Continuity, Convexity, Centrality: Axiomatizing and Comparing Location Indices

Stefania Merone

University of Oxford
Nuffield College

Big Picture

- Ultimately interested in the Fourfold Pattern of Risk Attitudes (FFP):
  - Risk-seeking on low chance gains, high chance losses; risk-averse otherwise
- Several related empirical puzzles:
  - Why does the FFP arises?
  - Why is it related to evaluation tasks only? (Harbaugh et al., 2010)
  - Why does it arise in riskless environments? (Oprea, 2024)
- Prospect theory can only explain the first point
- My solution: modelling CEs as location indices
  - Location index: single value best representing numerical data
  - Mean, median, quantiles, minimum, maximum...
  - Measures the position of a distribution on the real line

Certainty Equivalents as Location Indices

- The relation between means and certainty equivalents has deep roots:
  - Huygens (1657) set the fair price of a lottery equal to its expected value
  - Bernoulli (1738) moved beyond nominal values ("moral expectation")
  - Laplace (1820) linked moral expectations and individual preferences
  - von Neumann and Morgenstern (1944) acknowledged that an axiomatic treatment of utility is fundamentally a choice regarding the appropriate "mean"

- Duality between Kolmogorov-Nagumo (1930) and vNM axioms:
  - de Finetti and Emanuelli (1967) noticed the equivalence between axioms on quasi-linear means by Kolmogorov (1930) - Nagumo (1930) and EU axioms
  - Direct mathematical mapping between the Associativity Axiom in the theory of means and the Independence Condition in decision theory

- Since then, the relation has been further generalized:
  - Hong (1983) moved beyond quasi-linear means to address income inequality and decision-theoretic anomalies like the Allais Paradox
  - Fishburn (1986) formally unified certainty equivalence and mean values, providing a foundation for non-EU representations

What I do

1. Axiomatic foundation of location indices obtained by minimizing continuous and (strictly) convex loss functions
   - Statistics: OLS, WLS, maximum log-likelihood principle...
   - ML: Hedge algorithm, OGD, FTRL, SVM, convex surrogates...

2. Analysis of geometrical properties of the loss function
   - Asymmetry: uneven penalization of over- vs under-prediction
   - Convexity: sensitivity to outliers (Hampel, 1974)

3. Behavioral interpretation of such features
   - Origin asymmetry: first order risk preferences (Segal and Spivak, 1997)
   - Convexity: centrality preferences

4. FFP as interaction between risk and centrality preferences

What I do (in this talk)

1. Axiomatic foundation of location indices obtained by minimizing continuous and (strictly) convex loss functions ✓
   - Statistics: OLS, WLS, maximum log-likelihood principle...
   - ML: Hedge algorithm, OGD, FTRL, SVM, convex surrogates...

2. Analysis of geometrical properties of the loss function ✗
   - Asymmetry: uneven penalization of over- vs under-prediction
   - Convexity: sensitivity to outliers (Hampel, 1974)

3. Behavioral interpretation of such features ✗
   - Origin asymmetry: first order risk preferences (Segal and Spivak, 1997)
   - Convexity: centrality preferences

4. FFP as interaction between risk and centrality preferences ✗

Setting (1)

- $D = \{x_1, \ldots, x_n; p_1, \ldots, p_n\}$ is a *dataset*
  - $x_i \in \mathbb{R}$ are *observations* s.t. $x_1 < \cdots < x_n$
  - $p_i > 0$ are *relative frequencies* i.e. $\sum_{i=1}^{n} p_i = 1$
  - n is the number of observations
- The space of all D is $\mathcal{D}$
  - The space of all D with $n \leq 2$ is $\ddot{\mathcal{D}}$
  - $\delta_x \in \mathcal{D}$ denotes the degenerate dataset $\{x; 1\}$
- The primitive is the **location mapping** $m : \mathcal{D} \to \mathbb{R}$
  - Coarser yet easier description of the data
  - Information loss is costly
  - The cost of representing x with y is $c(x, y)$

Setting (2)

