Prospect Theory for Intertemporal Choice

Immanuel Lampe Matthias Weber

University of St.Gallen

Outline

Motivation

- Prospect Theory (PT) in general describes decisions under risk better than Expected Utility Theory (EUT)
- Daniel Kahneman won the Nobel Prize for his work on PT
- PT can explain several phenomena that EUT cannot explain

  - In atemporal settings (= outcomes materialize at one point in time)
  - PT in atemporal settings is well understood

Motivation

- However, many (most?) important decisions in economics and finance involve a risk and a time dimension:
  - Saving and consumption (retirement savings)
  - Asset allocation
  - Buying a house vs. renting
  - Insurance
  - Etc.
- Still unclear how to apply prospect theory when outcomes materialize at different points in time (intertemporal contexts)
  - In particular, two potential application methods mentioned in the literature

Motivation

What we do, in a nutshell:

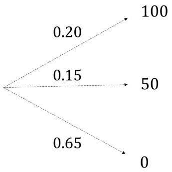
- Conduct an experiment on a representative sample
- Subjects evaluate intertemporal lotteries
- Find out which application method describes risky choices best (out-of-sample prediction performance)
- Deliver a calibration for intertemporal PT

Notes

- This is about how probability weighting should be applied, not about reference points
  - Reference points of zero by design
- We use PT with neutral propbabilities (lotteries)
  - We provide the benchmark: adaptations to specific environments are possible
  - Paper is also relevant for other theories with non-objective probabilities

Prospect Theory

- A prospect/lottery consists of outcomes arising with given probabilities, $(x_1 : p_1; ...; x_n : p_n)$, e.g., $(100 : 0.2;\ 50 : 0.15;\ 0 : 0.65)$



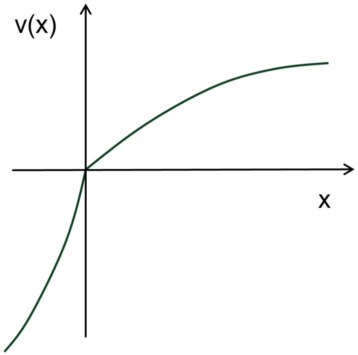
- Value of the prospect under EUT, in utility terms:

$$(x_1 : p_1; ...; x_n : p_n) \stackrel{EUT}{\rightarrow} \sum_{i=1}^{n} p_i u(x_i)$$

- Two differences in PT
  - Different utility/value function (incl. reference dependence)
  - Probability weighting

PT Value Function

- Gains and losses with respect to a reference point $R = 0$
- Kink around the reference point (loss aversion)
- Often slightly concave for gains, slightly convex for losses

