Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 414

Risk and Rationality: Uncovering Heterogeneity in Probability Distortion

Adrian Bruhin, Helga Fehr-Duda and Thomas Epper

May 2009

Risk and Rationality: Uncovering Heterogeneity in Probability Distortion

Adrian Bruhin Helga Fehr-Duda Thomas Epper

May 25, 2009

Abstract

It has long been recognized that there is considerable heterogeneity in individual risk taking behavior but little is known about the distribution of risk taking types. We present a parsimonious characterization of risk taking behavior by estimating a finite mixture regression model for three different experimental data sets, two Swiss and one Chinese, over a large number of real gains and losses. We find two distinct types of individuals: In all three data sets, the choices of roughly 80% of the subjects exhibit significant deviations from linear probability weighting, consistent with prospect theory. 20% of the subjects weight probabilities near linearly and behave essentially as expected value maximizers. Moreover, individuals are cleanly assigned to one type with probabilities close to unity. The reliability and robustness of our classification suggest using a mix of preference theories in applied economic modeling.

KEYWORDS: Individual Risk Taking Behavior, Latent Heterogeneity, Finite Mixture Regression Models, Prospect Theory
JEL CLASSIFICATION: D81, C49

1 Introduction

Risk is a ubiquitous feature of social and economic life. Many of our everyday choices, and often the most important ones, such as what trade to learn and where to live, involve risky consequences. While it has long been recognized that individuals differ in their risk taking attitudes, comparatively little is known about the distribution of risk preferences in the population.[1] Since preferences are one of the ultimate drivers of behavior, knowledge of the composition of risk attitudes is paramount to predicting economic behavior. Economic models often allow for heterogeneity, but this heterogeneity is usually defined by the boundaries of the standard model of preferences, expected utility theory (EUT). The empirical evidence, however, reveals that heterogeneity in risk taking behavior is of a substantive kind, i.e. some people evaluate risky prospects consistently with EUT, whereas other people depart substantially from expected utility maximization (Hey and Orme, 1994). Moreover, it seems to be the case that rational decision makers revealing EUT-preferences constitute only a minority of the population (Lattimore, Baker, and Witte, 1992).

To improve descriptive performance a plethora of alternative theories have been developed. Unfortunately, no single best fitting model has been identified so far (Harless and Camerer, 1994; Starmer, 2000) and, depending on the individual, one or the other model fits better. This finding poses a serious problem for applied economics. What the modeler needs is a *parsimonious* representation of risk preferences that is empirically well grounded and robust, and not a host of different functionals. Providing such a parsimonious characterization of heterogeneity in risk taking behavior is the objective of this paper.

Our method is based on a literature on classifying individuals which has recently emerged in the social sciences. On the basis of statistical classification

[1] Exceptions include Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2005); Eckel, Johnson, and Montmarquette (2005); Harrison, Lau, and Rutström (2007).

procedures, such as finite mixture regression models, investigators have tried to discover which decision rules people actually apply when playing games or dealing with complex decision situations (El-Gamal and Grether, 1995; Stahl and Wilson, 1995; Houser, Keane, and McCabe, 2004; Houser and Winter, 2004). The finite mixture regression approach does not require fitting a model for each individual, which is - given the usual quality of choice data - frequently impossible. Instead, our method reveals latent heterogeneity by estimating the proportions of distinct behavioral types in the population and assigning each individual to one endogenously defined behavioral type, characterized by a unique set of parameter values.

We apply such a finite mixture regression model to choice data from three different experiments, two of which were conducted in Zurich, Switzerland. The third experiment took place in Beijing, People's Republic of China. We analyze 448 subjects' decisions over real monetary gains and losses, which comprise a total of nearly 18,000 observations. All three experiments were designed in a similar manner and served to elicit certainty equivalents for binary lotteries. Using a flexible sign-dependent functional as basic behavioral model, we show the following main results for two- and three-component mixture models.

First, the estimation procedure renders a robust classification of risk taking behavior across all three data sets. Moreover, the proportions of these distinct types in their respective populations are very similar.

Second, almost all the experimental subjects are unambiguously assigned to one distinct type. Measuring the quality of classification by the *Normalized Entropy Criterion* (Celeux and Soromenho, 1996), ambiguity of assignments turns out to be extremely low. Thus, we observe hardly any "ambiguous" types, i.e. individuals with a high probability (of say 0.4) of being one type *and* a high probability (of say 0.6) of being another type. This clean segregation suggests that the classification procedure is able to capture the distinctive characteristics of each behavioral type.

Third, without restricting parameter values *a priori*, we find that, in all three data sets the minority types, who constitute about 20% of the population, weight probabilities and value monetary outcomes near linearly. Consequently, this group of individuals can essentially be characterized as expected value maximizers. This result is particularly interesting in the light of Rabin's calibration theorem (Rabin, 2000), which shows that expected utility maximizers should be approximately risk neutral for small stakes typically encountered in laboratory experiments if behavior under high stakes is to remain within a plausible range of risk aversion. Therefore, we label subjects belonging to this group of nearly risk neutral people as "EUT types". Moreover, the EUT group remains totally robust to increasing the number of types in the mixture.

Fourth, the majority of individuals, labeled as "CPT types", are characterized by significant departures from linear probability weighting, consistent with prospect theory. As three-group classifications show, this group's behavior can be characterized as a mixture of two different types: In all three data sets a proportion of approximately 30% of the subjects display pronounced departures from linear probability weighting, whereas the relative majority of 50% differ less radically from linear probability weighting.

Finally, within the class of CPT types, we find major differences between Swiss and Chinese behavior. Sensitivity to changes in probabilities is generally lower for the Chinese subjects than for the Swiss. While in both countries the majority CPT groups' probability weighting curves do not differ dramatically, the minority groups display diametrically opposed patterns of probability weighting. In particular, the minority Chinese CPT group weight probabilities extremely favorably rendering them risk seeking over a considerable range of probabilities. The minority Swiss CPT group, however, is characterized by the opposite behavior. Thus, our analysis provides a deeper understanding for the finding that, on average, the Chinese tend to be more risk seeking than Westerners (Kachelmeier and Shehata, 1992).

Our results show that the classification procedure successfully uncovers latent heterogeneity in the population. If there is heterogeneity of a substantive kind, as the data suggest, basing predictions on a single preference theory is inappropriate and may lead to biased results. EUT preferences should be taken account of alongside prospect theory preferences, even if rational EUT individuals constitute only a minority in the population. As the literature on the role of bounded rationality under strategic complementarity and substitutability has shown (Haltiwanger and Waldman, 1985, 1989; Fehr and Tyran, 2005; Camerer and Fehr, 2006), the mix of rational and irrational actors may be decisive for aggregate outcomes. Depending on the nature of the strategic interdependence the behavior of even a minority of players may drive the aggregate outcome. Therefore, the mix of types in the population is a crucial variable in predicting market outcomes. Since the finite mixture regressions provide a robust and reliable classification of individuals, the resulting estimates of group sizes and group-specific parameters may serve as valuable inputs for applied economics.

To the best of our knowledge, there is no previous study showing a nearly identical classification of risk preference types for three independent data sets. Related work by Harrison and colleagues (Harrison and Rutström, forthcoming; Harrison, Humphrey, and Verschoor, 2005; Andersen, Harrison, and Rutström, 2006) estimated risk taking behavior by finite mixture regressions, but decisively distinguishes itself from our analysis. Their estimation procedure is based on the *a priori* assumption that *choices*, irrespective by whom they were taken, may be EUT consistent or CPT consistent, i.e. it sorts choices by pre-defined decision model. In contrast, we aim at classifying *individuals* by endogenously defined type. Therefore, if there is a group of people whose behavior can best be described by EUT they should get identified by the classification procedure. Furthermore, in certain decision situations the behaviors of EUT individuals and CPT individuals do not differ substantially from one other. Consequently, the classification of EUT- and CPT-consistent decisions differs markedly from the percentages of decision makers classified as being one or the other type.

The paper is structured as follows. Section 2 describes the experimental design and procedures of the three experiments. The functional specification of the behavioral model and the finite mixture regression model are discussed in Section 3. Section 4 presents descriptive statistics of the data and the results of the classification procedure. Section 5 concludes.

2 Experimental Design

In the following section we describe the experimental setup and procedures. The experiments took place in Zurich in 2003 and 2006 as well as in Beijing in 2005. In Zurich, all subjects were recruited from the subject pool of the Institute for Empirical Research in Economics, which consists of students of all fields of the University of Zurich and the Swiss Federal Institute of Technology Zurich. In Beijing, subjects were recruited by flier distributed at the campuses of Peking University and Tsinhua University. Since all three experiments are based on the same design principles, we will present the prototype experiment Zurich 2003 in detail and describe to what extent the other two experiments deviate from the prototype. The main distinguishing features of the different experiments are summarized in Table 1.

We elicited certainty equivalents for a large number of two-outcome lotteries. One half of the lotteries were framed as choices between risky and certain gains ("gain domain"), the other half were presented as choices between risky and certain losses ("loss domain"). For each decision in the loss domain, subjects were endowed with a specific monetary amount, which served to cover potential losses and equalized expected payoffs of corresponding gain and loss lotteries. In the Zurich 2003 and the Beijing experiments, 50% of the subjects were confronted with decisions framed in the standard gamble format. The other 50% of the subjects had to make choices framed in contextual terms, i.e.

Table 1: **Differences in Experimental Design**

	Zurich 03	**Zurich 06**	**Beijing 05**
Number of:			
Subjects	179	118	151
Lotteries	50	40	28
Observations	8,906	4,669	4,225
Procedure	computerized	computerized	paper and pencil
Framing	abstract and contextual	contextual	abstract and contextual

gains were represented as risky or sure investment gains, losses as repair costs and insurance premiums, respectively. The Zurich 2006 experiment was based on contextually framed lotteries only. In Zurich, outcomes x_1 and x_2 ranged from zero Swiss Francs to 150 Swiss Francs[2]. The payoffs in the Beijing 2005 experiment were commensurate with the compensation in Zurich and varied between 4 and 55 Chinese Yuan[3]. Expected payoffs per subject amounted to approximately 31 Swiss Francs and 20 Chinese Yuan, respectively, which was considerably more than a local student assistant's hourly compensation, plus a show up fee of 10 Swiss Francs and 20 Chinese Yuan, thus generating salient incentives. Probabilities p of the lotteries' higher gain or loss x_1 varied from 5% to 95%. The gain lotteries for Zurich 2003 are presented in Table 2. The other two experiments essentially included a subset of these. The lotteries appeared in random order on a computer screen[4], in Beijing on paper.

[2]At the time of the experiments, one Swiss Franc equaled about 0.76 and 0.84 U.S. Dollars, respectively.

[3]At the time of the experiment, one Chinese Yuan equaled about 0.12 U.S. Dollars.

