Can Prospect Theory Be Used To Predict Investor's Willingness To Pay?*

Version: 2008-08-20

Carsten Erner
Finance Center Münster, DIA Research Group for Retirement Savings, University of Münster,
Universitätsstraße 14-16, 48143 Münster, Germany
Carsten.Erner@wiwi.uni-muenster.de

Alexander Klos
Finance Center Münster, DIA Research Group for Retirement Savings, University of Münster,
Universitätsstraße 14-16, 48143 Münster, Germany
Alexander.Klos@wiwi.uni-muenster.de

Thomas Langer
Finance Center Münster, DIA Research Group for Retirement Savings, University of Münster,
Universitätsstraße 14-16, 48143 Münster, Germany
Thomas.Langer@wiwi.uni-muenster.de

Abstract: Cumulative prospect theory (CPT) is widely considered to be the most successful descriptive theory for decision making under risk and uncertainty. Recently, sophisticated methods have been developed to reliably elicit CPT parameters on an individual basis (e.g., Abdellaoui, Bleichrodt, and Paraschiv 2007). The aim of this paper is to analyze whether such methods are suited to be applied in real world situations, in particular, in the context of investment counseling. Specifically, we examine whether CPT parameters elicited via standardized computer tools are successful in predicting the individual's willingness to pay (WTP) for different investment products. In a two-stage computerized experiment, we determine the CPT parameters for 200 subjects. We then elicit WTPs for various investment products and compare those to the WTPs that the subjects should have stated based on their individual CPT parameters. Surprisingly, we find hardly any predictive power of the elicited CPT parameters on the willingness to pay. We discuss several possible explanations for this finding, including domain specificity, competence effects, and decision error propagation in the elicitation of CPT parameters and WTPs. These explanations can account for at most only part of the low predictive power, which thus remains a puzzle. We therefore conclude that state-of-the-art methods of CPT parameter elicitation are not suited to be applied in a context of real world investment counseling.

Keywords: Cumulative Prospect Theory, Preference Elicitation, Error Propagation, Willingness to Pay, Experimental Finance, Structured Financial Product

JEL-classification: C91, D11, D81, G23

* We are indebted to Enrico Diecidue, Craig R. Fox and participants in the Finance Center Münster research seminar, the FUR XIII-Meeting in Barcelona, the SFB504-Seminar in Mannheim, the SPUDM21-Meeting in Warsaw, and the 2007 ESA World Meeting in Rome for valuable comments and insights. Our thanks also go to Christian Beer, Sebastian Dyck, Sven Nolte, and Tobias Pfaff for their assistance in running the experiments. Financial support from the DIA Research Group for Retirement Savings is gratefully acknowledged.

Can Prospect Theory Be Used To Predict Investor's Willingness To Pay?

Version: 2008-08-20

Abstract: Cumulative prospect theory (CPT) is widely considered to be the most successful descriptive theory for decision making under risk and uncertainty. Recently, sophisticated methods have been developed to reliably elicit CPT parameters on an individual basis (e.g., Abdellaoui, Bleichrodt, and Paraschiv 2007). The aim of this paper is to analyze whether such methods are suited to be applied in real world situations, in particular, in the context of investment counseling. Specifically, we examine whether CPT parameters elicited via standardized computer tools are successful in predicting the individual's willingness to pay (WTP) for different investment products. In a two-stage computerized experiment, we determine the CPT parameters for 200 subjects. We then elicit WTPs for various investment products and compare those to the WTPs that the subjects should have stated based on their individual CPT parameters. Surprisingly, we find hardly any predictive power of the elicited CPT parameters on the willingness to pay. We discuss several possible explanations for this finding, including domain specificity, competence effects, and decision error propagation in the elicitation of CPT parameters and WTPs. These explanations can account for at most only part of the low predictive power, which thus remains a puzzle. We therefore conclude that state-of-the-art methods of CPT parameter elicitation are not suited to be applied in a context of real world investment counseling.

Keywords: Cumulative Prospect Theory, Preference Elicitation, Error Propagation, Willingness to Pay, Experimental Finance, Structured Financial Product

JEL-classification: C91, D11, D81, G23

1. Introduction

Consider an investment advisor who sells financial products to retail investors. It is in her interest to recommend products that fit the risk preferences of her clients well. First, doing so might establish a comparative advantage over her competitors. Second, legal requirements such as those defined in the MiFID, the new Markets in Financial Instruments Directive (The European Parliament and the European Council 2004), may be fulfilled. Focusing on the second consideration, while the MiFID requires that banks collect "information on the length of time for which the client wishes to hold the investment, his preferences regarding risk taking, his risk profile, and the purposes of the investment." (Article 36(4), The European Parliament and the European Council 2006), the directive is silent about how such an ambitious goal can be achieved. The common practice among financial advisors is to use simple questionnaires to gain insight into their customer's preferences, with risk aversion and investment experience self-reported on a five- or seven-point scale.

From an academic perspective, such a procedure leaves much to be desired. Extent research shows that risk behavior follows complex patterns that cannot be described by a simple five-point scale. Furthermore, empirical evidence suggests that non-expected utility theories, such as Cumulative Prospect Theory (CPT) (Kahneman and Tversky 1979, Tversky and Kahneman 1992), explain risk behavior better than the traditional expected utility approach does (see Starmer 2000 for a general survey, and Camerer 2000 for a survey of field evidence on the descriptive validity of prospect theory). Accordingly, over the last few years, elaborate elicitation mechanisms for CPT have been developed (see, e.g., Wakker and Deneffe 1996, Abdellaoui 2000, Bleichrodt and Pinto 2000, and Abdellaoui, Bleichrodt, and Paraschiv 2007). These methods allow for a non-parametric elicitation of the value function and the probability weighting function. The purpose of this paper is to analyze whether such methods are suited to be applied in real world situations, in particular, in the context of investment counseling.

Specifically, we examine whether CPT parameters elicited via standardized computer tools, especially without time-consuming investor-advisor interactions, are successful in predicting the individual's willingness to pay (WTP) for different investment products. Our experiment can be considered an out-of-task evaluation of CPT. Out-of-task tests are of interest because they help determine the applicability of decision theories in practice. The academic community devotes great effort to the development of proper elicitation procedures. If such methods are to be applied, it is important to know when they work, that is, when the predictions based on preference measures correlate with actual behavior, and when they do not.

To pursue our research question, we bring the same subjects to the lab twice. In the first visit, we elicit CPT preference parameters on an individual basis by applying a modified version of the recently introduced elicitation procedure of Abdellaoui, Bleichrodt, and Paraschiv (2007). In the second visit, subjects state their willingness to pay (WTP) for various investment products with different risk profiles. Our analysis compares theoretical predictions based on the CPT parameters from the first part with the actually stated WTPs from the second part.

One reason to use the method of Abdellaoui, Bleichrodt, and Paraschiv (2007) is that it minimizes the influence of decision errors. Chained elicitation procedures propagate any decision error that is made. Blavatskyy (2006) has shown that the method of Abdellaoui, Bleichrodt, and Paraschiv (2007) is efficient with regard to error propagation. A second nice property of this method is that it is a completely parameter-free elicitation. While this is an interesting feature, which can make our elicitation data beneficial for a general analysis of the appropriateness of standard CPT assumptions, we do not use it for our main research goal, the prediction of WTPs. Instead, for the purpose of the prediction, we have to impose a parametric form. For the value function, we choose power utility, as it has been shown that this parametric form fits the data well (Abdellaoui, Bleichrodt, and Paraschiv 2007). For the probability weighting function, we choose the popular one-parameter form (see, e.g., Tversky and Kahneman 1992) and the two-parameter form (see, e.g., Gonzalez and Wu 1999).

To cover the full range of complex risk profiles, we use structured financial products as investment opportunities. Structured financial products allow issuers to form almost any type of tailor-made payout profile to serve clients with specific preference structures. This goal is achieved by combining an underlying, typically a stock or a stock index, with one or more options on that underlying. CPT, designed to capture complex patterns of risk behavior, should be especially suited to explain differences in the WTPs of these financial products.

The remainder of the paper is structured as follows. In Section 2, we introduce the procedure that is used in Part I of the study to elicit individual CPT parameters. In Section 3, we describe the structured products that we use in Part II of the study to determine WTPs. The overall experimental design is described in Section 4. In Section 5, we present as the main finding the result that the CPT parameters elicited in the first part have virtually no predictive power in forecasting the WTPs in the second part: for the majority of products, the predicted WTPs are not significantly correlated with the stated WTPs. Given this result, we examine several possible reasons for this finding. Some of these explanations (domain specificity, competence effects, and decision errors) can be analyzed based on our data. All of these reasons can explain at most only part of the low predictive power, which therefore remains a puzzle. We summarize our insights in Section 6, concluding that state-of-the-art methods of prospect theory parameter elicitation are not suited to be applied in a real world investment counseling context.

2. Cumulative Prospect Theory

2.1 Overview

Cumulative prospect theory is widely considered to be the most successful descriptive theory for decision making under risk and uncertainty (Kahneman and Tversky 1979, Tversky and Kahneman 1992). Distinguishing between gains and losses and thereby applying reference dependency is a main feature. With respect to this distinction, risk attitude within the CPT framework comprises three components, namely basic utility (curvature), probability weighting, and loss aversion (e.g., Köbberling and Wakker 2005, Wu and Markle 2008). Cumulative prospect theory enables the valuation of n-outcome lotteries. We treat the payoff structures of the investment products as discrete distributions. They can therefore be interpreted as n-outcome prospects.

Consider a prospect P with n outcomes x_i with probability p_i: $P = (x_1, p_1; \dots; x_n, p_n)$. In our case, the outcome represents the gain or loss yielded from an investment in the product. More formally, the outcome is defined by the difference of the product's payoff at maturity $\left(P_{Product,T,i}\right)$ and the product price at the time of the investment $\left(P_{Product,t}\right)$: $x_{Product,T,i} = P_{Product,T,i} - P_{Product,t}$. The monotonicity of the product's payoff function ensures ascending order of the outcomes ($x_n \geq \cdots \geq x_{k+1} \geq 0 \geq x_k \geq \cdots \geq x_1$), with $n-(k+1)$ gains and k losses. The reference point is assumed to be at a zero outcome, that is, one or more state(s) where the product payoff offsets the product price. Defining $v(\cdot)$ as the value function and $w(\cdot)$ as the probability weighting function, the CPT utility of a prospect (and hence of an investment product) is given by:

$$CPT(Product) = \sum_{i=1}^{k} \left[w^{-}\left(\sum_{j=1}^{i} p_j \right) - w^{-}\left(\sum_{j=1}^{i-1} p_j \right) \right] \cdot v^{-}(x_i)$$

$$+ \sum_{i=k+1}^{n} \left[w^{+}\left(\sum_{j=i}^{n} p_j \right) - w^{+}\left(\sum_{j=i+1}^{n} p_j \right) \right] \cdot v^{+}(x_i).$$

2.2 Elicitation Procedure

Several elicitation procedures have been put forward in the literature (see, e.g., Blavatskyy 2006 for an overview). We use a modified version of the recently presented non-parametric procedure for value function elicitation by Abdellaoui, Bleichrodt, and Paraschiv (2007). Blavatskyy (2006) proposes a three-stage approach, which is efficient with regard to the propagation of decision errors in chained elicitation procedures. The method of Abdellaoui, Bleichrodt, and Paraschiv (2007) is consistent with the first two stages and thus minimizes the influence of decision errors. We extend the approach of Abdellaoui, Bleichrodt, and Paraschiv (2007) by adding stage three of Blavatskyy (2006) to elicit different probabilities with given decision weights. Figure 1 provides an overview of the three stages of Blavatskyy (2006) and the four original steps of Abdellaoui, Bleichrodt, and Paraschiv (2007) together with steps 5 and 6 of our extension.

Figure 1: Structure of the Elicitation Procedure

Step 1 probabilities for $w^{\pm}(p^{\pm}_{0.5})=0.5$	*Step 2* value function for losses	*Step 3* linking gains and losses	*Step 4* value function for gains	*Step 5* probabilitiy weighting function for losses	*Step 6* probabilitiy weighting function for gains
Stage 1	*Stage 2*			*Stage 3*	

Note: Shown are the steps and stages of the elicitation procedure. Steps 1-4 are taken from Abdellaoui, Bleichrodt, and Paraschiv (2007). Steps 5 and 6 represent our extension with regard to eliciting the probability weighting function. The three stages reflect the connection to the three-stage approach by Blavatskyy (2006).

In step 1, we seek the probabilities $p^-_{0.5}$ and $p^+_{0.5}$ that give a decision weight of one-half: $w^-(p^-_{0.5}) = 0.5$ and $w^+(p^+_{0.5}) = 0.5$. These decision weights are elicited using the tradeoff method with $(L_Y, p'; L_M) \sim (L_X, p'; L_N)$ and $(L_Z, p'; L_M) \sim (L_Y, p'; L_N)$ followed by a probability-equivalent query $(L_Y) \sim (L_Z, p^-_{0.5}; L_X)$. L denotes a loss outcome with $L_M; L_N; L_X = -1; -6; -10$ and $p' = \frac{1}{3}$. The same is done for the gain domain. In step 2, the loss value function is elicited through chained certainty-equivalent queries using utility midpoints. L_1 as the maximum loss outcome is fixed to be $L_1 = -100$ €, with $U(L_1) = -1$. Amounts are chosen to be within the interval of $[0; -100]$ since the gains and losses from the products are similarly scaled. A sequence of outcomes is elicited through queries $L_r \sim (L_A, p^-_{0.5}; L_B)$ with $U(L_r) = -r = 0.5 \cdot U(L_A) + 0.5 \cdot U(L_B)$. We derive fourteen points with $r \epsilon \left\{\frac{1}{64}, \frac{1}{32}, \frac{1}{16}, \frac{3}{32}, \frac{1}{8}, \frac{3}{16}, \frac{7}{32}, \frac{15}{64}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{7}{8}\right\}$ and $m = 6$ iterations (see Appendix A). This is a slightly amended array compared to the original sequence, as it provides more utility points for the probability elicitations described below. In step 3, losses and gains are linked using three indifference statements $L_{0.25} \sim (l, 0.5; 0)$, $0 \sim (g, 0.5; l)$, and $G_{0.25} \sim (g, 0.5; 0)$, with G denoting a gain outcome. We follow Abdellaoui, Bleichrodt, and Paraschiv (2007) and use $r = 0.25$ as the linkage point. In step 4, eight points for the gain value function are elicited using the same sequence for r as for the losses, but with the constraint that $r < 0.25$. In steps 5 and 6, we implement the probability elicitations according to stage three of Blavatskyy (2006) and employ the probability-equivalent method to determine $w^-(p^-_s) = s$ using $L_B \sim (L_A, p^-_s; 0)$, with $U(L_B) = s \cdot U(L_A)$ and $s \epsilon \left\{\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}\right\}$, for losses. The same applies for the gain domain. In total, the procedure utilizes 43 indifference statements to elicit both value and probability weighting functions.