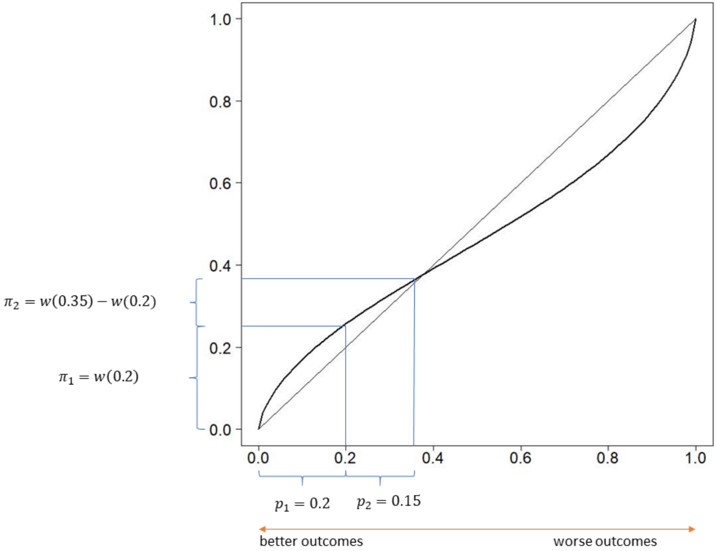


Probability Weighting

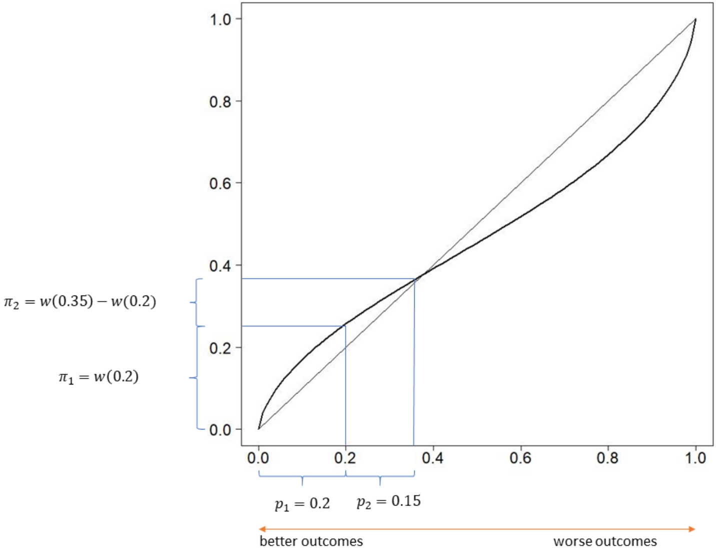
- Value of the prospect under PT, in utility/value terms:

$$(x_1 : p_1; ...; x_n : p_n) \overset{PT}{\to} \sum_{i=1}^{n} \pi_i v(x_i)$$

- How does this probability weighting work?
  - There is a weighting function w
  - The weighting is not $\pi_i = w(p_i)$



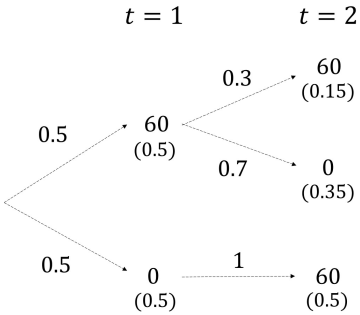
- Done separately for gains and losses



- Done separately for gains and losses

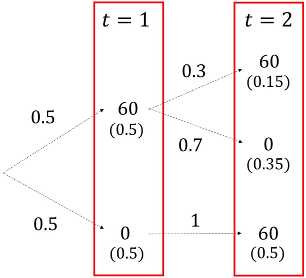
Intertemporal Prospects

- An intertemporal prospect yields (uncertain) payouts at different points in time.



Additional example

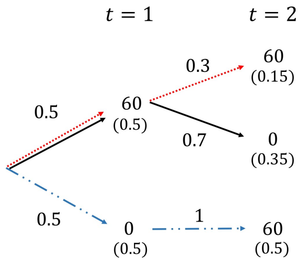
Time-Separation Method



$$PT_1 = w(0.5)v(60) + (1 - w(0.5))v(0)$$
$$PT_2 = w(0.65)v(60) + (1 - w(0.65))v(0)$$

$$PT = \delta(1)PT_1 \quad + \quad \delta(2)PT_2$$

Risk-Separation Method



$$PV_1 = \delta(1)v(60) + \delta(2)v(60)$$
$$PV_2 = \delta(1)v(60) + \delta(2)v(0)$$
$$PV_3 = \delta(1)v(0) + \delta(2)v(60)$$

$$PT = \pi_1 PV_1 + \pi_2 PV_2 + \pi_3 PV_3$$

Application Methods

- These two methods have been proposed in the literature
  - Time-separation method: e.g., Andreoni and Sprenger (2012), Krause et al. (2020)
  - Risk-separation method: e.g., Halevy (2008), Epper and Fehr-Duda (2015)
- Note: without (= with linear) probability weighting, both methods give the same results

Research Questions

1. Which way of applying PT describes risky choices best?
   - Papers with similar scope are scarce and very different (Andreoni et al., 2017; Rohde and Yu, 2024), with inconclusive results
2. What are good calibrations to apply prospect theory to intertemporal contexts?
   - Intertemporal applications usually use parametric specifications from atemporal contexts
   - Good reasons to assume that calibrations should be different in intertemporal contexts (e.g., Abdellaoui et al., 2013)

