# Characterizing, optimizing and backtesting metrics of risk

by

Qiuqi Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2023

**Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

| | |
|---|---|
| External Examiner: | Roger Laeven<br>Professor, Department of Quantitative Economics,<br>University of Amsterdam |
| Supervisor: | Ruodu Wang<br>Professor, Department of Statistics and Actuarial Science,<br>University of Waterloo |
| Internal Member: | Jun Cai<br>Professor, Department of Statistics and Actuarial Science,<br>University of Waterloo |
| Internal Member: | Alexander Schied<br>Professor, Department of Statistics and Actuarial Science,<br>University of Waterloo |
| Internal-External Member: | Kateryna Tatarko<br>Assistant Professor, Department of Pure Mathematics,<br>University of Waterloo |

## Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Measures of risk and riskmetrics were proposed to quantify the risks people are faced with in financial, statistical, and economic practice. They are widely discussed and studied by literature in the context of financial regulation, insurance, operations research, and statistics. Several major research topics on riskmetrics remain to be important in both academic study and industrial practice. First, characterization, especially axiomatic characterization of riskmetrics, lays essential theoretical foundation of specific classes of riskmetrics about why they are widely adopted in practice and research. It usually involves challenging mathematical approaches and deep practical insights. Second, riskmetrics are used by researchers in optimization as the objective functionals of decision makers. This links riskmetrics to the literature of operations research and decision theory, and leads to wide applications of riskmetrics to portfolio management, robust optimization, and insurance design. Third, relevant statistical models of estimation and hypothesis tests for riskmetrics need to be established to serve for practical risk management and financial regulation. In particular, risk forecasts and backtests of different riskmetrics are always the main concern and challenge for risk managers and financial regulators. In this thesis, we investigate several important questions in characterization, optimization, and backtest for measures of risk with different focuses on establishing theoretical framework and solving practical problems.

To offer a comprehensive theoretical toolkit for future study, in Chapter 2, we propose the class of distortion riskmetrics defined through signed Choquet integrals. Distortion riskmetrics include many classic risk measures, deviation measures, and other functionals in the literature of finance and actuarial science. We obtain characterization, finiteness, convexity, and continuity results on general model spaces, extending various results in the existing literature on distortion risk measures and signed Choquet integrals.

To explore deeper applications of distortion riskmetrics in optimization problems, in Chapter 3, we study optimization of distortion riskmetrics with distributional uncertainty. One of our central findings is a unifying result that allows us to convert an optimization of a non-convex distortion riskmetric with distributional uncertainty to a convex one, leading to practical tractability. A sufficient condition to the unifying equivalence result is the novel notion of closedness under concentration, a variation of which is also shown to be necessary for the equivalence. Our results include many special cases that are well studied in the optimization literature, including but not limited to optimizing probabilities, Value-at-Risk, Expected Shortfall, Yaari's dual utility, and differences between

distortion risk measures, under various forms of distributional uncertainty. We illustrate our theoretical results via applications to portfolio optimization, optimization under moment constraints, and preference robust optimization.

In Chapter 4, we study characterization of measures of risk in the context of statistical elicitation. Motivated by recent advances on elicitability of risk measures and practical considerations of risk optimization, we introduce the notions of Bayes pairs and Bayes risk measures. Bayes risk measures are the counterpart of elicitable risk measures, extensively studied in the recent literature. The Expected Shortfall (ES) is the most important coherent risk measure in both industry practice and academic research in finance, insurance, risk management, and engineering. One of our central results is that under a continuity condition, ES is the only class of coherent Bayes risk measures. We further show that entropic risk measures are the only risk measures which are both elicitable and Bayes. Several other theoretical properties and open questions on Bayes risk measures are discussed.

In Chapter 5, we further study characterization of measures of risk in insurance design. We study the characterization of risk measures induced by efficient insurance contracts, i.e., those that are Pareto optimal for the insured and the insurer. One of our major results is that we characterize a mixture of the mean and ES as the risk measure of the insured and the insurer, when contracts with deductibles are efficient. Characterization results of other risk measures, including the mean and distortion risk measures, are also presented by linking them to different sets of contracts.

In Chapter 6, we focus on a larger class of riskmetrics, cash-subadditive risk measures. We study cash-subadditive risk measures without quasi-convexity. One of our major results is that a general cash-subadditive risk measure can be represented as the lower envelope of a family of quasi-convex and cash-subadditive risk measures. Representation results of cash-subadditive risk measures with some additional properties are also examined. The notion of quasi-star-shapedness, which is a natural analogue of star-shapedness, is introduced and we obtain a corresponding representation result.

In Chapter 7, we discuss backtesting riskmetrics. One of the most challenging tasks in risk modeling practice is to backtest ES forecasts provided by financial institutions. To design a model-free backtesting procedure for ES, we make use of the recently developed techniques of e-values and e-processes. Model-free e-statistics are introduced to formulate e-processes for risk measure forecasts, and unique forms of model-free e-statistics for VaR and ES are characterized using recent results on identification functions. For a given model-free e-statistic, optimal ways of constructing the e-processes are studied. The proposed method can be naturally applied to many other risk measures and statistical quantities. We conduct extensive simulation studies and data analysis to illustrate the advantages of the model-free backtesting method, and compare it with the ones in the literature.

## Acknowledgements

I would like to thank all people that accompanied, helped, encouraged, and cared about me during the four-year time of my Ph.D. at the University of Waterloo.

First and foremost, I would like to express my sincere gratitude to my Ph.D. supervisor, Dr. Ruodu Wang, who is the most supportive, considerate, and conscientious talent I have ever met in my life. During the four years from my Ph.D. admission to my job seeking, he offered me selfless support of my academic career and personal life. His wisdom, patience, guidance, and commitment make me a better researcher and make my academic dream possible. He is a rigorous researcher reinforcing my academic skills, a responsible teacher leading my career directions, a kind senior listening to my obstacles and confusions, and an approachable friend enriching my recesses. I feel honored and fortunate to have him as my supervisor for my Ph.D., probably the most important period of time for my whole career.

I wish to convey my special gratitude to my M.Phil. supervisor, Dr. Yue Kuen Kwok, whom I regard and respect as one of the most important mentors in my life. His greatest kindness and help lead my way out of the darkest time of my college life, and to the brilliant world of academic research. His knowledge, attitude and personality will always be a good example of my academic career.

I would like to thank all my proposal and defense committee members, Dr. Roger Laeven, Dr. Jun Cai, Dr. Alexander Schied, and Dr. Kateryna Tatarko for spending their valuable time reading my thesis and providing insightful comments. I always find their expertise and suggestions inspiring for enhancing my current research and spotting my drawbacks.

I am always grateful for all the guidance, discussion, and support I received from faculty members in or outside Waterloo during my Ph.D. life. I would like to thank my co-authors, Dr. Paul Embrechts, Dr. Xia Han, Dr. Tiantian Mao, Dr. Silvana Pesenti, Dr. Yunran Wei, Dr. Jianming Xia, Dr. Johanna Ziegel, and Dr. Ričardas Zikitis, for their supportive instruction and inspiring discussion. Special thanks go to Dr. Qihe Tang for his generous support and suggestion for my academic career. I am grateful to meet with and get supported by Dr. Phelim Boyle, Dr. Mary Hardy, Dr. David Landriault, Dr. Bin Li, Dr. Mario Ghossoub, Dr. Fangda Liu, Dr. Haiyan Liu, Dr. Peng Liu, Dr. Hong Li, Dr. Zhenzhen Fan, and Dr. Peijun Sang. They are all kind and supportive mentors and seniors who played an important role of my Ph.D. time. I would like to thank Ms. Mary Lou Dufton, Mr. Greg Preston, and Mr. Steve Van Doormaal for their great patience and help with my administrative issues, conference organizing, and teaching duties.

I would like send my thanks and wishes to all my friends during my Ph.D. for their accompanying: Zhanyi Jiao, Mingren Yin, Yimiao Zhao, Yang Liu, Liyuan Lin, Peiqi Wang, Xiyue Han, Bingzhen Geng, Yiping Guo, Zhiqiao Song, Yumin Wang, Zhiyi Shen, Feiyu Zhu, Wenyuan Li, Yuyu Chen, Meng Yuan, Meixi Chen, Michael Boyuan Zhu, Jiayue Zhang, Yuanyuan Zhang, Xintong Li, Ziyue Shi, Benxuan Shi, Yechao Meng, Chi-Kuang Yeh, Jingyue Huang, Kecheng Li, Zhenzhen Huang, Zhenyuan Zhang, Ziyu Chi, Dongxu Ma, Qinyu Wu, Yuyu He, Zachary John Van Oosten, Jean-Gabriel Lauzier, Qinghua Ren, Yuxuan Guo, Yixuan Fan, Ying Wu, Hongjie He, ...... They live in almost all of my delightful memories during the four years. Special thanks go to Dr. Zhiwei Tong for his friendly and insightful connection and help during his visit and my job seeking period.