- $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ is:
  1. continuous in each argument
  2. strictly convex in its second argument
  3. s.t. $c(x, y) = 0 \Leftrightarrow x = y$
  4. s.t. $c(x, y) = c(x + k, y + k)$ for any $k \in \mathbb{R}$

- The set of functions satisfying these conditions is $\mathcal{C}$

- On each dataset D, n different errors are committed

- The cost of representing D with $\{y; 1\}$ is:

$$C(D, y) := \sum_{i=1}^{n} p_i c(x_i, y)$$

- A loss function $c(x, y)$ **rationalizes** the map $m(D)$ if:

$$m(D) = \arg\min_{y} C(D, y)$$

Axioms

Axiom 1 (Certainty Matching): $m(\{x;1\}) = x$

Axiom 2 (Continuity): For any $\{D_k\}_{k\in\mathbb{N}} \subseteq \ddot{\mathcal{D}}$ and $D \in \ddot{\mathcal{D}}$

$$D_k \to D \;\Rightarrow\; m(D_k) \to m(D)$$

Axiom 3 (Dominance): $\forall\, D, D' \in \ddot{\mathcal{D}},\; D \succ_{FOSD} D' \to m(D) > m(D')$

Axiom 4 (Translation Invariance): $\forall\, k \in \mathbb{R}, D \in \ddot{\mathcal{D}},\; m(D+k) = m(D) + k$

Axiom 5 (Odds Independence): For any $x_1 < x_2 < y < x_3 < x_4$, $\{x_2, x_3; p\}$,

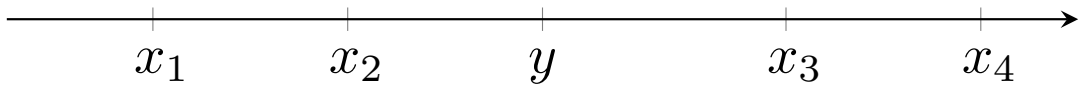
$\{x_2, x_4; q\}, \{x_1, x_3; r\}, \{x_1, x_4; s\}$ valued $y \Rightarrow$

$$\frac{q}{1-q} \cdot \frac{r}{1-r} = \frac{s}{1-s} \cdot \frac{p}{1-p}$$

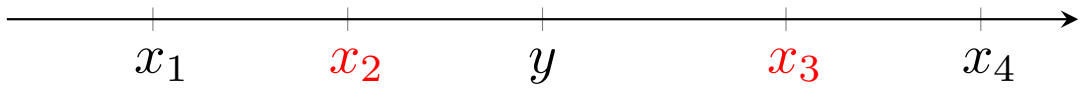
Axiom 6 (Reduction): $\forall\, D \in \mathcal{D}$ s.t. $m(D) = y$ and $n > 2$, $\exists\, D'$ s.t. $m(D') = y$ and $n' = n - 1$. The construction of D' follows one of two rules:

- Reduction by Cancellation
- Reduction by Aggregation

Odds Independence — Visualization

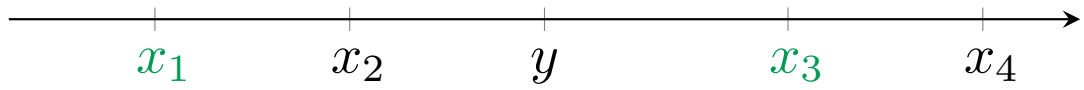


Odds Independence — Visualization



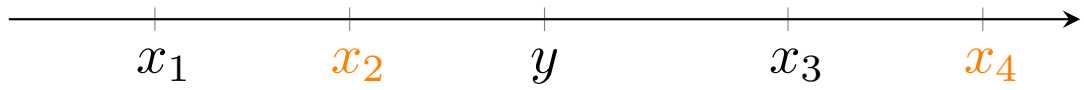
- $D = \{x_2, x_3; p, 1-p\}$

Odds Independence — Visualization



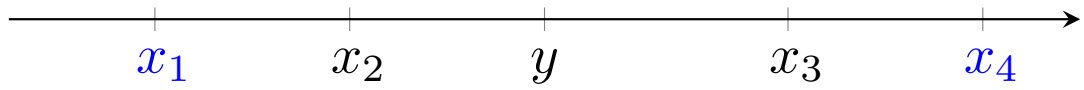
- $D = \{x_2, x_3; p, 1-p\}$
- $D' = \{x_1, x_3; r, 1-r\}$