2.3 Functional Forms and Fitting

To derive predictions over the whole outcome domain, we apply different functional forms for the value function and the probability weighting function and fit them using non-linear least squares estimations.

Value Function. Abdellaoui, Bleichrodt, and Paraschiv (2007) report that the results for different functional specifications like the exponential family or the expo-power family do not substantially deviate from the results for the standard power family. Based upon this finding, we use the two-part power function of Tversky and Kahneman (1992): $v^+(G) = G^{\alpha+}$ for gains and $v^-(L) = -k \cdot (-L)^{\alpha-}$ for losses. We allow the curvature parameter α to be different for gains and for losses. Scaling of the value axis is essentially arbitrary. The elicitation method therefore employs normalized maximum values of $v_{norm}^-(L_1) = -1$ and $v_{norm}^+(G_{0.25}) = 0.25$. To match this normalization, two scaling factors s need to be applied: $v_{norm}^+(G) = s^+ \cdot G^{\alpha+}$ and $v_{norm}^-(L) = -s^- \cdot (-L)^{\alpha-}$. Linking gains and losses through these scaling factors leads to the loss aversion coefficient k. Summarizing, we apply the following basic structure:

$$v^+(G) = \frac{v_{norm}^+(G)}{s^+} = G^{\alpha+} \text{ and } v^-(L) = \frac{v_{norm}^-(L)}{s^+} = -k \cdot (-L)^{\alpha-} \text{ with } k = \frac{s^-}{s^+}.$$

With regard to the scaling factors, we use two different approaches. The first one (called VF1) follows Abdellaoui, Bleichrodt and Paraschiv (2007). The scaling factors are pre-determined by normalizing the outcomes for gains and losses separately: $G_{norm} = \frac{G}{G_{0.25}}, G\epsilon[0, G_{0.25}]$ and $L_{norm} = -\frac{L}{L_1}, L\epsilon[-L_1, 0]$. Next, $v_{norm}^+(G) = 0.25 \cdot (G_{norm})^{\alpha+}$ for gains and $v_{norm}^-(L) = -(-L_{norm})^{\alpha-}$ for losses are fitted. Inserting the normalization rules, the scaling factors can be derived directly:

$$v_{norm}^+(G) = \underbrace{\frac{1}{4 \cdot (G_{0.25})^{\alpha+}}}_{s^+} \cdot (G)^{\alpha+} \text{ and } v_{norm}^-(L) = -\underbrace{\frac{1}{(-L_1)^{\alpha-}}}_{s^-} \cdot (-L)^{\alpha-}.$$

The normalization forces the fitted function to run through the outer points $(L_1; -1)$ and $(G_{0.25}; 0.25)$, which is a restriction that need not necessarily be applied. A more flexible approach should lead to better fit results. The second approach (called VF2) is identical to VF1 with the exception that $v_{norm}^{+}(G) = s^{+} \cdot G^{\alpha+}$ and $v_{norm}^{-}(L) = -s^{-} \cdot (-L)^{\alpha-}$ are directly fitted. Hence, the scaling factors can be flexibly chosen to maximize the fit.

Probability Weighting Function. Various different parametric forms have been discussed in the literature (see, e.g., Bleichrodt and Pinto 2000 for an overview). In this study, we apply the two most common forms. Tversky and Kahneman (1992) suggest a one-parameter approach (called PWF1):

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}.$$

A two-parametric form (called PWF2) is proposed by Gonzalez and Wu (1999), based on previous work by Goldstein and Einhorn (1987) and Lattimore, Baker, and Witte (1992). Curvature (discriminability) is controlled by γ and elevation (attractiveness) is controlled by δ:

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

Similar to the value function, we allow the parameters γ and δ, both for PWF1 and PWF2, to be different across the gain and the loss domains.

3. Investment Products

3.1 Overview

Even retail investors no longer engage only in plain stock or bond positions, but rather exploit the differentiated risk/return profiles offered through the usage of derivatives. Accordingly, investment companies have enlarged their product offerings in the area of financial engineering, bundling stocks with derivatives and selling them as separate securities. Since the market in Germany for such products has grown dynamically and is little regulated, there exists no comprehensive or standardized product classification.[1] As of the end of 2007, the open interest, including exchange-based as well as over-the-counter transactions, was

[1] Terms like “structured financial product”, “investment certificate”, “retail derivative”, or “securitized derivative” are used depending on the special context (see, e.g., Stoimenov and Wilkens 2005 or Muck 2006).

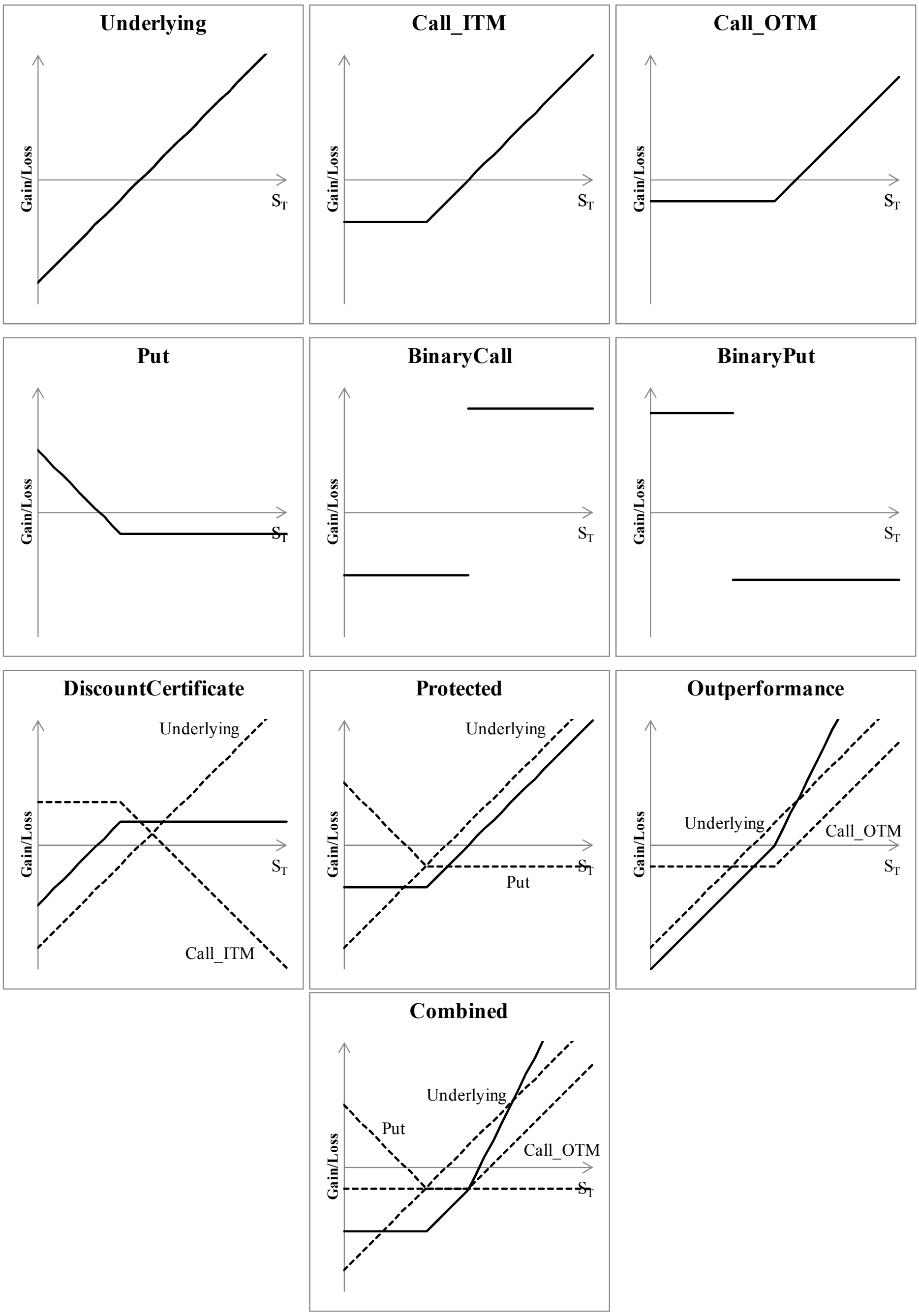
estimated to be approximately 135.1 billion €, rising by roughly 17 % compared to 2006 (Deutscher Derivate Verband 2008). While Europe and especially Germany are already large markets, there is also a growing U.S. market, which shows comparable growth rates to the German market and is also reflected in the literature. For example, Henderson and Pearson (2007) examine the payoff profiles of structured products in the U.S. market, while Bethel and Ferrel (2007) discuss emerging policy issues.

When directly asked about their reasons for investing in structured investment products, people mainly bring up “rational” motives like diversification or hedging (Fischer 2007). Branger and Breuer (2008) show, however, that the utility gain for CRRA-investors with different levels of risk aversion is, if positive at all, low and not sufficient to explain the empirically observed high demand for such products. This result points to a behavioral explanation. Shefrin and Statman (1993) illustrate how the relative attractiveness of different financial products with otherwise identical cash flows might be driven by different framings in a prospect theory framework. Breuer and Perst (2007) apply cumulative prospect theory and hedonic framing to evaluate two types of structured investment products and examine the dependencies of the products’ demands from different investors’ preferences. Their focus lies in describing parameter combinations that make a structured financial product attractive for a CPT investor.

3.2 List of Products

Structured investment products comprise a combination of a long position in the stock with one or more long or short option position(s). For valuation purposes, therefore, one can apply the pricing by duplication principle, where, in an arbitrage-free market, the value of a product equals the sum of the values of its components. In our experiment, we look at ten different investment products. The products are chosen to reflect a variety of different risk/return profiles, thereby enabling analyses of different preference structures. Gain/loss profiles of all ten products are shown in Figure 2.

Figure 2: Gain/Loss Profiles of all Ten Investment Products.



Note: This figure shows the gain/loss profiles (at maturity) of all ten investment products. Gains and losses are calculated relative to the products' arbitrage-free fair values (Black and Scholes (1973) framework). For the structured investment products, the diagram also shows the duplication strategies, where the components are marked with dotted lines while the composite product is displayed as a solid line.

As an apparent origin, the underlying itself is offered as an investment product. The product's payoff (P) at maturity (T) is given by the price of the underlying (S_T): $P_{Underlying,T} = S_T$. Next, plain vanilla options are considered. Two different strike prices are needed for the structured investment products; one lying above the current underlying price $\left(K_{high} > S_T\right)$ and one lying below $(K_{low} < S_T)$. An in-the-money (out-of-the-money) call delivers a payoff if the underlying price ends up above the low (high) strike price at maturity: $P_{Call\_ITM/Call\_OTM,T} = max\left(S_T - K_{low/high}; 0\right)$. Similarly, an out-of-the-money put option delivers a payoff if the underlying price lies below the low strike price: $P_{Put,T} = max(K_{low} - S_T; 0)$. An out-of-the-money binary call (binary put) pays the amount X if the underlying price ends up above (below) the strike price: $P_{BinaryCall,T} = \begin{cases} X, if\ S_T > K_{high} \\ 0, otherwise \end{cases}$ and $P_{BinaryPut,T} = \begin{cases} X, if\ S_T < K_{low} \\ 0, otherwise \end{cases}$, respectively.[2]

A discount certificate exhibits a covered call strategy, where a long position in the underlying is combined with a short call. Shorting the call leads to a capped payoff profile, limiting the maximum payoff to the strike price of the option: $P_{DiscountCertificate,T} = S_T - P_{Call\_ITM,T}$. Next, the protected product exhibits a protective put strategy, combining a long position in the underlying with a long position in a put option, thereby guaranteeing a minimum payoff of the strike price: $P_{Protected,T} = S_T + P_{Put,T}$. Double participation of any underlying price movement above the strike price is offered by an outperformance product that combines a long position in the underlying with a long position in an out-of-the-money call: $P_{Outperformance,T} = S_T + P_{Call\_OTM,T}$. A combined product joins the features of the protected and the outperformance product. Downside protection as well as double participation is provided through the combination of a long underlying position with a long put and a long out-of-the-money call: $P_{Combined,T} = S_T + P_{Put,T} + P_{Call\_OTM,T}$.

[2] Except for the in-the-money plain vanilla call options, the other options (plain vanilla puts, binary calls, binary puts) are only used in their out-of-the-money variants, which is why we omit the distinction "ITM/OTM" for them.

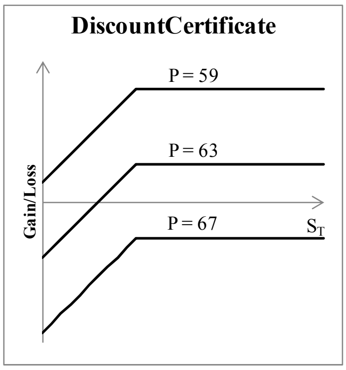
Several distribution-free arbitrage constraints referring to the relation of the values among the different products might be derived from the duplication principle. Those constraints not only hold at maturity, but at any given point in time (t). Consistency of the stated WTPs might be explored using those arbitrage constraints (see Appendix B for a list).

3.3 Deriving Predictions for WTPs

Cumulative prospect theory non-linearly aggregates the different preference sources, thereby hindering separate analyses of their individual influences. As a consequence, CPT prediction can only be based on a combined measure. We therefore derive the marginal product price, that is, the willingness to pay at which the investor is indifferent between buying and not buying the investment product. Indifference is reflected by a CPT value of zero.