Last but not least, I would like to thank my parents, Jinming Wang and Wei Fu, for their endless love, and their understanding and support to my study and career. I hope I could spend more time accompanying them in the future.

## Dedication

*To my family*

# Table of Contents


**List of Figures** **xvi**

**List of Tables** **xviii**

**1 Introduction and preliminary** **1**

1.1 Introduction . . . . . . . . . . 1

1.2 Preliminary . . . . . . . . . . 4

**2 Distortion riskmetrics on general spaces** **7**

2.1 Introduction . . . . . . . . . . 7

2.2 Distortion riskmetrics and their characterization . . . . . . . . . . 11

2.2.1 Notation and definition . . . . . . . . . . 11

2.2.2 Quantile representation and finiteness of signed Choquet integrals . . . . . . 12

2.2.3 Characterization and basic properties . . . . . . . . . . 13

2.3 Convexity, convex order consistency and mixture concavity . . . . . . . . . . 14

2.4 Continuity of distortion riskmetrics . . . . . . . . . . 17

2.5 Multi-dimensional distortion riskmetrics . . . . . . . . . . 18

2.6 Proofs of all results . . . . . . . . . . 20

**3 Optimizing distortion riskmetrics with distributional uncertainty** **31**

3.1 Introduction . . . . . . . . . . 31

3.2 Distortion riskmetrics with distributional uncertainty . . . . . . . . . . . . . . . . . . . 34
3.2.1 Problem formulation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.2.2 Notation and preliminaries . . . . . . . . . . . . . . . . . . . . . . . . . . 36
3.3 Equivalence between non-convex and convex riskmetrics . . . . . . . . . . . . . . . . . 38
3.3.1 Concentration and the main equivalence result . . . . . . . . . . . . . . . . . . 38
3.3.2 Some examples of distortion riskmetrics . . . . . . . . . . . . . . . . . . . . 40
3.3.3 Closedness under concentration for all intervals . . . . . . . . . . . . . . . . . 43
3.3.4 Examples of closedness under concentration within $\mathcal{I}$ but not for all intervals 45
3.3.5 Atomic probability space . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
3.4 Multi-dimensional setting . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.5 One-dimensional uncertainty set with moment constraints . . . . . . . . . . . . . . . 53
3.6 Related optimization problems . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.6.1 Portfolio optimization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.6.2 Preference robust optimization . . . . . . . . . . . . . . . . . . . . . . . . . 56
3.7 Applications and numerical illustrations . . . . . . . . . . . . . . . . . . . . . . . 57
3.7.1 Difference of risk measures under moment constraints . . . . . . . . . . . . . . 57
3.7.2 Preference robust portfolio optimization with moment constraints . . . . . . 58
3.7.3 Portfolio optimization with marginal constraints . . . . . . . . . . . . . . . . 60
3.8 Concluding remarks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
3.9 Omitted technical details from the chapter . . . . . . . . . . . . . . . . . . . . . . 65
3.9.1 Proofs of claims in some Examples . . . . . . . . . . . . . . . . . . . . . . . 65
3.9.2 A few additional technical remarks mentioned in the chapter . . . . . . . . . 67
3.10 Proofs of all technical results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
3.10.1 Proof of results in Section 3.2 . . . . . . . . . . . . . . . . . . . . . . . . . 68
3.10.2 Proofs of results in Section 3.3 . . . . . . . . . . . . . . . . . . . . . . . . . 69
3.10.3 Proofs of results in Section 3.4 . . . . . . . . . . . . . . . . . . . . . . . . . 76
3.10.4 Proofs of results in Section 3.5 and related lemmas . . . . . . . . . . . . . . . 77

**4 Bayes risk, elicitability, and the Expected Shortfall** **80**

4.1 Introduction . . . 80

4.2 Bayes pairs and Bayes risk measures . . . 82

4.2.1 Risk measures . . . 82

4.2.2 Bayes pairs and Bayes risk measures . . . 83

4.2.3 Examples of Bayes pairs . . . 86

4.3 Characterizing ES as a Bayes risk measure . . . 87

4.4 Elicitability of Bayes risk measures . . . 91

4.5 Other properties of Bayes risk measures . . . 92

4.6 Elicitable Bayes risk measures . . . 93

4.7 Concluding remarks . . . 95

4.8 Lemmas and proofs of several results . . . 97

4.8.1 Lemmas in the proof of Theorem 4.2 . . . 97

4.8.2 Proof of Proposition 4.1 . . . 102

4.8.3 Proof of Theorem 4.4 . . . 103

**5 Risk measures induced by efficient insurance contracts** **109**

5.1 Introduction . . . 109

5.2 Preliminaries on risk measures . . . 111

5.3 Optimal insurance contract design . . . 112

5.4 Risk measures implied by Pareto-optimal contracts . . . 115

5.4.1 Main characterization results . . . 115

5.4.2 Designing insurance menus . . . 118

5.5 Concluding remarks . . . 119

5.6 Proofs of main results and related technical lemmas . . . 120

5.6.1 Technical lemmas . . . 120

5.6.2 Proofs of all results . . . 126

**6 Cash-subadditive risk measures without quasi-convexity** **131**

6.1 Introduction . . . 131

6.2 Cash-subadditive risk measures . . . 134

6.3 Quasi-star-shapedness, quasi-normalization, and Lambda VaR . . . 138

6.3.1 Quasi-star-shapedness and quasi-normalization . . . 138

6.3.2 A new representation of Lambda VaR . . . 141

6.3.3 A few useful technical results . . . 143

6.4 Representation results on cash-subadditive risk measures . . . 145

6.5 Cash-subadditive risk measures with further properties . . . 148

6.5.1 Normalized and quasi-star-shaped cash-subadditive risk measures . . . 148

6.5.2 SSD-consistent cash-subadditive risk measures . . . 151

6.5.3 Risk sharing with SSD-consistent cash-subadditive risk measures . . . 154

6.6 Conclusion . . . 155

6.7 Additional results and technical discussions . . . 156

6.7.1 Connection to a representation of monetary risk measures . . . 156

6.7.2 Comonotonic quasi-convexity . . . 157

6.7.3 Writing cash-subadditive risk measures via monetary ones . . . 158

6.7.4 Law-invariant cash-subadditive risk measures and VaR . . . 160

6.7.5 Certainty equivalents of RDEU with discount factor ambiguity . . . 162

6.7.6 An additional result on the inf-convolution in Section 6.5.3 . . . 164

**7 E-backtesting** **166**

7.1 Introduction . . . 166

7.1.1 Related literature . . . 169

7.2 E-values and model-free e-statistics . . . 170

7.2.1 Definition and examples . . . 170

7.3 E-backtesting risk measures with model-free e-statistics . . . . . . . . . . . . . . . . 173

7.4 E-backtesting Value-at-Risk and Expected Shortfall . . . . . . . . . . . . . . . . . . . 175

7.5 Choosing the betting process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 178

7.5.1 GRO, GREE, GREL and GREM methods . . . . . . . . . . . . . . . . . . . . 178

7.5.2 Optimality of betting processes . . . . . . . . . . . . . . . . . . . . . . . . 181

7.5.3 Proof of Theorem 7.3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 185

7.6 Characterizing model-free e-statistics . . . . . . . . . . . . . . . . . . . . . . . . 187

7.6.1 Necessary conditions for the existence of model-free e-statistics . . . . . . . . 187

7.6.2 Characterizing model-free e-statistics for common risk measures . . . . . . . . 188

7.7 Simulation studies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 190

7.7.1 Backtests via stationary time series . . . . . . . . . . . . . . . . . . . . . . 191

7.7.2 Monitoring structural change of time series . . . . . . . . . . . . . . . . . . 192

7.8 Financial data analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 194

7.8.1 The NASDAQ index . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 194

7.8.2 Optimized portfolios . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 197

7.9 Concluding remarks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 199

7.10 Taylor approximation formulas for GREE and GREL . . . . . . . . . . . . . . . . . 201

7.11 Link between model-free e-statistics and identification functions . . . . . . . . . . . 201

7.12 Omitted proofs of all results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 202

7.13 Supplementary simulation and data analysis . . . . . . . . . . . . . . . . . . . . . 209

7.13.1 Forecasting procedure for stationary time series data . . . . . . . . . . . . . . 209

7.13.2 Comparing GREE and GREL methods for stationary time series . . . . . . . 210

7.13.3 Forecasting procedure for time series with structural change . . . . . . . . . . 214

7.13.4 Detailed setup of data analysis for optimized portfolios . . . . . . . . . . . . . 215

**8 Concluding remarks and future work** **217**

8.1 Concluding remarks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 217

8.2 Future work and open questions . . . . . . . . . . . . . . . . . . . . . . . . . . . 219

8.2.1 On characterization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 219

8.2.2 On optimization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 221

8.2.3 On elicitability and backtesting . . . . . . . . . . . . . . . . . . . . . . . . 223

**Bibliography** **227**

# List of Figures


3.1 An example of $h$ (left) and $\hat{h}$ (right) with the set of discontinuity points $\{t_1, t_2, t_3, t_4, t_5\}$ excluding 0 and 1; the dashed lines represent $h^*$ and $(\hat{h})^*$, which are identical by Proposition 3.1 . . . . . . . . . . 37