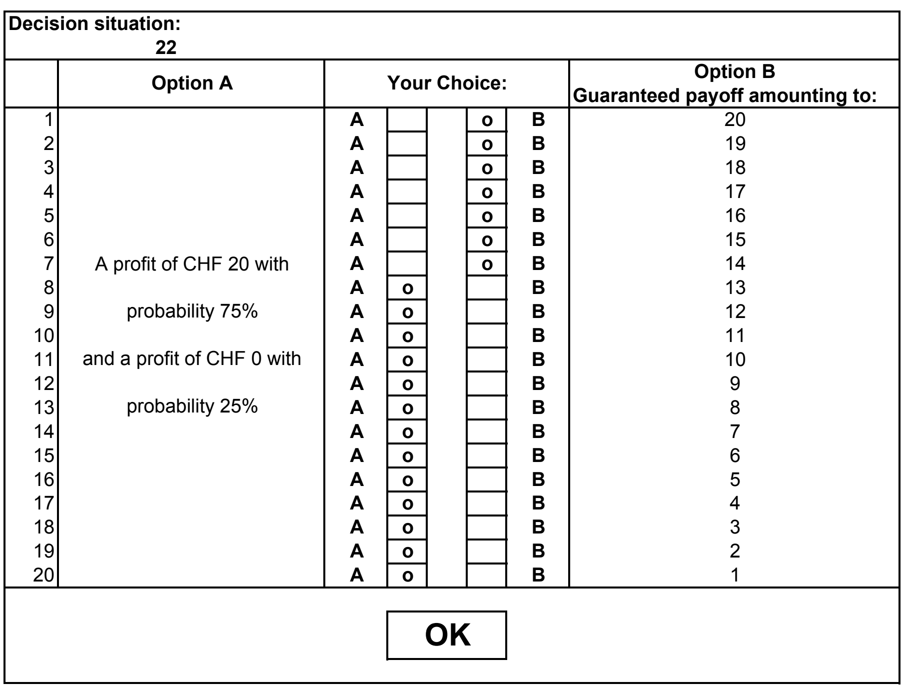
[4]The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

Table 2: **Gain Lotteries $(x_1, p; x_2)$, Zurich 2003**

p	x_1	x_2	p	x_1	x_2	p	x_1	x_2
0.05	20	0	0.25	50	20	0.75	50	20
0.05	40	10	0.50	10	0	0.90	10	0
0.05	50	20	0.50	20	10	0.90	20	10
0.05	150	50	0.50	40	10	0.90	50	0
0.10	10	0	0.50	50	0	0.95	20	0
0.10	20	10	0.50	50	20	0.95	40	10
0.10	50	0	0.50	150	0	0.95	50	20
0.25	20	0	0.75	20	0			
0.25	40	10	0.75	40	10			

Outcomes x_1 and x_2 are denominated in Swiss Francs (CHF).

Figure 1: **Design of the Decision Sheet**



Decision situation: **22**

	Option A	Your Choice:				Option B Guaranteed payoff amounting to:
1		A		o	B	20
2		A		o	B	19
3		A		o	B	18
4		A		o	B	17
5		A		o	B	16
6		A		o	B	15
7	A profit of CHF 20 with	A		o	B	14
8		A	o		B	13
9	probability 75%	A	o		B	12
10		A	o		B	11
11	and a profit of CHF 0 with	A	o		B	10
12		A	o		B	9
13	probability 25%	A	o		B	8
14		A	o		B	7
15		A	o		B	6
16		A	o		B	5
17		A	o		B	4
18		A	o		B	3
19		A	o		B	2
20		A	o		B	1

OK

In the computerized experiments, the screen displayed a decision sheet containing the specifics of the lottery under consideration and a list of 20 equally spaced certain outcomes, ranging from the lottery's maximum payoff to the lottery's minimum payoff, as shown in Figure 1.[5] The subjects had to indicate whether they preferred the lottery or the certain payoff for each row of the decision sheet. The lottery's certainty equivalent was calculated as the arithmetic mean of the smallest certain amount the subject preferred to the lottery and the subsequent certain amount on the list, when the subject had, for the first time, reported preference for the lottery. For example, if the subject had decided as indicated by the small circles in Figure 1, her certainty equivalent would amount to 13.5 Swiss Francs.

Before subjects were permitted to start working on the real decisions, they had to correctly calculate the payoffs for two hypothetical choices. In the computerized experiments, there were two trial rounds to familiarize the subjects with the procedure. At the end of the experiment, one row number of one decision sheet was randomly selected for each subject, and the subject's choice in that row determined her payment. Subjects were paid in private afterward. The subjects could work at their own speed, the vast majority of them needed less than an hour to complete the experimental tasks as well as a socio-economic questionnaire.

3 Econometric Model

This section discusses the specification of the finite mixture regression model, which allows controlling for latent heterogeneity in risk taking behavior in a parsimonious way. For the purpose of classifying subjects according to risk taking type, we need to specify three ingredients of the mixture model: the

[5]The format of the decision sheet for the Beijing experiment was identical to the Zurich one.

basic theory of decision under risk, the functional form of the decision model, and the specification of the error term.

The underlying theory of decision under risk should be able to accommodate a wide range of different behaviors. Sign- and rank-dependent models, such as cumulative prospect theory (CPT), capture reference dependence and nonlinear probability weighting. Therefore, a flexible approach, such as proposed by CPT, lends itself to describing risk taking behavior. Moreover, CPT nests EUT as special case. If there is a group of people, whose behavior can best be described by EUT, these individuals should be identified by the finite mixture regression as a unique group exhibiting the predicted behavior.

Suppose that there are C different types of individuals in the population. According to CPT, an individual belonging to a certain group $c \in \{1, \ldots, C\}$ values any binary gamble $\mathcal{G}_g = (x_{1g}, p_g; x_{2g}),\ g \in \{1, \ldots, G\}$, where $|x_{1g}| > |x_{2g}|$, by

$$v\left(\mathcal{G}_g\right) = v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g)).$$

The function $v(x)$ describes how monetary outcomes x are valued, whereas the function $w(p)$ assigns a subjective weight to every outcome probability p. The gamble's certainty equivalent $\hat{ce}_g$ can then be written as

$$\hat{ce}_g = v^{-1}\left[v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g))\right].$$

In order to make CPT operational, we have to assume specific functional forms for the value function $v(x)$ and the probability weighting function $w(p)$. A natural candidate for $v(x)$ is a sign-dependent power functional

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -(-x)^{\beta} & \text{otherwise,} \end{cases}$$

which can be conveniently interpreted and has turned out to be the best compromise between parsimony and goodness of fit in the context of prospect theory

(Stott, 2006).[6]

A variety of functional forms for modeling probability weights $w(p)$ have been proposed in the literature (Quiggin, 1982; Tversky and Kahneman, 1992; Prelec, 1998). We use the two-parameter specification suggested by Goldstein and Einhorn (1987) and Lattimore, Baker, and Witte (1992):

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}, \;\delta \geq 0,\; \gamma \geq 0.$$

We favor this specification because it has proven to account well for individual heterogeneity (Wu, Zhang, and Gonzalez, 2004) and the parameters are nicely interpretable. The parameter γ largely governs the slope of the curve and measures sensitivity towards changes in probability. The smaller the value of γ, the more strongly the probability weighting function departs from linear weighting. The parameter δ largely governs curve elevation and measures the relative degree of optimism. The larger the value of δ for gains, the more elevated is the curve, the higher is the weight placed on every probability and, consequently, the more optimistically the prospect is valued, ceteris paribus. For losses, the opposite holds. Linear weighting is characterized by $\gamma = \delta = 1$. In a sign-dependent model, the parameters may take on different values for gains and for losses.

We now turn to the third step of model specification. In the course of the experiments, we measured risk taking behavior of individual $i \in \{1, \ldots, N\}$ by her certainty equivalents ce_{ig} for a set of different lotteries. Since CPT explains *deterministic* choice, we have to add an error term ϵ_{ig} in order to estimate the parameters of the model based on the elicited certainty equivalents. The observed certainty equivalent ce_{ig} can then be written as $ce_{ig} = \hat{ce}_g + \epsilon_{ig}$. There

[6]Loss aversion, interpreted as difference between risk aversion with respect to nonmixed and *mixed* lotteries, i.e. lotteries with both positive and negative outcomes, is not identifiable in our data, as the lottery designs did not contain any mixed lotteries (Köbberling and Wakker, 2005).

may be different sources of error, such as carelessness, hurry or inattentiveness, resulting in accidentally wrong answers (Hey and Orme, 1994). The Central Limit Theorem supports our assumption that the errors are normally distributed and simply add white noise.

Furthermore, we allow for three different sources of heteroskedasticity in the error variance. First, for each lottery, subjects had to consider 20 certain outcomes, which are equally spaced throughout the lottery's range $|x_{1g}-x_{2g}|$. Since the observed certainty equivalent ce_{ig} is calculated as the arithmetic mean of the smallest certain amount preferred to the lottery and the subsequent amount on the list, the error is proportional to the lottery range. Second, as the subjects may be heterogeneous with respect to their previous knowledge, their attention span as well as their ability of finding the correct certainty equivalent, we expect the error variance to differ by individual. Third, lotteries in the gain domain may be evaluated differently from the ones in the loss domain. Therefore, we allow for domain-specific variance in the error term. This yields the form $\sigma_{ig} = \xi_i |x_{1g} - x_{2g}|$ for the standard deviation of the error term distribution, where ξ_i denotes an individual domain-specific parameter. Note that the model allows to test for both individual-specific and domain-specific heteroskedasticity by either imposing the restriction $\xi_i = \xi$, or by forcing all the ξ_i to be equal in both decision domains. Both types of restrictions are rejected by their corresponding likelihood ratio tests in all three samples with p-values close to zero. Therefore, we control for all three types of heteroskedasticity in the estimation procedure.

Having discussed all the necessary ingredients, we now turn to the specification of the finite mixture regression model. The basic idea of the mixture model is assigning an individual's risk-taking choices to one of C types of behavior, each characterized by a distinct vector of parameters $\theta_c = (\alpha_c, \beta_c, \gamma_c', \delta_c')'$[7]. We

[7]The vectors γ_c and δ_c contain the domain-specific parameters for the slope and the elevation of the probability weighting functions.

denote the proportions of these different types in the population by π_c. Given our assumptions on the distribution of the error term, the density of type c for the i-th individual can be expressed as

$$f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right) = \prod_{g=1}^{G} \frac{1}{\sigma_{ig}} \, \phi\left(\frac{ce_{ig} - \hat{ce}_g}{\sigma_{ig}}\right),$$

where $\phi(\cdot)$ denotes the density of the standard normal distribution. Since we do not know *a priori* to which group a certain individual belongs to, the proportions π_c are interpreted as probabilities of group membership. Therefore, each individual density of type c has to be weighted by its respective mixing proportion π_c, which, of course, is unknown and has to be estimated as well. Summing over all C components yields the individual's contribution to the model's likelihood L. The log likelihood of the finite mixture regression model is then given by

$$\ln L\left(\Psi; ce, \mathcal{G}\right) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c \, f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right),$$

where the vector $\Psi = (\theta_1', \ldots, \theta_C', \pi_1, \ldots, \pi_{C-1}, \xi_1, \ldots, \xi_N)'$ summarizes all the parameters of the model.

The parameters are estimated by the iterative Expectation Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977), which provides an additional feature: In each iteration, the algorithm calculates by Bayesian updating an individual's posterior probability τ_{ic} of belonging to group c. The final posterior probabilities represent a particularly valuable result of the estimation procedure. Not only do we obtain the probabilities of individual group membership, but we also have a method of judging the quality of classification at our disposal. If all the τ_{ic} are either close to zero or one, all the individuals are unambiguously assigned to one specific group. The τ_{ic} can be used to calculate a summary measure of ambiguity in order to gage the extent of dubious assignments. If classification has been successful, i.e. if genuinely distinct types have been identified, we should observe a low measure of entropy.

Various problems may be encountered when maximizing the likelihood function of a finite mixture regression model and, therefore, a customized estimation procedure was used that can adequately deal with these problems. Details of the estimation procedure, written in the *R* environment (R Development Core Team, 2006), are discussed in Appendix A.

4 Results

In the following section we describe observed risk taking behavior before presenting the results of the finite mixture regressions. As the finite mixture regression model is defined over a pre-specified number of groups, we need to assess the correct number of groups and, therefore, discuss model selection first. Second, we document the cleanness and robustness of individuals' segregation to types. Furthermore, each distinct type of individual is characterized by her estimated behavioral parameter values, whereby we address the issue of cross-cultural differences. We also discuss type-specific differences in observed behavior and relate them to demographic variables. Finally, we comment on the stability of classification with respect to model specification.