More technically, the marginal willingness to pay $(WTP^{CPT=0})$ can be derived by choosing the product price such that $CPT\left(Product \middle| P_{Product,t} = WTP^{CPT=0}\right) = 0$. This method is illustrated in Figure 3. Shown are the gain/loss profiles for a discount certificate with three different product prices $P_{Product,t}$. For example, the discount certificate generates losses in any state for $P_{Product,t} = 67$. Accordingly, $CPT\left(Product \middle| P_{Product,t} = 67\right)$ is negative, regardless of the preference parameter combination. Decreasing $P_{Product,t}$ means shifting the gain/loss profile upwards and thereby generating a gain/loss profile that stochastically dominates the distribution with the higher $P_{Product,t}$. Since CPT ensures monotonicity in comparing statistically dominated distributions, there exists a unique product price for which the condition $CPT\left(Product \middle| P_{Product,t} = WTP^{CPT=0}\right) = 0$ holds. A further upward shift of the distribution via choosing a product price below $WTP^{CPT=0}$ would make $CPT\left(Product \middle| P_{Product,t}\right)$ strictly positive. $WTP^{CPT=0}$ therefore defines the predicted marginal WTP for the considered CPT agent. A bisection approach is used to determine this number.

Figure 3: Gain/loss Profiles for different Product Prices.



Note: Shown are the gain/loss profiles of a discount certificate with three different product prices $P_{DiscountCertificate,t}$. A change in the product price vertically shifts the complete profile upwards or downwards.

4. Experimental Design

4.1 Overview

The experiment was split into two parts, which were run separately and tested in pilot sessions prior to the actual runs. The elicitation of CPT parameters according to the aforementioned elicitation procedure was conducted in part I. Part II comprised the presentation of the investment products together with the elicitation of the willingness to pay. Breaking the experiment into two parts splits the duration of the overall experiment, facilitating better concentration by the participants. It was made clear that there are no right or wrong answers and that the participants could ask for help at any time during the experiment. Both parts were fully computer-based including the presentation of the instructions.[3] Demographic questionnaires concluded the experiment. Exactly 200 subjects completed both parts of the experiment. Participants were advanced students from the business school of the University of Münster, Germany. The median completed years of study was three and 23 % of the participants were female.

[3] By clicking “Help”, the instructions could be accessed at any time during the experiment. Instructions are available from the authors upon request. Part I also contained further subtasks that are not analyzed in this study.

4.2 Part I

As we point out above, the elicitation procedure contained 43 indifference statements. Following Abdellaoui, Bleichrodt, and Paraschiv (2007), we did not directly ask the participants to state the outcome or probability that would lead to indifference. Instead, a series of binary choices was presented, converging towards indifference by a bisection approach. This was done to avoid the well-known problems associated with the direct request for an indifference value (see, e.g., Luce 2000).

With regard to the interface, we applied a modified version of the original method of Abdellaoui, Bleichrodt, and Paraschiv (2007). Prospects were presented via pie charts, with the size of the pieces reflecting the prospect's respective probabilities. In addition, exact information about probabilities and outcomes was presented within the pie. Each indifference query was elicited by three binary choices, each representing an iteration step in a bisection procedure. Participants had to choose which one of the prospects they preferred by clicking on it (alternative A in Appendix C, Panel I – top screenshot). The value of the variable was then changed according to the bisection procedure. Consequently, the interval for the variable narrowed in a stepwise fashion.[4]

After the bisection procedure, a final matching task was presented. In this step, a slider could be used to exactly choose the indifference value within the previously selected interval (alternative B in Appendix C, Panel I – bottom screenshot). In this final screen, the participant could mark a checkbox called "No indifference possible". In such cases we concluded that a response error occurred in an earlier step so the procedure started anew for this indifference query. Subjects received 30 € flat for their participation in part I. We did not use an incentive-compatible payment mechanism in this part. This was done both for opera-

[4] For some of the intermediate steps with unbounded intervals, a different approach had to be taken. A sequence of choices with increasing outcomes was presented (up to a maximum of 15 choices) until the subject changed her preference and thereby established a bounded interval. Then the bisection process started. In the case of no reversal, the last number was taken as the variable value. For both approaches, the initial value of the variable was determined to equal the expected values of the prospects. The alignment of the prospects to the left or to the right side was kept constant within one indifference query, but was randomly altered over queries and subjects.

tional reasons (many loss lotteries were included in the procedure), and for reasons of external validity. In real life, bank clients will probably not be paid for providing information about their preferences in regular advisory talks.

4.3 Part II

A list of ten investment products was presented. All of the products were based on the German stock index DAX as the underlying. Therefore, the historical discrete return distribution of the DAX was explained first, using a discrete histogram. The histogram was set up using the last 120 monthly returns, calculated back from the actual day of the experiment. Monthly returns were taken, as the products' time-to-maturity was defined to be one month. To increase comprehensibility we aggregated the single returns and formed 25 return classes. Payment for part II was incentive-compatible using the method of Becker, DeGroot, and Marschak (1964) (BDM), which was explained in detail. Each product was comprehensively described (see Appendix C, Panel II – top screenshot). At the top of the screen, a verbal explanation was given. The diagram consisted of the historical DAX return distribution in the back and the payoff profile of the product as well as the one of the underlying. Below the diagram, there was a slider with which the current level of the underlying could be moved within the interval given by the historical distribution. By moving the level of the underlying, the numbers in the scenario calculator to the right of the diagram, showing expected returns from the product, were adjusted. Participants could use the slider to get acquainted with the payoff structure. No time restrictions were applied. To test for the understanding, test quizzes were asked after the explanation of the histograms, after the BDM mechanism, and after each product. Participants could only advance with the experiment after correctly answering all quizzes. In the event of difficulties, the experimenter might have been asked for help.

It is clear that the valuation of an investment product depends on the expected distribution of the risky underlying. With regard to these expectations, we implemented two different scenarios. In scenario "real", the actual development of the DAX in the next four weeks, counting from the day of the experiment, was used. The subjective expectation about the future development of the underlying can of course differ for each subject. Therefore, to judge the stated WTPs, we had to elicit the individual expectation. We did so using a three-point elicitation (Pearson and Tukey 1965, Keefer and Bodily 1983) (see Appendix C, Panel II – middle screenshot). For this scenario, “real”, preferences might be influenced by ambiguity about the unknown probabilities. To avoid distortions due to ambiguity, we additionally run a scenario “fictional” that was based on the historical return distribution with known probabilities. Here, we told the subjects that we will randomly draw one of the historical outcomes, so subjects had unambiguous information about the possible future outcomes and their likelihood of occurrence. Shortly before each experiment run, the current DAX level was imported, ensuring no participant had an informational advantage about the current DAX level.[5]

The main objective of Part II was to elicit the willingness to pay for the investment products. For each of the two scenarios, all ten products were elicited, giving a total of twenty WTPs per subject. Similar to the preference elicitation, a combination of binary choice tasks based on a bisection procedure together with a final matching task was implemented. Each elicitation started with a product price of 0.01 € to make sure that subjects did not think the starting price conveys any information about reasonable prices. In each step, the subject was asked whether she is willing to buy the product at the current price. If the answer was yes, the price increased and the subject was asked again until she either declined to buy at the current price or a maximum price of 200 € was reached. Then, the bisection procedure started with the number of iterations depending on the size of the interval. In the final step, a slider could be used again to marginally adjust the product price. Similar to the preference elicitation, a checkbox “No indifference possible” could be marked, resulting in a restart of the procedure. The interface showed the gain and loss profile of the investment product based on the currently asked price (see Appendix C, Panel II – bottom screenshot). The presentation order of the two scenarios as well as of the ten products was completely randomized.

[5] The strike prices of the products were defined to be 2 % above (high strike) or below (low strike) the current DAX level, rounded to 50 cents.

Payment for Part II was incentive-compatible. One out of the twenty elicited WTPs was randomly determined for every subject. Payment was then based on this product. The randomly drawn product price had to be paid in exchange for the payoff of this investment product. The gain or loss from this investment was offset against the flat payment from the first part of the study. We assume that decision makers did not integrate the payments from both scenarios, so this procedure should have induced loss aversion even though hardly any overall real losses could occur by this mechanism.[6]

5. Results

5.1 Prediction Quality

To calculate the CPT value of an investment product, the possible states of the underlying need to be known. For scenario “fictional”, the 25 possible states were fixed and given to the subjects. A different approach had to be taken for scenario “real”, where no return distribution was given and where the subjects’ product valuation depended on their subjective expectations. We therefore elicited the individual expectations (mean and standard deviation) through a three-point approximation for continuous random variables (see Appendix D). Using the individually elicited mean and standard deviation, the subjectively expected return distribution was then approximated with a 75-state binomial tree (Cox, Ross, and Rubinstein 1979).[7] On an annualized basis, the subjective expectations about the return for scenario “real” were higher (mean: 0.1066, median: 0.1160, sd: 0.3264) in comparison to the historically determined mean return that was used in scenario “fictional” (mean: 0.0855). In contrast, the expected standard deviation (mean: 0.1911, median: 0.1933, sd: 0.0808) was smaller than the historical one (mean: 0.2534).

[6] And in fact, no subject finished the experiment with an overall loss. The mean (median) payment was 31.13 € (30.00 €) for both parts (including the 30 € flat payment for Part I), the minimum (maximum) payment was 24.11 € (42.26 €).

[7] The number of 75 terminal states is a reasonable choice (Hull 2008). Simulating more states, thereby generating smaller probabilities for the boundary states, could lead to precision problems in the probability weighting.

While stating WTPs, a subject could violate first-order stochastic dominance (FOSD) if she stated a WTP that was lower (higher) than the minimum (maximum) product payoff. For every subject i, scenario, and product combination, the respective data is flagged as violating FOSD if $WTP_{i,scenario,product}^{stated} \leq P_{i,scenario,product,T}^{min}$ or $WTP_{i,scenario,product}^{stated} \geq P_{i,scenario,product,T}^{max}$ does not hold. If one of these relations is violated for a product in scenario "fictional" the respective data in scenario "real" for the same subject is also flagged, and vice versa. A subject showing a violation for a certain product is assumed to have difficulty in understanding that special product and not with stating WTPs in general. We therefore do not exclude the subject completely, but only the data for the product with the violation. In total, 610 out of 2.880 records (21.18 %) are excluded.[8] Further dominance violations could be identified if one compares the stated WTPs for different products. Distribution-free constraints (see Appendix B) must hold. However, because these conditions can only be verified "across decision situations", we refrain from excluding such cases (Vossmann and Weber 2005).

Different measures are applied to assess the predictive quality. First, simple OLS regression analyses are run for every scenario-product combination with the stated WTP as the dependent variable and the predicted marginal WTP as the independent variable. More formally: $WTP_{i,scenario,product}^{stated} = \alpha_{scenario,product} + \beta_{scenario,product} \cdot WTP_{i,scenario,product}^{CPT=0}$. Perfect prediction would result in an intercept of $\alpha_{scenario,product} = 0$ and a slope coefficient of $\beta_{scenario,product} = 1$. Second, the correlation between the stated and predicted WTPs is measured by the Spearman rank correlation coefficient $(\rho_{scenario,product})$. A high ρ would indicate good predictive power.

[8] After exclusion of FOSD violations during the elicitation phase (see 5.2), 144 subjects remained in the sample for the predictions, leading to 2.880 stated WTPs (144 subjects, 2 scenarios, 10 products).

Table 1: Quality of Prediction – Overview.

	Regression (R^2)		Spearman (ρ)	
	fictional	real	fictional	real
VF1 / PWF1	0.0063 (0.0081)	0.0139 (0.0941)	0.0053 (0.0396)	0.0883 (0.1085)
VF1 / PWF2	0.0048 (0.0104)	0.0151 (0.0734)	0.0254 (0.0399)	0.0867 (0.0892)
VF2 / PWF1	0.0048 (0.0073)	0.0130 (0.0873)	0.0106 (0.0438)	0.0842 (0.1068)
VF2 / PWF2	0.0039 (0.0096)	0.0130 (0.0681)	0.0312 (0.0472)	0.0783 (0.0887)

Note: Reported are the median values (mean values in parentheses) over all ten products for the regression analysis (R^2) and the correlation analysis (ρ).

The very low R^2 and the similarly low Spearman correlation coefficients, reported in Table 1, show that $WTP_{CPT=0}$ has very poor predictive power. Random guessing would result in $R^2 = 0$ and $\rho = 0$, which is almost the case (especially in scenario "fictional"). Although it does not make any difference at this low level of prediction, the four different fitting approaches do not provide systematically different prediction results. Since the VF2 / PWF2 approach shows the best fit and therefore should in principle lead to the best forecasts, only the numbers for this approach are examined and reported in more detail. Table 2 presents the results for each product. Highly significant predictions, according to both measures, only occur for the group of discount certificates. It is difficult to find a reasonable explanation for this exception. Interestingly, negative slope coefficients obtain in three cases.

