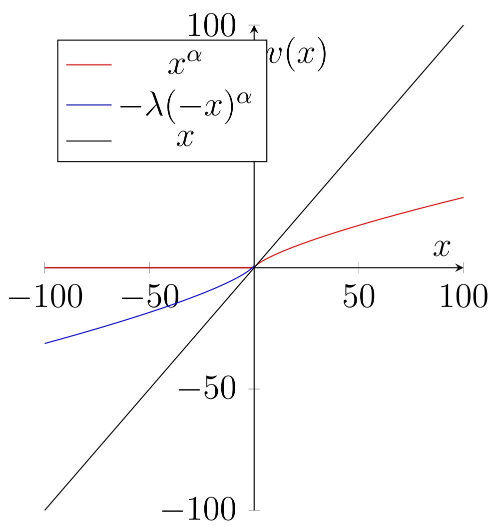


(a)

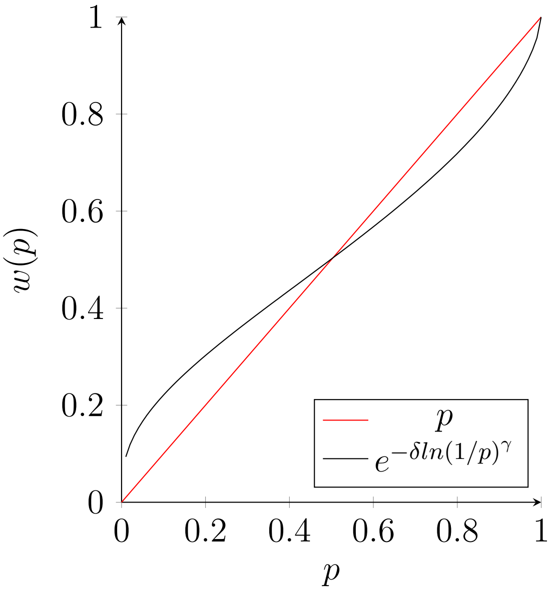


(b)

Figure 4.3: RDU_2.(a) Utility function with $\alpha = 0.9357$, $\beta = 0.0014$; (b) weighting function with $\delta = 0.9524$, $\gamma = 0.7257$.



(a)



(b)

Figure 4.4: CPT: (a) Value function with $\alpha = 0.9357$, $\lambda = 1.1129$; (b) weighting function with $\delta = 0.8771$, $\gamma = 0.6514$.

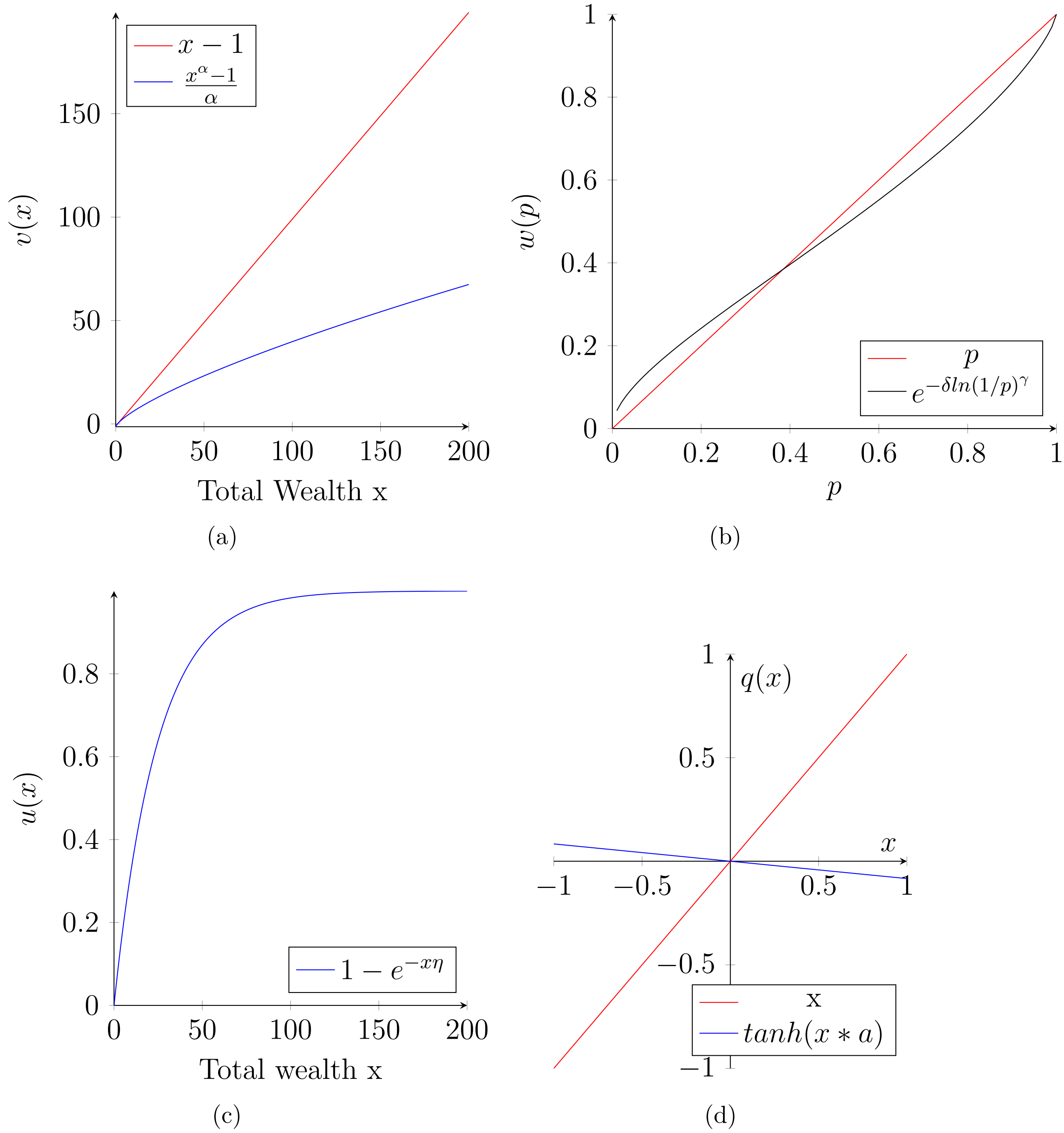


Figure 4.5: QDT_{RDU1}. (a) Utility function with $\alpha = 0.7423$; (b) weighting function with $\delta = 0.9894$, $\gamma = 0.7572$; (c) CARA utility function with $\eta = 0.0408$; (d) attraction factor with $a = -0.0839$.

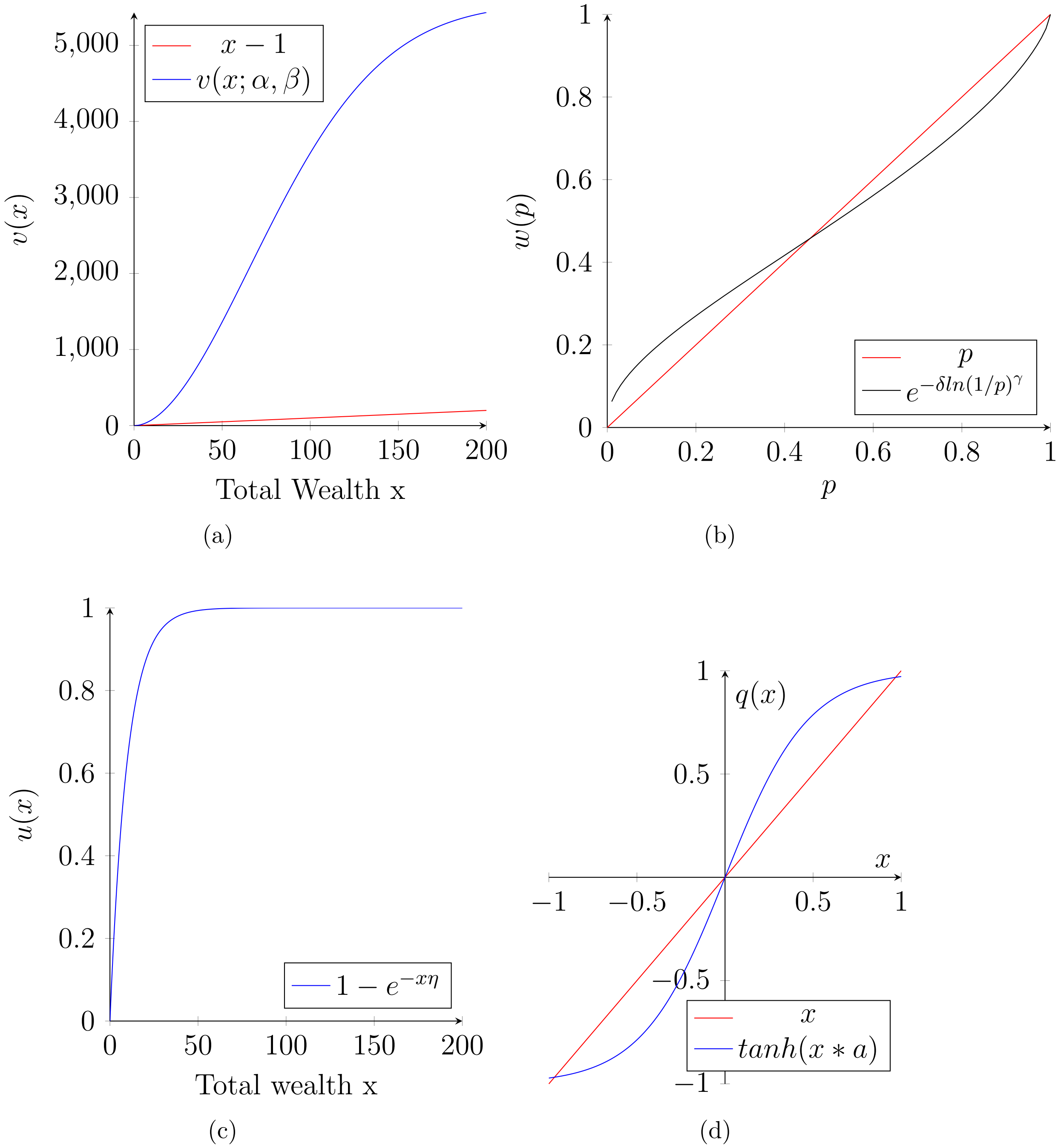
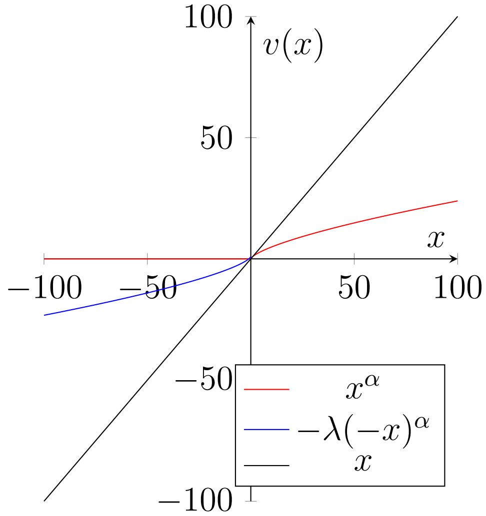
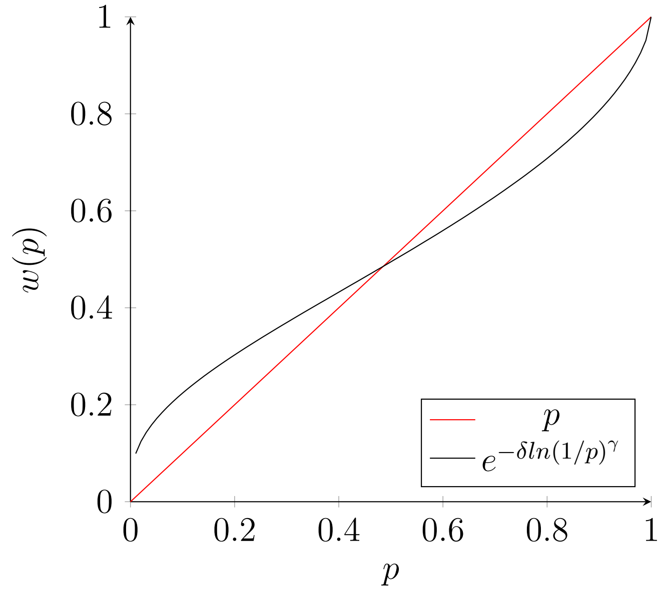