Preview of the Results

- Risk-separation method performs much better than time-separation method
- Calibration:
  - almost linear value functions (in loss and gain domains)
  - loss aversion parameter close to one
  - inverse-s shaped probability weighting functions (as in atemporal PT)
  - moderate discounting of all quarters (exponential) or distinction between now and future (quasi-hyperbolic; both versions predict equally well)

Experiment

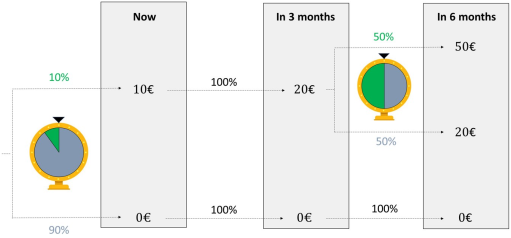
- Experiment on a sample representative for the Dutch population
  - LISS panel (Centerdata, Tilburg University)
- Study was pre-registered (data analysis follows pre-analysis plan)

Details

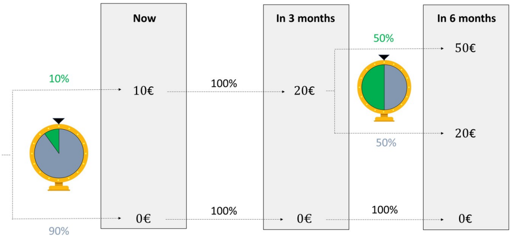
Experiment

- A total of 48 decision tasks.
- In each task subjects see a lottery ("risky option") with three uncertain payouts (today, in three months, in six months).
- We elicit the switching point from the lottery to a safe option that yields three certain and identical payouts (multi-period certainty equivalent; CE).
- For 75% of subjects hypothetical choices (T1), for the rest (T2) part of the choices incentivized
  - T1: 15 EUR for participation
  - T2: 15 EUR for participation, on average 84 EUR in addition
  - Incentivized and hypothetical choices do not differ [no strategic interaction, social image, self image]

Decision Screen



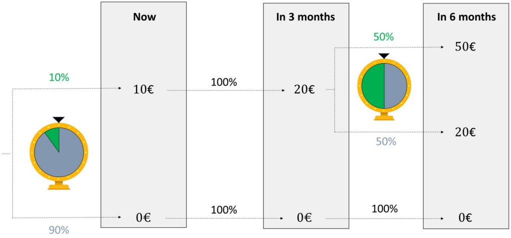
Decision Screen



I choose the **risky option** if the payout amount of the safe option that I receive three times (now, in 3 months and in 6 months) lies between:	I choose the **safe option** if the payout amount of the safe option that I receive three times (now, in 3 months and in 6 months) lies between:
0€ and **7€**	**8€** and 27€

0€ 27€

Decision Screen



I choose the **risky option** if the payout amount of the safe option that I receive three times (now, in 3 months and in 6 months) lies between:	I choose the **safe option** if the payout amount of the safe option that I receive three times (now, in 3 months and in 6 months) lies between:
0€ and **7€**	**8€** and 27€

0€ 27€

↑ CE: 7-8

Lotteries

- 6 sets of 8 lotteries each

	Gains	Losses	Mixed
Small Stakes	Set 1	Set 2	Set 3
Large Stakes	Set 4	Set 5	Set 6

48 Lotteries

36 Calibration Lotteries
For any combination of the value, weighting and discount function TS = RS
$\Rightarrow$ Estimation leads to same parameters for both methods

12 Test Lotteries (2 from each set)
1-6: TS $>$ Linear Probability Weighting $>$ RS
7-12: TS $<$ Linear Probability Weighting $<$ RS

Parametric Specification

- For each application method, we use 12 combinations of value, probability weighting, and time-discount functions

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Value	Power	x	x	x	x	x	x						
	Exponential							x	x	x	x	x	x
Weighting	T+K (1992)	x	x					x	x				
	Prelec (1998)			x	x					x	x		
	G+E (1987)					x	x					x	x
Discount	Exponential	x		x		x		x		x		x	
	Quasi hyp.		x		x		x		x		x		x