Table 2: Quality of Prediction – By product.

	fictional vs. real	Regression (R^2) fictional	Regression (R^2) real	Spearman (ρ) fictional	Spearman (ρ) real
Underlying (N = 105)	0.448	0.0006 (0.06) [66.17/-0.022]	0.0050 (0.52) [59.21/0.084]	0.0525 (0.60)	0.0777 (0.43)
BinaryCall (N = 133)	0.979	0.0010 (0.13) [3.11/0.022]	0.0046 (0.61) [2.81/0.048]	-0.0051 (0.95)	0.0620 (0.48)
BinaryPut (N = 131)	3.457 ***	0.0049 (0.01) [2.21/0.005]	0.0203 (2.67) [1.50/0.082]	-0.0075 (0.93)	0.1004 (0.25)
Call_ITM (N = 136)	1.033	0.0019 (0.26) [2.32/0.026]	0.3605 (75.52) *** [0.48/0.848]	-0.0436 (0.61)	-0.0637 (0.46)
Call_OTM (N = 130)	0.411	0.0184 (2.40) [1.85/-0.112]	0.0027 (0.35) [1.74/0.067]	-0.0803 (0.36)	-0.0752 (0.39)
Put_OTM (N = 124)	3.870 ***	0.0073 (0.90) [1.49/0.052]	0.0056 (0.69) [1.13/0.068]	0.0279 (0.76)	0.0712 (0.43)
DiscountCertificate (N = 93)	-1.535	0.0262 (2.45) [49.31/0.189]	0.1658 (18.08) *** [31.27/0.482]	0.2179 (0.04) **	0.3599 (0.00) ***
Protected (N = 87)	0.349	0.0002 (0.01) [67.18/0.008]	0.0813 (7.52) *** [46.17/0.322]	0.0802 (0.46)	0.1024 (0.35)
Outperformance (N = 107)	-1.259	0.0344 (3.74) * [54.14/0.167]	0.0323 (3.50) * [51.73/0.203]	0.1953 (0.04) **	0.1738 (0.07) *
Combined (N = 89)	0.898	0.0060 (0.52) [65.84/0.031]	0.0025 (0.22) [68.79/-0.015]	0.0345 (0.75)	0.0790 (0.46)

Note: Reported are the results for the regression analysis (R^2) and the correlation analysis (ρ) by product. Below the results are the F-values (Regression) / p-values (Spearman) in parentheses. For the regressions, the coefficients $[\alpha_{task,product}/\beta_{task,product}]$ are also reported. The column "fictional vs. real" shows the results of Wilcoxon matched-pair signed-rank tests (z-value) for testing the equality of the stated WTPs between the two scenarios (two-sided). *, **, and *** denote significance at the 1 %, 5 %, and 10 % level, respectively.

Comparing scenario "fictional" with scenario "real", the predictions for the latter seem to be slightly better. Wilcoxon matched-pair signed-rank tests reveal that the stated WTPs do not significantly differ between the two scenarios, except for only two products. These results taken together lead to the conjecture that the subjects did not fully ignore their subjective expectations when valuing the products for scenario "fictional". Hence, the experimental control over expectations probably did not work (see Loewenstein 1999 for a discussion of the doctrine of context-free experiments in economics).

For the research question we address, the results for the median investor are less important as we want to examine prediction quality at the individual level. However, taking a look at the median investor might reveal some interesting insights. Taking the individual results as a benchmark, the predictions of $WTP_{median}^{CPT=0}$ in forecasting WTP_{median}^{stated} are slightly better (see Appendix E). However, especially for the single options, the aggregated view shows a high degree of deviation. In scenario "fictional" ("real"), $WTP_{median}^{CPT=0}$ underestimates WTP_{median}^{stated} in six (five) cases, and thus does not follow any systematic pattern regarding under-/overestimation. The median values exhibit internally consistent WTPs, as all arbitrage constraints for the product values hold.

So far, we have shown that CPT does not perform well in predicting the WTP for the task at hand. We now turn to a discussion on the possible factors influencing the low predictive power. In particular, we examine two groups of explanations, one that we can address with our data and the other that can only be discussed theoretically.

5.2 Preference Parameters

We analyze the elicited preference parameters. A high number of inconsistencies or a low fitting quality of our functional forms might lead to CPT parameters that do not reflect the subjects' preferences and thereby play a role in the low prediction quality.

We begin by examining the degree of error that subjects made during the elicitation, measured by the number of first-order stochastic dominance violations. Since both the value function and the probability weighting function are monotonic, FOSD violations occur if the monotonicity for the sequence of elicited outcomes and probabilities does not hold. Considering the value function (step 2 and step 4), we flag a subject as violating FOSD if $L_1 < L_{7/8} < \cdots < L_0$ for losses or $G_{1/4} > G_{15/64} > \cdots > G_0$ for gains is violated. Since the values of L_r and G_r are elicited subsequently, violations for the value function can only occur "within decision situations". This is not valid for the probability weighting function (step 5 and step 6), where the probabilities are elicited independent of each other, meaning that the monotonicity can only be violated "across decision situations" (Vossmann and Weber 2005). We therefore follow the literature and only flag those subjects as violating FOSD for the probability weighting function, where $0 < p_s < 1 \ \forall \ s$ does not hold for gains or losses.[9]

In total, 52 subjects (26 %) have to be excluded due to FOSD violations. Additionally, four subjects have to be excluded due to technical reasons (no convergence of the fitting algorithm). These filters lead to a final data set consisting of 144 subjects. A high number of inconsistencies is not uncommon in individual decision making. For example, Camerer (1989) conducts an individual decision making experiment where subjects are confronted with binary choices between lotteries. To measure the degree of preference inconsistencies, some of the choices are answered repeatedly. If a subject were just choosing randomly, one would expect that 50% of all choices would be inconsistent. Camerer (1989) reports that 31.6 % of all answers are inconsistent, pointing to a high degree of noise. Starmer and Sugden (1989) and Wu (1994) find similar error rates. Other elicitation studies mention inconsistency problems, but do not report numbers (e.g., Bleichrodt and Pinto 2000).

Next we consider the effect of fitting quality. Note that for both the value function and the probability weighting function, the estimation procedure leads to a high R^2_{adj} (median over individual R^2_{adj}). For the value function we get 0.9891 (losses), and 0.9917 (gains) for VF2, and 0.9871 (losses), and 0.9896 (gains) for VF1. The fitting for VF2 is significantly better than that for VF1 at the 1 % level (two-sided Wilcoxon matched-pair signed-rank test), supporting our approach for the parametric function. For the probability weighting function the fitting for PWF2 (median R^2_{adj} for losses (gains) of 0.9739 (0.9693)) is significantly better than that for PWF1 (median R^2_{adj} for losses (gains) of 0.9376 (0.9394)) at the 1 % level. A high R^2_{adj} is not surprising, however, since the elicitation procedure and the FOSD tests ensure

[9] In addition, violations could also occur for the intermediate steps (step 1 and step 3), which are also checked.

monotonicity (at least for the value function), thereby easing the estimation (similar R_{adj}^2 are obtained by Abdellaoui, Bleichrodt, and Paraschiv 2007).

Table 3 summarizes the results of the fitting for the value function. Overall, we find low curvature, both for losses and for gains. Abdellaoui, Bleichrodt, and Paraschiv (2007) find substantially higher curvatures. Our results are in accordance with existing results for small outcome amounts, which we apply for the prospects (see, e.g., Wakker and Deneffe 1996). Wilcoxon tests, however, reveal that the curvature parameters significantly differ from one. Additionally, the parameters for gains and losses are significantly different from each other (at the 1 % level). The loss aversion parameter is similar to that in existing studies (e.g., Tversky and Kahneman 1992, Abdellaoui, Bleichrodt, and Paraschiv 2007). Of note is the convex curvature for gains.

Table 3: Fitted Parameters – Value Function.

	VF1		VF2	
	Losses	Gains	Losses	Gains
α	0.9031	1.0717	0.9424	1.1466
	$\alpha^- = 1$ | -2.473 **	$\alpha^+ = 1$ | 4.435 ***	$\alpha^- = 1$ | -1.940 *	$\alpha^+ = 1$ | 4.858 ***
	$\alpha^- = \alpha^+$ | -4.389 ***		$\alpha^- = \alpha^+$ | -4.411 ***	
k	2.4214		2.3214	

Note: Reported are the median values of the fitted preference parameters α and k according to the two definitions VF1 and VF2. Below the parameters are the results of Wilcoxon matched-pair signed-rank tests (H_0 (two-sided), z-value). *, **, and *** denote significance at the 1 %, 5 %, and 10 % level, respectively.

Table 4 summarizes the fitting for the probability weighting function. Similar to the results for the value function we find low curvature (and elevation). Parameters are (mostly) significantly different from one, which is in line with Tversky and Kahneman (1992) and Abdellaoui (2000). Histograms, shown in Appendix F, reveal substantial variance for the individual data for all preference parameters.

Table 4: Fitted Parameters – Probability Weighting Function.

	PWF1		PWF2	
	Losses	Gains	Losses	Gains
γ	0.9553	0.9851	0.8692	0.9332
	$\gamma^- = 1$ | -1.964 **	$\gamma^+ = 1$ | 1.436	$\gamma^- = 1$ | -4.764 ***	$\gamma^+ = 1$ | -2.856 ***
	$\gamma^- = \gamma^+$ | -1.139		$\gamma^- = \gamma^+$ | -0.696	
δ			1.0834	0.7552
			$\delta^- = 1$ | 3.468 ***	$\delta^+ = 1$ | -2.036 **
			$\delta^- = \delta^+$ | 3.394 ***	

Note: Reported are the median values of the fitted parameters γ and δ according to the two definitions PWF1 and PWF2. Below the parameters are the results of Wilcoxon matched-pair signed-rank tests (H_0 (two-sided), z-value). *, **, and *** denote significance at the 1 %, 5 %, and 10 % level, respectively.

Further insights can be drawn from the median values to the indifference queries (see Appendix G). For probability weighting, they show persistent underweighting in the gain domain ($w(p_r) < p_r \; \forall \, r$) and generally persistent overweighting in the loss domain. A convex probability weighting function for gains was also obtained by van de Kuilen, Wakker, and Zou (2006). Risk behavior in CPT is determined by the probability weighting function and the value function. The persistent underweighting of all probabilities together with the convex value function imply that on average subjects are approximately risk neutral for gains in an expected utility sense.

5.3 Domain-specific Preferences

It might be argued that the difference in the context between part I (preference elicitation based on simple prospects) and part II (WTP elicitation based on discrete return distributions) plays a role, as research has shown that risk preferences might be context dependent (see, e.g., Weber, Blais, and Betz 2002). With respect to financial decisions, Dohmen et al. (2006) and Nosić and Weber (2008) show that predictive measures of risk-taking behavior (i.e., preferences, risk perception, and beliefs) are domain-specific, that is, elicitations from the lottery domain perform worse compared to elicitations from the financial domain in predicting financial decision making.

In order to investigate domain-specificity, we included ten simple lottery choices in part II for a consistency check. This task closely resembles the prospect-based elicitation procedure applied in part I and therefore relies on the same domain (see Appendix H for the prospects). Based on the elicited preference parameters from part I, the subject's choices for the ten lotteries are predicted. If somebody is just guessing randomly, one would expect five out of ten correct predictions. On average, 6.08 (median: 6, sd: 1.88) choices were forecasted correctly for the whole sample. Although this is significantly better than random guessing (significantly different from five at the 1 % level, two-sided Wilcoxon matched-pair signed-rank test), the predictive power is not as high as one would have expected for such simple lotteries. More strikingly, the prediction simply based on expected values is even better (mean: 6.54, median: 7, sd: 2.00). These results suggest that domain-specificity might play some role, but does not explain the whole picture.

5.4 Competence Effects

Various studies have shown that competence influences financial decision making (see, e.g., Lusardi and Mitchell 2007 or Agarwal et al. 2008). Motivated by these findings, we also examine competence effects as structured financial products might be classified as complex investment opportunities requiring some knowledge to be evaluated adequately. We hypothesize that subjects with higher competence show more consistent answers in the task at hand.

Of the participants, more than 30 % had previously attended lectures in behavioral finance, decision analysis, or capital market theory. Almost half of the participants had experience (yes/no question) investing in underlyings (stocks, funds, indexes), almost 30 % had experience investing in plain vanilla options, and only a few had experience investing in structured financial products. The median self-reported knowledge (scale from 1 (no) to 7 (high)) for "finance" was 4 (mean: 3.77, sd: 1.39), for "statistics" was 4 (mean: 4.27, sd: 1.19), for "structured financial products" was 3 (mean: 3.43, sd: 1.67), and for "pricing by duplication" was 2 (mean: 2.56, sd: 1.78). Overall, experience with and knowledge about structured financial products varied largely, making it plausible to examine competence effects.

We set up two different competence groups. The first one is based on the self-assessed knowledge in the areas “finance”, “statistics”, “structured financial products”, and “pricing by duplication”. A subject is assigned to this group if she answered 4 or higher for each of these four questions (28 subjects). The second group is based on the non-self-assessed performance in the test questions. Two quizzes were presented on the historical DAX distribution and the BDM mechanism, and one quiz each was presented for every product. A subject belongs to competence group 2 if she correctly answered the first two quizzes (DAX, BDM) in the first trial and needed at most a second trial for each product (51 subjects after additional exclusion of two subjects due to technical reasons).

For competence group 2, the prediction quality is slightly improved compared to the base case. The Spearman correlation coefficient is significant for two (four) products for scenario “fictional” (“real”) compared to one (two) products in the base case. Competence group 1 shows virtually no improvement (see Appendix I for the details). In addition, another argument against the influence of competence comes from the results of the CPT consistency check as described in the preceding section. If one reason for the low predictive power for the WTPs derives from lack of knowledge with respect to structured financial products, one would expect considerably better predictions for the simple choice tasks in the consistency check. One can assume that the subjects did not have problems in understanding the prospects. However, as the preceding section shows, the predictions even for those simple choices are weak. In addition, group 1 (mean: 6.43, median: 7, sd: 2.27) and group 2 (mean: 6.10, median: 6, sd: 1.99) show slightly better performance (no significance), which is counterintuitive, as no competence effect is expected with respect to the simple lottery questions. Predictions based on expected values are again better for both competence groups. In conclusion, competence effects can only explain a small fraction of our results.

5.5 Decision Errors

Chained elicitation procedures propagate any error that is made by the subject in any foregoing stage of the procedure. Blavatskyy (2006) shows theoretically that the elicitation approach we use is efficient with regard to decision errors. As an extension, we conduct a simulation study for the elicitation procedure in order to check this insight and assess the influence on the prediction quality.

The elicited parameter set $(\alpha^+, \alpha^-, k, \gamma^+, \gamma^-, \delta^+, \delta^-)^{true}$ from part I of our experiment is assumed to reflect the true parameters of each subject. Given these "true" preferences, we simulate how the subject would act in each step of the elicitation procedure. Let $v_{noerror}(P)$ denote the theoretical value of a prospect P without the incorporation of decision errors. In each indifference query of the elicitation procedure, $v_{noerror}(P)$ is calculated according to CPT given $(\alpha^+, \alpha^-, k, \gamma^+, \gamma^-, \delta^+, \delta^-)^{true}$.[10] Then, a decision error is applied to $v_{noerror}(P)$, thereby simulating distorted answers. Based on this set of distorted answers, a second parameter set is fitted $(\alpha^+, \alpha^-, k, \gamma^+, \gamma^-, \delta^+, \delta^-)^{error}$, representing preference parameters distorted by decision errors during the elicitation procedure.