Odds Independence — Visualization



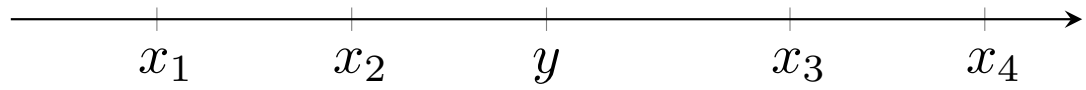
- $D = \{x_2, x_3; p, 1-p\}$
- $D' = \{x_1, x_3; r, 1-r\}$
- $D'' = \{x_2, x_4; q, 1-q\}$

Odds Independence — Visualization



- $D = \{x_2, x_3; p, 1-p\}$
- $D' = \{x_1, x_3; r, 1-r\}$
- $D'' = \{x_2, x_4; q, 1-q\}$
- $D''' = \{x_1, x_4; s, 1-s\}$

Odds Independence — Visualization



- $D = \{x_2, x_3; p, 1-p\}$
- $D' = \{x_1, x_3; r, 1-r\}$
- $D'' = \{x_2, x_4; q, 1-q\}$
- $D''' = \{x_1, x_4; s, 1-s\}$

$$m(D) = m(D') = m(D'') = m(D''') \Rightarrow \frac{\frac{p}{1-p}}{\frac{r}{1-r}} = \frac{\frac{q}{1-q}}{\frac{s}{1-s}}$$

Reduction by Cancellation

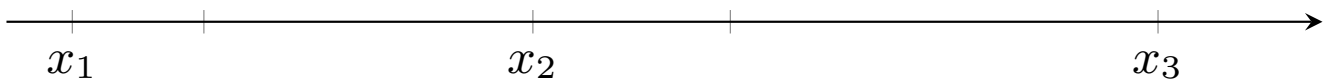
- If $y = x_k$ for some $x_k \in supp(D)$, then:

$$D' = \left\{ \{x_i\}_{i \neq k} \, ; \, \left\{ \frac{p_i}{1 - p_k} \right\}_{i \neq k} \right\}$$

- If $D = \{1, 2, 3; \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$ and $m(D) = 2$, then $D' = \{1, 3; \frac{1}{2}, \frac{1}{2}\}$

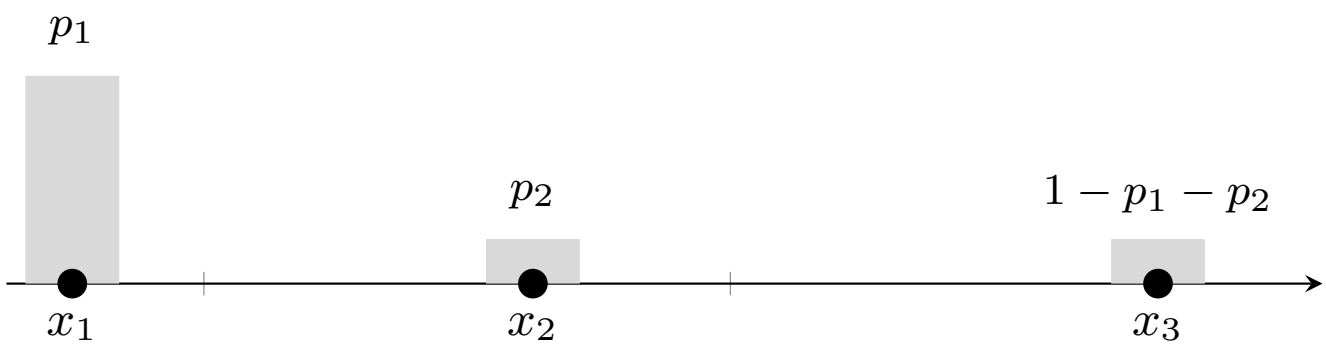
Reduction by Aggregation (1)