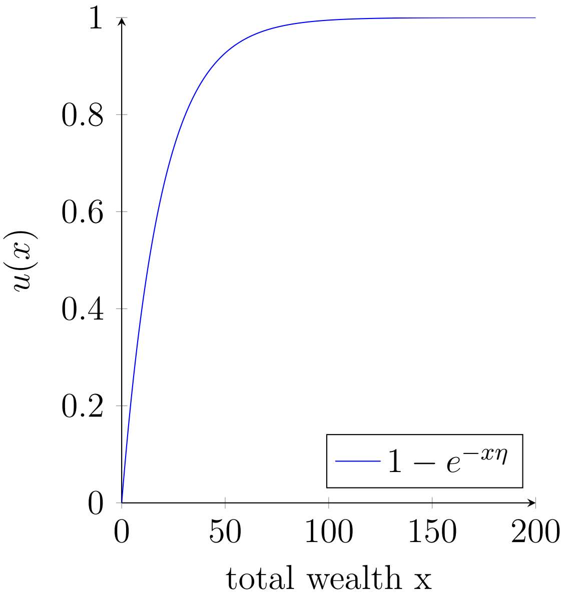
Figure 4.6: QDT_{RDU2}. (a) Utility function with $\alpha = 1.8888$, $\beta = 1.7809 \cdot 10^{-4}$; (b) weighting function with $\delta = 0.9318$, $\gamma = 0.7138$; (c) CARA utility function with $\eta = 0.1014$; (d) attraction factor with $a = 2.1271$.



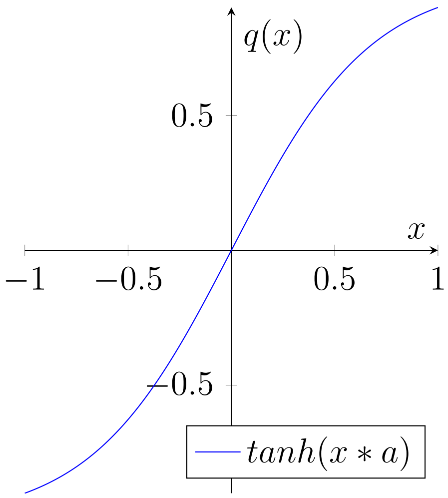
(a)



(b)



(c)



(d)

Figure 4.7: QDT_{CPT}. (a) Value function with $\alpha = 1.8888$, $\lambda = 1.0156$; (b) weighting function with $\delta = 0.8862$, $\gamma = 0.6280$; (c) CARA utility function with $\eta = 0.0.0523$; (d) attraction factor with $a = 1.4735$.

4.1.1 Hypothesis Testing at the aggregate level

Statistical tests give a more rigorous approach for the horse-race.
We first begin by comparing models within RDU_{type} models; let us recall the nesting relations among them:

- RDU_1 is nested in QDT_{RDU1} (sending η and a to 0)
- RDU_2 is nested in QDT_{RDU2} (sending η and a to 0)
- RDU_1 is nested in RDU_2 (sending β to 0)
- QDT_{RDU1} is nested in QDT_{RDU2} (sending β to 0)

Regarding the first two relations, we can safely use the Wilks' Theorem (subsection 3.3.3). We have to be more careful for the other two, as explained below.
If we precisely sticked to the definition of the two-parameter utility function (equation 2.11) provided by Danyang Xie in [25], we would have to impose $\beta >= 0$. Therefore, RDU_1 would be recovered from RDU_2 by putting β on the boundary of its domain. In such case, the test statistic does not follow a chi-2 with 1 degree of freedom but a mixed law such that the statistic is 0 with probability $\frac{1}{2}$ and follows a chi-2 with 1 degree of freedom with probability $\frac{1}{2}$ (H. Chernoff, 1954 [47]).
However, the condition $\beta >= 0$, together with $\alpha \leq 1$, was adopted in [25] so that the utility function always represents risk-averse behavior.

In our case instead, we do not impose these contraints, letting the estimation procedure decide if the utility function is risk-averse, risk-seeking, or both, as shown in the precedent subsection. Therefore, we can safely use the Wilks' theorem also for the last two comparisons.
For k=2, the cumulative distribution function of χ^2 is:

$$F(x;2) = P(\chi^2(2) \leq x) = 1 - \exp(-\frac{x}{2}) \tag{4.1}$$

$$p - value = 1 - F(D;2) = exp(-\frac{D}{2}) \tag{4.2}$$

For k=1:

$$F(x;1) = P(\chi^2(1) \leq x) = \frac{\gamma(\frac{1}{2}, \frac{x}{2})}{\Gamma(\frac{1}{2})} \tag{4.3}$$

$$p - value = 1 - F(D;1) = 1 - \frac{\gamma(\frac{1}{2}, \frac{D}{2})}{\Gamma(\frac{1}{2})} \tag{4.4}$$

where $\gamma(s,t)$ is the lower incomplete gamma function and $\Gamma(k)$ the ordinary gamma function.

In table 4.3 the p-values for all the comparisons are reported; they clearly show how QDT_{RDU} wins over RDU, and also that adding a parameter to the utility function (RDU_2 and QDT_{RDU2}) gives statistically significant improvement.

	$RDU_1 \in QDT_{RDU1}$	$RDU_2 \in QDT_{RDU2}$	$RDU_1 \in RDU_2$	$QDT_{RDU1} \in QDT_{RDU2}$
p-value	8.2830e-11	6.5085e-31	1.0325e-14	0

Table 4.3: p-values computed for all the comparisons among RDU_{type} models.

Next, we compare QDT_{RDU2}, the winner among RDU_{type} models, with QDT_{CPT} through Vuong Test and Clarke Test (table 4.4). Such results clearly show that QDT_{CPT} is selected (by far) at the aggregate level. This is not surprising, given the huge differences in MSE (table 4.2).

	Vuong Test	Clarke Test
p-value	5.6742e-6	6.1419e-6

Table 4.4: p-values for QDT_{RDU2} vs QDT_{CPT}

4.2 Individual level

In this section we are going to compare the results obtained at the individual level. We will first focus on the mean explained fraction (choices at time 1) and mean predicted fraction (choices at time 2), i.e. for each subject we calculate the fraction of gambles correctly explained/predicted, given the individual parameter estimation, and then take the average across the subjects. The parameters η and a, relative to the attraction term of QDT_{RDU1} and QDT_{CPT}, will be fixed using the values obtained at the aggregate level. This implies that RDU(CPT) is not nested in $QDT_{RDU}(QDT_{CPT})$ at this level.