We model decision errors according to the white noise error model (Hey and Orme 1994) with an absolute additive error term. The error term ε is assumed to be independently and identically distributed with $\varepsilon \sim N(0; \sigma)$. Blavatskyy (2006) applies three different standard deviations for the three different types of elicitation methods, namely the tradeoff method, the probability equivalent method, and the certainty equivalent method. We follow this approach and we differ between σ^{CE} for the certainty equivalent queries (step 2 and step 4) and σ^{PE} for the probability equivalent queries (step 5 and step 6). More formally, the certainty equivalent queries $L_r \sim (L_A, p_{0.5}^-; L_B)$ with $U(L_r) = 0.5 \cdot U(L_A) + 0.5 \cdot U(L_B)$ become

$$U(L_r) = 0.5 \cdot U(L_A) + 0.5 \cdot U(L_B) + \varepsilon^{CE} \text{ with } \varepsilon^{CE} \sim N\left(0; \sigma^{CE}\right)$$

and the probability equivalent queries $L_B \sim (L_A, p_s^-; 0)$ with $U(L_B) = s \cdot U(L_A)$ become

$$U(L_B) + \varepsilon^{PE} = s \cdot U(L_A) \text{ with } \varepsilon^{PE} \sim N\left(0; \sigma^{PE}\right).$$

For our research question, we want to examine how much distorted preference parameters influence the predictive power of CPT. As the benchmark, we calculate the true $WTP_{true}^{CPT=0}$ given the true set of preference parameters $(\alpha^+, \alpha^-, k, \gamma^+, \gamma^-, \delta^+, \delta^-)^{true}$ with $CPT\left(Product \middle| P_{Product,t} = WTP_{true}^{CPT=0}\right) = 0$. De-

[10] Throughout the simulations we use the VF2 / PWF2 combination as the functional form.

cision errors not only affect the elicited preference parameters but also distort the elicitation of the WTP. We therefore add another error term and calculate the distorted $WTP_{true}^{CPT=0}$ as $CPT\left(Product\middle|P_{Product,t} = WTP_{true}^{CPT=0}\right) + \epsilon^{WTP} = 0$ with $\epsilon^{WTP} \sim N(0;\sigma^{CE})$ and the distorted parameters $(\alpha^+,\alpha^-,k,\gamma^+,\gamma^-,\delta^+,\delta^-)^{error}$. Since the elicitation of the WTP is a certainty equivalent query we apply the respective standard deviation σ^{CE}.

Existing simulation studies do not calibrate their error model to empirically observed error quotas to obtain a proxy for σ^{CE} and σ^{PE}. Instead, the standard deviations are fixed by assumption. Bleichrodt and Pinto (2000), for example, assume a standard deviation of 0.05. Vossmann and Weber (2005) cover a broad range of possible values. The scale of the standard deviations is noted in terms of utility and is therefore dependent on the normalization procedure. We do not specify the error parameters exogenously, but instead calibrate our model endogenously based on the observed error quotas from part I of our experiment. Calibration is based on the number of FOSD violations. The probability equivalent queries from step 5 and step 6 utilize the outcomes already elicited in step 2 and step 4. This leads to a two-step calibration approach. First, σ^{CE} is calibrated to the empirically observed quota of FOSD violations (q^{CE}) for the value function queries from step 2 and step 4. Then, σ^{PE} is calibrated to the empirically observed quota (q^{PE}) for the probability weighting function queries from step 5 and step 6 but with given σ^{CE}.[11] As can be seen from Table 5, the error quotas decline for the competence groups compared to the base case. Table 5 also shows the calibrated standard deviations. There is no clear conclusion as to whether the certainty equivalent or the probability equivalent queries evoke the higher number.

[11] All simulations are conducted with 50 runs each. For step 1 and step 3 the elicitation procedure comprises only a few indifference queries. A reasonable calibration is thus not feasible for the tradeoff method. We consequently do not apply error terms during these two steps.

Table 5: Quotas of FOSD Violations and Calibration Results.

	q^{CE}	q^{PE}	σ^{CE}	σ^{PE}
All	19.50 %	13.50 %	0.450	0.400
Group 1	11.11 %	13.89 %	0.400	0.875
Group 2	11.29 %	8.06 %	0.350	0.325

Note: The first two columns show the empirical error quotas. If a subject shows a FOSD violation for both, the certainty equivalent and the probability equivalent queries, she is counted twice, since we sequentially calibrate the standard deviations. The last two columns show the calibrated standard deviations. See Appendix J for an overview of the grid search.

Table 6 presents the simulation results for the preference parameters and the WTP predictions. A high R^2 and ρ show that decision errors do not lead to heavily distorted preference parameters. If one were to rely on the distorted parameters, as we do in our experiment, one would still use a good proxy of the "true" parameters. The same applies for the WTP predictions. Except for the regression measure on the whole sample, all numbers indicate low distortions by decision errors.

Table 6: Simulation Results for Preference Parameters.

	CPT Parameters		WTP	
	Regression (R^2)	Spearman (ρ)	Regression (R^2)	Spearman (ρ)
All	0.9890	0.9923	0.7700	0.9713
Group 1	0.9860	0.9835	0.9423	0.9511
Group 2	0.9893	0.9904	0.9343	0.9680

Note 1: Reported are the mean values of the median R^2 and ρ over all seven preference parameters and all ten products, respectively. We run the regression $CPT^{true} = \alpha + \beta \cdot CPT^{error}$, with $CPT\varepsilon\{\alpha^+, \alpha^-, k, \gamma^+, \gamma^-, \delta^+, \delta^-\}$ and the regression $WTP^{true} = \alpha + \beta \cdot WTP^{error}$ as well as the corresponding correlation analyses. See Appendix K and Appendix L for the results by parameter and by product.

We conclude that decision errors cannot explain the low correlation observed in our experiment, even if the standard deviation of the error term is calibrated on the basis of a relatively high number of FOSD violations.[12]

[12] We also analyzed a relative error term. For the certainty equivalent queries we apply $U(L_r) = \left(0.5 \cdot U(L_A) + 0.5 \cdot U(L_B)\right) \cdot (1 + \varepsilon_{relative}^{CE})$ with $\varepsilon_{relative}^{CE} \sim N(0; \sigma_{relative}^{CE})$ and for the probability equivalent queries we use $U(L_B) \cdot (1 + \varepsilon_{relative}^{PE}) = s \cdot U(L_A)$ with $\varepsilon_{relative}^{PE} \sim N(0; \sigma_{relative}^{PE})$. Consequently, the WTP is distorted by $WTP_{relativeerror}^{CPT=0}$ as $CPT\left(Product \middle| P_{Product,t} = WTP_{relativeerror}^{CPT=0}\right) \cdot (1 + \epsilon_{relative}^{WTP}) = 0$ with

5.6 Other Factors

Besides the possible explanations above that can be addressed with our data, other factors may influence the results.

Incentive-compatible payment for preference elicitation. The payment for the first part was not incentive-compatible. The lack of incentives might lead to distorted answers or at least higher variance for the choices at hand (Camerer and Hogarth 1999). However, this was done on purpose. First, our choice follows the literature (especially Abdellaoui, Bleichrodt, and Paraschiv 2007), where subjects do not receive incentive-compatible payments. Second, high external validity considerations impact our design process: in everyday bank business, especially in advising high-net-worth individuals, (substantial) incentives for eliciting customers' preferences can hardly be implemented.

Personal interview versus fully computer-based elicitation. Previous studies on parameter-free elicitations mostly use personal interview sessions (see, e.g., Bleichrodt and Pinto 2000; Abdellaoui, Vossmann, and Weber 2005; Abdellaoui, Bleichrodt, and Paraschiv 2007; Abdellaoui, Bleichrodt, and L'Haridon 2008). There are at least two reasons for using personal interviews. First, this method can drive the subjects to think about their decisions more thoroughly. Second, the interaction between subject and interviewer makes it possible to resolve inconsistencies during the elicitation procedure. However, it is not clear to what extent answers might be biased due to the implicit or explicit influence of the experimenter (i.e., a demand effect). We therefore opt to employ a fully computer-based approach (like, e.g., Abdelloui 2000). A further reason is motivated by external validity concerns. Although personal interviews might well be feasible in advising high-net-worth individuals in highly personalized customer relationships, this would be too costly and therefore not advisable for selling products to retail investors. The question of whether preference parameters obtained from personal interview sessions lead to better out-of-task performance is left for further research.

$\epsilon_{relative}^{WTP} \sim N(0; \sigma_{relative}^{CE})$. The results do not deviate substantially from the results for the absolute error term (see Appendix M to Appendix O).

Subjects may use another reference point. The choice of reference point is indeed essential for the valuation. A discount certificate, for example, might be framed in a way that the investor perceives the discount compared to a direct underlying investment as a gain and thus adjusts the reference point (see Shefrin and Statman 1993). However, our product descriptions focus on objectively presenting the outcome distribution in reference to the product price, which represents the status quo. Framing should therefore be minimized in our experiment. The possibility that subjects have brought their reference point to the lab is an interesting topic for further research. There is little work on endogenous reference points. Every elicitation procedure and CPT application suffers from this problem. The problem is even more important for field data applications. In the lab, we can at least control the way the information is presented. In our experiment, subjects are shown a gain/loss profile that assigns a gain/loss of zero to the reference point we assumed for our analysis (see again the screenshots in Appendix C). Nevertheless, pinpointing the way reference points are formed, without yielding to the temptation of simply introducing too many free parameters that govern the reference point formation process, is an important research area in individual decision making.

Prospect theory is not an adequate descriptive theory of risk. As already mentioned in the introduction, CPT combines many attractive features of decision making under risk and uncertainty, which can potentially explain many phenomena in the lab and in the field. Although this does not necessarily mean that CPT performs well at the individual level, there is much reason to believe that CPT is at least best suited to do so. Nevertheless, there is a growing literature on violations of CPT in individual decision making (see, e.g., Birnbaum 2006 and Wu and Markle 2008).

Prospect theory is an adequate descriptive theory of risk, but the elicitation procedure is not adequate. There are many different methods to elicit CPT preferences, from the non-parametric method discussed here to the parametric fitting of the CPT functions based on choices or certainty equivalents in the lab (see, e.g., Gonzalez and Wu 1999) and in the field (see, e.g., Post et al. 2008 and Andersen et al. 2006). Comparing these methods, based on both actual data and simulations in the spirit of the one we carry out above, could answer the question of whether the elicitation procedure per se introduces a systematical bias. Work in this vein seems to be a promising field for further research (see Wilcox 2007 for a recent contribution addressing maximum likelihood estimation of choice data and its out-of-task performance). Our simulations show that, at least with respect to the propagation of errors, the employed elicitation procedure is robust.

Prospect theory is an adequate descriptive theory of risk, but risk preferences are temporarily unstable. Harrison et al. (2005) present evidence that the Holt and Laury (2002) measure of risk aversion is stable over a time horizon of approximately six months. Whether CPT parameters are stable over time, is an open question and an interesting topic for further research. If the parameters are not stable, then the temporal instability of preference may contribute to the low correlations in our experiment.

6. Summary

In this paper, we use a state-of-the-art method to estimate CPT preference parameters on an individual basis. Based on these estimates, we predict the behavior of subjects in a different decision context. The predictions and the actual stated preferences show almost no correlation, although a simulation study suggests that there should be substantial correlation if CPT is the appropriate model and if there is a reasonable error rate. We examine domain-specificity and competence effects as possible explanations. At best, all explanations can account for only a small proportion of the lack of correlation. Other possible explanations that cannot be investigated with our data are also discussed.

We hope that more research will be done on the out-of-task-performance of CPT and other decision theories to investigate the reasons for the virtually non-existing relationship between CPT predictions and stated WTPs. Such a research agenda is necessary to understand the usefulness of several decision theories for practical applications on an individual basis, as there is a good possibility that calibrating such models on an individual basis is of less value than previously thought.

References

Abdellaoui, M. 2000. Parameter-Free Elicitation of Utility and Probability Weighting Functions. *Management Science* **46** 1497-1512.

Abdellaoui, M. Bleichrodt, H. L'Haridon, O. 2008. A Tractable Method to Measure Utility and Loss Aversion under Prospect Theory. *Journal of Risk and Uncertainty* **36** 245-266.

Abdellaoui, M. Bleichrodt, H. Paraschiv, C. 2007. Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. *Management Science* **53** 1659–1674.

Abdellaoui, M. Vossmann, F. Weber, M. 2005. Choice-Based Elicitation and Decomposition of Decision Weights for Gains and Losses Under Uncertainty. *Management Science* **51** 1384-1399.

Agarwal, S. Driscoll, J. C. Gabaix, X. Laibson, D. 2008. The Age of Reason: Financial Decisions Over the Lifecycle, *Working Paper* Harvard University.

Andersen, S. Harrison, G. W. Lau, M. I. Rutström, E. E. 2006. Dynamic Choice Behavior in a Natural Experiment. *Working Paper* CEBR and University of Central Florida.

Becker, G. M. DeGroot, M. H. Marschak, J. 1964. Measuring Utility by a Single-response Sequential Method. *Behavioral Science* **9** 226-232.

Bethel, J. E. Ferrel, A. 2007. Policy Issues Raised by Structured Products. In: Fuchita, Y. Litan, R. E. (eds.) *New Financial Instruments and Institutions* Brookings Institution Press, Baltimore (MD), 167-191.

Birnbaum, M. H. 2006. Evidence Against Prospect Theories in Gambles With Positive, Negative, and Mixed Consequences. *Journal of Economic Psychology* **27** 737-761.

Black, F. S. Scholes, M. S. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81** 637-654.

Blavatskyy, P. 2006. Error Propagation in the Elicitation of Utility and Probability Weighting Functions. *Theory and Decision* **60** 315-334.

Bleichrodt, H. Pinto, J. L. 2000. A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis. *Management Science* **46** 1485-1496.

Branger, N. Breuer, B. 2008. The Optimal Demand for Retail Derivatives. *Working Paper* University of Münster.

Breuer, W. Perst, A. 2007. Retail Banking and Behavioral Financial Engineering: The Case of Structured Products. *Journal of Banking & Finance* **31** 827-844.