3.2 Left panel: quantile function of $F$; right panel: quantile function of $F^{\mathcal{I}}$ where $\mathcal{I} = \{(0, 1/3), (1/2, 2/3)\}$ . . . . . . . . . . 38

3.3 Left panel: $h$ and $h^*$ for the TK distortion riskmetric with $\gamma = 0.7$ in Example 3.2; right panel: $h$ and $h^*$ for the inter-quantile range in Example 3.3 . . . . . . . . . . 42

3.4 Left panel: inverse-S-shaped distortion functions $h_1$ and $h_2$ in Example 3.4; right panel: $h = h_1 - h_2$ and $h^*$ of the same example . . . . . . . . . . 42

3.5 An example of $h$ (left) and $\hat{h}$ (right); in this figure, $J_h = \{t_1, t_2, t_3, t_4, t_5\}$, $J_+ = \{t_1\}$, $J_- = \{t_2, t_3\}$, and $J_0 = \{t_5\}$. Moreover, the sets we use in the proof of (i) are $\hat{J} = \{t_1, t_2, t_3, t_4\}$, $\hat{J}_+ = \{t_1, t_4\}$, $\hat{J}_- = \{t_2, t_3\}$, $\hat{J}^0_+ = \{t_4\}$, and $\hat{J}^0_- = \{t_3\}$ . . . . . . . 70

4.1 A Venn diagram for three classes of law-invariant risk measures . . . . . . . . . . 89

5.1 Solid lines represent the ceded loss functions of deductible insurance with coinsurance (left-hand panel) and deductible insurance with policy limit (right-hand panel); dashed lines represent ceded loss function with direct deductible . . . . . . . . . . 115

7.1 Realized losses and ES forecasts with a linear extending business (left panel); average log-transformed e-processes obtained by different methods over $1,000$ simulations (right panel) . . . . . . . . . . 183

7.2 Realized losses and ES forecasts with a non-linear business cycle (left panel); average log-transformed e-processes obtained by different methods over $1,000$ simulations (right panel) . . . 184

7.3 Realized losses and ES forecasts with iid losses (left panel); average log-transformed e-processes obtained by different methods over $1,000$ simulations (right panel) . . . . 185

7.4 Percentage of detections (%) of $\mathrm{VaR}_{0.95}$ (left panels) and $\mathrm{ES}_{0.95}$ (right panels) forecasts over $10,000$ simulations of time series and 250 trading days with structural changes at $b^*$ (top panels); ARLs of backtesting procedures (bottom panels); black lines ("monitor") represent the results of the sequential monitoring method . . . . . 195

7.5 Negated percentage log-returns of the NASDAQ Composite index (left panel); $\mathrm{ES}_{0.875}$ and $\mathrm{ES}_{0.975}$ forecasts fitted by normal distribution (right panel) from Jan 3, 2000 to Dec 31, 2021 . . . 196

7.6 Log-transformed e-processes testing $\mathrm{ES}_{0.975}$ with respect to the number of days for the NASDAQ index from Jan 3, 2005 to Dec 31, 2021; left panel: GREE method, middle panel: GREL method, right panel: GREM method . . . 197

7.7 Log-transformed e-processes testing $\mathrm{ES}_{0.975}$ with respect to the number of days for portfolio data from Jan 3, 2005 to Dec 31, 2021; left panel: GREE method, middle panel: GREL method, right panel: GREM method . . . 200

7.8 Portfolio data fitted by different distribution from Jan 3, 2005 to Dec 31, 2021; left panel: negated percentage log-returns, right panel: $\mathrm{ES}_{0.975}$ forecasts . . . 216

# List of Tables


2.1 Some examples of one-dimensional distortion riskmetrics . . . . . . . . . . . . . . . . . . 9

2.2 Comparison with results in Wang et al. (2020) . . . . . . . . . . . . . . . . . . . . . . . . 10

3.1 Optimal results in (3.24) for difference between two TK distortion riskmetrics . . . . 59

3.2 Optimal values in (3.25) for TK distortion riskmetrics . . . . . . . . . . . . . . . . . . 60

3.3 Comparison of the numerical results of the two optimization problems (3.29) and (3.30) for $\mathrm{VaR}_{0.95}$ and $\mathrm{ES}_{0.95}$ with $a = 0$ and $b = 1$ . . . . . . . . . . . . . . . . . . . . 63

3.4 Comparison of the numerical results of the two optimization problems (3.29) and (3.30) for $\mathrm{VaR}_{0.95}$ and $\mathrm{ES}_{0.95}$ with $a = 1/(2n)$ and $b = 2/n$ . . . . . . . . . . . . . . . 64

3.5 Comparison of the numerical results of the two optimization problems (3.31) and (3.32) for TK distortion riskmetrics with $a = 0$ and $b = 1$ . . . . . . . . . . . . . . . . 65

3.6 Comparison of the numerical results of the two optimization problems (3.31) and (3.32) for TK distortion riskmetrics with $a = 1/(2n)$ and $b = 2/n$ . . . . . . . . . . . 66

5.1 Connections between sets of ceded loss functions and classes of risk measures . . . . 117

5.2 Connections between Pareto-optimal sets of contracts and the insurer's risk measures 119

6.1 Representation results related to this chapter, where monotonicity is always assumed; definitions of the properties are in Sections 6.2 and 6.3. CA stands for cash additivity and CS stands for cash subadditivity. . . . . . . . . . . . . . . . . . . . . . . . . . . 133

7.1 Comparison of existing backtesting methods for ES . . . . . . . . . . . . . . . . . . . . 167

7.2 Percentage of detections (%) for $\mathrm{VaR}_{0.99}$ forecasts over $1,000$ simulations of time series and 500 trading days using the GREM method . . . . . . . . . . . . . . . . . . 192

7.3 The average number of days taken to detect evidence against $\mathrm{VaR}_{0.99}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days using the GREM method; numbers in brackets are average final log-transformed e-values . 192

7.4 Percentage of detections (%) for $\mathrm{ES}_{0.975}$ forecasts over $1,000$ simulations of time series and 500 trading days using the GREM method . . . . . . . . . . . . . . . . . . 193

7.5 The average number of days taken to detect evidence against $\mathrm{ES}_{0.975}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days using the GREM method; "–" represents no detection; numbers in brackets are average final log-transformed e-values . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 194

7.6 Average $\mathrm{ES}_{0.975}$ forecasts (boldface in brackets) and the number of days taken to detect evidence against the forecasts for the NASDAQ index from Jan 3, 2005 to Dec 31, 2021; "–" means no detection is detected till Dec 31, 2021 . . . . . . . . . . 198

7.7 Average $\mathrm{ES}_{0.975}$ forecasts (boldface in brackets) and the number of days taken to detect evidence against the forecasts for portfolio data from Jan 3, 2005 to Dec 31, 2021; "–" means no detection is detected till Dec 31, 2021 . . . . . . . . . . . . . . . . 199

7.8 Average point forecasts of VaR and ES at different levels over $1,000$ simulations of time series and 500 trading days; values in boldface are underestimated by at least 10% compared with values in the last line . . . . . . . . . . . . . . . . . . . . . . . . . 210

7.9 Percentage of detections (%) for $\mathrm{VaR}_{0.99}$ forecasts over $1,000$ simulations of time series and 500 trading days . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 211

7.10 The average number of days taken to detect evidence against $\mathrm{VaR}_{0.99}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days; numbers in brackets are average final log-transformed e-values . . . . . . . . . . . . . 212

7.11 Percentage of detections (%) for $\mathrm{ES}_{0.975}$ forecasts over $1,000$ simulations of time series and 500 trading days . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 213

7.12 The average number of days taken to detect evidence against $\mathrm{ES}_{0.975}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days; "–" represents no detection; numbers in brackets are average final log-transformed e-values 214

7.13 22 selected stocks in S&P 500 sectors divided by GICS level 1 as of Jan 3, 2005 for the portfolio . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 215

# Chapter 1

# Introduction and preliminary

## 1.1 Introduction

Riskmetrics are common tools to represent preferences, model decisions under risks, and quantify different types of risks. To fix terms, we refer to *riskmetrics* as any mapping from a set of random variables to the real line, and *risk measures* as riskmetrics that are monotone in the sense of Artzner et al. (1999).