4.1 Descriptive Statistics

At the level of observed data, risk taking behavior can be conveniently summarized by relative risk premia $RRP = (ev - ce)/|ev|$, where ev denotes the expected value of a lottery's payoff and ce stands for its certainty equivalent. $RRP > 0$ indicates risk aversion, $RRP < 0$ risk seeking, and $RRP = 0$ risk neutrality. In the context of EUT, risk preferences are captured solely by the curvature of the utility function, which in turn determines the sign of relative risk premia. Hence, the sign of RRP should be independent of p, the probability of the more extreme lottery outcome. In Figures 2 through 4, median risk premia sorted by p show a systematic relationship between RRP and p, however: In all three data sets subjects' choices display a fourfold pattern, i.e. they are risk averse for low-probability losses and high-probability gains, and they are risk seeking for low-probability gains and high-probability losses. Therefore, at a first glance, average behavior is adequately described by a model such as CPT rather than EUT. As the following sections show, the median RRPs gloss over an important feature of the data as there is substantial latent heterogeneity in risk taking behavior.

Figure 2: **Median Relative Risk Premia, Zurich 2003**

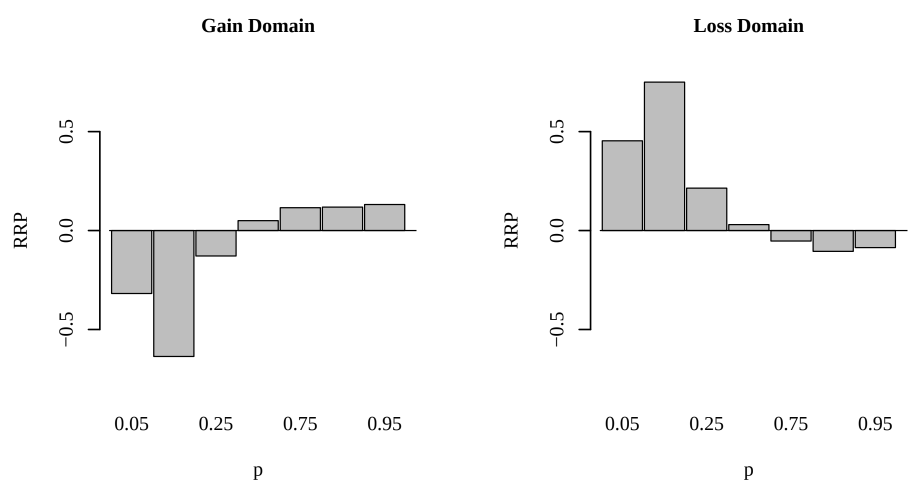


Figure 3: **Median Relative Risk Premia, Zurich 2006**

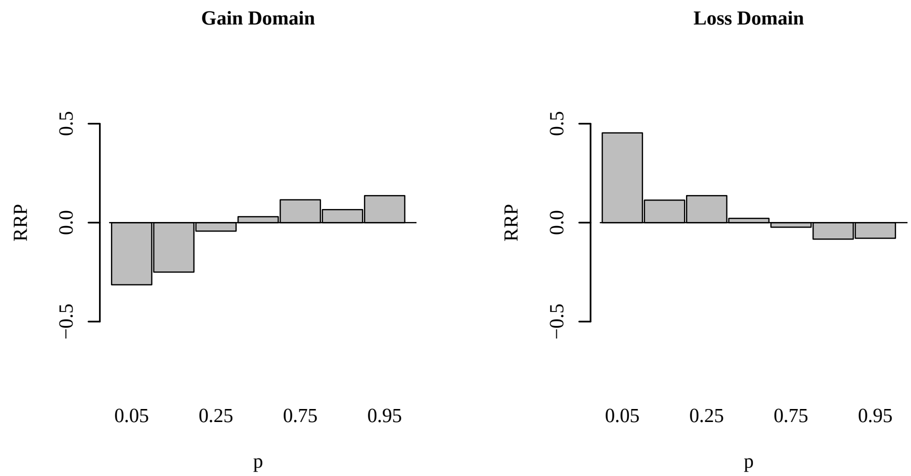
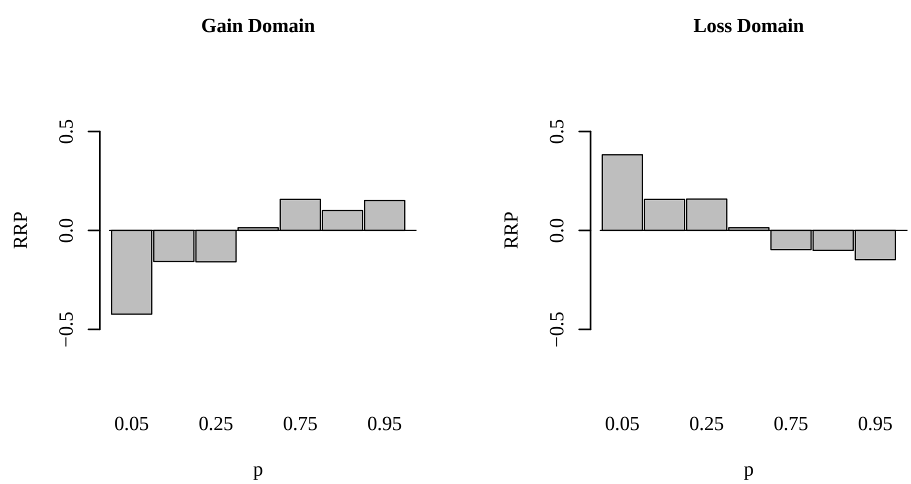


Figure 4: **Median Relative Risk Premia, Beijing 2005**



4.2 Model Selection

So far we have not addressed the issue whether a finite mixture regression model is actually to be preferred over a single-component model in the first place, and what the number of groups C in the mixture model, often termed *model size*, should be. In order to deal with these questions the researcher needs a criterion for assessing the correct number of mixture components. The literature on model selection in the context of mixture models is quite controversial, however, and there is no best solution.[8] For this reason, rather than relying on a single measure, we examine several criteria with differing characteristics to get a handle on the problem of model selection.

Obviously, the classical information criteria, the *Akaike Information Criterion AIC* and the *Bayesian Information Criterion BIC* are a natural starting point for our analysis. Unfortunately, the AIC criterion is order-inconsistent, i.e. the probability that it is minimized at the true size of the model does not approach unity with increasing sample size, and it tends to overfit models (Celeux and Soromenho, 1996). The BIC, on the other hand, has been proved to be consistent under suitable regularity conditions but may suffer from over- or underestimating the number of mixture components (Biernacki, Celeux, and Govaert, 2000).

Aside from these problems, both classical criteria share the principle of trading off model parsimony against goodness of fit, but do not directly measure the ability of the mixture to provide well separated and nonoverlapping components, which, ultimately, is the objective of estimating mixture models. Therefore, Celeux and Soromenho (1996) propose to use the *Normalized Entropy Criterion NEC*, which is based on the posterior probability of group membership τ_{ic}. Biernacki, Celeux, and Govaert (1999) argue that the NEC criterion appears to be less sensitive than AIC and BIC. However, the NEC focuses solely on the quality of classification and does not take model fit into account. Consequently, these authors argue that NEC is not adequate with respect to the choice of the relevant form of the mixture model.

[8] "The problem of identifying the number of classes is one of the issues in mixture modeling with the least satisfactory treatment." (Wedel (2002), p.364)

Ideally, what the researcher would like to have at her disposal is a criterion that delivers both an assessment of model fit, making allowance for parsimony, and the quality of classification. Biernacki, Celeux, and Govaert (2000) therefore suggest to modify the BIC criterion by factoring in a penalty for mean entropy. When the mixture components are well separated, mean entropy is close to zero and its weight in their proposed *Integrated Completed Likelihood Criterion ICL* is negligible. In the one-component case there is no entropy by definition, and therefore the ICL coincides with the BIC. While there is no theoretical justification for this approach, simulations show a superior performance compared to other heuristic criteria, such as the NEC (Biernacki, Celeux, and Govaert, 2000), as well as compared to AIC and BIC (McLachlan and Peel, 2000).

As different criteria may come up with conflicting results concerning the correct number of mixture components, model selection is a difficult problem. One way of dealing with this issue is to use one's central research question as a guide line. Our concern here is twofold: First, given the vast heterogeneity in individual risk taking behavior, it is doubtful whether a single-component model is adequate. Therefore, the crucial question is whether $C > 1$ should be preferred to $C = 1$.[9] Second, considering the heated dispute about the empirical validity of expected utility theory, another objective of our study is to find out whether a stable group of expected utility maximizers can be reliably identified in all three data sets. Bearing these objectives in mind, we calculated values for four different criteria, AIC, BIC, NEC as well as ICL, and three different model sizes, $C \in \{1, 2, 3\}$, presented in Table 3. According to these criteria, the model size which minimizes the respective criterion value should be preferred.

As AIC, BIC, and therefore also ICL, are highest at $C = 1$ for all three data sets, $C > 1$ is clearly favored over $C = 1$. As the NEC criterion is not defined for $C = 1$, Biernacki, Celeux, and Govaert (1999) argue in favor of a

[9] Parameter estimates for $C = 1$ are presented in Appendix B.

Table 3: **Model Selection Criteria**

Zurich 03	**AIC**	**BIC**	**NEC**	**ICL**
$C = 1$	-38,398	-35,815	n.a.	-35,815
$C = 2$	-39,629	-36,997	**0.0099**	-36,991
$C = 3$	**-40,504**	**-37,822**	0.0131	**-37,807**
Zurich 06	**AIC**	**BIC**	**NEC**	**ICL**
$C = 1$	-20,858	-19,297	n.a.	-19,297
$C = 2$	-22,173	-20,568	**0.0041**	-20,566
$C = 3$	**-22,622**	**-20,971**	0.0049	**-20,968**
Beijing 05	**AIC**	**BIC**	**NEC**	**ICL**
$C = 1$	-18,485	-16,529	n.a.	-16,529
$C = 2$	-19,585	-17,585	**0.0061**	-17,582
$C = 3$	**-19,965**	**-17,920**	0.0114	**-17,912**

multi-component model if there is a $C > 1$ with $NEC(C) \leq 1$, which is the case here. We therefore conclude that a finite mixture model is superior to a single-component model, given the unanimous recommendation by all four criteria.

With regard to the choice between $C = 2$ and $C = 3$, the three-group classifications seem to be favored by all criteria but NEC. As entropy is generally extremely low for both the two-group and three-group classifications, both model sizes seem quite sensible, however. Before we infer from these results that we should choose $C = 3$, we take a closer look at the difference between the two-group and three-group classifications.[10] What is of special interest here is whether one group remains fairly stable and the other group gets subdivided into two new ones when model size is increased, or whether the individuals get reshuffled to three new types. If the latter were the case, a two-group specification would clearly be misleading. In order to answer this question we examine relative group sizes and transition patterns of individuals' type assignment.

Table 4 displays the estimated relative group sizes of the behavioral types for model sizes $C = 2$ and $C = 3$. As the percentages reveal, all the *Type I* groups remain stable with respect to relative group size. Moreover, with a few exceptions, the *Type I* individuals remain *Type I* when model size is increased: Only a total of 2% of the individuals move into or out of *Type I* when an additional component is introduced into the two-group finite mixture model.[11] Increasing model size results in a decomposition of the original *Type II* groups, as there is still considerable heterogeneity within these groups. Thus, from the point of view of identifying *Type I* individuals, the two-group classifications

[10] In principle, it is possible to estimate mixture models with even more components, but problems of multimodality and potential unboundedness of the likelihood function tend to become more severe with increasing model size.