Camerer, C. F. 1989. An Experimental Test of Several Generalized Utility Theories. *Journal of Risk and Uncertainty* **2** 61-104.

Camerer, C. F. 2000. Prospect Theory in the Wild: Evidence from the Field. In: Kahneman, D. Tversky, A. (eds.) *Choices, values and frames*. Cambridge University Press, New York (NY), 288-300.

Camerer, C. F. Hogarth, R. M. 1999. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty* **19** 7-42.

Cox, J. C. Ross, S. A. Rubinstein, M. 1979. Option Pricing: A Simplified Approach. *Journal of Financial Economics* **7** 229-263.

Deutscher Derivate Verband 2008. *Markt für derivative Wertpapiere legt 2007 um 17 Prozent zu* (in German), Deutscher Derivate Verband, Frankfurt am Main.

Dohmen, T. J. Falk, A. Huffman, D. Schupp, J. Sunde, U. Wagner G. G. 2006. Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. *Working Paper* CEPR.

Fischer, R. 2007. Do Investors in Structured Products Act Rationally? *Working Paper* European Business School.

Goldstein, W. M. Einhorn, H. J. 1987. Expression Theory and the Preference Reversal Phenomena. *Psychological Review* **94** 236-254.

Gonzalez, R. Wu, G. 1999. On the Shape of the Probability Weighting Function. *Cognitive Psychology* **38** 129-166.

Harrison, G. W. Johnson, E. McInnes, M. M. Rutström, E. E. 2005. Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters* **1** 31-35.

Henderson, B. J. Pearson, N. D. 2007. Patterns in the Payoffs of Structured Equity Derivatives. *Working Paper* Seton Hall University and University of Illinois at Urbana-Champaign.

Hey, J. D. Orme, C. 1994. Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica* **62** 1291-1326.

Holt, C. A. Laury, S. K. 2002. Risk Aversion and Incentive Effects. *American Economic Review* **92** 1644-1655.

Hull, J. C. 2008. *Options, Futures, and Other Derivatives (Seventh Edition)*. Prentice Hall, Upper Saddle River (NJ).

Kahneman, D. Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47** 263-292.

Keefer, D. L. Bodily, S. E. 1983. Three-Point Approximations for Continuous Random Variables. *Management Science* **29** 595-609.

Köbberling, V. Wakker, P. P. 2005. An Index of Loss Aversion. *Journal of Economic Theory* **122** 119-131.

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

Loewenstein, G. 1999. Experimental Economics from the vantage-point of Behavioral Economics. *The Economic Journal* **109** F25-F34.

Luce, R. D. 2000. *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. Lawrence Erlbaum Associates, Mahwah (NJ).

Lusardi, A. Mitchell, O. S. 2007. Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education Programs. *Working Paper* University of Pennsylvania.

Muck, M. 2006. Where Should You Buy Your Options? The Pricing of Exchange-Traded Certificates and OTC Derivatives in Germany. *The Journal of Derivatives* **14** (Fall 2006) 82-96.

Nosić, A. Weber, M. 2008. How Risky Do I Invest: The Role of Risk Attitudes, Risk Perceptions and Overconfidence. *Working Paper* University of Mannheim.

Pearson, E. S. Tukey, J. W. 1965. Approximate Means and Standard Deviations based on Distances. *Biometrika* **52** 533-546.

Post, T. van den Assem, M. J. Baltussen, G. Thaler, R. H. 2008. Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show. *American Economic Review* **98** 38-71.

Shefrin, H. Statman, M. 1993. Behavioral Aspects of the Design and Marketing of Financial Products. *Financial Management* **22** 123-134.

Soll, J. B. Klayman, J. 2004. Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30** 299-314.

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

Starmer, C. Sugden, R. 1989. Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect. *Journal of Risk and Uncertainty* **2** 159-178.

Stoimenov, P. A. Wilkens, S. 2005. Are Structured Products 'Fairly' Priced? An Analysis of the German Market for Equity-linked Instruments. *Journal of Banking & Finance* **29** 2971-2993.

The European Parliament and the European Council 2004. *Markets in Financial Instruments Directive,* Directive 2004/39/EC.

The European Parliament and the European Council 2006. *Markets in Financial Instruments Directive: Implementing Directive,* Directive 2006/73/EC.

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

van de Kuilen, G. Wakker, P. P. Zou, L. 2006. A Midpoint Technique for Easily Measuring Prospect Theory's Probability Weighting. *Working Paper* University of Amsterdam.

Vossmann, F. Weber, M. 2005. Measuring Covariation between Preference Parameters: A Simulation Study. *Working Paper* University of Mannheim.

Wakker, P. P. Deneffe, D. 1996. Eliciting von Neumann-Morgenstern Utilities when Probabilities Are Distorted or Unknown. *Management Science* **42** 1131-1150.

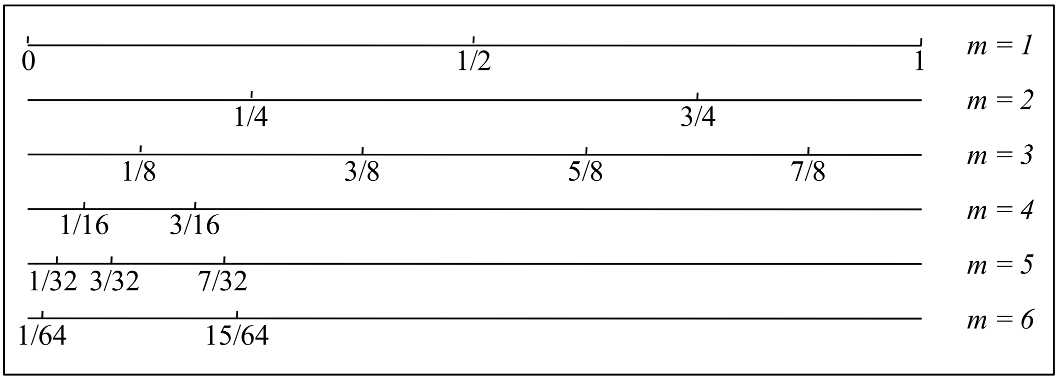
Weber, E. U. Blais, A.-R. Betz, N. E. 2002. A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making* **15** 263-290.

Wilcox, N. T. 2007. Predicting Risky Choices Out-of-Context: A Monte Carlo study. *Working Paper* University of Houston.

Wu, G. 1994. An Empirical Test of Ordinal Independence. *Journal of Risk and Uncertainty* **9** 39-60.

Wu, G. Markle, A. B. 2008. An Empirical Test of Gain-Loss Separability in Prospect Theory. *Management Science* **54** 1322-1335.

Appendix A: Elicited Points on the Value Function.



Note: This figure shows the chaining of the elicitation of the value function with m iterations. Numbers represent the utility r for which the respective outcome is elicited by a prospect consisting of the two adjacent utilities from the foregoing iteration(s).

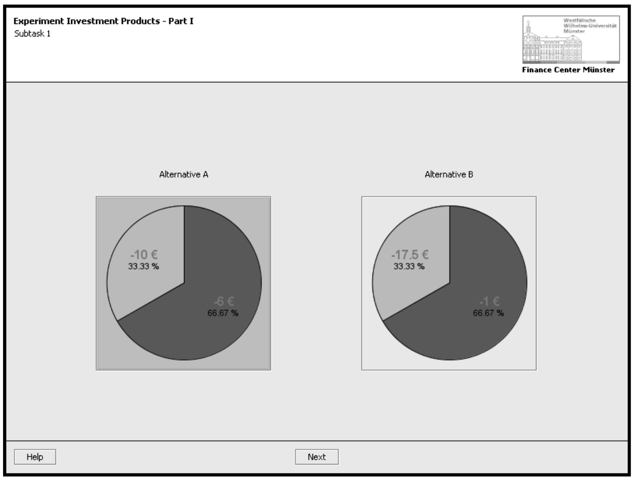
Appendix B: Arbitrage Constraints.

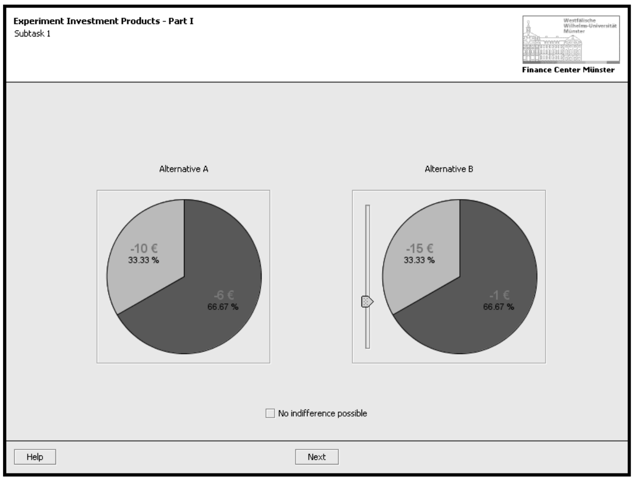
We apply the following arbitrage constraints on the values of different investment products:

- A discount certificate combines a long position in the underlying with a short position in a call. The price of a discount certificate therefore has to be smaller than the price of the underlying ($P_{Underlying,t} \geq P_{DiscountCertificate,t}$).
- A protected certificate combines the underlying with a long position in a put and therefore has to trade at a higher price than the underlying ($P_{Protected,t} \geq P_{Underlying,t}$).
- An outperformance certificate combines the underlying with a long position in a call and therefore has to trade at a higher price than the underlying ($P_{Outperformance,t} \geq P_{Underlying,t}$).
- A combined certificate combines a protected with an outperformance certificate and thus combines the underlying with long positions in a put and a call. It therefore has to trade at a higher price as the underlying ($P_{Combined,t} \geq P_{Underlying,t}$). The put in the combined certificate has the same strike as in the protected certificate, and because of the additional call in the combined certificate it has to trade at a higher price than the protected one ($P_{Combined,t} \geq P_{Protected,t}$). An analogous argument applies with regard to the outperformance certificate ($P_{Combined,t} \geq P_{Outperformance,t}$).
- An in-the money call must be more expensive than an out-of-the money call ($P_{Call\_ITM,t} \geq P_{Call\_OTM,t}$).

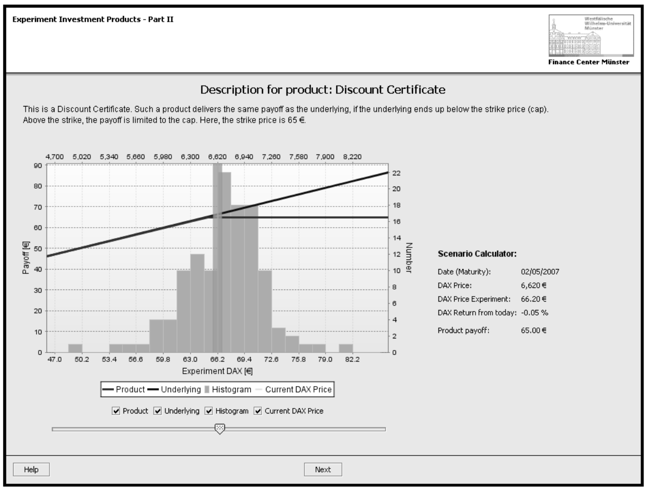
Appendix C: Screenshots from the Experiment.

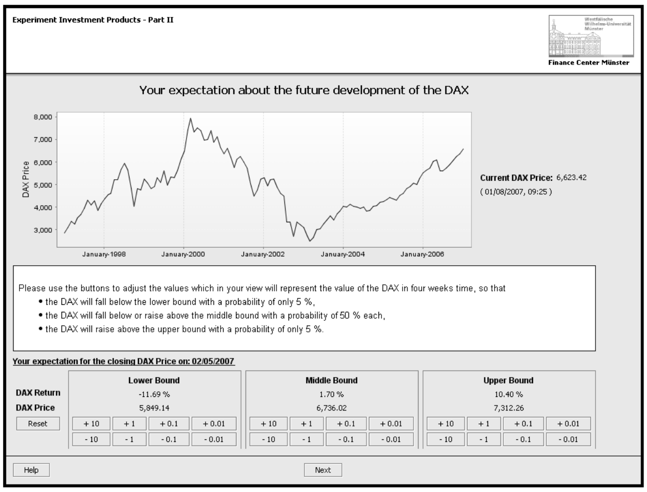
Panel I – Experiment Part I

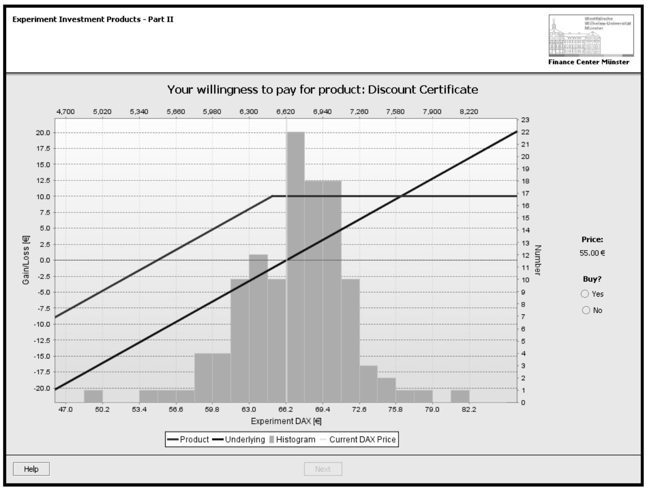




Panel II – Experiment Part II







Appendix D: Three-point Elicitation of Subjective Expectations.

Subjective expectations are elicited through a three-point elicitation method (Pearson and Tukey 1965, Keefer and Bodily 1983). Every subject i was asked to state her expectations for the 5 %, 50 %, and 95 % percentile ($x(p)_i$ with $p\epsilon\{0.05, 0.50, 0.95\}$) for the expected return distribution of the underlying over the products' time-to-maturity (ttm). The mean

$$\mu_i^{ttm} = 0.185 \cdot x(0.05)_i + 0.63 \cdot x(0.50)_i + 0.185 \cdot x(0.95)_i$$

and the standard deviation

$$\sigma_i^{ttm} = \sqrt{0.185 \cdot (x(0.05)_i)^2 + 0.63 \cdot (x(0.50)_i)^2 + 0.185 \cdot (x(0.95)_i)^2 - \left(\mu_i^{ttm}\right)^2}$$

can then be derived. The respective historical values were chosen as the starting values for the percentiles, thereby reducing the effect of overconfidence in estimating intervals (see, e.g., Soll and Klayman 2004 for an overview). This comes at the cost of potential anchoring and adjustment.