- Let $D = \{x_1, x_2, x_3; p_1, p_2\}$ with $m(D) = y \notin supp(D)$



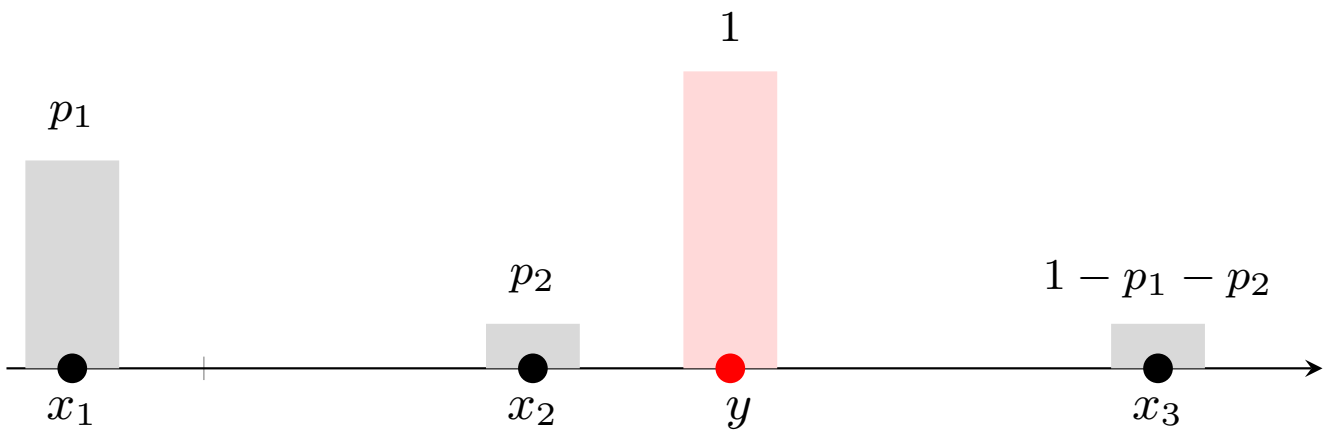
Reduction by Aggregation (1)

- Let $D = \{x_1, x_2, x_3; p_1, p_2\}$ with $m(D) = y \notin supp(D)$



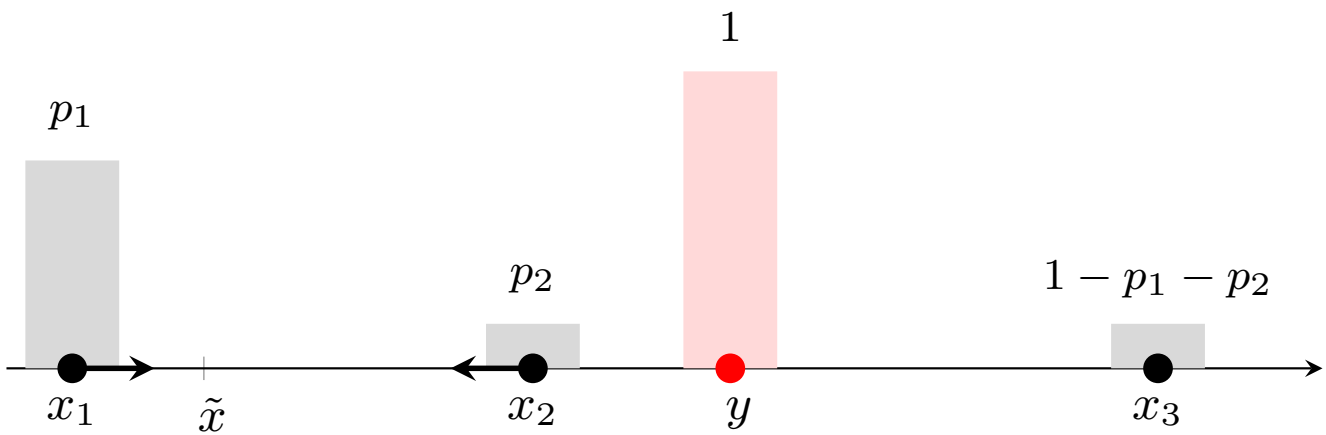
Reduction by Aggregation (1)