[11] Across all three data sets, only 2 individuals are newly assigned to *Type I* and 7 individuals leave *Type I*, when C is increased from 2 to 3.

Table 4: **Relative Group Sizes**

Zurich 03	*Type I*	*Type II*	*Type III*
$C = 2$	17.1 %	82.9 %	
$C = 3$	16.7 %	27.3 %	56.0 %
Zurich 06	*Type I*	*Type II*	*Type III*
$C = 2$	22.3 %	77.7 %	
$C = 3$	22.0 %	29.8 %	48.2 %
Beijing 05	*Type I*	*Type II*	*Type III*
$C = 2$	20.3 %	79.7 %	
$C = 3$	19.9 %	29.3 %	50.8 %

are informative by themselves whereas three groups render a more detailed description of the original *Type II* individuals. In the following, therefore, we will present the results both for $C = 3$ and $C = 2$.

4.3 Clean and Robust Segregation of Behavioral Types

In order to be of value to applied economics, a classification of risk taking behavior should meet two conditions: First, it should be clean, i.e. all the individuals should be clearly associated with one specific risk taking type. Second, the classification should be robust across different experiments based on the same design principles. Regarding the first condition, entropy criteria, based on the posterior probabilities of group membership, can be used to evaluate the quality of classification. One such measure is the *Normalized Entropy Criterion NEC*, introduced in the previous section. If all the individuals can be clearly assigned to one of the different behavioral groups, the posterior probabilities of group membership τ_{ic} are close to zero or one, and $NEC \approx 0$. As Table 3 shows, NEC always lies in the vicinity of zero, irrespective of the assumptions on model size

C.

The high quality of the classifications can also be inferred directly from the distributions of the individuals' posterior probabilities of group membership. In Figure 5, based on $C = 2$, τ_{EUT} denotes the posterior probability of belonging to the first type, which can indeed be characterized, as we will demonstrate below, as expected utility maximizers.[12] As the distributions of τ_{EUT} show, the individuals' posterior probabilities of behaving consistently with EUT are either close to one or close to zero for practically all the individuals in all three data sets, indicating an extremely clean segregation of subjects to types. Our result is quite remarkable as it substantiates that there are distinct types in the population, be they Swiss or Chinese. And it also shows that the underlying behavioral model provides a sound basis of discriminating between them.

With respect to the second criterion, robustness of classification, Figure 5 illustrates the probably most striking result of our study, namely similar distributions of types across all three data sets. In all three histograms of Figure 5, there are about four times as many individuals with τ_{EUT} close to zero, compared to individuals with τ_{EUT} close to one. This finding is mirrored by the estimates of the relative group sizes, displayed in Table 4, which shows a stable proportion of *Type I* of about 20%, irrespective of model size C. Moreover, it can be shown that the hypothesis that same distribution of types prevails in all three data sets cannot be rejected. Similarly, when model size is increased to $C = 3$, relative group sizes turn out to be of equal magnitudes in all three data sets and are statistically indistinguishable from one another. Therefore, classification is not only unambiguous, but also results in roughly equal mixing proportions, demonstrating that classification is robust across experiments.

This finding leads us to the next question. Do the respective types identified in each data set also exhibit similar patterns of behavior? This question will

[12] As group membership is stable, histograms of τ_{EUT} for $C = 3$ are qualitatively the same.

be addressed in the following sections, dedicated to the characterization of the endogenously defined types of behavior.

Figure 5: **Distribution of Posterior Probability of Assignment to EUT, τ_{EUT} ($C = 2$)**

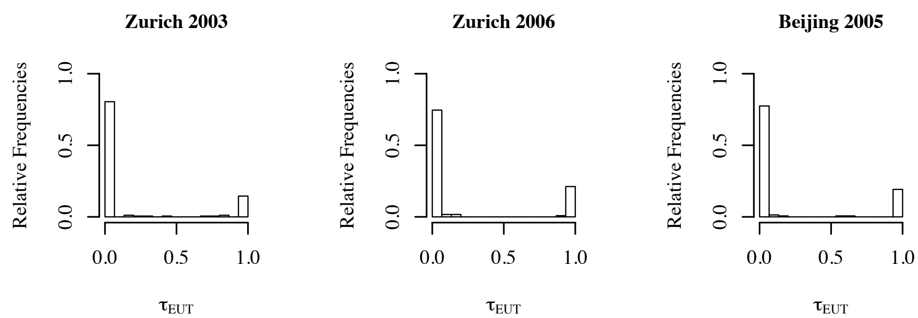


Figure 6: **Type-Specific Probability Weighting Functions, Zurich 2003**

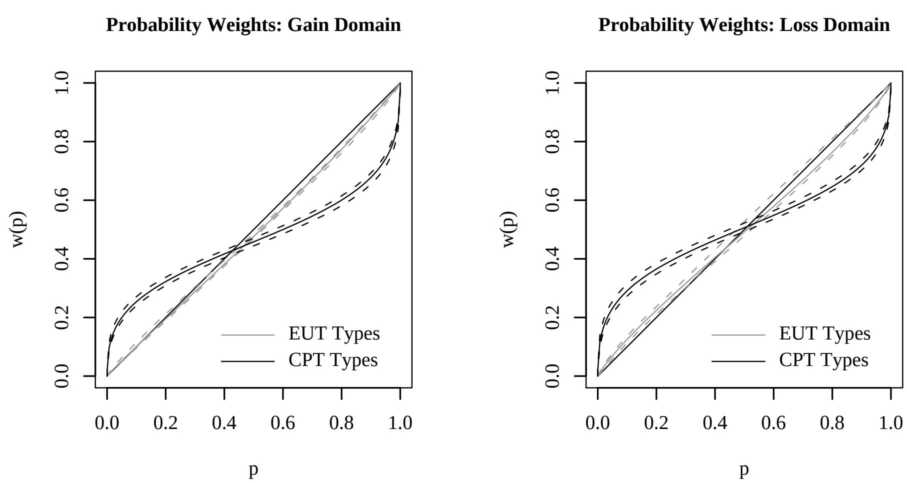


Table 5: **Classification of Behavior** ($C = 2$)

	EUT Types				CPT Types			
Parameters	ZH 03	ZH 06	BJ 05	Pooled	ZH 03	ZH 06	BJ 05	Pooled
π	0.171	0.223	0.203	0.193	0.829	0.777	0.797	0.807
	(0.026)	(0.025)	(0.020)	(0.013)				
Gains								
α	0.978	0.988	1.083	0.981	1.054	0.901	0.389	0.941
	(0.014)	(0.018)	(0.102)	(0.011)	(0.021)	(0.026)	(0.107)	(0.013)
γ	0.954	0.945	0.911	0.943	0.415	0.425	0.245	0.377
	(0.022)	(0.020)	(0.033)	(0.021)	(0.015)	(0.015)	(0.014)	(0.009)
δ	0.910	0.909	0.889	0.911	0.845	0.862	1.315	0.926
	(0.015)	(0.019)	(0.052)	(0.012)	(0.022)	(0.028)	(0.074)	(0.013)
Losses								
β	1.007	1.013	1.023	1.015	1.107	1.122	1.144	1.139
	(0.018)	(0.023)	(0.084)	(0.013)	(0.028)	(0.047)	(0.107)	(0.019)
γ	0.871	0.953	0.949	0.950	0.417	0.451	0.309	0.397
	(0.043)	(0.020)	(0.040)	(0.023)	(0.017)	(0.014)	(0.013)	(0.009)
δ	0.967	1.049	1.066	1.072	1.025	1.059	0.937	0.991
	(0.062)	(0.033)	(0.065)	(0.026)	(0.028)	(0.044)	(0.053)	(0.016)
$\ln L$	20,185	11,336	10,108	41,385				
Parameters	371	249	315	909				
Individuals	179	118	151	448				
Observations	8,906	4,669	4,225	17,800				

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications. Parameters include additional estimates for ξ_i for domain- and individual-specific error variances. ZH stands for Zurich, BJ for Beijing.

Figure 7: **Type-Specific Probability Weighting Functions, Zurich 2006**

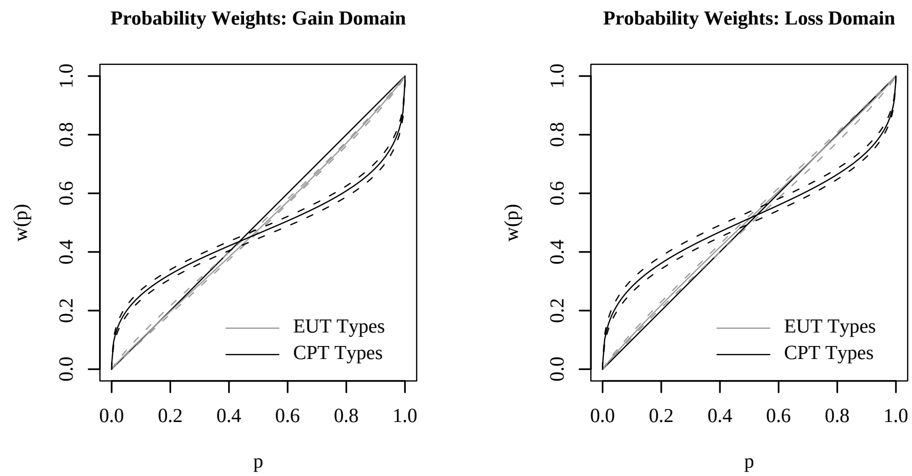
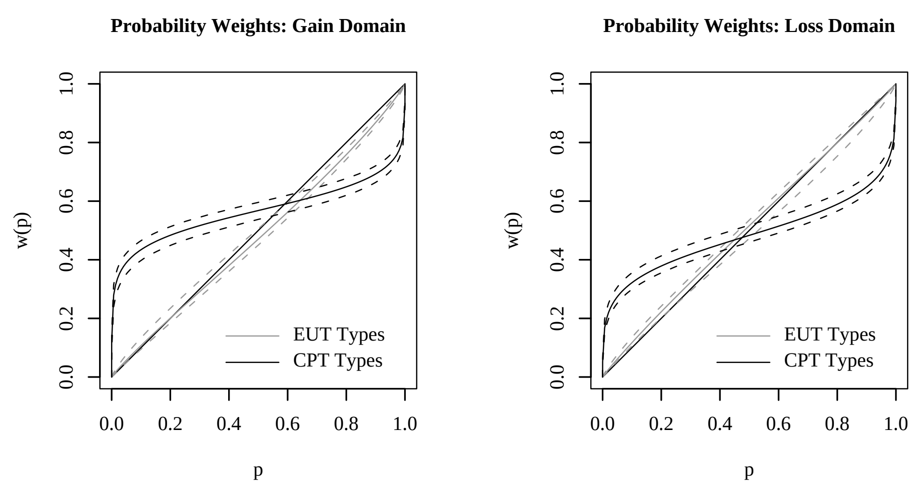


Figure 8: **Type-Specific Probability Weighting Functions, Beijing 2005**



4.4 Characterization of the Minority Type

Irrespective of model size, the first type of individuals encompasses about 20% of the subjects in all three data sets, thus constituting the minority type. In order to characterize risk taking behavior we examine the parameter estimates of the value functions and probability weighting functions. Table 5 displays, for $C = 2$, the type-specific parameter estimates of the finite mixture regression model and their standard errors, obtained by the bootstrap method with $4,000$ replications (Efron and Tibshirani, 1993). When model size is increased to three groups, parameter estimates, presented in Tables 9 to 12 in Appendix C, remain unchanged for the minority type, as group membership does not change substantially. Therefore, from the point of view of identifying this type of individuals, model size is not a crucial issue.