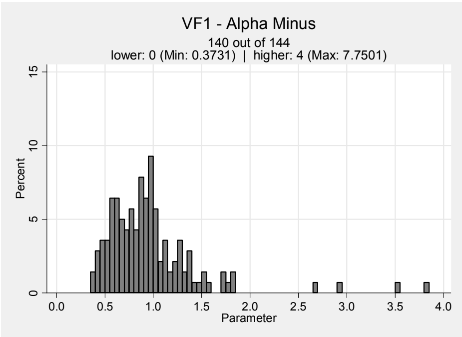
Appendix E: Quality of Prediction – By Product for Median Investor.

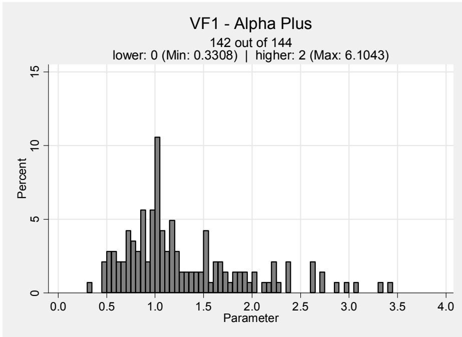
	fictional			real		
	WTP^{EV}_{median}	$WTP^{CPT=0}_{median}$	WTP^{stated}_{median}	WTP^{EV}_{median}	$WTP^{CPT=0}_{median}$	WTP^{stated}_{median}
Underlying	66.95	65.26	66.09	66.97	65.53	66.25
(N = 104)	(0.00) ***	(0.14)		(0.00) ***	(0.00) ***	
BinaryCall	4.50	2.94	3.00	4.03	2.71	2.50
(N = 133)	(0.00) ***	(0.86)		(0.00) ***	(0.60)	
BinaryPut	2.83	1.71	1.94	3.00	1.90	1.50
(N = 131)	(0.00) ***	(0.60)		(0.00) ***	(0.00) ***	
Call_ITM	3.01	2.05	2.00	2.64	1.72	1.90
(N = 135)	(0.00) ***	(0.30)		(0.00) ***	(0.06) *	
Call_OTM	1.48	0.89	1.01	1.14	0.67	1.10
(N = 131)	(0.02) **	(0.00) ***		(0.73)	(0.00) ***	
Put_OTM	1.11	0.64	1.04	0.69	0.38	0.91
(N = 126)	(0.42) *	(0.00) ***		(0.06) *	(0.00) ***	
DiscountCertificate	63.93	62.78	61.52	64.60	63.49	62.07
(N = 94)	(0.00) ***	(0.01) ***		(0.00) ***	(0.00) ***	
Protected	68.01	67.08	67.61	67.72	66.72	67.26
(N = 87)	(0.00) ***	(0.02) **		(0.39)	(0.00) ***	
Outperformance	68.43	66.54	66.50	68.16	66.53	66.50
(N = 106)	(0.00) ***	(0.92)		(0.00) ***	(0.77)	
Combined	69.49	68.20	67.75	68.84	67.64	67.30
(N = 89)	(0.00) ***	(0.20)		(0.00) ***	(0.14)	

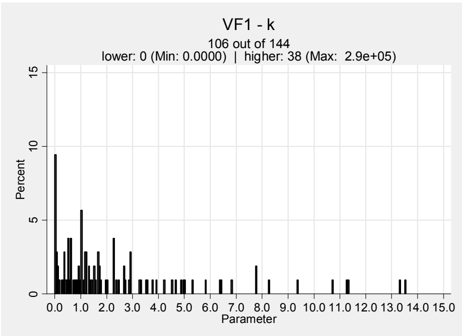
Note: Slightly different numbers of observations compared to the individual results obtain due to slightly different (aggregated) $PO^{min}_{median,task,product}$ and $PO^{max}_{median,task,product}$. For scenario "fictional", $WTP^{CPT=0}_{median}$ is based on the return distribution of the respective experiment run while for scenario "real", $WTP^{CPT=0}_{median}$ is based on the median expected μ_i^{ttm} and σ_i^{ttm}. The same applies for the expected values WTP^{EV}_{median}. P-values of a Binomial test (two-sided, null hypothesis stating $WTP^{stated}_{median} = WTP^{CPT=0}_{median}$ or $WTP^{stated}_{median} = WTP^{CPT=0}_{median}$) are reported in parentheses. *, **, and *** denote significance at the 1 %, 5 %, and 10 % level, respectively.

Appendix F: Histograms of CPT Parameters.

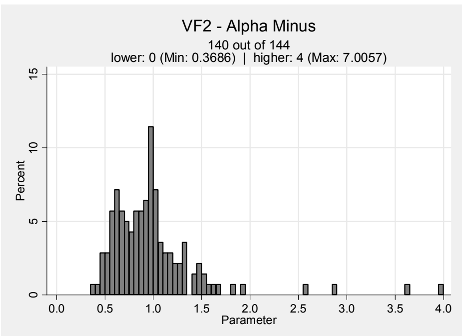
Panel I – VFI

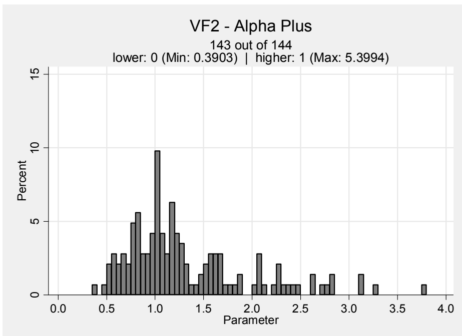


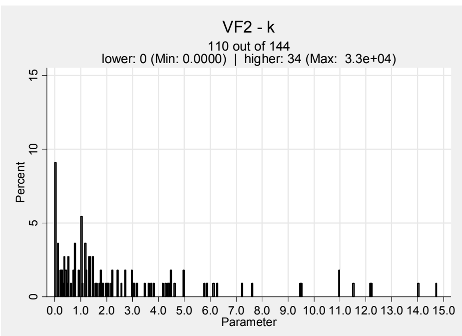




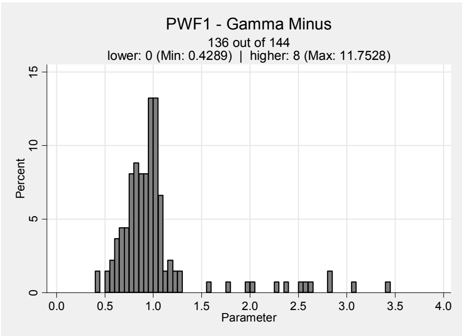
Panel II – VF2

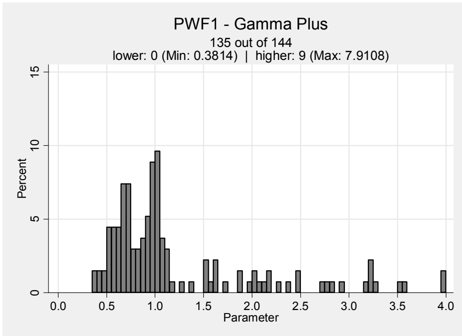




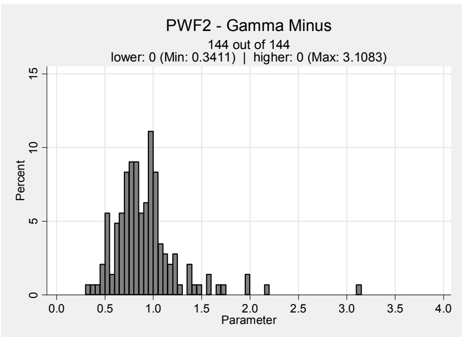


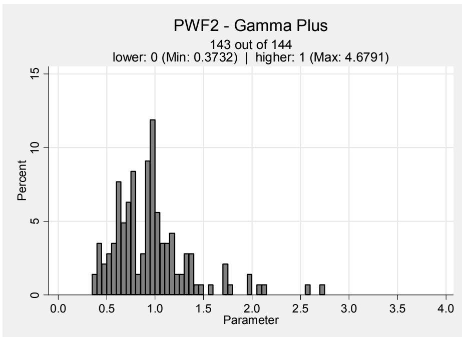
Panel III – PWF1

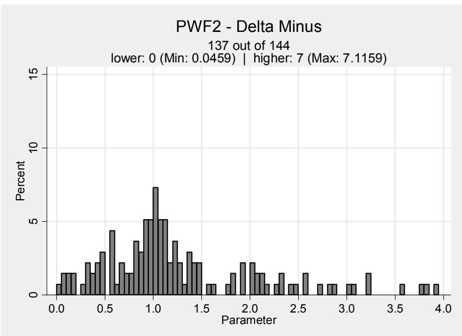


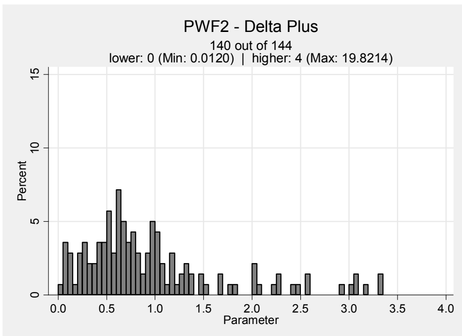


Panel IV – PWF2

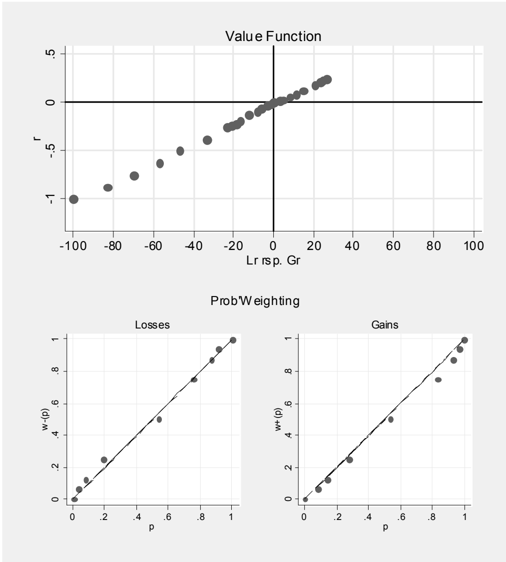








Appendix G: Elicited Values for the Median Investor.



Probability Weighting							
p	0.0625 (1/16)	0.1250 (1/8)	0.2500 (1/4)	0.5000 (1/2)	0.7500 (3/4)	0.8750 (7/8)	0.9375 (15/16)
$w^{-}(p)$	0.0369	0.0767	0.1912	0.5356	0.7573	0.8692	0.9147
$w^{+}(p)$	0.0753	0.1426	0.2738	0.5343	0.8275	0.9303	0.9663
Value Function – Losses							
r	-1.000 (1/1)	-0.8750 (7/8)	-0.7500 (3/4)	-0.6250 (5/8)	-0.5000 (1/2)	-0.3750 (3/8)	-0.2500 (1/4)
L_r	-1.000	-0.8300	-0.7000	-0.5698	-0.4683	-0.3339	-0.2323
r	-0.2344 (15/64)	-0.2188 (7/32)	-0.1875 (3/16)	-0.1250 (1/8)	-0.0938 (3/32)	-0.0625 (1/16)	-0.0313 (1/32)
L_r	-0.2087	-0.1891	-0.1677	-0.1228	-0.0807	-0.0603	-0.0300
r	-0.0156 (1/64)						
L_r	-0.0121						
Value Function – Gains							
r	0.0156 (1/64)	0.0313 (1/32)	0.0625 (1/16)	0.0938 (3/32)	0.1250 (1/8)	0.1875 (3/16)	0.2188 (7/32)
G_r	0.0291	0.0452	0.0817	0.1137	0.1493	0.2062	0.2351
r	0.2344 (15/64)	0.2500 (1/4)					
G_r	0.2476	0.2636					

Note: The diagram graphically shows the elicited values while the table reports the exact numbers for all points r on the value function and for all points s on the probability weighting function. The elicited values $L_r$and G_r for the value function are divided by 100 ($|L_1|$).

Appendix H: Lotteries from the Consistency Check.

1	50 % 100 € 50 % -50 €	50 % 50 € 50 % -30 €	**2**	50 % 100 € 50 % 50 €	50 % 150 € 50 % 15 €
3	100 % -30 €	90 % -20 € 10 % -80 €	**4**	50 % 50 € 50 % -100 €	10 % 100 € 90 % -50 €
5	10 % 100 € 90 % 50 €	100 % 60 €	**6**	50 % -10 € 50 % -100 €	50 % -20 € 50 % -80 €
7	100 % -25 €	10 % 100 € 90 % -50 €	**8**	50 % 150 € 50 % 40 €	90 % 100 € 10 % 50 €
9	10 % -20 € 90 % -80 €	50 % -35 € 50 % -100 €	**10**	90 % 100 € 10 % -50 €	100 % 50 €

Appendix I: Quality of Predictions – Results per Competence Group.