Looking at the results in table 4.5 we get the following conclusions:

- for the mean predicted fraction, we have essentially the same result, both with RDU_{type} models and with CPT_{type} ones
- for the mean explained fraction, QDT_{RDU1} gives a slightly better result than RDU_1.

Mean Fraction	RDU_1	QDT_{RDU1}	CPT	QDT_{CPT}
Explained	0.74 (0.09)	0.75 (0.08)	0.76 (0.08)	0.77 (0.08)
Predicted	0.73 (0.10)	0.73 (0.10)	0.73 (0.10)	0.74 (0.09)

Table 4.5: Mean Explained and Predicted Fraction (standard deviation is in brackets). For each subject we calculate the fraction of gambles correctly explained/predicted, given the individual parameter estimation, and then take the average across the subjects. Regarding the mean predicted fraction, we may say that RDU_1 performs quite well, considering it has one parameter less than CPT. QDT_{RDU1} gives a slightly better mean explained fraction than RDU_1.

The similar performances of the models may be regarded as a further confirmation of an intrinsic limit of predictability due to randomness of choice, as suggested in [16].

4.2.1 "Naive" Subject Classification

It is also interesting to compare performance of the models for specific subjects, to account somehow for heterogeneity in the population. In a rough way, we record for each subject which model better fits his choices (both at time 1 and at time 2), having in this way a pseudo-classification of the subjects. Figures 4.8 and 4.9 present such analysis, where we that:

- QDT_{CPT} describes the majority of decision-makers; RDU_1 and CPT have the same subject percentage. (time 1)

- Majority (33 %) of subjects is better described by RDU_1 (time2)

The main drawback of this analysis is that it does not give an exact measure of how better one model is than another, but it is a good start.
However, we can try to understand how much these predictions actually differ. We form a vector taking the difference between explained (predicted) fraction of two models, then we sort it in descendant order and we plot it. Figures from 4.10 to 4.15 show this analysis for six pairs of models.
Useful for this discussion is also the sum of the elements of the difference vector, to see if the differences compensate each other.(Table 4.6) Here we report the main conclusions at this level:

- if we had to classify the subjects into RDU_1 and QDT_{RDU1} the order of magnitude of the differences (time 1 and 2) is too small to think at an heterogeneity of the sample.

- if we had to classify subjects into RDU_1 and CPT, at time 1 the vast majority would be CPT, while at time 2 the division would be more symmetric.

- if we had to classify subjects into QDT_{RDU1} and QDT_{CPT}, at time 1 the vast majority would be QDT_{CPT}, while at time 2 the division would be more symmetric.

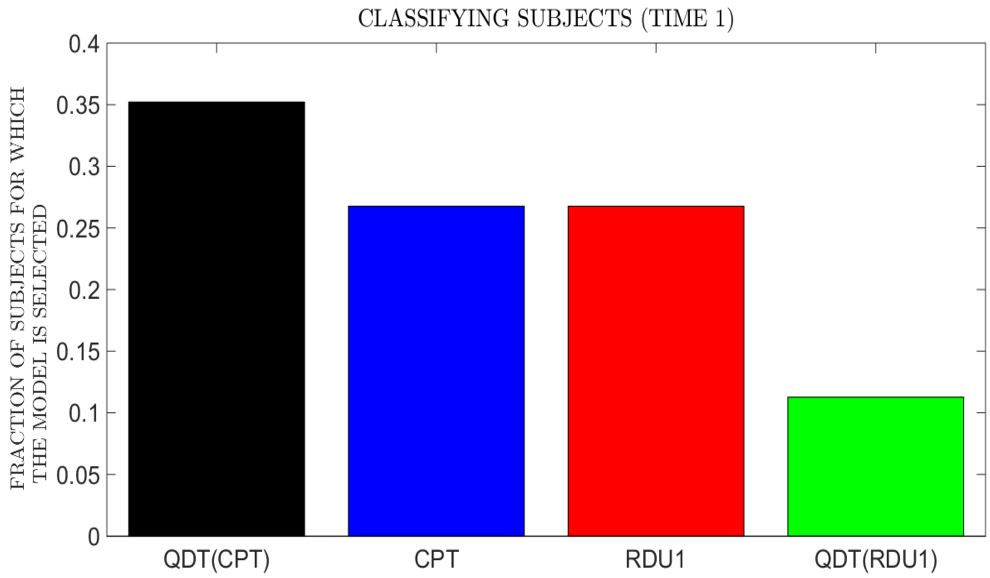


Figure 4.8: Subject Classification according to to the model that best predicts their individual choices at time 1. RDU_1 and CPT have the same subject percentage.

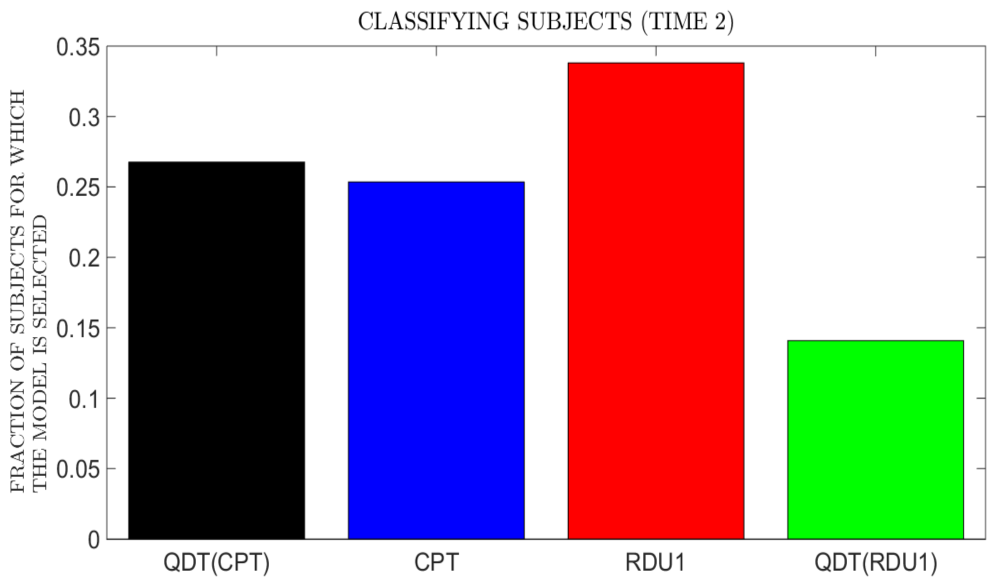


Figure 4.9: Subject Classification according to the model that best predicts their individual choices at time 2. 33 % of subject is better described by RDU_1.

	RDU_1-QDT_{RDU1} time 1/2	RDU_1-CPT time 1/2	QDT_{RDU1}-QDT_{CPT} time 1/2
Sum of differences	-0.90/0.20	-2.89/-0.05	-2.82/ -1.02

Table 4.6: Sum of the elements of the difference vector for several pairs of models at time 1 and time 2.

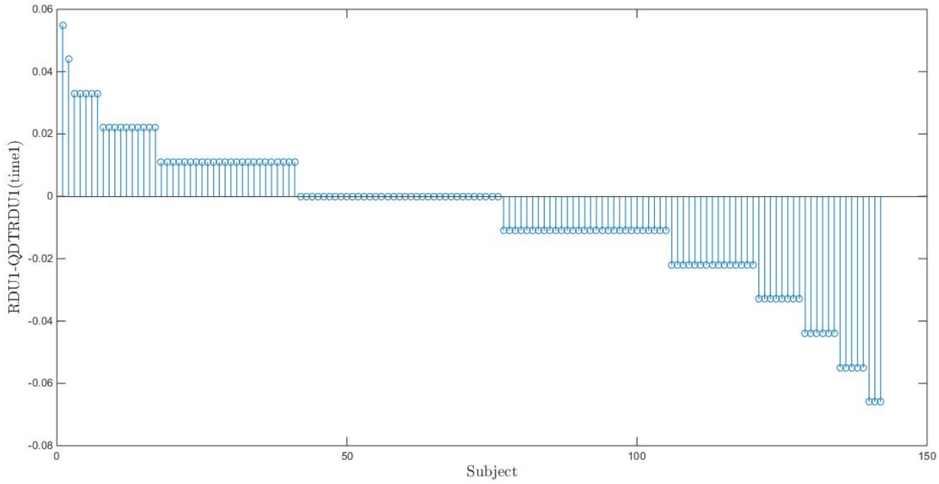


Figure 4.10: $RDU_1 - QDT_{RDU1}$ (time 1)

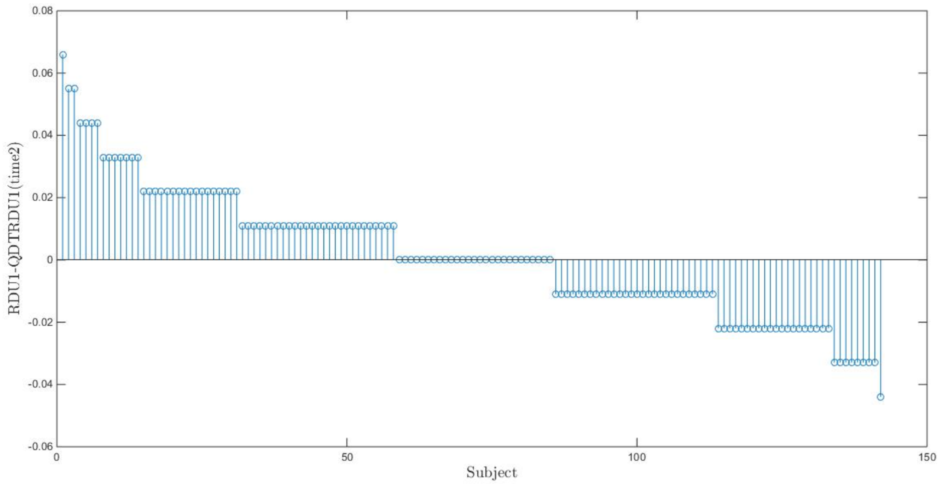


Figure 4.11: $RDU_1 - QDT_{RDU1}$ (time 2)