Concerning probability weighting, Table 5 displays almost identical parameter estimates across all three data sets as well as the pooled data. Without having imposed any restrictions on the parameters, we find that the minority types' probability weighting functions are roughly linear, as the parameter estimates for both γ and δ are close to one. Since the probability weights are a nonlinear combination of these parameters, inference needs to be based on γ and δ jointly. Therefore, we constructed the 95%-confidence bands for the probability weighting curves by the percentile bootstrap method. Figures 6, 7, and 8 contain the graphs of the type-specific probability weighting functions for each decision domain. The gray solid lines correspond to the estimated curves for the first type, referred to as "EUT type", and the gray dashed lines delimit their respective confidence bands. For both gains and losses, the confidence bands for the first type in fact include the diagonal over a wide range of probabilities, demonstrating high congruence with linear probability weighting. Where the confidence bands do not include the diagonal, the curves still lie extremely close to linear weighting. In sum, in all three data sets, we find the first behavioral type to exhibit near linear probability weighting.

With respect to the valuation of monetary outcomes, the second element of the decision model, the estimated parameters α and β also display a high degree of conformity. As can be inferred from the bootstrapped standard errors in Table 5, the 95%-confidence intervals of each single curvature estimate contains unity, implying that the hypothesis of linear value functions cannot be rejected. Together with near linear probability weighting, this result justifies regarding the first type of individuals as largely consistent with expected value maximization, and therefore EUT.

4.5 Characterization of the Majority Types

In contrast to the EUT groups, model size makes a difference in characterizing the majority types. In the two-group classification, average behavior of the majority group can be interpreted as a mixture of two different subtypes. We will present results for $C = 2$ first and then comment on the $C = 3$ results.

The majority types' probability weighting curves for $C = 2$ are pictured as black lines in Figures 6, 7, and 8. The solid lines correspond to the estimated curves, the dashed lines mark the corresponding 95%-confidence bands. For both gains and losses, all three figures show inverted S-shaped probability weighting functions. Consequently, we label these individuals as "CPT types". However, CPT individuals do not display as uniform a behavior across cultures as do the EUT individuals: Whereas the Swiss probability weighting curves in Figures 6 and 7 are quite similar, the Chinese ones in Figure 8 differ markedly from the Swiss. First, the Chinese probability weighting functions are generally flatter in the middle part than the Swiss curves, which indicates a lower sensitivity towards changes in probabilities (for gains estimated $\gamma = 0.245$ for Chinese subjects versus 0.415 and 0.425 for Swiss subjects, and for losses estimated $\gamma = 0.309$ versus 0.417 and 0.452; see Table 5). Second, the Swiss probability weighting curves for gains exhibit the familiar shape, i.e. intersection with the diagonal at probabilities of about 0.4, whereas the Chinese probability weighting function is much more elevated than the Swiss ones, implying substantially more optimistic weighting of gain probabilities (estimated

$\delta = 1.316$ versus 0.845 and 0.862, respectively). Similarly, cultural differences are evident in the estimated value function parameters over gains. Contrary to the slightly concave or convex Swiss curvatures, the Chinese value function is distinctly concave (estimated $\alpha = 0.387$).

As already noted in the section on model selection, model fit improves with increasing model size. When a third group is allowed for, we find that the original CPT groups get subdivided into two different CPT types, each characterized by a specific variety of nonlinear probability weighting. Estimates are presented in Tables 9 to 12 in Appendix C. The difference between CPT I and CPT II types manifests itself predominantly in their relative strength of optimism: the elevation of the probability weighting curves, measured by δ, differ substantially between CPT I and CPT II as the respective Figures 12 to 14 in Appendix C show. CPT II individuals, who constitute the relative majority of approximately 50% in all three data sets, exhibit S-shaped probability weighting curves with δ in the vicinity of one. Swiss CPT I individuals, however, are systematically less optimistic than CPT II types, whereas the Chinese CPT I type encompasses highly optimistic individuals, explaining the prevalence of optimism in the Chinese population. The high degree of optimism of Chinese CPT I types manifests itself not only in the gain domain but also in the loss domain, which does not get uncovered by the two-group classification. This evidence may constitute a valuable piece of information when more disaggregate estimates of risk taking behavior are called for. When the focus of research lies on a parsimonious characterization of risk taking types with respect to EUT versus CPT, two-group classifications are sufficiently informative due to the stability of EUT group membership.

4.6 Observed Behavior by Type

So far we have characterized the different behavioral types by their estimated parameter values. The obvious question arises whether the discriminatory power of the classification procedure can also be traced at the behavioral level. After assigning the subjects to one of the two major types, EUT and CPT, based on their posterior probability of group membership τ_{ic}, the observed relative risk premia can be broken down by type as depicted in Figure 9, exemplary for the Chinese data set. As can be seen, median RRP of the Chinese EUT type are close to zero, reflecting near risk neutral behavior in accordance with expected value maximization. A similar picture can be shown to emerge for the Zurich 2003 and Zurich 2006 data sets.

When tracing behavior of the CPT types at the level of observed RRP in Figure 9, we find a pronounced fourfold pattern of Chinese risk attitudes, with more extreme departures from risk neutrality than the aggregate risk premia in Figure 4. As before, a similar picture can be shown to emerge for the Zurich 2003 and Zurich 2006 data. These findings document that individuals' type assignment is highly congruent with observed behavioral differences.

Obviously, the type-specific median relative risk premia do not differ greatly at $p = 0.5$. In decision situations when the more extreme reward materializes with a 50% chance, the typical CPT individual will not over- or underweight its probability significantly, and therefore her behavior will often not be distinguishable from a typical EUT type's. This consideration can be illustrated by means of Figure 10, which displays the departures of average CPT behavior from EUT, measured by the type-specific differences in median normalized certainty equivalents. Each circle in Figure 10 corresponds to one specific lottery played in any of the three experiments, encompassing a total of 59 gain and 59 loss lotteries, ordered by the probability of the more extreme lottery outcome. At a gain probability of 25%, for instance, CPT lottery evaluations, on average, exceed EUT ones by up to 30% of their corresponding expected values. The dashed lines in the graphs represent the case when median CPT behavior does not differ from median EUT behavior. Positive values in the graphs indicate that, on average, CPT types are relatively more risk seeking than EUT types. The opposite holds for negative values. As the graphs show, zero differences occur solely at the 0.5 probability level where, in some cases, average CPT behavior is totally indistinguishable from EUT behavior. The bulk of type-specific differences in lottery evaluations lie in the range of about +/- 20% of expected values, but there are also a few observations with up to +/- 300% of expected value, where the more extreme outcomes materialize with a low probability. In these cases, CPT types tend to overreact pronouncedly to stated probabilities. In order to provide an overall measure we conducted two-sided Mann-Whitney tests which indicate significant differences (at the 5% level) in the type-specific distributions of the certainty equivalents for 75% of the lotteries.

Figure 9: **Median Relative Risk Premia by Type, Beijing 2005**

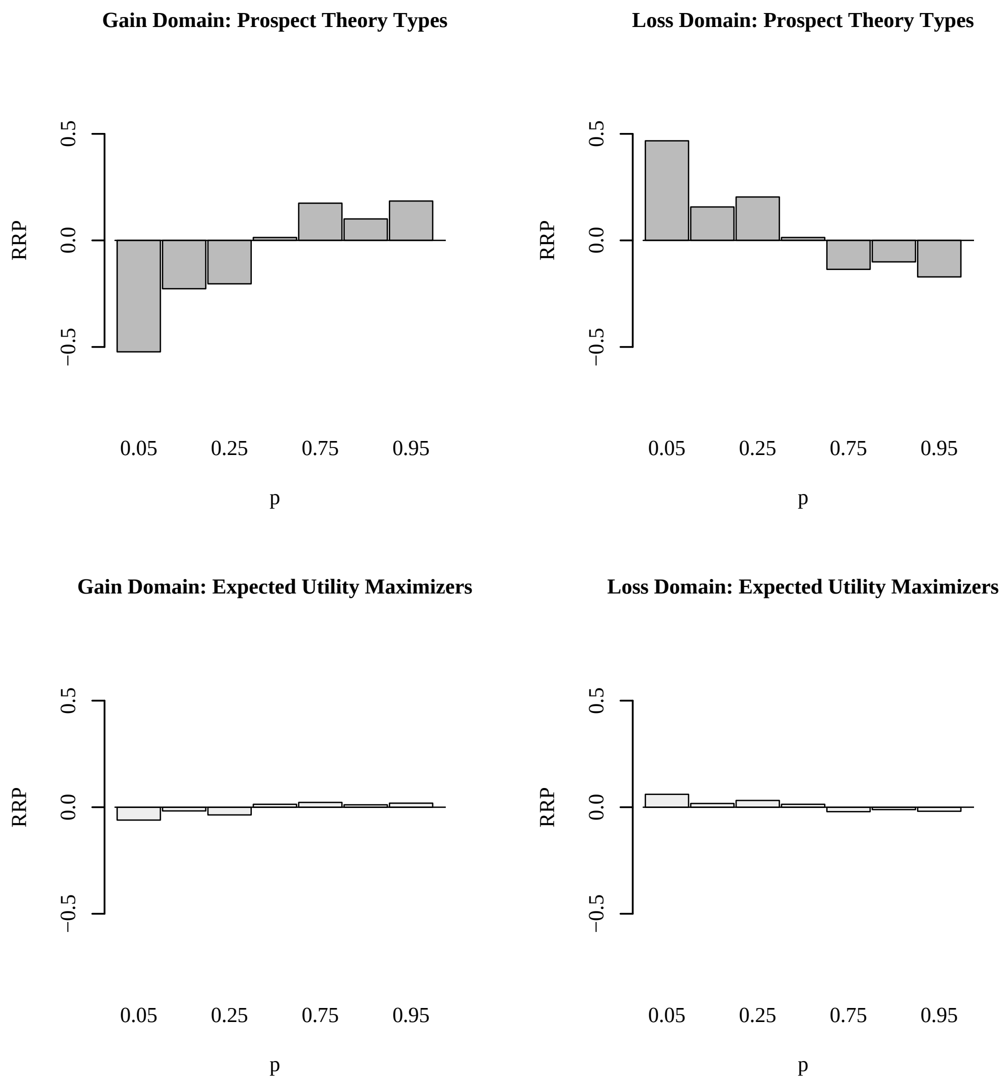
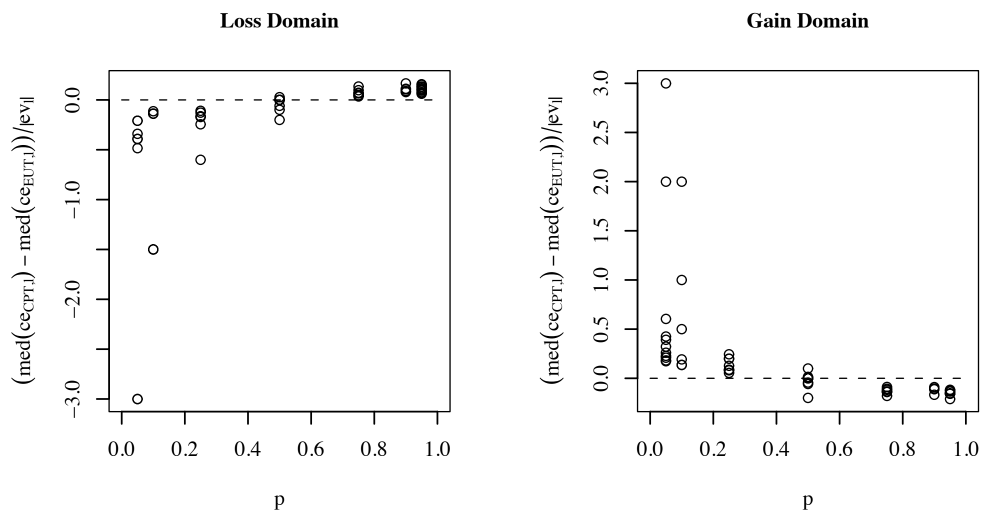


Figure 10: **Differences in Median Normalized Certainty Equivalents, Pooled**



4.7 Demographics and Group Membership

The finite mixture regression model is a powerful tool to uncover latent heterogeneity in behavior. Given our clean and robust classification of types, it is an interesting question whether we can characterize the composition of the different groups by demographic variables. In particular, can we explain who the EUT types are? In order to answer this question we conducted two kinds of analysis. First, we estimated a single-component model with demographic variables as covariates. This procedure uncovers systematic behavioral differences among groups defined by observable socio-economic characteristics. Second, we ran a regression of the posterior probability of EUT group membership τ_{EUT} on a number of demographic variables available in all three data sets. For both approaches we included the following variables: a gender dummy *female*, the number of semesters enrolled at university *semester*, and a binary variable *high-income* for incomes above a certain threshold. Summary statistics for these variables are included in Appendix D.