▸ Details

Maximum Likelihood Estimation (Calibration Set)

Assumptions (standard) for the log-likelihood function:

- Stated certainty equivalents are affected by noise $\epsilon_{i,j}$, with $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma_{i,j}^2)$.
- Standard deviation $\sigma_{i,j} = \epsilon_i w_j$
  - subject-specific ϵ_i
  - proportional to lotteries payout range w_j

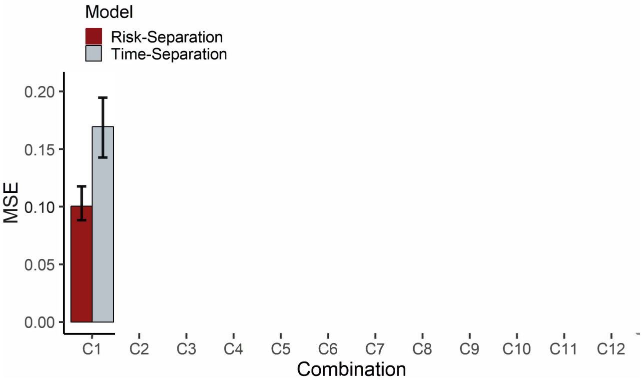
Details

Measurement of Prediction Performance (Test Set)

- **Main outcome variable:** $MSE = \frac{1}{n}\sum_{i=1}^{n} MSE_i$.
- $MSE_i = \frac{1}{12}\sum_{j=1}^{12}\left(\frac{1}{w_j}\left(CE_{i,j} - \widehat{ce_j}\right)\right)^2$
  - $\widehat{ce_j}$: predicted certainty equivalent for lottery j by given model
  - $CE_{i,j}$: certainty equivalent reported by player i for lottery j
  - w_j: payout range of lottery j

- Standard errors are bootstrapped (at the participant level)
- Tests are paired bootstrap tests

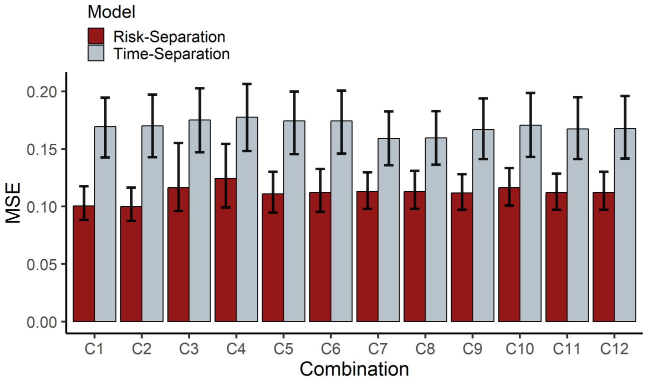
Main Result



- For any combination (C1,C2,...,C12), the risk-separation method predicts decisions better than the time-separation method [even holds for all lotteries!]

Illustration I
Illustration II

Main Result

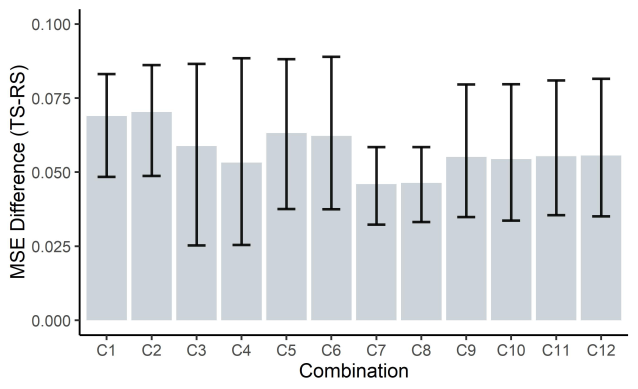


- For any combination (C1,C2,...,C12), the risk-separation method predicts decisions better than the time-separation method [even holds for all lotteries!]