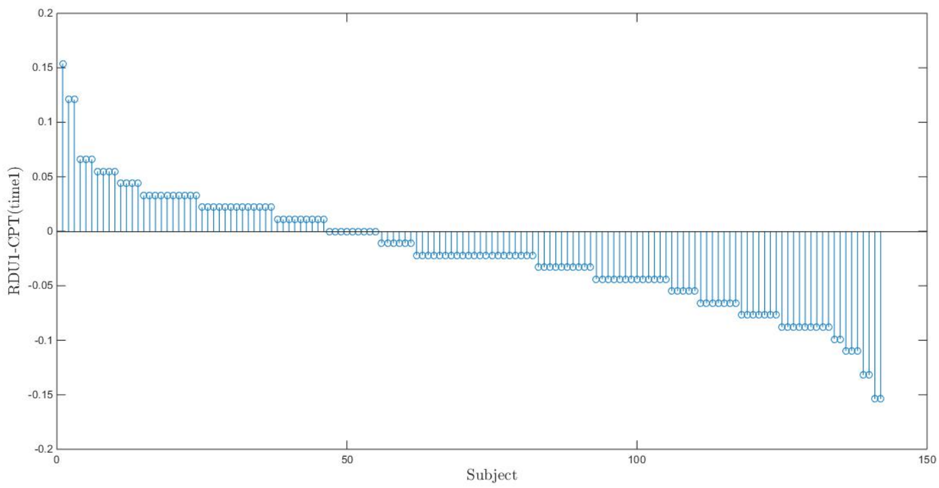


Figure 4.12: $RDU_1 - CPT$ (time 1)

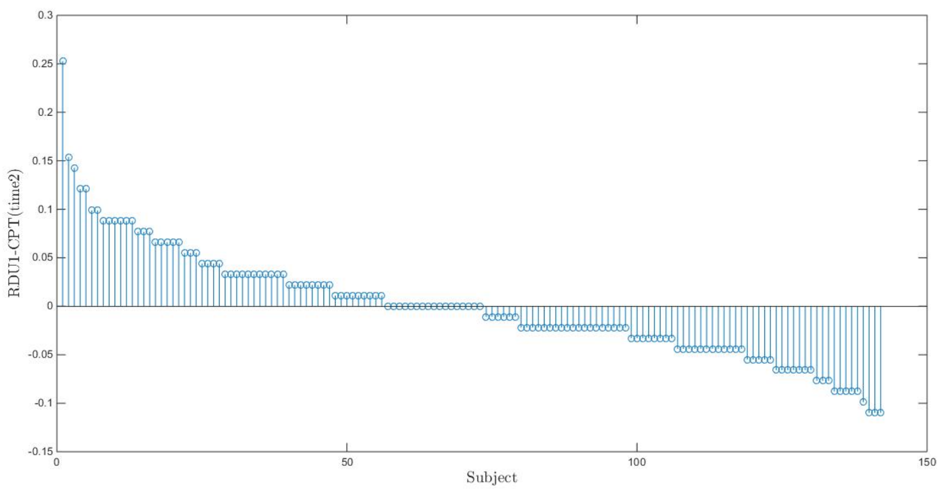


Figure 4.13: $RDU1 - CPT$ (time 2)

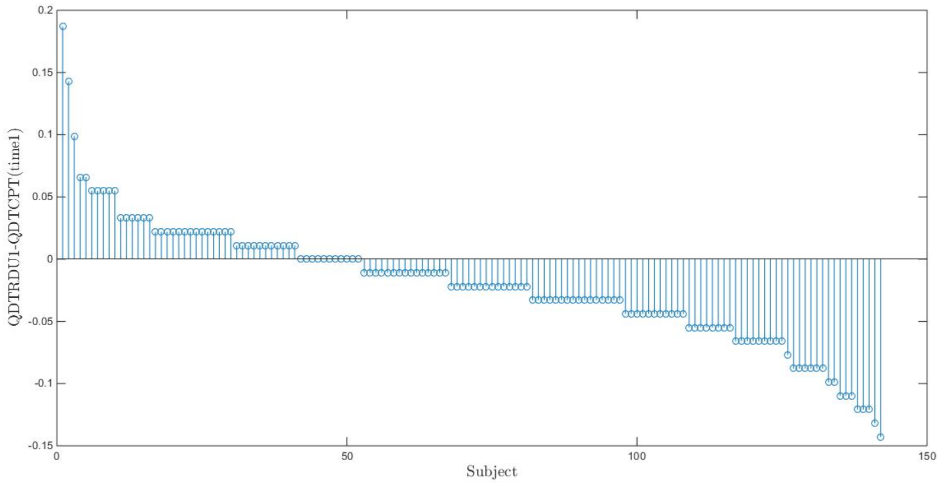


Figure 4.14: $QDT_{RDU1} - QDT_{CPT}$ (time 1)

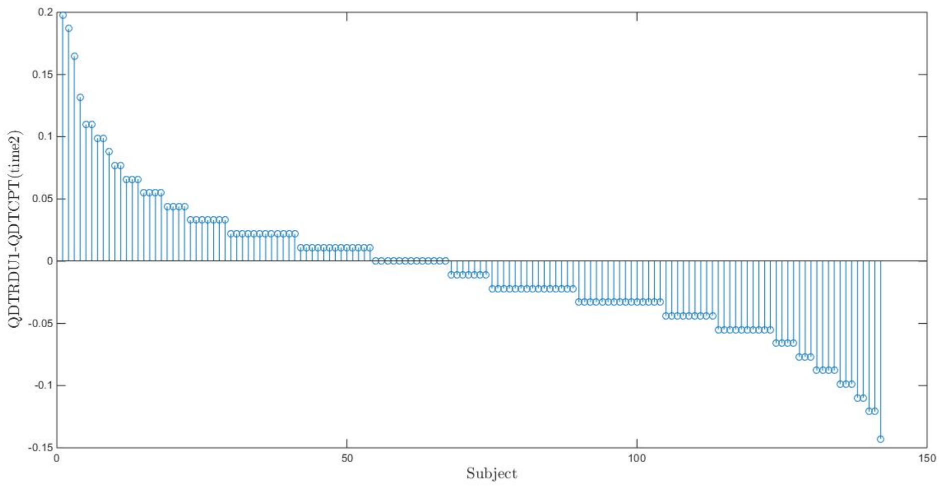


Figure 4.15: $QDT_{RDU1} - QDT_{CPT}$ (time 2)

4.2.2 Subject classification with hypothesis testing

We now shift to a more rigorous classification of the subjects, using appropriate statistical tests. As already said, the precedent model selection procedure does not take into account how much the performance of the theories differ from each other; therefore, for each subject we are "forced" to choose one winner, without allowing for heterogeneity within the subject himself.
The main feature of the following analysis is to relax the assumption that one of the competing models has to be chosen over the other, allowing non-discrimination as a possible outcome.

To classify subjects as $RDU1$, QDT_{RDU1}, CPT or QDT_{CPT}, the procedure is the following:

1. Likelihood based "quasi-nested" model selection (subsection 3.3.2) within each "type" of model:
   - Model selection(Likelihood based) between RDU_1 and QDT_{RDU1} (let us call M_R the selected model)
   - Model selection(Likelihood based) between CPT and QDT_{CPT} (let us call M_C the selected model)
2. Non-nested hypothesis testing for the selection between M_R and M_C

As for the first step, at the individual level, the two formulations include the same number of parameters (parameters of the q-factor, a and η, are fixed). Thus, the model selection can be done according to the log-likelihoods: the preferred model is the one which has the biggest log-likelihood. Tables 4.7 and 4.8 show the results.

	RDU_1	QDT_{RDU1}
Proportion of subjects for which the model is selected (time 1)	30%	70 %
Proportion of subjects for which the model is selected (time 2)	45%	55 %
Mean of the loglikelihood (time 1):	-91.50	-90.74
Mean of the loglikelihood (time 2):	-97.48	-97.22

Table 4.7: Model selection according to the likelihood. The likelihood is bigger with QDT_{RDU1} for most subjects and average, so the QDT_{RDU1} model is the preferred one.

	CPT	QDT_{CPT}
Proportion of subjects for which the model is selected (time 1)	34%	65 %
Proportion of subjects for which the model is selected (time 2)	36%	64 %
Mean of the loglikelihood (time 1):	-86.53	-85.77
Mean of the loglikelihood (time 2):	-99.18	-98.34

Table 4.8: Model selection according to the likelihood. The likelihood is bigger with QDT_{CPT} for most subjects and average, so the QDT_{CPT} model is the preferred one.

As already explained in subsection 3.3.4, for non-nested hypothesis testing we use two different tests: the Vuong Test and the Clarke Test. The results of our analysis are shown in figures 4.16 and 4.17.
The main conclusions are the following:

1. For a big fraction of subjects (in Vuong test for 70-80% of subjects) we can not clearly distinguish between CPT_{type} and RDU_{type} models (the "draw" category)

2. Less conservative Clarke test allows to select a better suited decision model for a majority of subjects (the "draw" category is reduced to 40-45%); the tendency is in favor of RDU_{type} models (both RDU_1 and QDT_{RDU1})

3. Quantum factor plays an important role: models with quantum factor are selected for the majority of subjects, as shown in figure 4.17.