- Let $D = \{x_1, x_2, x_3; p_1, p_2\}$ with $m(D) = y \notin supp(D)$



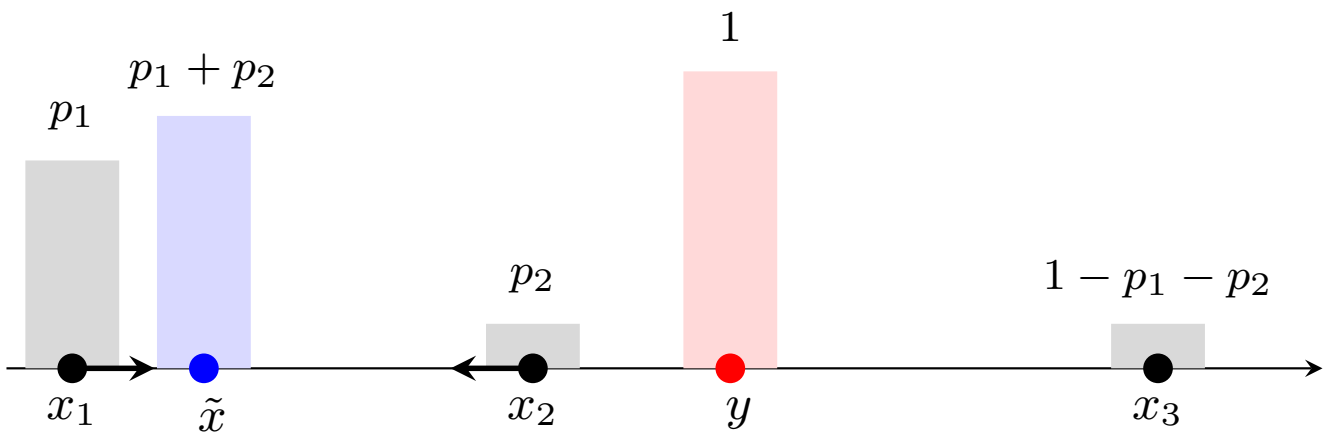
Reduction by Aggregation (1)

- Let $D = \{x_1, x_2, x_3; p_1, p_2\}$ with $m(D) = y \notin supp(D)$



Reduction by Aggregation (1)

- Let $D = \{x_1, x_2, x_3; p_1, p_2\}$ with $m(D) = y \notin supp(D)$



Reduction by Aggregation (2)

If $y \notin \text{supp}(D)$, two adjacent values $(x_i, x_{i+1}) \in supp(D)$ **lying on the same side** of y are replaced by a unique value $\tilde{x}$ with frequency $(p_i + p_{i+1})$

To determine $\tilde{x}$:

- Fix an anchor k strictly on the opposite side of y from the pair
- for each x on the pair's side, let $s_y(x;k) \in (0,1)$ be the (unique) weight on x such that $m\{s_y(x;k)\,\delta_x + [1 - s_y(x;k)]\,\delta_k\} = y$
- Then, $\tilde{x}$ is the unique value satisfying:

$$\frac{1 - s_y(\tilde{x};k)}{s_y(\tilde{x};k)} = \frac{p_i}{p_i + p_{i+1}} \cdot \frac{1 - s_y(x_i;k)}{s_y(x_i;k)} + \frac{p_{i+1}}{p_i + p_{i+1}} \cdot \frac{1 - s_y(x_{i+1};k)}{s_y(x_{i+1};k)} \quad (1)$$

Reduction by Aggregation (3)