Riskmetrics are important and intensively studied in both financial regulations and insurance. As one of the popular standard measures of risk in academic research and industrial practice, Value-at-Risk (VaR) has been adopted to measure capital requirement and insurance risk for decades. Let $\mathcal{X}$ be a set of random variables and let $F_X$ represent the cumulative distribution function of a random loss $X \in \mathcal{X}$. The Value-at-Risk (VaR) of $X$, using the sign convention of McNeil et al. (2015), is defined by

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : F(x) \geqslant \alpha\}, \quad \alpha \in (0, 1]. \tag{1.1}$$

Faced with some shortcomings of VaR, including its failure to incorporate tail risk and non-subadditivity, Artzner et al. (1999) studied and characterized coherent measures of risk. Based on results of this seminal paper, convex risk measures were further studied by Föllmer and Schied (2002a) and Frittelli and Rosazza Gianin (2002). Other classes of riskmetrics include but are not limited to deviation measures (Rockafellar et al., 2006), systemic risk measures (Chen et al., 2013), and Yaari's dual utility (Yaari, 1987), leading to distortion risk measures (West, 1996) and spectral risk measures (Acerbi, 2002). Among other riskmetrics, the Expected Shortfall (ES, also called CVaR/AVaR/TVaR), similarly to VaR, has drawn more attention in banking and insurance as a standard risk measure; see the Fundamental Review of the Trading Book in BCBS (2016, 2019) in the context of financial regulation. The Expected Shortfall (ES) of $X \in \mathcal{X}$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_t(X)\,\mathrm{d}t, \quad \alpha \in (0,1). \tag{1.2}$$

Most of the fundamental work on the specific classes of riskmetrics we mention above is about their characterization, especially axiomatic characterization. Axiomatic approaches for riskmetrics has been playing an essential role in economic decision theory; see Gilboa et al. (2019) for a discussion. Characterizing common classes of riskmetrics lays foundations from a theoretical perspective about why such riskmetrics are adopted and applied in practice, thus helps researchers and practitioners deepen the understanding of them. Moreover, characterization studies of riskmetrics also contribute to evaluating the risk attitudes of decision makers given their behaviors or given some specific properties are satisfied.

Besides characterization, optimization of riskmetrics remains to be a hot topic in economic decision theory and finance. When riskmetrics are optimized subject to some Knightian uncertainty, decisions should be taken based on the worst or best case riskmetrics. One of the popular types of Knightian uncertainty taken into account in optimization problems is uncertainty on the distribution of underlying risk. Optimization research on this topic generally belongs to *distributionally robust optimization*; see e.g., Popescu (2007) on moment constraints, Delage and Ye (2010) on parameter uncertainty, Wiesemann et al. (2014) on probability constraints, and Blanchet and Murthy (2019) on uncertainty of distributional distance. Worst-case riskmetrics under uncertainty is also related to the robust risk aggregation problems as one of the classical problems in quantitative risk management research; see e.g., Embrechts et al. (2015) for marginal constraints with dependence uncertainty. Beyond distributional uncertainty, other types of certainty are also of great interest in the literature of robust optimization problems; see e.g., Armbruster and Delage (2015) and Guo and Xu (2020) for preference robust optimization. Moreover, optimization of riskmetrics leads to the optimal (re)insurance design problem, which initiated from the seminal work of Arrow (1963) and is still a popular topic in the field of actuarial science.

Backtesting risk measures, especially backtesting ES, has long been a crucial topic interested by both the industry and the academy. Backtesting (Christoffersen, 2011; McNeil et al., 2015, see e.g.,) is known as the process of monitoring the risk measurement performance over time by comparing realized looses with conditional forecasts of risk measures, which is important for risk management of financial institutions. Therefore, backtestability is one of the essential criteria for a risk measure to be a good choice adopted by financial institutions. For example, the risk measures VaR and expectile (see e.g., Bellini et al., 2014; Ziegel, 2016) are known to be *backtestable* and *elicitable*. However, another popular risk measure, ES, turns out to be not backtestable with solely the information of itself (Acerbi and Szekely, 2017). This restricts the applications of ES in the banking practice, although it satisfies other nice mathematical properties such as coherence. As a result, explicit and model-free model-free methods of backtesting ES and other risk measures are of great interest and demand. We investigate this problem and find a solution by adopting the new concept of e-values and e-tests introduced by Vovk and Wang (2021).

This thesis contains results generally on characterizing, optimizing and backtesting different classes of riskmetrics. In the following paragraphs, we give a map of the remaining chapters in this thesis.

Chapter 2 introduces the class of functionals called "*distortion riskmetrics*". A distortion riskmetric is a real-valued functional $\rho$ with the following form

$$\rho(X)=\int_{-\infty}^{0}\left(h(\mathbb{P}(X\geqslant x))-h(1)\right)\,\mathrm{d}x+\int_{0}^{\infty}h(\mathbb{P}(X\geqslant x))\,\mathrm{d}x, \tag{1.3}$$

where $h$ is a function of bounded variation on $[0,1]$ with $h(0)=0$ and $X$ is a random variable in the domain of $\rho$; a precise definition is given in Definition 2.1 of Chapter 2 below. Distortion riskmetrics contain many common measures of risk and deviation and is allowed to be not monotone, not normalized and nonconvex. We study characterization and properties of distortion riskmetrics on general spaces beyond bounded random variables. Although most results in Chapter 2 are not surprising and similar to those in the literature, this chapter serves as a useful toolkit for later chapters related to distortion riskmetrics.

Chapter 3 studies much more deepened work on distributionally robust optimization of distortion riskmetrics introduced in Chapter 2. One of the central findings in this chapter is a unifying result that allows us to convert an optimization of a non-convex distortion riskmetric with distributional uncertainty to a convex one, leading to great tractability. The key to the unifying equivalence result is the novel notion of closedness under concentration of sets of distributions. The end of this chapter illustrates the theoretical results via applications to portfolio optimization, optimization under moment constraints, and preference robust optimization.

Chapters 4–6 are basically about characterization of riskmetrics. Chapter 4 characterizes the Expected Shortfall (ES) in the context of stastistical elicitation. We show in this chapter that ES is the only class of coherent Bayes risk measures under some continuity assumption, where we propose

Bayes risk measures as the counterpart of elicitable risk measures extensively studied in the recent literature. Chapter 5 further investigates the role of ES in actuarial science. We show that, under the framework of the optimal insurance problem, ES can be characterized as the only risk measure chosen by the insured and the insurer given that some set of contracts with deductible forms are Pareto optimal. Chapter 6 focuses instead on a much more general class of riskmetrics, namely the cash-subadditive risk measures. Cash-subadditive risk measures have attracted more and more attention in mathematical finance to address possible defaultability of underlying bonds. One of the major results in this chapter is that we can represent a cash-subadditive risk measure as the lower envelope of a family of quasi-convex cash-subadditive risk measures. Other explicit representation results together with other properties are also given. Especially, we propose a new property called quasi-star-shapedness, which turns out to have sound economic interpretations and be a good fit with cash-subaditivity in terms of obtaining the representation result.

Given the importance of ES in both banking and insurance, Chapter 7 explores the problem of backtesting ES, following the study of ES in Chapters 4 and 5 but through a different perspective. This chapter aims for producing a model-free backtesting procedure of ES and other risk measures based on the newly developed notion of e-values. Model-free e-statistics testing the mean, variance, quantile and ES are characterized respectively. Several methods of obtaining the process of e-values are introduced together with their theoretical asymptotic optimality results. Detailed procedures for our backtesting approach are demonstrated through extensive simulation study and real data analysis.