The estimates for the single-component model are presented in Table 14 in Appendix D. The only variable that consistently affects behavioral parameters across experiments is *female*[13]: Being female is associated with a substantially lower value of γ, the slope of the probability weighting function. This finding implies that women tend to be less sensitive to changes in probability than are men, in line with the evidence in Fehr-Duda, de Gennaro, and Schubert (2006).

Consistent with these findings, women are also less likely to belong to the EUT groups, as the estimates of the linear probability model in Table 6 show: The coefficients for *female* are significantly negative for all three data sets. There are no consistent effects of *highincome* on the probability of EUT group membership. Neither does *semester* have a significant effect in all data sets. Since the driving force of classification is individuals' proneness to nonlinear probability weighting, observable demographic characteristics would have to be systematically associated with this propensity in order to have any explanatory power for classification. In our experience, in student subject pools we generally do not find socio-economic characteristics, other than gender, that are systematically correlated with the curvature of the probability weighting function. Other factors than demographics may be more important here, but this question is still underresearched.

4.8 Robustness to Model Specification

The final part of our analysis concerns robustness of classification results with respect to alternative specifications of the value function. For instance, people may not evaluate gambles in isolation, but integrate prospective outcomes with their wealth or consumption spending. To account for the possibility that subjects integrate prospective outcomes with some background variable, we re-

[13]Note that the percentage of females is approximately 50% in all three data sets.

Table 6: **EUT Group Membership and Demographics**

Zurich 03	**Coefficient**	**Std. Err.**
intercept	0.131**	0.038
female	**-0.127****	**0.030**
semester	0.024**	0.005
highincome	0.027	0.039
R^2	0.056	
Zurich 06	**Coefficient**	**Std. Err.**
intercept	0.292**	0.045
female	**-0.207****	**0.036**
semester	0.003	0.006
highincome	0.259**	0.085
R^2	0.087	
Beijing 05	**Coefficient**	**Std. Err.**
intercept	0.212**	0.038
female	**-0.069***	**0.031**
semester	0.017*	0.008
highincome	-0.090*	0.041
R^2	0.020	

Linear probability model with dependent variable τ_{EUT}, $C = 2$. Bootstrapped standard errors based on 4'000 replications. Bootstrapping accounts for dependent variable being an estimated quantity. ** Significant at 1%, * significant at 5%. *highincome* equals one if disposable income per month exceeds CHF 1,500 or CHN 1,000, respectively.

estimated the model with the value function being defined over the sum of the prospective lottery outcome and an additional type-specific background parameter k, such that $v(x) = (x + k)^{\alpha}$ over gains and *mutatis mutandis* over losses, i.e. $v(x) = (x + \omega + k)^{\beta}$, where ω stands for the initial endowment.[14]

Extending the model, for $C = 2$, in such a manner yields the following insights: First, as Figure 11 shows, the stability of classification is not affected by the alternative model specification: For all three data sets, the distribution of the posterior probability of belonging to EUT is almost unchanged when background consumption is introduced into the model. The stability of group assignment is also reflected in the estimated relative group sizes π_{EUT}. Table 7 clearly shows that these values practically do not change when background consumption k is included. Moreover, not a single subject out of 448 is assigned to a different group, defined by $\tau_{ic} \geq 0.5$, after allowing for integration with background consumption. Finally, the estimated probability weighting functions for both the EUT types and the CPT types remain stable as well when background consumption is introduced into the model, as Figures 15 to 17, presented in Appendix E, confirm. In sum, our analysis attests that the distribution of types, individuals' type affiliations, as well as the estimated probability weighting functions are robust to inclusion of background consumption. This robustness result represents further evidence that decision makers' tendency to weight probabilities nonlinearly is the driving force of classification.

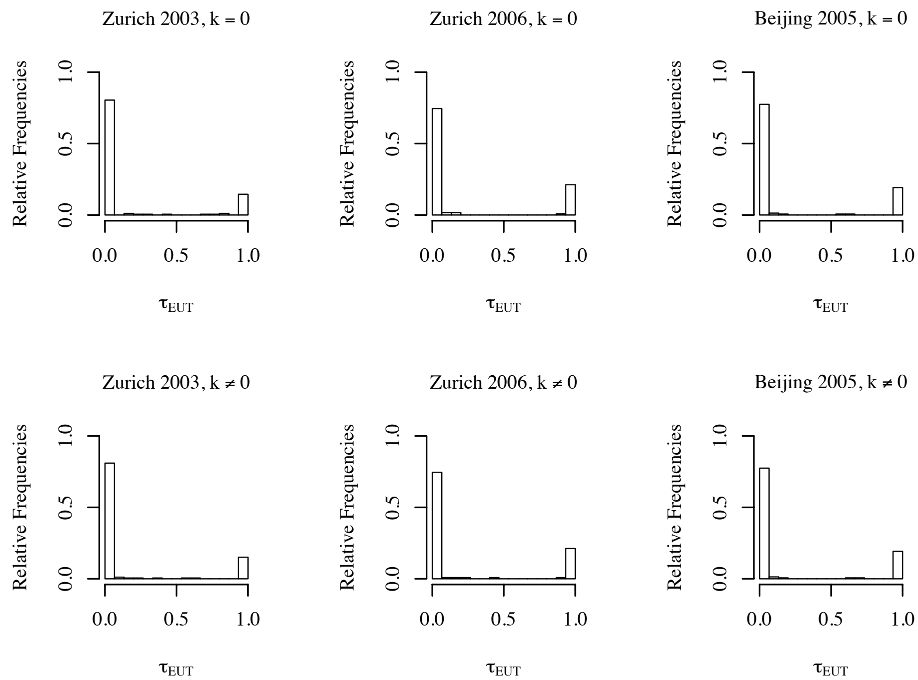
Table 7: **Estimated Model-Specific Proportions of EUT Types,** π_{EUT}

Consumption	**Zurich 03**	**Zurich 06**	**Beijing 05**
$k = 0$	0.171	0.223	0.203
k *endogenous*	0.163	0.227	0.203

[14] Note that k is not identifiable when functions v are near linear.

Figure 11:

Distribution of Posterior Probability of Assignment to EUT τ_{EUT}



5 Concluding Remarks

We conducted three experiments based on the same design principles and applied a finite mixture regression model to the choice data. Our results provide novel insights: In all three data sets the procedure renders a parsimonious characterization of risk taking behavior. Across experiments, we find an equal mix of distinct types, each characterized by a specific pattern of probability distortion. Almost every single individual is identified as one specific type, rendering segregation extremely clean. 20% of the population, by and large, adhere to linear probability weighting and behave essentially as expected value maximizers, whereas majority preferences are more suitably represented by a model such as prospect theory's, which can accommodate nonlinear probability weighting. In each data set, the overall CPT group is composed of a smaller group of 30%, displaying substantial departures from linear probability weighting, and a relative majority of 50%, departing less radically from linear probability weighting. Moreover, classification is robust to an alternative model specification.

Whereas the distribution of types is the same in the Swiss and the Chinese data sets, there are substantial cultural differences in CPT type behavior, the most prominent being the existence of a pronouncedly optimistic group of Chinese subjects who distort small probabilities much more strongly than do the Swiss. This prevalence of Chinese optimism in lottery valuation may explain previous findings that, on average, Chinese respondents are relatively more risk seeking than Western ones (Kachelmeier and Shehata, 1992; Hsee and Weber, 1999). We also identify a gender difference in risk taking behavior: Women are less likely to belong to the EUT group as they generally depart more strongly from linear probability weighting than do men. This finding corroborates previous research (Fehr-Duda, de Gennaro, and Schubert, 2006; Harrison and Rutström, forthcoming).

Our findings demonstrate that the finite mixture regression approach is a powerful tool to identify and characterize the distribution of risk taking types in the population. In this study, the individual is the unit of classification, i.e. our objective was to assign the *entirety* of an individual's choices to one distinct type of behavior. As the low measures of entropy demonstrate, this endeavor was successful and almost every individual got unambiguously assigned to one endogenously defined behavioral type. Previous work by Harrison and Rutström (forthcoming) tried to accomplish a different goal: They estimated the probability that any one lottery choice, irrespective of the identity of the decision maker, was consistent with EUT or CPT, respectively, and found that "each [specification] is equally likely for [these] data" (section 4.1, second paragraph). Such a different perspective on classification is bound to entail results markedly divergent from ours, since behavior of EUT and CPT individuals may be quite similar in certain decision situations. Our results indicate, for example, that for probabilities in the neighborhood of 0.5, CPT individuals' behavior may be hardly distinguishable from EUT behavior.

When we started this project we were quite confident that we would find a considerable percentage of expected utility maximizers. What really surprised us is the robust percentage of EUT types, even across two so different cultures as the Swiss and Chinese. Since the subject pools in all three experiments consisted of students, further research is needed to see whether the proportion of near rational EUT types changes significantly in a representative sample and whether the complexity of decision tasks greatly alters type assignment. If it can be ascertained that near rational actors constitute a non-negligible proportion of the population, their behavior, depending on the nature of the strategic environment, may be decisive for aggregate outcomes. The existence of a robust share of near rational actors suggests using a mix of preference theories for modeling behavior rather than a single theory, which would yield systematically biased results. In our data, prospect theory adequately describes behavior of the majority of subjects but the parameter estimates exhibit culture- as well as type-specific values. Researchers should take this evidence into account when constructing, estimating, and applying models of choice under risk.

References

Andersen, S., G. W. Harrison, and E. E. Rutström (2006): "Choice Behavior, Asset Integration and Natural Reference Points," *Department of Economics, College of Business Administration, University of Central Florida, Working Paper*, 06-07.

Biernacki, C., G. Celeux, and G. Govaert (1999): "An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Model," *Pattern Recognition Letters*, 20, 267–272.

——— (2000): "Assesing a Mixture Model for Clustering with the Integrated Completed Likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7), 719–725.

Camerer, C. F., and E. Fehr (2006): "When Does Economic Man Dominate Social Behavior?," *Science*, 311.

Celeux, G., D. Chauveau, and J. Diebolt (1995): "On Stochastic Version of the EM Algorithm," *Penn State University, Working Paper*, (RR-2514), 22 p.