Even if we wanted to look only at Clarke Test results, we would however note that a relevant fraction of the population is equally described by RDU_{type} and CPT_{type} models. This result may have several reasons:

- Small sample size (weak test).
- No model is correctly specified, the true model is a different one.
- Some choices are better "caught" by one decision making mechanism, while others by another one, and so on.

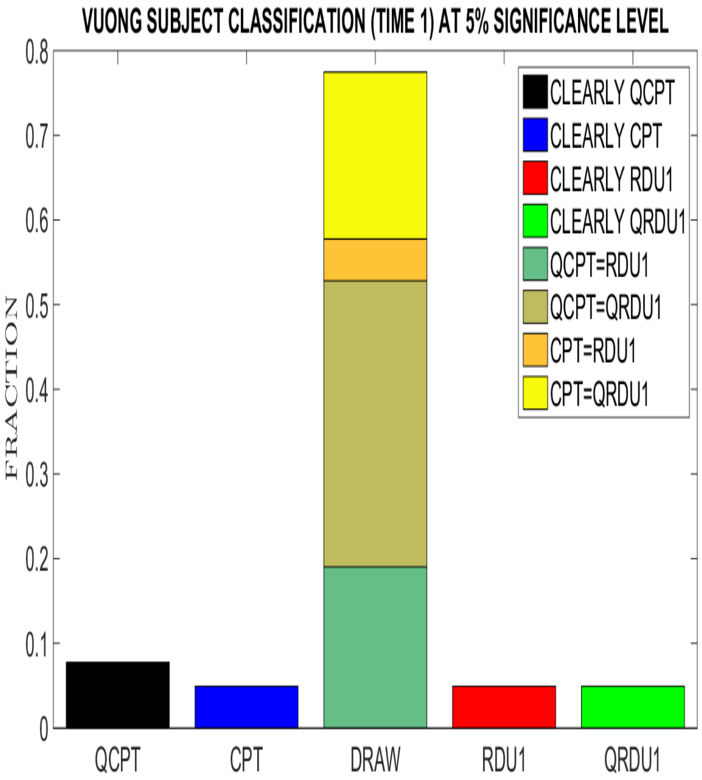
The last point is the one we want to investigate further, i.e if a subsample of the population can be significantly better described by a mixture of these models (final chapter 5).

Another interesting aspect to analyze is the stability of subjects' classification over time. In other words, we investigate whether the selected choice model that is best suited for an individual remains unchanged between times 1 and 2. Referring to the Clarke results, we record the fraction of subjects whose classification shifted from one class to another between time 1 and time 2. To make things simple, we exclude the "DRAW" category, taking into account only the "clear" shifts. The results are reported in the (symmetric) table 4.9.

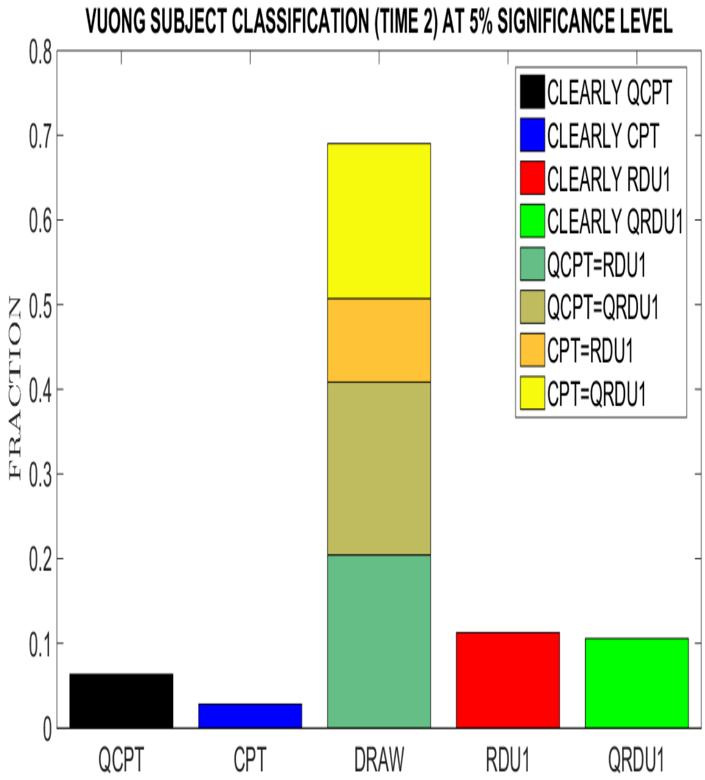
Our analysis of the experimental dataset shows that none of the decision makers that are clearly classified as either RDU_{type} or CPT_{type} have switched to the other "type" of decision making mechanism. In other words, for all clearly classified subjects the best suited model remained within the same type: CPT or RDU based.
Broadly speaking, it indicates that these subjects have a unique behavior with respect to asset integration: either they asset integrate both at time 1 and at time 2 (RDU_{type}), or they don't (CPT_{type}).

	RDU_1	QDT_{RDU1}	CPT	QDT_{CPT}
RDU_1	X	19%	0	0
QDT_{RDU1}	19%	X	0	0
CPT	0	0	X	4%
QDT_{CPT}	0	0	4%	X

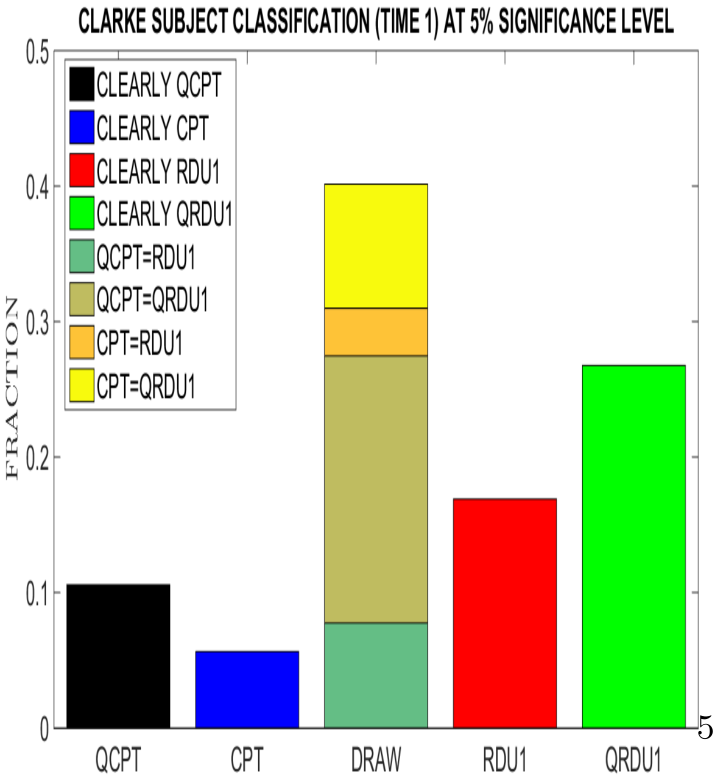
Table 4.9: Subject classification switching. Each entry contains the percentage of subjects who "clearly" shifted from one category to another. We do not take into account the direction of the shift, so the matrix is symmetric. The main result is that there is no clear switch at all between RDU_{type} and CPT_{type}, and vice-versa.



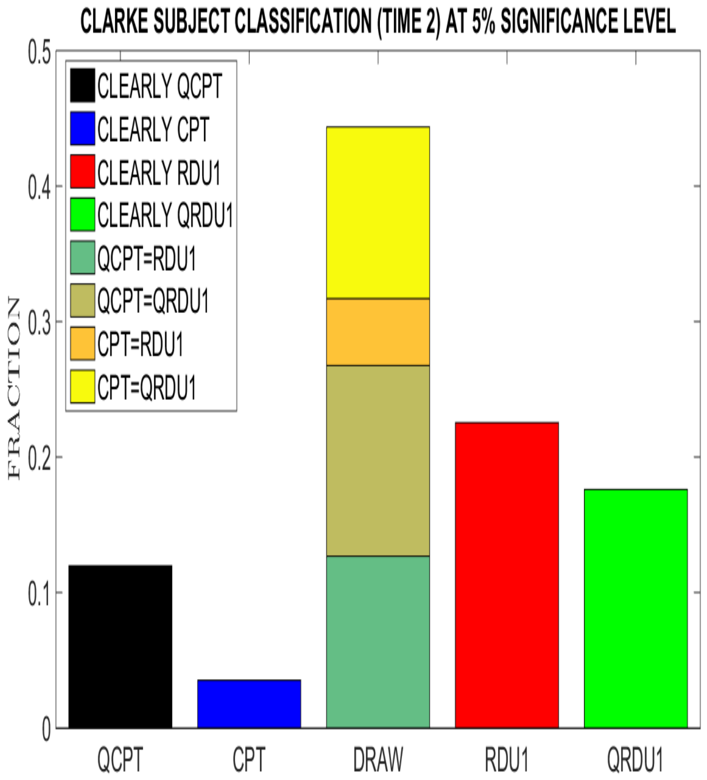
(a) Vuong Test (Time 1)



(b) Vuong Test (Time 2)