Chapter 8 concludes the thesis and discusses potential open questions that are of great interest in the field.

## 1.2 Preliminary

Throughout the thesis, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two random variables $X$ and $Y$ have the same distribution under $\mathbb{P}$ is denoted by $X \stackrel{\mathrm{d}}{=} Y$. For $x, y \in \mathbb{R}$, we write $x \vee y = \max\{x, y\}$, $x \wedge y = \min\{x, y\}$, $x_+ = x \vee 0$ and $x_- = (-x) \vee 0$. Throughout the thesis, "increasing" and "decreasing" are in the nonstrict sense. For $p \in [1, \infty)$, $\mathcal{L}^p$ is the space of random variables with finite $p$-th moment; $\mathcal{L}^\infty$ is that of essentially bounded random variables; and $\mathcal{L}^0$ represents the space of all random variables. For a distribution $F$ and a random variable $X$, $X \sim F$ means that $X$ has distribution $F$. Denote by $F_X$ the distribution function of the random variable $X$. We define the left-continuous generalized inverse of $F$ (left-quantile) as

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geqslant t\}, \quad t \in (0,1], \tag{1.4}$$

while the right-continuous generalized inverse of $F$ (right-quantile) is defined as

$$F^{-1+}(t) = \inf\{x \in \mathbb{R} : F(x) > t\}, \quad t \in [0,1), \tag{1.5}$$

where we adopt the convention that $\inf(\emptyset) = \infty$. For an event $A \in \Omega$, we denote its complement by $A^c$.

Random variables $X$ and $Y$ are *comonotonic* if there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that for each $\omega, \omega' \in \Omega_0$,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0.$$

As an equivalent definition, a random vector $(X_1, \dots, X_n)$ is called comonotonic if there exists a random variable $Z$ and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ almost surely for all $i = 1, \dots, n$. A random variable $X$ is said to *first-order stochastically dominate* $Y$, denoted by $X \succeq_1 Y$, if $\mathbb{E}[f(X)] \geqslant \mathbb{E}[f(Y)]$ for all increasing functions $f : \mathbb{R} \to \mathbb{R}$; the random variable $X$ is said to be larger than a random variable $Y$ in *convex order* (or *second stochastic order*), denoted by $X \geqslant_{\mathrm{cx}} Y$ (or $X \succeq_2 Y$), if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for all convex $\phi : \mathbb{R} \to \mathbb{R}$, provided that both expectations exist.

A functional $\rho$, mapping from some space of random variables to $\mathbb{R}$, may satisfy the following properties, where the statements hold for all random variables $X, Y$ in the domain of $\rho$.

(a) *Monotonicity*: $\rho(X) \leqslant \rho(Y)$ for $X \leqslant Y$.[1]

(b) *Translation invariance*: $\rho(X + c) = \rho(X) + c$ for all $c \in \mathbb{R}$.[2]

(c) *Positive homogeneity*: $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda \geqslant 0$.

(d) *Subadditivity*: $\rho(X + Y) \leqslant \rho(X) + \rho(Y)$.

(e) *Convexity*: $\rho(\lambda X + (1 - \lambda)Y) \leqslant \lambda \rho(X) + (1 - \lambda)\rho(Y)$ for all $\lambda \in [0, 1]$.

(f) *Convex order consistency (SSD-consistency)*: $\rho(X) \leqslant \rho(Y)$ for $X \leqslant_{\mathrm{cx}} Y$.

[1]A functional $\rho$ is said to be *increasing* (or *decreasing*) if $X \leqslant Y$ almost surely implies $\rho(X) \leqslant \rho(Y)$ (or $\rho(X) \geqslant \rho(Y)$, respectively). The terms "increasing" and "decreasing" are always in the non-strict sense.

[2]Translation invariance is also called cash invariance or cash additivity.

(g) *Law invariance:* $\rho(X) = \rho(Y)$ for $X \stackrel{\mathrm{d}}{=} Y$.

(h) *comonotonic-additivity*: $\rho(X + Y) = \rho(X) + \rho(Y)$ if $X$ and $Y$ are comonotonic.

Following the terminology of Föllmer and Schied (2016), a *monetary risk measure* is a functional that is monotone and translation invariant; a *coherent risk measure* is a functional that is monotone, translation invariant, positively homogeneous, and subadditive.

# Chapter 2

# Distortion riskmetrics on general spaces

## 2.1 Introduction

In this chapter we study *distortion riskmetrics* on general model spaces. Let us first explain our somewhat unusual choice of terminology, "distortion riskmetrics". Clearly, the term "*distortion*" addresses the dominating role played by the (not necessarily monotone) distortion function $h$ in (1.3), whereas the term "*riskmetrics*" is chosen to distinguish $\rho$ from the classic notions of risk measures and deviation measures. For instance, risk measures are often required to be monotone and translation-invariant in the sense of Artzner et al. (1999), and deviation measures are required to be convex in the sense of Rockafellar et al. (2006). Insurance risk measures and premium principles are typically assumed to be monotone with some other properties as in e.g., Gerber (1974) or Wang et al. (1997). Our notion of distortion riskmetrics does not require monotonicity, translation-invariance or convexity, and it unifies risk measures, deviation measures, and many other functionals in the literature of finance and insurance.

This chapter is not the first to study functionals in (1.3) in risk management. Historically, such functionals, assuming monotonicity, were studied by Yaari (1987) in the economic literature and by Denneberg (1994) and Wang et al. (1997) in the actuarial literature. More recently, for non-monotone $h$, Wang et al. (2020) called the functional in (1.3) a *signed Choquet integral* on the space $\mathcal{L}^\infty$ of bounded random variables. To be precise, a signed Choquet integral refers to the right-hand side of (1.3). We note that a signed Choquet integral should be interpreted as an "integral", thus a mathematical operation, and not a functional. Mathematically, signed Choquet integrals can be formulated for any random variable, leading to a finite, infinite or undefined value in (1.3), whereas a distortion riskmetric is defined on a domain of financial relevance. The difference is negligible if the study is confined to $\mathcal{L}^\infty$, but it becomes delicate in the case of a larger space such as an $\mathcal{L}^p$-space; see Section 2.2. Moreover, the term "distortion riskmetric" better describes the practical purpose of these functionals in risk management. For the above reasons, we decided to invent the term "distortion riskmetrics", which will hopefully be the standard term for the object in (1.3) in the future.

As hinted above, monotone (increasing) distortion riskmetrics have been studied for decades under different names: the L-functionals (Huber and Ronchetti, 2009) in statistics, Yaari's dual utilities (Yaari, 1987) in decision theory, distorted premium principles (Denneberg, 1994; Wang et al., 1997; Denuit et al., 2005) in insurance, and distortion risk measures (Kusuoka, 2001; Acerbi, 2002) in finance. Some specific examples of distortion risk measures include the Value-at-Risk (VaR), the Expected Shortfall (ES, or TVaR/CVaR), the performance measures in Cherny and Madan (2009), the GlueVaR in Belles-Sampera et al. (2014), and the economic risk measures in Kou and Peng (2016). Non-monotone examples of signed Choquet integrals include the mean-median deviation, the Gini deviation, the inter-quantile range, some deviation measures of Rockafellar et al. (2006), and the Gini Shortfall of Furman et al. (2017). We collect some examples of one-dimensional distortion riskmetrics in Table 2.1.

Moreover, distortion riskmetrics serve as the building block of law-invariant convex risk functionals in the sense that any law-invariant convex risk functional can be written as a supremum of signed Choquet integrals plus constants (Liu et al., 2020) and this includes all law-invariant convex risk measures in Föllmer and Schied (2016) and all law-invariant deviation measures in Grechuk et al. (2009), as well as the classic mean-variance and mean-standard deviation principles in insurance.