▸ Illustration I

▸ Illustration II

Statistical Tests

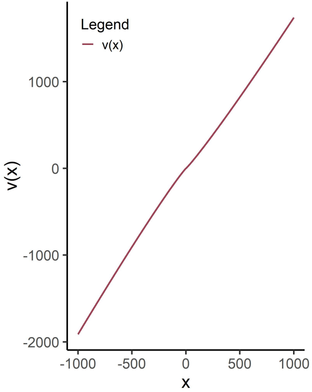


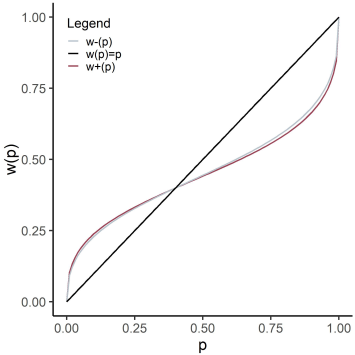
Intuition of the Main Result

- It seems more natural for people to think along outcome streams than separating time periods
  - Different outcome streams are disjoint probabilistic events, while different time periods are not
  - Separating time periods (where outcomes depend on other time periods) may be not natural
- Note that it is not necessary that people make calculations to determine their certainty equivalents (in general, they do not)

Example Calibration

- C6: i) power value function, ii) G+E (1987) probability weighting functions, iii) quasi-hyperbolic discounting)

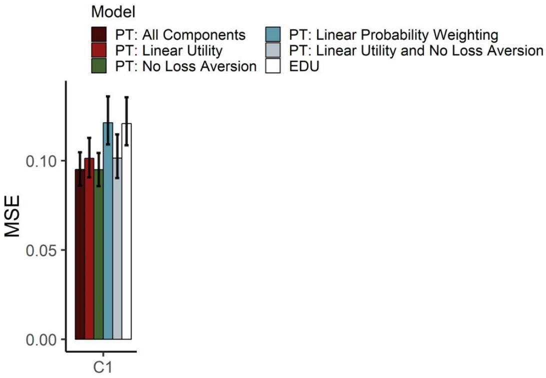




- quasi-hyperbolic disc.: $\delta(t) = k \exp(-rt)$, $k = 0.885$, $r = 0.001$

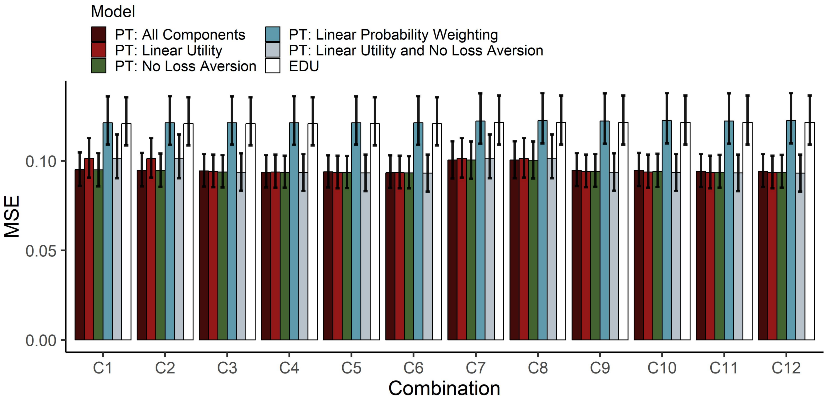
MPV

PT Components and Comparison to EDU



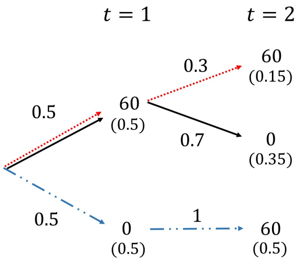
- It's all about probability weighting

PT Components and Comparison to EDU



- It's all about probability weighting

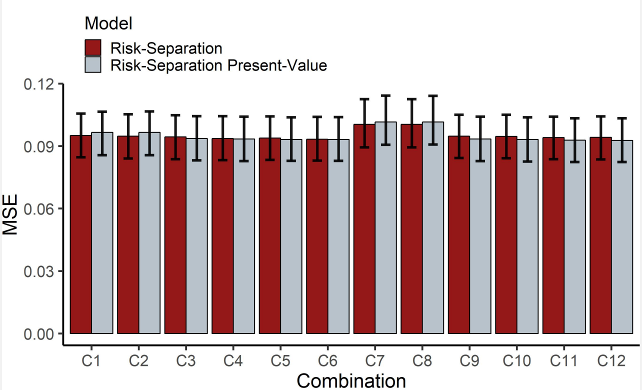
Risk-Separation Present-Value Method



$$MPV_1 = \delta(1)60 + \delta(2)60$$
$$MPV_2 = \delta(1)60 + \delta(2)0$$
$$MPV_3 = \delta(1)0 + \delta(2)60$$