- If $m(D) \notin supp(D)$, Reduction is a *similarity-preserving* principle
- $s_y(x;k)$ measures the similarity of x and y: the more weight x must carry to drag a two-point location index onto y, the closer x and y are
- The corresponding odds $\frac{1-s_y(x;k)}{s_y(x;k)}$ are a dissimilarity index
- Equation (1) requires the similarity of $\tilde{x}$ and y to be the weighted average of the one of x_i and y and x_{i+1} and y, respectively

Representation Theorem

Theorem

The following are equivalent:

1. *m satisfies Axioms 1-6*
2. *$\exists\, c \in \mathcal{C}$ such that:*

$$m(D) = \arg\min_{y} \sum_{i=1}^{n} p_i \, c(x_i, y) \quad \forall\, D \in \mathcal{D}$$

Also, c is unique up to any positive multiplicative constant

- Axiomatic foundation of translation-invariant M-estimators (Huber, 1964)
- $m(D)$ is the optimal forecast for a given scoring function $c(x, y)$
- Asymmetric version of Fishburn (1986) (not nested)

Proof Sketch — Sufficiency (1)

1. Define frequencies:

- $\forall\, x^- < 0$, find p^- s.t. $m(\{x^-, 1, p^-\}) = 0$
- $\forall\, x^+ > 0$, find p^+ s.t. $m(\{\hat{x}^-, x^+, p^+\}) = 0$ for a $\hat{x}^-$ (irrelevant by *Odds Independence*)
- *Continuity* and *Dominance* guarantee existence and uniqueness

2. Constructing c' around 0:

- Normalize $c'(1, 0) = -1$
- Using FOC: $c'_l(x, 0) = \frac{1-p^-}{p^-}$
- Using FOC: $c'_r(x, 0) = -\frac{p^+}{1-p^+} \cdot \frac{1-\hat{p}^-}{\hat{p}^-}$

3. Generalize to any center:

- Use *Translation Invariance* to shift the centre from 0 to any value y

Proof Sketch — Sufficiency (2)

4. Prove c' is well-behaved:

- We show that c' is:
  - continuous in each argument
  - strictly decreasing in x
  - strictly increasing in y

5. Get total loss function:

- Integrate $c'(x, y)$ w.r.t. y to get $c(x, y)$
- Prove that $c \in \mathcal{C}$
- In particular, *Certainty Matching* ensures $c(x, y) \geq 0$, with equality iff $x = y$

6. Induction step:

- If c rationalizes m for a dataset with n entries, *Reduction* guarantees it also rationalizes m for a dataset with $n + 1$ entries

Conclusions

- I provide an axiomatic characterization of location indices as minimizers of continuous, (strictly) convex loss functions
- Why this matters:
  1. Provides a unified framework connecting decision theory, statistics, and machine learning
  2. Establishes that the choice of location index is fundamentally a choice about how to penalize prediction errors
  3. Grounds the behavioral study of risk preferences in the geometrical features of loss functions
- Anomalies in decision making under risk may reflect departures from the implicit loss function assumed by classical EU theory

Related Literature

- *Means*: Kolmogorov (1930), Nagumo (1930), Nagumo (1931), De Finetti (1931), Fishburn (1986), Hong (1983), Muliere and Parmigiani (1993), Chambers (2009), De Carvalho (2016), Singpurwalla and Lai (2020), Bellini and Peri (2022)

- *Robust Estimators*: Huber (1964), Breckling and Chambers (1988), van der Vaart and Wellner (1996), Chen (1996), Newey and Powell (1987), Ziegel (2016)

- *Quantiles and Expectiles*: Koenker and Bassett Jr (1978), Pennecchi and Callegaro (2006), Rostek (2010), Ehm et al. (2016), Bellini and Di Bernardino (2017), Daouia et al. (2020), Philipps (2022), Bellini et al. (2025)

- *Certainty Equivalents*: Huygens (1657), Bernoulli (1738), **?**, von Neumann and Morgenstern (1944), de Finetti and Emanuelli (1967), Kahneman et al. (1979) Tversky and Kahneman (1992), Harbaugh et al. (2010), Oprea (2024), Roorda and Joosten (2025)

Thank you!

stefania.merone@economics.ox.ac.uk

References I

F. Bellini and E. Di Bernardino. Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506, 2017.

F. Bellini and I. Peri. An axiomatization of λ-quantiles. *SIAM Journal on Financial Mathematics*, 13(1):SC26–SC38, 2022.