We already mentioned that characterization and various properties of distortion riskmetrics are studied on $\mathcal{L}^\infty$ by Wang et al. (2020). As a follow-up of the previous work, the main purpose of this chapter is to extend the domain of distortion riskmetrics to more general spaces, including $\mathcal{L}^p$-spaces for $p \in [1, \infty)$. In many applications, risk measures such as the industry standard VaR and ES are defined on spaces beyond $\mathcal{L}^\infty$ to include unbounded loss distributions, e.g., normal, Pareto or t-distributions. Furthermore, for many convex risk measures, their natural domains on which key properties are preserved are Banach spaces much larger than $\mathcal{L}^\infty$; see e.g., Filipović and Svindland (2012), Pichler (2013) and Liebrich and Svindland (2017). Indeed, there is an extensive

| name (notation) | formula for $X \in \mathcal{X}$ and parameters | distortion function for $t \in [0,1]$ | domain $\mathcal{X}$ | convex? | monotone? |
|---|---|---|---|---|---|
| mean ($\mathbb{E}$) | $\mathbb{E}[X]$ | $t$ | $\mathcal{L}^1$ | yes | yes |
| Value-at-Risk ($\mathrm{VaR}_\alpha$) | $F_X^{-1}(\alpha), \quad \alpha \in (0,1)$ | $\mathbb{1}_{\{t>1-\alpha\}}$ | $\mathcal{L}^0$ | no | yes |
| ES/TVaR/CVaR ($\mathrm{ES}_\alpha$) | $\frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(t)\,\mathrm{d}t, \quad \alpha \in (0,1)$ | $\frac{t}{1-\alpha} \wedge 1$ | $\mathcal{L}^{0,1}$ | yes | yes |
| Gini deviation | $\frac{1}{2}\mathbb{E}[\lvert X^* - X^{**}\rvert]$ | $t - t^2$ | $\mathcal{L}^1$ | yes | no |
| mean-median deviation | $\min_{x \in \mathbb{R}} \mathbb{E}[\lvert X - x\rvert]$ | $t \wedge (1-t)$ | $\mathcal{L}^1$ | yes | no |
| essential supremum (ess-sup) | $F_X^{-1}(1)$ | $\mathbb{1}_{\{0<t\leqslant 1\}}$ | $\mathcal{L}^{0,\infty}$ | yes | yes |
| essential infimum (ess-inf) | $F_X^{-1+}(0)$ | $\mathbb{1}_{\{t=1\}}$ | $\mathcal{L}^{\infty,0}$ | no | yes |
| range | $F_X^{-1}(1) - F_X^{-1+}(0)$ | $\mathbb{1}_{\{0<t<1\}}$ | $\mathcal{L}^\infty$ | yes | no |
| inter-quantile range ($\mathrm{IQR}_\alpha$) | $F_X^{-1+}(\alpha) - F_X^{-1}(1-\alpha), \quad \alpha \in [1/2, 1)$ | $\mathbb{1}_{\{1-\alpha\leqslant t\leqslant \alpha\}}$ | $\mathcal{L}^0$ | no | no |
| inter-ES range ($\mathrm{IER}_\alpha$) | $\mathrm{ES}_\alpha(X) + \mathrm{ES}_\alpha(-X), \quad \alpha \in (0,1)$ | $\frac{t}{1-\alpha} \wedge 1 + \frac{\alpha - t}{1-\alpha} \wedge 0$ | $\mathcal{L}^1$ | yes | no |
| Range Value-at-Risk ($\mathrm{RVaR}_{\alpha,\beta}$) | $\frac{1}{\beta-\alpha}\int_\alpha^\beta F_X^{-1}(t)\,\mathrm{d}t, \quad 0<\alpha<\beta<1$ | $\frac{(t-1+\beta)_+}{\beta-\alpha} \wedge 1$ | $\mathcal{L}^1$ | no | yes |
| Gini Shortfall ($\mathrm{GS}_\alpha^\lambda$) | $\mathrm{ES}_\alpha(X) + \lambda\mathbb{E}[\lvert X_\alpha^* - X_\alpha^{**}\rvert]$ <br> $\alpha \in (0,1),\ \lambda \geqslant 0$ | $\frac{t}{1-\alpha} \wedge 1 + \frac{2\lambda t(1-t-\alpha)_+}{(1-\alpha)^2}$ | $\mathcal{L}^{0,1}$ | $\lambda \leqslant 1/2$ | $\lambda \leqslant 1/2$ |
| proportional hazard principle/MAXVAR | $\frac{1}{\alpha}\int_0^1 (1-t)^{(1-\alpha)/\alpha} F_X^{-1}(t)\,\mathrm{d}t, \quad \alpha \geqslant 1$ | $t^{1/\alpha}$ | $\cup_{p>\alpha}\mathcal{L}^{1,p} \subseteq \mathcal{X}$ | yes | yes |
| dual power principle/MINVAR | $\alpha\int_0^1 t^{\alpha-1} F_X^{-1}(t)\,\mathrm{d}t, \quad \alpha \geqslant 1$ | $1-(1-t)^\alpha$ | $\cup_{q>1/\alpha}\mathcal{L}^{q,1} \subseteq \mathcal{X}$ | yes | yes |
| GlueVaR ($\mathrm{GlueVaR}_{\beta,\alpha}^{\omega_1,\omega_2}$) | $\omega_1\mathrm{ES}_\alpha(X) + \omega_2\mathrm{ES}_\beta(X) + \omega_3\mathrm{VaR}_\alpha(X)$ <br> $0<\alpha\leqslant\beta<1,\ (\omega_1,\omega_2,\omega_3) \in \Delta_3$ | $\omega_1(\frac{t}{1-\alpha}\wedge 1) + \omega_2(\frac{t}{1-\beta}\wedge 1) + \omega_3\mathbb{1}_{\{t>1-\alpha\}}$ | $\mathcal{L}^{0,1}$ | no | yes |

Table 2.1: Some examples of one-dimensional distortion riskmetrics

*Notation.* $F_X^{-1}(\alpha) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant \alpha\}$ for $\alpha \in (0,1]$ and $F_X^{-1+}(\alpha) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) > \alpha\}$ for $\alpha \in [0,1)$. $\mathcal{L}^{p,q} = \{X \in \mathcal{L}^0 : X_- \in \mathcal{L}^p,\ X_+ \in \mathcal{L}^q\}$ for $p, q \geqslant 0$. $\Delta_n = \{(x_1,\dots,x_n) \in (0,1)^n : x_1 + \cdots + x_n = 1\}$ is the interior of the standard $n$-simplex. $X^*, X^{**}$ are iid copies of $X$ and $X_\alpha^*, X_\alpha^{**}$ are iid copies of $F_X^{-1}(U_\alpha)$ where $U_\alpha \sim \mathrm{U}[\alpha, 1]$.

literature on risk measures defined on general spaces (e.g., Delbaen, 2002; Föllmer and Schied, 2002a; Ruszczyński and Shapiro, 2006) and in particular on $\mathcal{L}^p$-spaces (Frittelli and Rosazza Gianin, 2002) or Orlicz spaces (Cheridito and Li, 2009). Different from the previous literature, we consider many functionals that are not necessarily monotone or convex. Notably, as a special example, the inter-quantile range (see Table 2.1) is not monotone, convex, or $\mathcal{L}^p$-continuous, but it is a popular measure of dispersion in statistics, and it belongs to the class of distortion riskmetrics. Finally, we extend distortion riskmetrics to a multi-dimensional setting, where the concepts of elicitability and convex level sets has been popular recently; see Fissler and Ziegel (2016), Frongillo and Kash (2021) and Wang and Wei (2020).

Most results in this chapter are similar to those in the literature in terms of both statements and proofs, and our findings that these results hold on general spaces are not surprising. However, most of the results in previous literature on $\mathcal{L}^\infty$, especially those in Wang et al. (2020), may not be convenient to directly use in practice where most applications require results on more general spaces of random variables. As such, more general results are in need, and this chapter can be viewed as a convenient toolkit for future studies and applications of distortion riskmetrics. Nevertheless, there are several additions to the existing literature. The similarity of this chapter with Wang et al. (2020) and the new results of this chapter are summarized in Table 6.1.

| corresponding results | | | new results |
|---|---|---|---|
| this chapter (on general spaces) | | Wang et al. (2020) (on $\mathcal{L}^\infty$) | this chapter (on general spaces) |
| Theorem 2.1 | $\longleftrightarrow$ | Theorem 1 | Proposition 2.1 |
| Theorem 2.2 | $\longleftrightarrow$ | Theorem 2 | Proposition 2.3 |
| Proposition 2.2 | $\longleftrightarrow$ | Lemmas 2 and 3 | Theorem 2.4 |
| Theorem 2.3 | $\longleftrightarrow$ | Theorem 3 | Theorem 2.5 |
| Theorem 2.6 | $\longleftrightarrow$ | Theorem 4 | Proposition 2.4 |

Table 2.2: Comparison with results in Wang et al. (2020)

Below we briefly explain the new results. First, an ES-based representation of convex distortion riskmetric $\rho$ in Theorem 2.5 is new to the literature. Four other new results, all requiring the considered domain to be larger than $\mathcal{L}^\infty$, are the finiteness condition in Proposition 2.1, the domain of convex distortion riskmetrics in Proposition 2.3, the existence of dominating convex functionals in Theorem 2.4, and the $\mathcal{L}^p$-continuity in Proposition 2.4. Moreover, the condition in Theorem 2.6

is slightly weakened compared to a similar result on $\mathcal{L}^\infty$ in Wang et al. (2020).