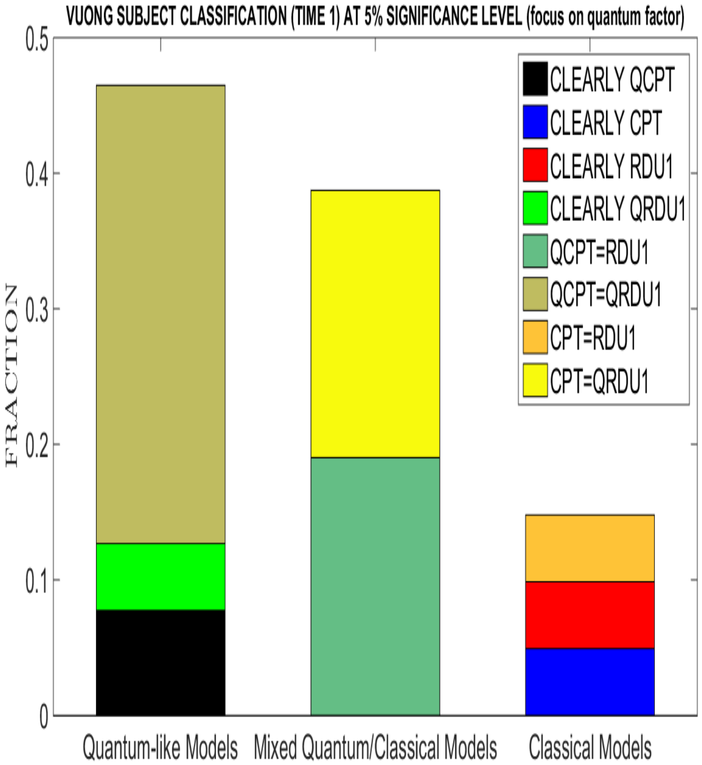


(c) Clarke Test (Time 1)

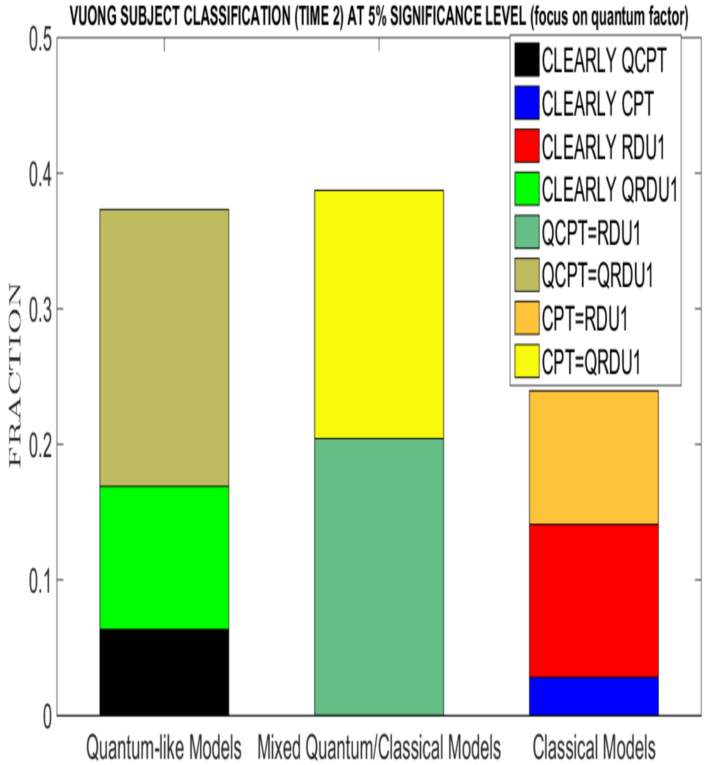


(d) Clarke Test (Time 2)

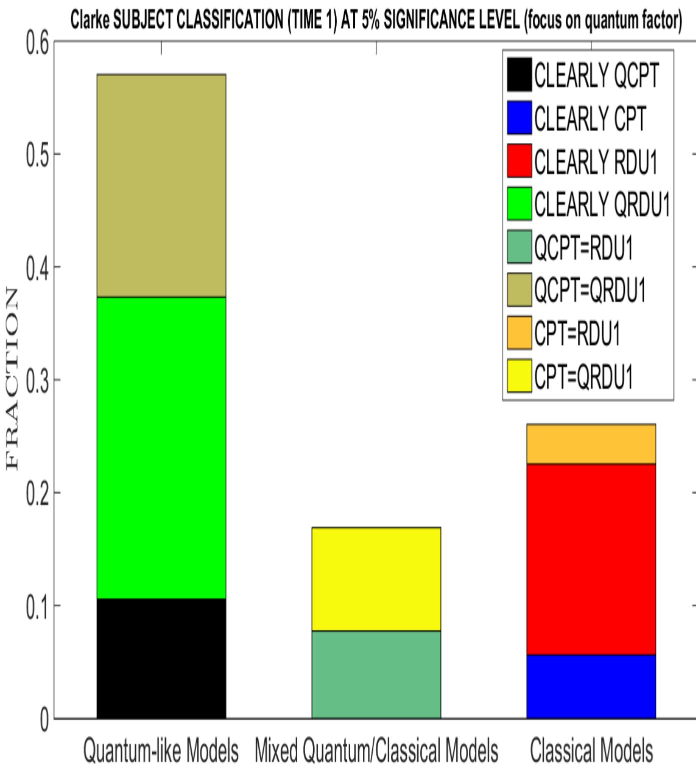
Figure 4.16: Subject classification. The Draw bar indicates no discrimination between the "winner" of RDU_{type} models and the "winner" of CPT_{type} models. For example, CPT=RDU1 indicates no discrimination between CPT(which won against QDT_{CPT}) and RDU1(which won against QDT_{RDU1}). QDT always describes a significant fraction of subjects. However, a large fraction of subjects falls in the "Draw" bar.



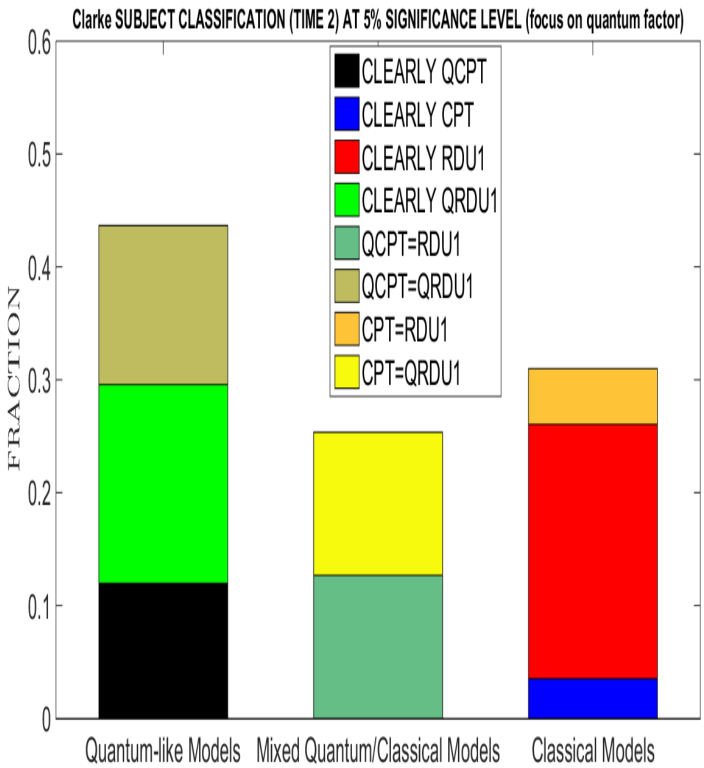
(a) Vuong Test (Time 1)



(b) Vuong Test (Time 2)



(c) Clarke Test (Time 1)



(d) Clarke Test (Time 2)

Figure 4.17: Subject classification: focus on the relevance of the quantum factor. By looking from this point of view, we see that the quantum models describes the majority of subjects for both tests and both times.

Chapter 5

Conclusions and possible extensions

In this paper, we have presented another parametrization of Quantum Decision Theory based on Rank-Dependent Utility Theory. We then compared RDU, RDU-based QDT, CPT and CPT-based QDT. CPT had already been studied on the same experiment by R.O Murphy and R. ten Brincke [19], while CPT-based QDT was proposed by S.Vincent et al. in [16].
RDU and RDU-based QDT do not need to rely on a definition of reference point, a very important feature in out-of-laboratory applications.

At the aggregate level, assuming homogeneous preferences across subjects (representative agent approach), we found that QDT_{CPT} outperformed the others, according to MSE measure and hypothesis testing. Within RDU_{type} models, quantum factor has been shown fundamental to improve performances. It could be argued that adding two parameters obviously gives a better model, but nested hypothesis testing confirmed the authentic significance of such improvements.

At the individual level, following [19], we chose an estimation method "halfway" between homogeneous preferences and treating each individual as unique, resulting in a *regularization* of the likelihood function to avoid overfitting. In terms of mean explained/predicted fraction of choices, similar results were reached by the four theories analyzed (74-76%), confirming an intrinsic limit of predictability due to randomness of choice, as suggested in [16].