Celeux, G., and G. Soromenho (1996): "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model," *Journal of Classification*, 13, 195–212.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005): "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," *Institute for the Study of Labor (IZA), Bonn (Germany), Discussion Paper*, 1730.

Eckel, C., C. Johnson, and C. Montmarquette (2005): "Savings Decisions of the Working Poor: Short and Long-Term Horizons," .

Efron, B., and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

El-Gamal, M. A., and D. M. Grether (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90(432), 1137–1145.

Fehr, E., and J. R. Tyran (2005): "Individual Irrationality and Aggregate Outcomes," *Journal of Economic Perspectives*, 19(4).

Fehr-Duda, H., M. de Gennaro, and R. Schubert (2006): "Gender, Financial Risk, and Probability Weights," *Theory and Decision*, 60, 283–313.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*.

Goldstein, W., and H. Einhorn (1987): "Expression Theory and the Preference Reversal Phenomena," *Psychological Review*, 94, 236–254.

Haltiwanger, J. C., and M. Waldman (1985): "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," *American Economic Review*, 75(3), 326–340.

——— (1989): "Limited Rationality and Strategic Complements: The Implications for Macroeconomics," *Quarterly Journal of Economics*, 104(3), 463–483.

Harless, D. W., and C. F. Camerer (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1290.

Harrison, G. W., S. J. Humphrey, and A. Verschoor (2005): "Choice Under Uncertainty in Developing Countries," *Department of Economics, College of Business Administration, University of Central Florida, Working Paper*, 0529.

Harrison, G. W., M. I. Lau, and E. Rutström (2007): "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109(2).

Harrison, G. W., and E. E. Rutström (forthcoming): "Expected Utility Theory and Prospect Theory: One Wedding and A Decent Funeral," *Experimental Economics*.

Hey, J. D., and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), 1291–1326.

Houser, D., M. Keane, and K. McCabe (2004): "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence using a Bayesian Type Classification Algorithm," *Econometrica*, 72(3), 781–822.

Houser, D., and J. Winter (2004): "How do behavioral assumptions affect structural inference?," *Journal of Business and Economic Statistics*, 22(1), 64–79.

Hsee, C. K., and E. U. Weber (1999): "Cross-National Differences in Risk Preferences and Lay Predictions," *Journal of Behavioral Decision Making*, 12, 165–179.

Kachelmeier, S. J., and M. Shehata (1992): "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review*, 82(5), 1120–1141.

Köbberling, V., and P. Wakker (2005): "An Index of Loss Aversion," *Journal of Economic Theory*, 122.

Lattimore, P. K., J. R. Baker, and A. D. Witte (1992): "The Influence of Probability on Risky Choice," *Journal of Economic Behavior and Organization*, 17, 377–400.

McLachlan, G., and D. Peel (2000): *Finite Mixture Models*. Wiley Series in Probabilities and Statistics.

Prelec, D. (1998): "The Probability Weighting Function," *Econometrica*, 3, 497–527.

Quiggin, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

R Development Core Team (2006): "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing (Vienna, Austria)*.

Rabin, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68(5), 1281–1292.

Render, R. A., and X. Walker (1984): "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, 26(2), 195–239.

Stahl, D. O., and P. W. Wilson (1995): "On Players' Models of other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254.

Starmer, C. (2000): "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38(2), 332–382.

Stott, H. P. (2006): "Cumulative Prospect Theory's Functional Menagerie," *Journal of Risk and Uncertainty*, 32, 101–130.

Tversky, A., and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Wedel, M. (2002): "Concomitant variables in finite mixture models," *Statistica Neerlandica*, 56(3), 362–375.

Wu, G., J. Zhang, and R. Gonzalez (2004): *Decision under Risk*. D. Koehler and N. Harvey (Eds.), The Blackwell Handbook of Judgment and Decision Making, Oxford University Press.

A Estimation of the Finite Mixture Regression Model

As it is generally the case in finite mixture models, direct maximization of the log likelihood function

$$\ln L(\Psi; ce, \mathcal{G}) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c \, f(ce_i, \mathcal{G}; \theta_c, \xi_i)$$

may encounter several problems, even if it is in principle feasible (for a general treatise see for example McLachlan and Peel (2000)). First, the highly non-linear form of the log likelihood causes the optimization algorithm to be rather slow or even incapable of finding the maximum. Second, the likelihood of a finite mixture model is often multimodal and therefore we have no guaranty that a standard optimization routine will converge towards the global maximum rather than to one of the local maxima.

However, if individual group-membership were observable and indicated by $t_{ic} \in \{0, 1\}$ the individual contribution to the likelihood function would be given by

$$\tilde{\ell}(\Psi_i; ce_i, \mathcal{G}, t_i) = \prod_{c=1}^{C} \left[\pi_c \, f(ce_i, \mathcal{G}; \theta_c, \xi_i)\right]^{t_{ic}}$$

By using the above formulation and taking logarithms, the complete-data log likelihood function

$$\ln \tilde{L}(\Psi; ce, \mathcal{G}, t) = \sum_{i=1}^{N} \sum_{c=1}^{C} t_{ic} \left[\ln \pi_c + \ln f(ce_i, \mathcal{G}; \theta_c, \xi_i)\right]$$

would follow directly. As relative group sizes sum up to one, their maximum likelihood estimates, $\hat{\pi}_c = 1/N \sum_{i=1}^{N} t_{ic}$, would be given analytically by the relative number of individuals in the respective group. Furthermore, the maximum likelihood estimates of the group-specific parameters could be obtained separately in each group by numerically maximizing the corresponding joint density function which would simplify the optimization problem considerably.

The EM algorithm proceeds iteratively in two steps, E and M, while it treats the unobservable t_{ic} as missing data. In the E-step of the $(k+1)$-th iteration the expectation of the complete-data log likelihood $\tilde{L}$, given the actual fit of the data $\Psi^{(k)}$, is computed. This yields, according to Bayes' law, the posterior probabilities of individual group-membership

$$\tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right) = \frac{\pi_c^{(k)} f\left(ce_i, \mathcal{G}; \theta_c^{(k)}, \xi_i^{(k)}\right)}{\sum_{m=1}^{C} \pi_m^{(k)} f\left(ce_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\right)}$$

which replace the unknown indicators of individual group-membership, t_{ic}. Given $\tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right)$, the complete-data log likelihood, $\tilde{L}$, is maximized in the following M-step which yields the updates of the model parameters,

$$\pi_c^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right),$$

and

$$\left(\theta_1^{(k+1)}, \ldots, \theta_C^{(k+1)}, \xi_1^{(k+1)}, \ldots, \xi_N^{(k+1)}\right) = \\ \underset{\theta_1,\ldots,\theta_C,\xi_1,\ldots,\xi_N}{\arg\max} \sum_{i=1}^{N} \sum_{m=1}^{C} \tau_{im}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right) \ln f\left(ce_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\right).$$

As Dempster, Laird, and Rubin (1977) show, the likelihood never decreases from one iteration to the next, i.e. $L\left(\Psi^{(k+1)}; ce, \mathcal{G}\right) \geq L\left(\Psi^{(k)}; ce, \mathcal{G}\right)$, which makes the EM algorithm converge monotonically towards the nearest maximum of the likelihood function regardless whether this maximum is global or just local. In the Zurich 2003 data set, we therefore needed to apply a stochastic extension, the Simulated Annealing Expectation Maximization (SAEM) algorithm proposed by Celeux, Chauveau, and Diebolt (1995), in order to overcome the EM algorithm's tendency to converge towards local maxima. In each iteration, there is a non-zero probability that the SAEM algorithm leaves the current optimization path and starts over in a different region of the likelihood function which results in much higher chances of finding the global maximum. But this robustness against multimodality of the objective function comes at the cost of much higher computational demands.

As the EM algorithm is computationally highly demanding, even in its basic form, and tends to become tediously slow when close to convergence our estimation routine relies on a hybrid estimation algorithm (Render and Walker, 1984): It first uses either the EM or the SAEM algorithm and takes advantage of their robustness before it switches to the direct maximization of the log likelihood by the much faster BFGS algorithm. The estimation routine in this form turned out to be efficient and robust as it reliably converged towards the same maximum likelihood estimates regardless of the randomly chosen start values.

B Aggregate Behavior

Table 8: **Single-Component Models**

	Gains			**Losses**		
Parameters	ZH 03	ZH 06	BJ 05	ZH 03	ZH 06	BJ 05
α/β	1.041	0.916	0.443	1.077	1.093	1.131
	(0.021)	(0.021)	(0.116)	(0.025)	(0.036)	(0.123)
γ	0.482	0.519	0.318	0.487	0.579	0.383
	(0.010)	(0.017)	(0.016)	(0.012)	(0.027)	(0.015)
δ	0.869	0.886	1.296	1.030	1.039	0.944
	(0.020)	(0.022)	(0.081)	(0.026)	(0.033)	(0.062)
$\ln L$	19,563	10,671	9,550			
Parameters	364	242	308			
Individuals	179	118	151			
Observations	8,906	4,669	4,225			

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications. ZH stands for Zurich, BJ for Beijing.