The chapter is organized as follows. In Section 2.2, we collect basic definitions needed for this chapter, and present a functional characterization of distortion riskmetrics. In Section 2.3, results related to convexity, convex order consistency, and mixture concavity are presented. Section 2.4 contains results on continuity properties of distortion riskmetrics and and Section 3.4 extends the discussions to the multi-dimensional setting. To facilitate the main purpose of the chapter as a toolkit, most proofs are self-contained and are relegated to the last section.

## 2.2 Distortion riskmetrics and their characterization

### 2.2.1 Notation and definition

We assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless. Throughout this chapter, the set $\mathcal{X} \supseteq \mathcal{L}^\infty$ is a law-invariant convex cone, that is, for all random variables $X$ and $Y$,

(i) if $X \in \mathcal{X}$ and $X \stackrel{\mathrm{d}}{=} Y$, then $Y \in \mathcal{X}$;

(ii) if $X \in \mathcal{X}$, then $\lambda X \in \mathcal{X}$ for all $\lambda > 0$;

(iii) if $X, Y \in \mathcal{X}$, then $X + Y \in \mathcal{X}$.

Let $\mathcal{M}$ be the set of distribution functions of random variables in $\mathcal{X}$. For simplicity, in the definitions of the left- and right-quantile in (1.4) and (1.5), we let $F^{-1}(0) = F^{-1+}(0)$ and $F^{-1+}(1) = F^{-1}(1)$.

Next, corresponding to (1.3), we give a formal definition of the *distortion riskmetric* using the signed Choquet integral (Choquet, 1954) on a general space. Denote by

$$\mathcal{H} = \{h : h \text{ maps } [0,1] \text{ to } \mathbb{R},\ h \text{ is of bounded variation}, h(0) = 0\}.$$

**Definition 2.1.** A functional $\rho_h : \mathcal{X} \to \mathbb{R}$, whose domain $\mathcal{X} \supseteq \mathcal{L}^\infty$ is a law-invariant convex cone, is a *distortion riskmetric* if there exists $h \in \mathcal{H}$ such that $\rho_h(X) = \int X \,\mathrm{d}h \circ \mathbb{P}$, where $\int X \,\mathrm{d}h \circ \mathbb{P}$ is a signed Choquet integral defined by

$$\int X \,\mathrm{d}h \circ \mathbb{P} = \int_{-\infty}^{0} \left(h(\mathbb{P}(X \geqslant x)) - h(1)\right) \mathrm{d}x + \int_{0}^{\infty} h(\mathbb{P}(X \geqslant x)) \,\mathrm{d}x. \tag{2.1}$$

The function $h$ is called the *distortion function* of $\rho_h$.

Generally, the two integrals in (2.1) may not be finite, and hence $\int X \,\mathrm{d}h \circ \mathbb{P}$ may be infinite or even not well-defined (i.e., $\infty - \infty$). We emphasize that according to our definition, a distortion riskmetric $\rho_h : \mathcal{X} \to \mathbb{R}$ is only defined when $\int X \,\mathrm{d}h \circ \mathbb{P}$ is finite (i.e., both integrals are finite), and hence the two terms "distortion riskmetrics" and "signed Choquet integrals" are no longer interchangeable, in contrast to the case of $\mathcal{L}^\infty$ studied by Wang et al. (2020). In other words, $\mathcal{X}$ and $h$ have to be compatible, making (2.1) finite. In Section 2.2.2 below we will give a sufficient condition for (2.1) to be finite. The notion of a distortion function $h$ we use in this chapter is broader than the classical sense in which $h$ is assumed increasing with $h(1) = 1$.

For a given distortion riskmetric $\rho_h : \mathcal{X} \to \mathbb{R}$, the distortion function $h \in \mathcal{H}$ is unique. To see this, suppose that $\rho_{h_1}(X) = \rho_{h_2}(X)$ for all $X \in \mathcal{X}$. Choose a random variable $X \sim \text{Bernoulli}(p)$ with a fixed $p \in [0, 1]$. It follows that

$$\rho_{h_i}(X) = h_i(p) + \int_1^\infty h_i(0) \,\mathrm{d}x = h_i(p), \quad i = 1, 2.$$

Since $p$ is arbitrary, we get $h_1 = h_2$ on $[0, 1]$.

*Remark* 2.1. A distortion riskmetric $\rho_h$ can be equivalently expressed as

$$\rho_h(X) = \int_{-\infty}^0 \left(h(\mathbb{P}(X > x)) - h(1)\right) \mathrm{d}x + \int_0^\infty h(\mathbb{P}(X > x)) \,\mathrm{d}x. \tag{2.2}$$

Indeed, since $\mathbb{P}(X > x) = \mathbb{P}(X \geqslant x)$ almost everywhere on $\mathbb{R}$, we know $h(\mathbb{P}(X > x)) = h(\mathbb{P}(X \geqslant x))$ almost everywhere on $\mathbb{R}$.

### 2.2.2 Quantile representation and finiteness of signed Choquet integrals

The quantile representation of signed Choquet integrals is obtained in Lemma 3 of Wang et al. (2020) on $\mathcal{L}^\infty$ and Theorems 4 and 6 of Dhaene et al. (2012) for increasing $h$. Combining the above results, we have the following quantile representation of signed Choquet integrals on a general space with distortion functions not necessarily increasing.

**Lemma 2.1.** *For $h \in \mathcal{H}$ and $X \in \mathcal{L}^0$ such that $\int X \,\mathrm{d}h \circ \mathbb{P}$ is well-defined (it may take values $\pm\infty$),*

*(i) if $h$ is right-continuous, then $\int X \,\mathrm{d}h \circ \mathbb{P} = \int_0^1 F_X^{-1+}(1-t) \,\mathrm{d}h(t)$;*

*(ii) if $h$ is left-continuous, then $\int X \,\mathrm{d}h \circ \mathbb{P} = \int_0^1 F_X^{-1}(1-t) \,\mathrm{d}h(t)$;*

*(iii) if $F_X^{-1}$ is continuous on $(0, 1)$, then $\int X \,\mathrm{d}h \circ \mathbb{P} = \int_0^1 F_X^{-1}(1-t) \,\mathrm{d}h(t) = \int_0^1 F_X^{-1+}(1-t) \,\mathrm{d}h(t)$.*

Now we focus on $\mathcal{L}^p$-spaces for $p \in [1, \infty]$. Define a set of distortion functions $\mathcal{H}_1$ as

$$\mathcal{H}_1 = \{h \in \mathcal{H} : h \text{ is absolutely continuous on } [0, \epsilon) \cup (1 - \epsilon, 1] \text{ for some } \epsilon \in (0, 1)\}.$$

Note that $\mathcal{H}_1$ excludes only special examples such as the essential supremum, the essential infimum, and the range in Table 2.1. Moreover, noticing that $h$ is differentiable almost everywhere on $[0, 1]$ due to bounded variation, we let

$$\mathcal{H}_q = \left\{h \in \mathcal{H}_1 : h' \in \mathcal{L}^q((0, \epsilon) \cup (1 - \epsilon, 1)) \text{ for some } \epsilon \in (0, 1)\right\},$$

where $h'$ is (in a.e. sense) the derivative of $h$ and $q$ is the conjugate of $p \in [1, \infty]$ (i.e., $1/p+1/q = 1$). Next, we give a sufficient condition for $\rho_h$ to be well defined, which is almost necessary in case that $h$ is concave.

**Proposition 2.1.** *For $p \in [1, \infty)$, $q$ being its conjugate,*

*(i) $\int X \, \mathrm{d}h \circ \mathbb{P}$ is finite for all $X \in \mathcal{L}^p$ if $h \in \mathcal{H}_q$;*

*(ii) if $h \in \mathcal{H}$ is concave and $\int X \, \mathrm{d}h \circ \mathbb{P}$ is finite for all $X \in \mathcal{L}^p$, then $h \in \mathcal{H}_r$ for all $r < q$.*

As a consequence of Proposition 2.1, if $h \in \mathcal{H}$ is absolutely continuous and $\int_0^1 |h'(t)|^q \, \mathrm{d}t < \infty$, then $\int X \, \mathrm{d}h \circ \mathbb{P}$ is finite for all $X \in \mathcal{L}^p$. In particular, the case $p = q = 2$ gives a sufficient condition for the finiteness of $\int X \, \mathrm{d}h \circ \mathbb{P}$ for $X \in \mathcal{L}^2$.