Next, we proceeded to a subject classification, identifying each decision-maker as RDU, CPT, QDT_{RDU} or QDT_{CPT}. A first "naive" classification, based on the percentage of gambles explained/predicted, was used; the results clearly show the heterogeneity of behavior in the sample. Since this method neglected *how much better* is a theory with respect to another, more sophisticate criteria have been adopted, such as hypothesis testing. This analysis confirmed the existence of substiantial heterogeneity across the sample, weakening the consistence of the representative agent approach. The majority of subjects was best described by quantum models, showing again the key-role of QDT in characterizing decision-makers attitudes.
Moreover, this analysis showed that, both with Vuong Test and with Clarke Test, a relevant fraction of subjects was equally described by RDU_{type} and CPT_{type} models, resulting in an unclear classification. Such result may suggest the possibility of heterogeneity of choices within the subject himself. For istance, decision maker i can asset integrate for some choices (RDU_{type}) while for other gambles she can evaluate utilities neglecting initial wealth (CPT_{type}). Following this argument, a natural extension of the present work would be to characterize each subject by mixture models, so as to capture different psychological decision-making mechanism within the subject himself.
Finally, we analyzed the *stability* of the classification from time 1 to time 2. Our results show that none of the decision makers (clearly classified as either RDU_{type} or CPT_{type}) have switched to the other "type" of decision making mechanism, suggesting that the difference between decision makers make difficult to have a "universal" theory.

Bibliography

[1] D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 1738.

[2] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 1979.

[3] John Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 1982.

[4] D. Kahneman and A. Tversky. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 1992.

[5] F. Mosteller and P. Nogee. An experimental measurement of utility. *Journal of Political Economy*, 1951.

[6] Luce and R. Duncan. Conditional logit analysis of qualitative choice behavior. *New York: John Wiley and Sons.*, 1959.

[7] J.D. Hey and C.Orme. Investigating generelizations of the expected utility theory using experimental data. *Econometrica*, 1994.

[8] E. Carbone and J.D. Hey. A comparison of the estimates of eu and non-eu preferences functionals using data from pairwise choice and complete ranking experiments. *Geneva Papers on Risk Insurance Theory*, 1995.

[9] M.H. Birnbaum and A.Chavez. Test of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior of Human Decision Processes*, 1997.

[10] Ennio Cascetta. Transportation systems analysis: Models and applications. *Mathematics Subject Classification*, 2000.

[11] V.I. Yukalov and D. Sornette. Quantum decision theory as quantum theory of measurement. *Physics Letters*, 2008.

[12] L. J. Savage. The foundations of statistics. *John Wiley and Sons Inc., New York.*, 1954.

[13] A. Tversky and D. Kahneman. Judgments of and by representativeness. *Cambridge, UK*, 1982.

[14] Daniel Ellsberg. Risk, ambiguity, and the savage axioms. *Quarterly*

Journal of Economics, 75, 1961.

[15] V.I. Yukalov and D. Sornette. Preference reversal in quantum decision theory. *Frontiers in Psychology*, 2015.

[16] V.I. Yukalov S. Vincent, T. Kovalenko and D. Sornette. Calibration of quantum decision theory, aversion to large losses and predictability of probabilistic choices. *Swiss Finance Institute Research Paper*, 2016.

[17] Nicholas C. Barberis. Thirty years of prospect theory in economics: A review and assessment. *JOURNAL OF ECONOMIC PERSPECTIVES*, 27:173–196, 2013.

[18] R.O. Murphy M. Schulte-Mecklenbeck, T. Pachur and R. Hertwig. Does prospect theory capture psychological processes. *ETH Zurich, Chair of Decision Theory and Behavioral Game Theory and the Max Planck Institute for Human Development*, 2013.

[19] R.O. Murphy and R.H.W. ten Brincke. Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates. *ETH Zurich working paper*, 2014.

[20] M. Allais. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21, 1953.

[21] K. J. Arrow. Aspects of the theory of risk bearing. *The Theory of Risk Aversion. Helsinki: Yrjo Jahnssonin Saatio.*, 1965.

[22] Carl Simon and Lawrence Blume. Mathematics for economists (student ed.). *Viva Norton. p. 363.*

[23] ENRICO DIECIDUE and PETER P. WAKKER. On the intuition of rank-dependent utility. *The Journal of Risk and Uncertainty*, 23, 2001.

[24] D. Prelec. The probability weighting function. *Econometrica*, 1998.

[25] Danyang Xie. Power risk aversion utility functions. *Annals of Economics and Finance*, 1:265–282, 2000.

[26] D. R. Cox. Simple regression. *Analysis of Binary Data*, 1970.

[27] C. I. Bliss. The method of probits. *Science*, 79, 1934.

[28] DR Cox. The regression analysis of binary sequences (with discussion). *J Roy Stat Soc B*, 20:215, 1958.

[29] V.I. Yukalov and D. Sornette. Mathematical structure of quantum decision theory. *Advances in Complex Systems*, 2010.

[30] V.I. Yukalov and D. Sornette. Manipulating decision making of typical agents. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 44:1155, 2014.

[31] D. R. Burnham, K. P.; Anderson. Model selection and multimodel inference (2nd ed.). *Springer-Verlag.*, 2002.
[32] Sara Bühlmann, Peter; Van De Geer. Statistics for high-dimensional data. *Springer Series in Statistics*, 9, 2011.
[33] J. Rieskamp H. Nilsson and E. J. Wagenmakers. Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 2011.
[34] S. Farrell and C: J. H Ludwig. Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bullettin and Review*, 2008.
[35] Sebastiano Pilati. Random numbers: Non-uniform distributions. *Lecture Notes*, 2017.
[36] C. Gourieroux and A. Monfort. Testing non-nested hypotheses. *Elseuier Science B.V.*, 4, 1994.
[37] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 1974.
[38] G. SCHWARZ. Estimating the dimension of a model. *Annals of Statistics*, 6:461, 1978.
[39] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9:60, 1938.
[40] D.R. Cox. Test of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:105, 1961.
[41] A.C. Atkinson. A test for discriminating between models. *Biometrika*, 56:337, 1969.
[42] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 1982.
[43] Quang H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *The Econometric Society*, 1989.
[44] Glenn W. Harrison and J. Todd Swarthout. Cumulative prospect theory in the laboratory: A reconsideration. 2016.
[45] Kevin A. Clarke. Nonparametric model discrimination in international relations. *The Journal of Conflict Resolution*, 2003.
[46] Kevin A. Clarke. A simple distribution-free test for nonnested model selection. *Political Analysis*, 2007.
[47] Chernoff H. On the distribution of the likelihood ratio. *Ann. Math. Statist.*, 25:573, 1954.

Appendices

.1 Details of Statistical Tests used

.1.1 Vuong Test

In his paper [43], Vuong developed a simple likelihood-ratio test using the Kullback Leibler Information Criterion (KLIC). The null hypothesis is that we cannot discriminate between the competing models, while the alternative hypothesis is that one model is closer to the true data generating process.

If $\mathbf{F}_\theta = \{f(y|z;\vec{\theta}), \vec{\theta} \in \vec{\Theta} \subset R^p\}$ is a conditional model, its distance from the true distribution $h^0(y|z)$, as measured by the minimum KLIC, is

$$E^0[logh^0(y|z)] - E^0[logf(y|z;\vec{\theta}^*)] \tag{1}$$

where $E^0[]$ indicates the expectation value taken with respect to the true joint distribution (y,z) (we recall that in our case the z are constant variables representing the lotteries) and $\vec{\theta}^*$ is the pseudo-true value of $\vec{\theta}$ (the estimate of $\vec{\theta}$ when $f(y|z)$ is not the true model).
The smaller the KLIC in equation 1, the closer is the model to the truth. Therefore, since the first term of 1 is the same for each competing model, we can choose the "best" model on the base of $E^0[logf(y|z;\vec{\theta}^*)]$; the model with the largest one is the best one, because it minimizes the KLIC. Given two conditional models $\mathbf{F}_\theta$ and $\mathbf{G}_\gamma = \{g(y|z;\vec{\gamma}), \vec{\gamma} \in \vec{\Gamma} \subset R^q\}$, which may be nested, non-nested, or overlapping, consider the following hypotheses:

$$H_0 : E^0\left[log\frac{f(y|z;\vec{\theta}^*)}{g(y|z;\vec{\gamma}^*)}\right] = 0 \quad \mathbf{F}_\theta \sim \mathbf{G}_\gamma \tag{2}$$

$$H_f : E^0\left[log\frac{f(y|z;\vec{\theta}^*)}{g(y|z;\vec{\gamma}^*)}\right] > 0 \quad \mathbf{F}_\theta \succeq \mathbf{G}_\gamma \tag{3}$$

$$H_g : E^0\left[log\frac{f(y|z;\vec{\theta}^*)}{g(y|z;\vec{\gamma}^*)}\right] < 0 \quad \mathbf{F}_\theta \preceq \mathbf{G}_\gamma \tag{4}$$

H_0 does not require that either of the competing models have to be correctly specified(i.e neither has to "contain" the true density). It is important to note it because we will not assume that any model is correctly specified, since our goal will be to stress the need for a mixture model.