$$PT = \pi_1 v(MPV_1) + \pi_2 v(MPV_2) + \pi_3 v(MPV_3)$$

Risk-Separation Present-Value Method



Sketch of an Application Example (Insurance)

Does it matter for real-world applications which method is used?

- Think of (private) long-term care insurance
- Small probabiliy that an insurable event appears in any single time period
  - Overweighting by time-separation method (overinsurance)
- Relatively large probability that an event appears at some point
  - Underweighting by risk-separation method (underinsurance)
- Underinsurance is in line with empirical evidence (e.g., Brown and Finkelstein, 2011)

Conclusion

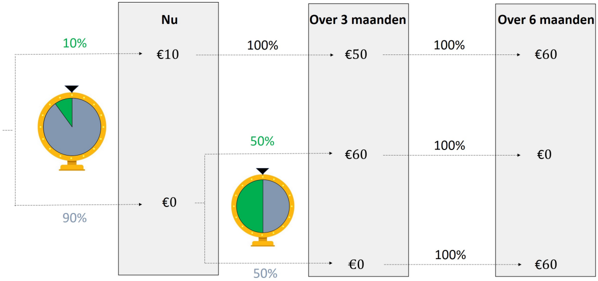
- Risk-separation method performs much better than time-separation method
- Risk-separation present-value method as good as regular risk-separation method
- Calibration:
  - almost linear value functions (in loss and gain domains)
  - loss aversion parameter close to one
  - inverse-s shaped probability weighting functions (as in atemporal PT)

Conclusion

- Does this matter in applications?
    - Yes, e.g. explaining long-term care insurance choices
- Does this matter for other theories with non-objective probabilities?
    - Yes, with both time and risk, there is always the matter time-first vs. risk-first

Thank you for your attention!

Example Lottery (Same Evaluation)



Return

Number of Subjects

- 378 subjects completed the experiment
- Data exclusion is strict (ensures that results are not driven by carelessness or misunderstandings) and follows pre-registration:
  - subjects stated comprehension difficulties or low attention in at least one post-experimental question
  - short median decision times
- Left with 100 subjects
- Main result (comparison of the methods) identical when conducted with all subjects
- Demographic variables between excluded and general subject pool very similar

Return

Function Specifications

Specification	Parameters
Utility functions	
Power utility: $v(x) = \mathbb{1}_{x \geq 0} x^{\alpha} - \mathbb{1}_{x<0} \lambda \, (-x)^{\alpha}$	α, λ
Exponential utility: $v(x) = \mathbb{1}_{x \geq 0} \frac{1-\exp(-\alpha x)}{\alpha} - \mathbb{1}_{x<0} \lambda \frac{1-\exp(\beta x)}{\beta}$	α, β, λ
Probability weighting functions (gains and losses)	
Tversky and Kahneman (1992): $w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{1/\gamma}}$	γ^+, γ^-
Prelec (1998): $w(p) = \exp(-\nu(-ln(p))^{\gamma})$	$\gamma^+, \gamma^-, \nu^+, \nu^-$
Goldstein Einhorn (1987): $w(p) = \frac{\nu p^{\gamma}}{\nu p^{\gamma}+(1-p)^{\gamma}}$	$\gamma^+, \gamma^-, \nu^+, \nu^-$
Time-discount functions	
Exponential discounting: $\delta(t) = \exp(-rt)$	r
Quasi-Hyperbolic discounting: $\delta(t) = \mathbb{1}_{t=0} + \mathbb{1}_{t>0} k \exp(-rt)$	k, r

Maximum Likelihood Procedure (Details)