Competence Group 1	Regression (R^2)		Spearman (ρ)	
	fictional	real	fictional	real
Underlying	0.0000	0.0011	0.0715	0.1335
(N = 27)	(0.00)	(0.03)	(0.72)	(0.51)
BinaryCall	0.0864	0.0970	0.2248	0.2658
(N = 28)	(2.46)	(2.79)	(0.25)	(0.17)
BinaryPut	0.0161	0.1023	0.1081	0.3224
(N = 27)	(0.41)	(2.85)	(0.59)	(0.10)
Call_ITM	0.0124	0.0780	-0.2179	-0.4702
(N = 28)	(0.33)	(2.20)	(0.27)	(0.01) **
Call_OTM	0.0036	0.2362	0.0225	-0.2576
(N = 28)	(0.09)	(0.40)	(0.91)	(0.19)
Put_OTM	0.0104	0.2092	0.1891	0.3925
(N = 28)	(0.27)	(6.88) **	(0.34)	(0.04) **
DiscountCertificate	0.0278	0.0364	0.2287	0.2583
(N = 24)	(0.63)	(0.83)	(0.28)	(0.22)
Protected	0.1092	0.0969	0.2951	0.1086
(N = 26)	(2.94) *	(2.57)	(0.14)	(0.60)
Outperformance	0.0004	0.0068	0.0162	-0.2264
(N = 25)	(0.01)	(0.16)	(0.94)	(0.28)
Combined	0.1032	0.1022	0.1542	0.0322
(N = 22)	(2.30)	(2.28)	(0.49)	(0.89)
Competence Group 2	Regression (R^2)		Spearman (ρ)	
	fictional	real	fictional	real
Underlying	0.0022	0.0146	0.0805	0.1074
(N = 46)	(0.10)	(0.65)	(0.59)	(0.48)
BinaryCall	0.1103	0.1389	0.2462	0.4050
(N = 49)	(5.82) **	(7.58) ***	(0.09) *	(0.00) ***
BinaryPut	0.1029	0.0753	0.2356	0.2535
(N = 48)	(5.28) **	(3.75) *	(0.11)	(0.08) *
Call_ITM	0.2792	0.1327	0.4027	0.1021
(N = 49)	(18.21) ***	(7.19) **	(0.00) ***	(0.49)
Call_OTM	0.0093	0.0001	0.1624	-0.0094
(N = 48)	(0.43)	(0.01)	(0.27)	(0.95)
Put_OTM	0.1533	0.0642	0.1499	0.2762
(N = 49)	(8.51) ***	(3.23) *	(0.30)	(0.05) *
DiscountCertificate	0.0213	0.2458	0.1685	0.4561
(N = 41)	(0.85)	(12.71) ***	(0.29)	(0.00) ***
Protected	0.0000	0.1549	0.1963	0.1801
(N = 42)	(0.00)	(7.33) ***	(0.21)	(0.25)
Outperformance	0.0019	0.0099	0.0057	0.1197
(N = 44)	(0.08)	(0.42)	(0.97)	(0.44)
Combined	0.0187	0.0231	0.1511	0.1570
(N = 37)	(0.67)	(0.83)	(0.37)	(0.35)

Note: Reported are the results for the regression analysis (R^2) and the correlation analysis (ρ) by product per competence group. Also reported are the F-values (Regression) / p-values (Spearman) in parentheses. *, **, and *** denote significance at the 1 %, 5 %, and 10 % level, respectively.

Appendix J: Calibration Results for the Error Model.

	Certainty Equivalent			Probability Equivalent		
All						
σ	0.425	0.450	0.475	0.375	0.400	0.425
Quota	17.83 %	19.39 %	19.97 %	13.43 %	13.47 %	14.42 %
Group 1						
σ	0.375	0.400	0.425	0.850	0.875	0.900
Quota	10.29 %	11.36 %	11.43 %	13.29 %	13.43 %	14.93 %
Group 2						
σ	0.325	0.350	0.375	0.300	0.325	0.350
Quota	10.16 %	11.69 %	12.98 %	7.53 %	8.04 %	8.59 %

Note: Reported are the mean quotas of FOSD violations for different standard deviations of the error for the certainty equivalent queries and the probability equivalent queries. The ones that best fit the empirical quota for the group are shaded and reported in the middle section together with the respective two adjacent values.

Appendix K: CPT Parameters – Error Model.

All	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Alpha Minus	0.9999	0.9988/1.0000	0.9987	0.9968/0.9994
Alpha Plus	0.9974	0.9948/0.9987	0.9922	0.9859/0.9959
k	1.0000	1.0000/1.0000	0.9975	0.9951/0.9985
Gamma Minus	0.9742	0.8683/0.9890	0.9818	0.9589/0.9906
Gamma Plus	0.9845	0.9580/0.9954	0.9870	0.9740/0.9931
Delta Minus	0.9827	0.8894/0.9965	0.9930	0.9872/0.9966
Delta Plus	0.9844	0.9008/0.9991	0.9956	0.9895/0.9985
Group 1	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Alpha Minus	0.9991	0.9964/0.9998	0.9960	0.9896/1.0000
Alpha Plus	0.9959	0.9872/0.9988	0.9820	0.9452/0.9974
k	1.0000	1.0000/1.0000	0.9954	0.9800/1.0000
Gamma Minus	0.9815	0.6404/0.9958	0.9819	0.9383/0.9955
Gamma Plus	0.9578	0.5831/0.9847	0.9635	0.7052/0.9925
Delta Minus	0.9903	0.2655/0.9975	0.9750	0.9187/0.9962
Delta Plus	0.9774	0.7868/0.9962	0.9910	0.9643/0.9974
Group 2	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Alpha Minus	0.9996	0.9989/0.9999	0.9987	0.9958/0.9997
Alpha Plus	0.9980	0.9936/0.9987	0.9905	0.9660/0.9964
k	1.000	0.9999/1.0000	0.9971	0.9920/0.9992
Gamma Minus	0.9776	0.8717/0.9889	0.9857	0.9322/0.9966
Gamma Plus	0.9718	0.7659/0.9935	0.9808	0.9529/0.9950
Delta Minus	0.9934	0.9599/0.9974	0.9873	0.9701/0.9956
Delta Plus	0.9844	0.7548/0.9980	0.9929	0.9859/0.9975

Note: Reported are the median, minimum, and maximum for the three different competence groups. Very few records had to be eliminated due to technical reasons (no convergence).

Appendix L: WTP – Error Model.

All	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Underlying	0.8973	0.5843/0.9636	0.9745	0.9565/0.9838
BinaryCall	0.6968	0.4079/0.9748	0.9903	0.9825/0.9948
BinaryPut	0.7643	0.3794/0.9807	0.9880	0.9594/0.9937
Call_ITM	0.7823	0.5207/0.9657	0.9838	0.9738/0.9896
Call_OTM	0.7215	0.3835/0.9668	0.9707	0.9458/0.9834
Put_OTM	0.7226	0.4500/0.9616	0.9658	0.9173/0.9802
DiscountCertificate	0.6058	0.2429/0.9281	0.9500	0.9222/0.9740
Protected	0.7183	0.4763/0.9209	0.9427	0.9201/0.9605
Outperformance	0.9337	0.7587/0.9826	0.9792	0.9602/0.9844
Combined	0.8573	0.6347/0.9761	0.9679	0.9553/0.9759

Group 1	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Underlying	0.9724	0.9447/0.9864	0.9663	0.9200/0.9838
BinaryCall	0.9697	0.9360/0.9883	0.9856	0.9504/0.9987
BinaryPut	0.9613	0.9021/0.9889	0.9850	0.9078/0.9991
Call_ITM	0.9587	0.8879/0.9846	0.9644	0.9159/0.9883
Call_OTM	0.9332	0.8130/0.9845	0.9565	0.9000/0.9932
Put_OTM	0.9307	0.5701/0.9802	0.9557	0.8565/0.9887
DiscountCertificate	0.9317	0.9001/0.9599	0.9506	0.9069/0.9826
Protected	0.8766	0.7960/0.9410	0.8858	0.8091/0.9454
Outperformance	0.9719	0.9537/0.9832	0.9615	0.9154/0.9792
Combined	0.9171	0.8227/0.9472	0.8999	0.8600/0.9461

Group 2	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Underlying	0.9508	0.8880/0.9784	0.9751	0.9634/0.9858
BinaryCall	0.9445	0.8068/0.9874	0.9858	0.9678/0.9941
BinaryPut	0.9618	0.8218/0.9903	0.9869	0.9710/0.9938
Call_ITM	0.9516	0.7760/0.9784	0.9786	0.9584/0.9927
Call_OTM	0.9248	0.6418/0.9698	0.9674	0.9355/0.9858
Put_OTM	0.9185	0.5348/0.9782	0.9665	0.9383/0.9850
DiscountCertificate	0.8278	0.6208/0.9536	0.9517	0.9192/0.9745
Protected	0.9240	0.6827/0.9620	0.9478	0.9183/0.9687
Outperformance	0.9737	0.9422/0.9846	0.9737	0.9584/0.9842
Combined	0.9655	0.9068/0.9823	0.9466	0.9302/0.9634

Note: Reported are the median, minimum, and maximum for the three different competence groups. Very few records had to be eliminated due to technical reasons (no convergence).

Appendix M: Calibration Results for the Error Model – Relative Error.

	Certainty Equivalent			Probability Equivalent		
All						
σ	0.035	0.036	0.037	0.0170	0.0180	0.0190
Quota	17.32 %	19.76 %	21.43 %	12.74 %	13.33 %	14.13 %
Group 1						
σ	0.0290	0.0300	0.0310	0.0300	0.0310	0.0320
Quota	10.21 %	10.57 %	12.07 %	12.29 %	14.14 %	17.36 %
Group 2						
σ	0.0290	0.0300	0.0310	0.0140	0.0150	0.0160
Quota	7.45 %	10.98 %	12.08 %	6.90 %	8.11 %	8.12 %

Note: Reported are the mean quotas of FOSD violations for different standard deviations of the error for the certainty equivalent queries and the probability equivalent queries. The ones that best fit the empirical quota for the group are shaded and reported in the middle section together with the respective two adjacent values.

Appendix N: CPT Parameters – Error Model – Relative Error.

All	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Alpha Minus	0.9980	0.9950/0.9991	0.9925	0.9863/0.9958
Alpha Plus	0.9938	0.9870/0.9962	0.9943	0.9910/0.9959
k	0.9882	0.8626/0.9999	0.9950	0.9919/0.9969
Gamma Minus	0.9711	0.9472/0.9827	0.9706	0.9501/0.9872
Gamma Plus	0.9833	0.9665/0.9907	0.9788	0.9657/0.9855
Delta Minus	0.9664	0.9236/0.9829	0.9904	0.9855/0.9955
Delta Plus	0.9802	0.9060/0.9972	0.9960	0.9926/0.9976
Group 1	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Alpha Minus	0.9902	0.9812/0.9957	0.9867	0.9624/0.9977
Alpha Plus	0.9915	0.9816/0.9973	0.9837	0.9416/0.9977
k	0.9979	0.8419/1.0000	0.9921	0.9824/0.9987
Gamma Minus	0.9891	0.9641/0.9963	0.9780	0.9503/0.9931
Gamma Plus	0.9805	0.9426/0.9941	0.9584	0.9097/0.9819
Delta Minus	0.9724	0.8608/0.9965	0.9786	0.9325/0.9951
Delta Plus	0.9958	0.8848/0.9997	0.9926	0.9729/0.9977
Group 2	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Alpha Minus	0.9960	0.9870/0.9976	0.9938	0.9860/0.9975
Alpha Plus	0.9951	0.9901/0.9971	0.9919	0.9787/0.9966
k	0.9221	0.5807/0.9967	0.9953	0.9895/0.9982
Gamma Minus	0.9709	0.9433/0.9828	0.9762	0.9561/0.9930
Gamma Plus	0.9804	0.9534/0.9881	0.9754	0.9535/0.9858
Delta Minus	0.9779	0.9519/0.9900	0.9893	0.9764/0.9968
Delta Plus	0.9928	0.9356/0.9986	0.9932	0.9861/0.9970

Note: Reported are the median, minimum, and maximum for the three different competence groups. Very few records had to be eliminated due to technical reasons (no convergence).

Appendix O: WTP – Error Model – Relative Error.

All	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Underlying	0.9823	0.9786/0.9864	0.9798	0.9712/0.9862
BinaryCall	0.9897	0.9861/0.9935	0.9914	0.9854/0.9951
BinaryPut	0.9937	0.9903/0.9957	0.9939	0.9906/0.9956
Call_ITM	0.9948	0.9918/0.9970	0.9935	0.9902/0.9957
Call_OTM	0.9954	0.9914/0.9976	0.9925	0.9892/0.9952
Put_OTM	0.9953	0.9918/0.9973	0.9935	0.9909/0.9957
DiscountCertificate	0.9502	0.9411/0.9622	0.9569	0.9376/0.9725
Protected	0.9725	0.9616/0.9785	0.9559	0.9454/0.9667
Outperformance	0.9883	0.9836/0.9913	0.9828	0.9757/0.9896
Combined	0.9868	0.9780/0.9898	0.9741	0.9653/0.9820
Group 1	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Underlying	0.9748	0.9578/0.9930	0.9695	0.9368/0.9887
BinaryCall	0.9833	0.9584/0.9916	0.9879	0.9615/0.9977
BinaryPut	0.9711	0.9281/0.9951	0.9921	0.9802/0.9974
Call_ITM	0.9841	0.9536/0.9925	0.9880	0.9609/0.9961
Call_OTM	0.9816	0.9380/0.9947	0.9795	0.9325/0.9948
Put_OTM	0.9853	0.9338/0.9953	0.9826	0.9571/0.9948
DiscountCertificate	0.9405	0.9059/0.9921	0.9620	0.9130/0.9929
Protected	0.8873	0.8032/0.9436	0.8972	0.8677/0.9714
Outperformance	0.9774	0.9499/0.9929	0.9625	0.9333/0.9880
Combined	0.9280	0.8502/0.9672	0.9022	0.7870/0.9500
Group 2	Regression (R^2)		Spearman (ρ)	
	Median	Min/Max	Median	Min/Max
Underlying	0.9827	0.9769/0.9895	0.9825	0.9735/0.9902
BinaryCall	0.9934	0.9876/0.9966	0.9916	0.9819/0.9962
BinaryPut	0.9927	0.9868/0.9969	0.9918	0.9833/0.9962
Call_ITM	0.9936	0.9874/0.9968	0.9905	0.9827/0.9951
Call_OTM	0.9924	0.9840/0.9966	0.9901	0.9816/0.9944
Put_OTM	0.9950	0.9879/0.9976	0.9924	0.9861/0.9966
DiscountCertificate	0.9494	0.9350/0.9654	0.9588	0.9362/0.9778
Protected	0.9758	0.9690/0.9855	0.9613	0.9403/0.9810
Outperformance	0.9875	0.9829/0.9927	0.9795	0.9704/0.9869
Combined	0.9854	0.9801/0.9914	0.9566	0.9420/0.9748

Note: Reported are the median, minimum, and maximum for the three different competence groups. Very few records had to be eliminated due to technical reasons (no convergence).

.