The quantity of interest, $E^0[logf(y|z;\vec{\theta}^*)] - E^0[logg(y|z;\vec{\gamma}^*)]$, is unknown, but, under weak general conditions, can be consistently estimated by

$$\frac{1}{n}\sum_{t=1}^{n} log\left[\frac{f(Y_t|Z_t;\hat{\vec{\theta}}_n)}{g(Y_t|Z_t;\hat{\vec{\gamma}}_n)}\right] = \frac{1}{n}LR_n(\hat{\vec{\theta}}_n,\hat{\vec{\gamma}}_n) \tag{5}$$

where n is the number of observations while $\hat{\vec{\theta}}_n$ and $\hat{\vec{\gamma}}_n$ are the ML estimators of $\vec{\theta}^*$ and $\vec{\gamma}^*$, respectively.
The LR statistic is asymptotically distributed as a normal or a weighted sum of chi-squares, depending on whether $f(-|-;\vec{\theta}^*) = g(-|-;\vec{\gamma}^*)$ or not.
More formally, under general assumptions we have:

- if $f(-|-;\vec{\theta}^*) = g(-|-;\vec{\gamma}^*)$, then

$$2LR_n(\hat{\vec{\theta}}_n,\hat{\vec{\gamma}}_n) \xrightarrow{D} M_{p+q}(-;\vec{\lambda}^*) \tag{6}$$

  where $M_m(-;\vec{\lambda})$ is the cumulative distribution function (CDF) of

$$\sum_{i=1}^{m}\lambda_i Z_i^2 \quad \lambda_i \in R \quad Z_i \sim N(0,1) \tag{7}$$

  and $\vec{\lambda}^*$ is the vector of p+q (possibly negative) eigenvalues of

$$W = \begin{bmatrix} -B_f(\vec{\theta}^*)A_f^{-1}(\vec{\theta}^*) & -B_{fg}(\vec{\theta}^*,\vec{\gamma}^*)A_g^{-1}(\vec{\gamma}^*) \\ B_{gf}(\vec{\gamma}^*,\vec{\theta}^*)A_f^{-1}(\vec{\theta}^*) & B_g(\vec{\gamma}^*)A_g^{-1}(\vec{\gamma}^*) \end{bmatrix}$$

$$A_f(\vec{\theta}) = E^0\left[\frac{\partial^2 logf(Y_t|Z_t;\vec{\theta})}{\partial\theta_i\partial\theta_j}\right], B_f(\vec{\theta}) = E^0\left[\frac{\partial logf(Y_t|Z_t;\vec{\theta})}{\partial\theta_i}\frac{\partial logf(Y_t|Z_t;\vec{\theta})}{\partial\theta_j}\right] \quad i,j = 1...p \tag{8}$$

  ,

$$B_{fg}(\vec{\theta},\vec{\gamma}) = B'_{gf}(\vec{\gamma},\vec{\theta}) = E^0\left[\frac{\partial logf(Y_t|Z_t;\vec{\theta})}{\partial\theta_l}\frac{\partial logg(Y_t|Z_t;\vec{\gamma})}{\partial\gamma_m}\right] \quad l = 1..p, m = 1..q \tag{9}$$

- if $f(-|-;\vec{\theta}^*) \neq g(-|-;\vec{\gamma}^*)$, then

$$\frac{LR_n(\hat{\vec{\theta}}_n,\hat{\vec{\gamma}}_n)}{\sqrt{n}} - \frac{E^0\left[log\frac{f(Y_t|Z_t;\vec{\theta}^*)}{g(Y_t|Z_t;\vec{\gamma}^*)}\right]}{\sqrt{n}} \xrightarrow{D} N(0,w_*^2) \tag{10}$$

In the most general case of possibly overlapping models (our case) we cannot tell it a-priori. Since these quantities are unknown we need another test for this condition.
It can be proven that:

$$f(-|-;\vec{\theta}^{*}) = g(-|-;\vec{\gamma}^{*}) \iff w_{*}^{2} = 0 \tag{11}$$

$$w_{*}^{2} = Var^{0}\left(log\left[\frac{f(y|z;\vec{\theta}^{*})}{g(y|z;\vec{\gamma}^{*})}\right]\right) \tag{12}$$

Thus, we can equivalently test that the variance w_{*}^{2} is equal to 0.

$$H_{0}^{w} : w_{*}^{2} = 0 \quad vs \quad H_{A}^{w} : w_{*}^{2} \neq 0 \tag{13}$$

A natural statistic is:

$$\hat{w}_{n}^{2} = \frac{1}{n}\sum_{t=1}^{n}\left[log\frac{f(Y_t|Z_t;\hat{\vec{\theta}}_n)}{g(Y_t|Z_t;\hat{\vec{\gamma}}_n)}\right]^{2} - \left[\frac{1}{n}\sum_{t=1}^{n}log\frac{f(Y_t|Z_t;\hat{\vec{\theta}}_n)}{g(Y_t|Z_t;\hat{\vec{\gamma}}_n)}\right]^{2} \tag{14}$$

Actually, if $w_{*}^{2} = 0$ we can already conclude that $\mathbf{F}_{\theta} \sim \mathbf{G}_{\gamma}$, without carrying out the LR test.
If $w_{*}^{2} \neq 0$ (and thus $f(-|-;\vec{\theta}^{*}) \neq g(-|-;\vec{\gamma}^{*})$), we may still have $E^{0}[logf(y|z;\vec{\theta}^{*})] = E^{0}[logg(y|z;\vec{\gamma}^{*})]$ so that a LR test of H_0 against H_f or H_g must still be carried out.

The variance test, under very general assumptions, is the following:

- under H_{0}^{w}, $\forall x \geq 0$, $Pr(n\hat{w}_{n}^{2} \leq x) - M_{p+q}(x;\hat{\vec{\lambda}}_{n}^{2}) \xrightarrow{a.s.} 0$,with $\hat{\vec{\lambda}}_{n}^{2}$ being the vector of the squares of $\hat{\vec{\lambda}}_{n}$ (eigenvalues of W_n, sample analog of W)
- under $H_{A}^{w}, n\hat{w}_{n}^{2} \xrightarrow{a.s.} +\infty$

Essentially, the variance test consists in choosing a x so that $M_{p+q}(x;\hat{\vec{\lambda}}_{n}^{2}) = 1-\alpha$ for some significance level α, and in rejecting H_{0}^{w} if $n\hat{w}_{n}^{2} > x$.

To sum up, our non-nested hypothesis test is a sequential procedure:

1. Test H_{0}^{w} vs H_{A}^{w} using the variance test based on $n\hat{w}_{n}^{2}$
   - if H_{0}^{w} is not rejected, conclude that "$F_{\theta} = G_{\gamma}$"

- if H_0^w is rejected, go to 2.

2. test H_0 against H_f or H_g using the normal model selection test.

This sequential procedure has a significance level which is asymptotically bounded above by the maximum between:

- α_1: asymptotic significance level for the variance test

- α_2: asymptotic significance level for the normal LR test

In order to correctly implement the test, we have to take into account that competing models can have different number of parameters. Indeed, this difference obviously affects the log-likelihoods. Vuong suggested, among others, to use Schwarz's (1978, [38]) information criteria, making the adjusted statistic:

$$L\tilde{R}_n(\hat{\vec{\theta}}_n, \hat{\vec{\gamma}}_n) = LR_n(\hat{\vec{\theta}}_n, \hat{\vec{\gamma}}_n) - K_n(F_\theta, G_\gamma) = LR_n(\hat{\vec{\theta}}_n, \hat{\vec{\gamma}}_n) - \left[\frac{p}{2}log(n) - \frac{q}{2}log(n)\right] \tag{15}$$

Vuong justifies the correction by noting that as long as the correction factor divided by the square root of n obeys:

$$\frac{K_n(F_\theta, G_\gamma)}{\sqrt{n}} = o_p(1) \tag{16}$$

the adjusted statistic has the same asymptotic properties of the unadjusted one.

.1.2 Clarke Test

The goal of the Vuong test is to assess if the *average* log-likelihood ratio is significantly different from zero, favouring one of the competing models. On the other hand, the Clarke test, first introduced by Clarke in [45], aims to determine whether the *median* log-likelihood ratio is statistically different from zero. Informally, this test counts the number of log-likelihood ratios above and below 0; if half of them is greater than 0 and half less than 0, we cannot discriminate between the models. If model f is "better" than model g, the number of positive ratios should be significantly higher than half (and

vice-versa).
The null hypothesis is therefore:

$$H_0 : Pr\left(log\left[\frac{f(Y_t|Z_t;\vec{\theta}^*)}{g(Y_t|Z_t;\vec{\gamma}^*)}\right] > 0\right) = Pr\left(log\left[\frac{f(Y_t|Z_t;\vec{\theta}^*)}{g(Y_t|Z_t;\vec{\gamma}^*)}\right] < 0\right) = \frac{1}{2} \iff M = 0 \tag{17}$$

where M is the median log-likelihood ratio.

Letting $X_t = log\Big[f(Y_t|Z_t;\hat{\vec{\theta}}_n)\Big] - log\Big[g(Y_t|Z_t;\hat{\vec{\gamma}}_n)\Big]$, and

$$\Psi_t = \begin{cases} 1, & \text{if } X_t > 0 \\ 0, & \text{if} X_t < 0 \end{cases} \tag{18}$$

The test statistic is

$$B = \sum_{t=1}^{n} \Psi_t \tag{19}$$

B is the number of positive differences, and it is distributed Binomial with parameters n and $p = \frac{1}{2}$.
If model F_θ is "better" than model G_γ, B will be significantly larger than its expected value under the null hypothesis $\frac{n}{2}$. We reject the null hypothesis if $B \geq c_\alpha$ OR $B \leq d_\alpha$, where c_α and d_α are the smallest integer such that:

$$\sum_{c=c_\alpha}^{n} \binom{n}{c} \frac{1}{2}^n \leq \alpha \tag{20}$$

$$\sum_{c=0}^{d_\alpha} \binom{n}{c} \frac{1}{2}^n \geq \alpha \tag{21}$$

This test, like the Vuong test, is sensitive to the difference in number of parameters of the competing models. Unfortunately, we cannot follow the Vuong's suggestion, because here we are not dealing with the average log-likelihood ratio. However, Clarke suggested to apply the *average* correction to each X_t. Therefore, we shift the individual log-likelihoods for model F_θ by a factor of $\frac{p}{2n}log(n)$ and the individual log-likelihoods for model G_γ by a factor of $\frac{q}{2n}log(n)$.