C Three-Group Classifications

Figure 12: **Probability Weights CPT I vs. CPT II, Zurich 2003**

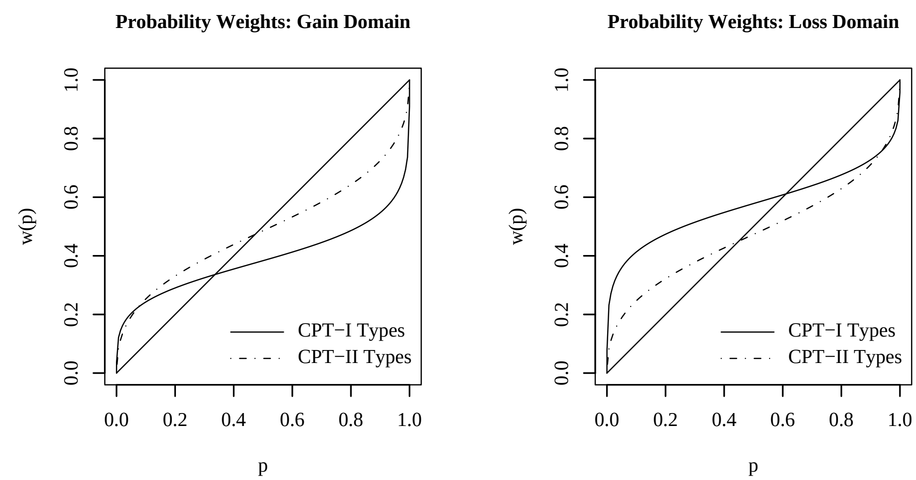


Figure 13: **Probability Weights CPT I vs. CPT II, Zurich 2006**

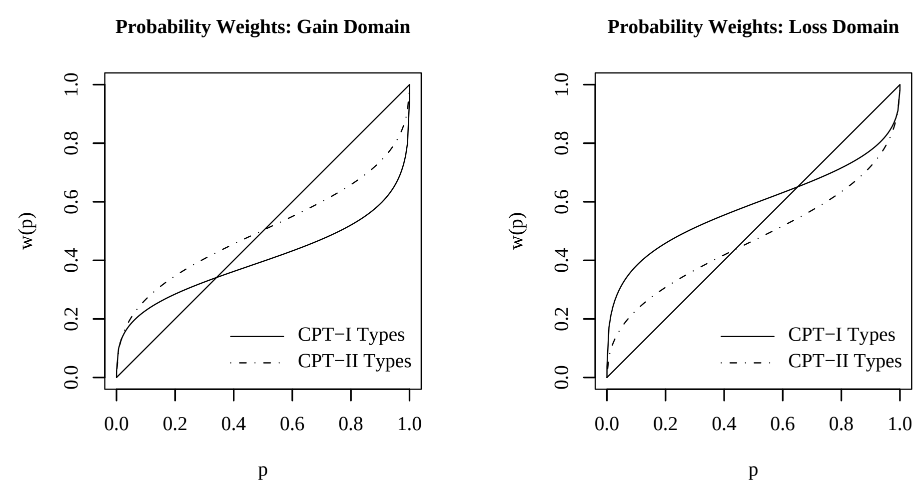


Figure 14: **Probability Weights CPT I vs. CPT II, Beijing 2005**

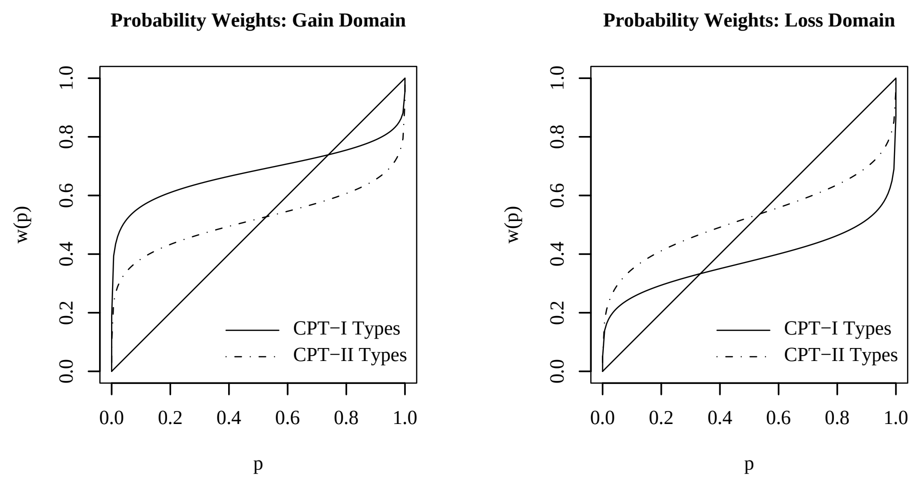


Table 9: **Classification of Behavior with $C = 3$, Zurich 2003**

	Gains				**Losses**		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.167	0.273	0.560				
	(0.016)	(0.015)	(0.022)				
α	0.954	1.007	1.075	β	1.006	1.237	1.091
	(0.013)	(0.016)	(0.015)		(0.020)	(0.044)	(0.015)
γ	0.944	0.302	0.467	γ	0.885	0.304	0.459
	(0.041)	(0.031)	(0.013)		(0.042)	(0.029)	(0.015)
δ	0.930	0.622	0.944	δ	1.024	1.371	0.897
	(0.020)	(0.023)	(0.017)		(0.043)	(0.075)	(0.016)
$\ln L$			20,630				
Parameters			378				
Individuals			179				
Observations			8,906				

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications.
Parameters include estimates of ξ_i for domain- and individual-specific error variances.

Table 10: **Classification of Behavior with $C = 3$, Zurich 2006**

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.220	0.298	0.482				
	(0.020)	(0.025)	(0.030)				
α	0.990	0.884	0.908	β	1.012	1.100	1.141
	(0.024)	(0.042)	(0.031)		(0.029)	(0.083)	(0.049)
γ	0.946	0.362	0.465	γ	0.952	0.393	0.491
	(0.084)	(0.081)	(0.022)		(0.081)	(0.078)	(0.023)
δ	0.905	0.658	1.012	δ	1.054	1.460	0.878
	(0.042)	(0.054)	(0.043)		(0.074)	(0.122)	(0.054)
$\ln L$			11,567				
Parameters			256				
Individuals			118				
Observations			4,669				

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications.
Parameters include estimates of ξ_i for domain- and individual-specific error variances.

Table 11: **Classification of Behavior with $C = 3$, Beijing 2005**

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.199	0.293	0.508				
	(0.017)	(0.026)	(0.027)				
α	1.083	0.032	0.489	β	1.023	1.348	1.111
	(0.098)	(0.155)	(0.113)		(0.070)	(0.149)	(0.102)
γ	0.911	0.244	0.254	γ	0.948	0.263	0.332
	(0.051)	(0.049)	(0.023)		(0.053)	(0.046)	(0.019)
δ	0.889	2.194	1.085	δ	1.062	0.600	1.106
	(0.094)	(0.241)	(0.113)		(0.057)	(0.093)	(0.075)
$\ln L$			10,304				
Parameters			322				
Individuals			151				
Observations			4,225				

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications.
Parameters include estimates of ξ_i for domain- and individual-specific error variances.

Table 12: **Classification of Behavior with $C = 3$, Pooled**

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.198	0.316	0.486				
	(0.010)	(0.011)	(0.013)				
α	0.960	0.901	0.957	β	1.019	1.250	1.139
	(0.009)	(0.009)	(0.010)		(0.008)	(0.010)	(0.009)
γ	0.915	0.309	0.451	γ	0.935	0.339	0.444
	(0.032)	(0.015)	(0.010)		(0.027)	(0.013)	(0.011)
δ	0.935	0.726	1.063	δ	1.055	1.230	0.878
	(0.009)	(0.012)	(0.010)		(0.013)	(0.013)	(0.011)
$\ln L$			42,105				
Parameters			916				
Individuals			448				
Observations			17,800				

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications.
Parameters include estimates of ξ_i for domain- and individual-specific error variances.

D Demographics

Table 13: **Summary Statistics for Demographic Variables**

Zurich 03		**Mean**	**Std. Err.**
Individuals	179		
female		0.430	0.037
semester		3.676	0.159
highincome		0.162	0.028
Zurich 06		**Mean**	**Std. Err.**
Individuals	118		
female		0.441	0.046
semester		3.551	0.240
highincome		0.051	0.020
Beijing 05		**Mean**	**Std. Err.**
Individuals	151		
female		0.483	0.041
semester		2.238	0.133
highincome		0.146	0.029

highincome: equals one if disposable income per month above is CHF 1,500 and CHN 1,000, respectively. Thresholds chosen by distributional considerations and relative students' hourly wages.

Table 14: **Effects of Socio-Economic Variables on Parameters**

	Gains			Losses		
Regressors	ZH 03	ZH 06	BJ 05	ZH 03	ZH 06	BJ 05
α/β						
constant	1.101**	0.935**	0.538**	1.075**	1.049**	1.553**
	(0.051)	(0.039)	(0.189)	(0.061)	(0.047)	(0.373)
female	-0.008	-0.041	-0.424	0.103	0.136	-0.347
	(0.042)	(0.044)	(0.325)	(0.069)	(0.069)	(0.351)
semester	-0.016	0.002	0.096	-0.009	-0.006	-0.095
	(0.012)	(0.006)	(0.091)	(0.013)	(0.008)	(0.106)
highincome	-0.024	-0.049	-0.436	0.078	0.064	-0.450
	(0.059)	(0.112)	(0.251)	(0.085)	(0.126)	(0.387)
γ						
constant	0.434**	0.562**	0.374**	0.472**	0.746**	0.454**
	(0.037)	(0.057)	(0.025)	(0.037)	(0.063)	(0.035)
female	**-0.143****	**-0.186****	**-0.113****	**-0.149****	**-0.324****	**-0.112****
	(0.022)	(0.057)	(0.031)	(0.026)	(0.054)	(0.036)
semester	0.031**	0.023	0.001	0.019	0.011*	0.001
	(0.012)	(0.010)	(0.009)	(0.011)	(0.005)	(0.015)
highincome	0.204**	-0.110	-0.007	0.002	-0.051	-0.046
	(0.079)	(0.098)	(0.034)	(0.071)	(0.070)	(0.033)
δ						
constant	0.848**	0.945**	1.295**	1.008**	0.990**	0.754**
	(0.051)	(0.042)	(0.125)	(0.068)	(0.047)	(0.176)
female	-0.147**	-0.134**	0.195	0.091	0.021	0.186
	(0.041)	(0.045)	(0.227)	(0.074)	(0.065)	(0.172)
semester	0.021	-0.001	-0.062	-0.001	0.008	0.038
	(0.013)	(0.006)	(0.063)	(0.014)	(0.006)	(0.053)
highincome	-0.072	-0.064	0.214	-0.059	-0.016	0.227
	(0.060)	(0.123)	(0.185)	(0.084)	(0.156)	(0.238)
$\ln L$	19,755	10,816	9,601			
Parameters	382	260	326			
Observations	8,906	4,669	4,225			

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications.
** Significant at 1%-level; * significant at 5%-level. ZH stands for Zurich, BJ for Beijing.

E Background Consumption

Figure 15: **Probability Weights Zurich 2003, *k endogenous***

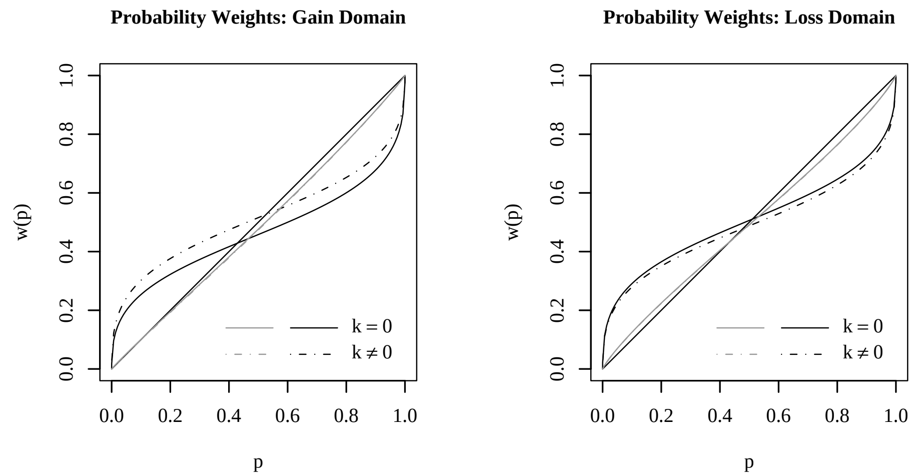


Figure 16: **Probability Weights Zurich 2006, *k endogenous***

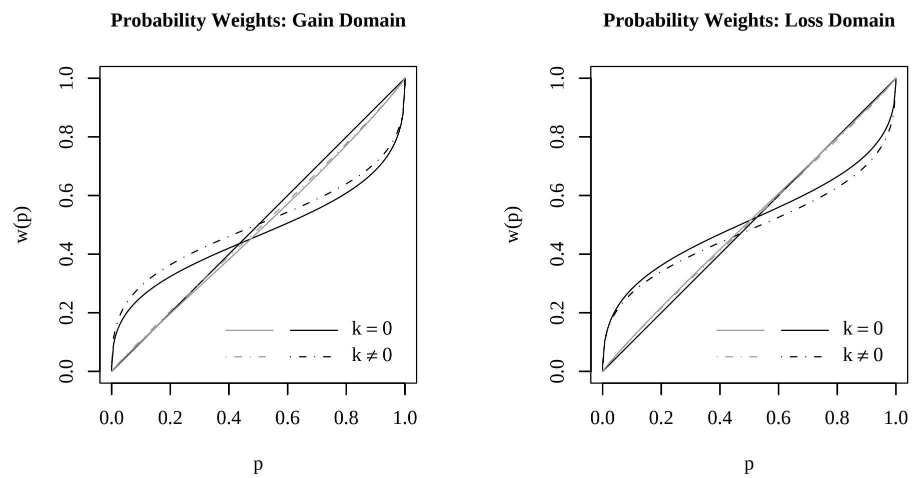


Figure 17: **Probability Weights Beijing 2005, *k endogenous***