F. Bellini, F. Maccheroni, T. Mao, R. Wang, and Q. Wu. Disappointment aversion and expectiles. *arXiv preprint arXiv:2508.05541*, 2025.

D. Bernoulli. Specimen theoriae novae de mensura sortis. 1738.

J. Breckling and R. Chambers. M-quantiles. *Biometrika*, 75(4):761–771, 1988.

C. P. Chambers. An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(2):335–342, 2009.

Z. Chen. Conditional lp-quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & probability letters*, 29(2):107–115, 1996.

A. Daouia, S. Girard, and G. Stupfler. Tail expectile process and risk assessment. 2020.

M. De Carvalho. Mean, what do you mean? *The American Statistician*, 70(3): 270–274, 2016.

B. De Finetti. *Sul concetto di media*. Istituto italiano degli attuari, 1931.

B. de Finetti and F. Emanuelli. Economia delle assicurazioni, trattato italiano di economia, vol. *XVI, UTET, Torino*, 1967.

References II

W. Ehm, T. Gneiting, A. Jordan, and F. Krüger. Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(3): 505–562, 2016.

P. C. Fishburn. Implicit mean value and certainty equivalence. *Econometrica: Journal of the Econometric Society*, pages 1197–1205, 1986.

F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393, 1974.

W. T. Harbaugh, K. Krause, and L. Vesterlund. The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, 120(545):595–611, 2010.

C. S. Hong. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. *Econometrica: Journal of the Econometric Society*, pages 1065–1092, 1983.

P. J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, Mar 1964. doi: 10.1214/aoms/1177703732. URL https://doi.org/10.1214/aoms/1177703732.

C. Huygens. *De ratiociniis in ludo aleae*. Ex officinia J. Elsevirii, 1657.

D. Kahneman, A. Tversky, et al. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391, 1979.

R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

References III

A. Kolmogorov. ksur la notion de la moyenne, latti accad. *Naz. Lincei Rend*, 12: 388, 1930.

P. S. Laplace. *Théorie analytique des probabilités*, volume 7. Courcier, 1820.

P. Muliere and G. Parmigiani. Utility and means in the 1930s. *Statistical Science*, pages 421–432, 1993.

M. Nagumo. Über eine klasse der mittelwerte. In *Japanese journal of mathematics: transactions and abstracts*, volume 7, pages 71–79. The Mathematical Society of Japan, 1930.

M. Nagumo. On mean values. *Tokyo Buturigakko-Zassi*, 40:520–527, 1931.

W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

R. Oprea. Decisions under risk are decisions under complexity. *American Economic Review*, 114(12):3789–3811, 2024.

F. Pennecchi and L. Callegaro. Between the mean and the median: the lp estimator. *Metrologia*, 43(3):213, 2006.

C. Philipps. Interpreting expectiles. *Available at SSRN*, 3881402, 2022.

B. Roorda and R. A. Joosten. Prospect theory in betweenness form for binary lotteries. 2025.

M. Rostek. Quantile maximization in decision theory. *The Review of Economic Studies*, 77(1):339–371, 2010.

U. Segal and A. Spivak. First-order risk aversion and non-differentiability. *Economic Theory*, 9(1):179–183, 1997.

References IV

N. D. Singpurwalla and B. Lai. What does the" mean" really mean? *arXiv preprint arXiv:2003.01973*, 2020.

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

A. W. van der Vaart and J. A. Wellner. M-estimators. In *Weak Convergence and Empirical Processes: With Applications to Statistics*, pages 284–308. Springer, 1996.

J. von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1944.

J. F. Ziegel. Coherence and elicitability. *Mathematical Finance*, 26(4):901–918, 2016.