- ce_j = the certainty equivalent of lottery j resulting from an evaluation under one model specification.
- $CE_{i,j}$ = certainty equivalent player j reports for lottery i.
- Assumptions (as Hey et al., 2009; Bruhin et al., 2010)
  - Noise: $CE_{i,j} = ce_j + \epsilon_{i,j}$, with $\sim \mathcal{N}(0, \sigma^2_{i,j})$.
  - SD subject specific and payout range dependent $\sigma_{i,j} = \epsilon_i w_j$.
- contribution of participant i:

$$f(\theta, \epsilon_i | CE_i) = \prod_{j=1}^{36} \frac{1}{\sigma_{i,j}} \phi \left(\frac{CE_{i,j} - ce_j(\theta)}{\sigma_{i,j}} \right)$$

- All n participants

$$log\ L(\theta, \epsilon | CE) = \sum_{i=1}^{n} log\ f(\theta, \epsilon_i | CE_i) = \sum_{i=1}^{n} \sum_{j=1}^{36} log\ \left[\frac{1}{\sigma_{i,j}} \phi \left(\frac{CE_{i,j} - ce_j(\theta)}{\sigma_{i,j}} \right) \right]$$

Return

Mean Absolute Prediction Error (By Lottery)

	Low-stake lotteries						High-stake lotteries					
	L7	L8	L15	L16	L23	L24	L31	L32	L39	L40	L47	L48
Payout range	23	20	23	20	30	50	467	400	467	400	600	1000
Mean error TS	6	8.1	7.3	5.6	7.9	16.6	144.7	148.6	146.4	124.6	180.3	319.5
Mean error RS	5.2	5.6	6.1	5.6	7.2	14.6	108.7	118.1	114.4	115.1	150.7	276.7

Notes: The mean absolute prediction error of Lottery j is calculated as mean$(|CE_{i,j} - \widehat{ce}_j|)$, with $CE_{i,j}$ denoting the certainty equivalent subject j reported for lottery j and $\widehat{ce}_j$ denoting the predicted certainty equivalent resulting from the parameters estimated on the calibration set.

Return

Participant Types

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Time-separation types	10	5	1	1	1	1	5	5	2	1	2	2
Risk-separation types	81	85	84	81	89	88	87	87	86	89	88	88
Unclassified	9	10	15	18	10	11	8	8	12	10	10	10

Notes: Subject i is classified as time-separation type if $MSE_i^{RS} - SE(\triangle MSE) > MSE_i^{TS}$ or as risk-separation type if $MSE_i^{TS} - SE(\triangle MSE) > MSE_i^{RS}$. $SE(\triangle MSE)$ denotes the standard error of the difference in MSE between the two methods.

Return

References I

Abdellaoui, M., Bleichrodt, H., l'Haridon, O., and Paraschiv, C. (2013). Is there one unifying concept of utility? An experimental comparison of utility under risk and utility over time. *Management Science*, 59(9):2153–2169.

Andreoni, J., Feldman, P., and Sprenger, C. (2017). A stream of prospects or a prospect of streams: On the evaluation of intertemporal risks. NBER Working Paper No. 24075.

Andreoni, J. and Sprenger, C. (2012). Risk preferences are not time preferences. *American Economic Review*, 102(7):3357–76.

Brown, J. R. and Finkelstein, A. (2011). Insuring long-term care in the United States. *Journal of Economic Perspectives*, 25(4):119–42.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4):1375–1412.

References II

Epper, T. and Fehr-Duda, H. (2015). Risk preferences are not time preferences: Balancing on a budget line: Comment. *American Economic Review*, 105(7):2261–71.

Halevy, Y. (2008). Strotz meets Allais: Diminishing impatience and the certainty effect. *American Economic Review*, 98(3):1145–62.

Hey, J. D., Morone, A., and Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty*, 39(3):213–235.

Krause, J., Ring, P., and Schmidt, U. (2020). Hyperopic loss aversion. SSRN Working Paper No. 3582050.

Rohde, K. I. M. and Yu, X. (2024). Intertemporal correlation aversion—a model-free measurement. *Management Science*, 70(6):3493–3509.