### 2.2.3 Characterization and basic properties

Before we further characterize distortion riskmetrics, we define some properties of functionals.

(a) *Continuity at infinity*: $\lim_{M\to\infty} \rho((X \wedge M) \vee (-M)) = \rho(X)$.

(b) *Uniform sup-continuity:* For any $\epsilon > 0$, there exists $\delta > 0$, such that $|\rho(X)-\rho(Y)| < \epsilon$ whenever ess-sup$|X - Y| < \delta$, where ess-sup$(\cdot)$ is the essential supremum in Table 2.1.

The above two properties, together with law invariance and comonotonic-additivity, are satisfied by distortion riskmetrics, and moreover, they indeed characterize distortion riskmetrics, similarly to the case of bounded random variables studied by Wang et al. (2020) and the case of increasing Choquet integrals in Wang et al. (1997) and Kou and Peng (2016), all based on a classic result of Schmeidler (1986).

**Theorem 2.1.** *A functional $\rho : \mathcal{X} \to \mathbb{R}$ is law invariant, comonotonic-additive, continuous at infinity and uniformly sup-continuous if and only if $\rho$ is a distortion riskmetric.*

*Remark* 2.2. From the proof of necessity part of Theorem 2.1 in Section 2.6, we can see a distortion riskmetric $\rho_h : \mathcal{X} \to \mathbb{R}$ is, in fact, Lipschitz continuous with respect to $\mathcal{L}^\infty$-norm with Lipschitz constant $\mathrm{TV}_h$, the total variation of $h$ on $[0, 1]$. This continuity is stronger than uniform sup-continuity. This point will be further developed in Section 2.4.

Below we present some basic properties of distortion riskmetrics which are useful in later sections. They are well-established for random variables in $\mathcal{L}^\infty$ and $h \in \mathcal{H}$.

**Proposition 2.2.** *For $h, h_1, h_2 \in \mathcal{H}$,*

*(i) if $h_1(1) = h_2(1)$, then $h_1 \leqslant h_2$ on $[0, 1] \Leftrightarrow \rho_{h_1} \leqslant \rho_{h_2}$ on $\mathcal{X}$. In particular, $h_1 = h_2$ on $[0, 1] \Leftrightarrow \rho_{h_1} = \rho_{h_2}$ on $\mathcal{X}$;*

*(ii) $\rho_h$ is increasing (resp. decreasing) if and only if $h$ is increasing (resp. decreasing);*

*(iii) for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$, $\rho_h(X + c) = \rho_h(X) + ch(1)$;*

*(iv) for all $\lambda > 0$ and $X \in \mathcal{X}$, $\rho_h(\lambda X) = \lambda \rho_h(X)$;*

*(v) for all $X \in \mathcal{X}$, $\rho_h(-X) = \rho_{\hat{h}}(X)$, where $\hat{h} : [0, 1] \to \mathbb{R}$ is given by $\hat{h}(x) = h(1 - x) - h(1)$ for all $x \in [0, 1]$.*

## 2.3 Convexity, convex order consistency and mixture concavity

In this section, we study the important class of convex distortion riskmetrics and their related properties. As shown in Theorem 3 of Wang et al. (2020), the following properties: convexity, convex order consistency, and mixture concavity, on $\mathcal{L}^\infty$, are equivalent to concavity of the distortion function. We establish a similar result on general spaces, as well as a few new results on convex distortion riskmetrics.

We first justify that for a convex distortion riskmetric, if its domain $\mathcal{X}$ is a linear space, then it is contained in $\mathcal{L}^1$; hence, it makes sense to confine our study to subsets of $\mathcal{L}^1$. Note also that $\mathcal{L}^1$ is the canonical space for law-invariant convex risk measures (e.g., Filipović and Svindland, 2012).

**Proposition 2.3.** *Suppose that $\mathcal{X}$ is a linear space and $\rho_h : \mathcal{X} \to \mathbb{R}$ is a convex distortion riskmetric. Then $\mathcal{X} \subseteq \mathcal{L}^1$ unless $\rho_h = 0$ on $\mathcal{X}$.*

The assumption that $\mathcal{X}$ is a linear space in Proposition 2.3 is not dispensable. An important example is the Expected Shortfall (ES) in Table 2.1 at level $\alpha \in (0,1)$, defined in (1.2), where its domain $\mathcal{X}$ can be chosen as $\{X \in \mathcal{L}^0 : X_+ \in \mathcal{L}^1\}$, which is larger than $\mathcal{L}^1$. In addition, we let $\mathrm{ES}_0 = \mathbb{E}$ which is finite on $\mathcal{L}^1$ and $\mathrm{ES}_1$ be the essential supremum which is finite on the set of random variables bounded from above. For $\alpha \in [0,1]$, $\mathrm{ES}_\alpha$ is a convex distortion riskmetric with distortion function $h$ given by

$$h(t) = \frac{t}{1-\alpha} \wedge 1, \quad t \in [0,1],\ \alpha \in [0,1)$$

and $h(t) = \mathbb{1}_{\{t>0\}}$ if $\alpha = 1$. These facts will be useful later.

Next, we fix some terminology. For a functional $\rho : \mathcal{X} \to \mathbb{R}$ and all random variables $X, Y \in \mathcal{X}$, $\rho$ is *quasi-convex* if $\rho(\lambda X + (1-\lambda)Y) \leqslant \rho(X) \vee \rho(Y)$ for all $\lambda \in [0,1]$. For a law-invariant functional $\rho$, define $\tilde{\rho} : \mathcal{M} \to \mathbb{R}$ such that $\tilde{\rho}(F) = \rho(X)$ where $X \sim F$, and $\rho$ is *concave on mixtures* if $\tilde{\rho}$ is concave. The following result characterizes convex order using distortion riskmetrics. For a version of this result for increasing $h$, see Theorem 5.2.1 of Dhaene et al. (2006).

**Theorem 2.2.** *For all random variables $X, Y \in \mathcal{L}^1$, $X \leqslant_{\mathrm{cx}} Y$ if and only if $\rho_h(X) \leqslant \rho_h(Y)$ for all concave functions $h \in \mathcal{H}$ such that $X$ and $Y$ are in the domain of $\rho_h$.*

In the following theorem, we present six equivalent conditions about convexity of a distortion riskmetric on a general space, similar to Theorem 3 of Wang et al. (2020). Recall that $\mathcal{X}$ is a law-invariant convex cone containing $\mathcal{L}^\infty$. In the following result, we further assume $\mathcal{X} \subseteq \mathcal{L}^1$ as discussed above.

**Theorem 2.3.** *For a distortion riskmetric $\rho_h : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathcal{L}^1$, the following are equivalent: (i) $h$ is concave; (ii) $\rho_h$ is convex order consistent; (iii) $\rho_h$ is subadditive; (iv) $\rho_h$ is convex; (v) $\rho_h$ is quasi-convex; (vi) $\rho_h$ is concave on mixtures.*

A few well known characterization results in risk management can be directly obtained from Theorem 2.1 and 2.3. For a history of these results, see Föllmer and Schied (2016).

**Corollary 2.1.** *Suppose that $\mathcal{X} \subseteq \mathcal{L}^1$. A functional $\rho : \mathcal{X} \to \mathbb{R}$ is law invariant, increasing, cash invariant, continuous at infinity, and comonotonic-additive if and only if $\rho$ is a distortion riskmetric $\rho_h$ for an increasing $h$ with $h(1) = 1$. In addition, $\rho$ satisfies any of the properties (ii)-(vi) in Theorem 2.3 if and only if $h$ is concave, and in that case $\rho$ is a coherent risk measure.*

Note that in Corollary 2.1 we do not assume uniform sup-continuity as it is implied by monotonicity and cash invariance. In case $\mathcal{X} = \mathcal{L}^\infty$, continuity at infinity can also be removed from the statement. In Corollary 2.1, $\rho = \rho_h$ is a distortion risk measure or a dual utility (Yaari, 1987). If $h$ is concave, then $\rho = \rho_h$ is commonly known as a spectral risk measure; see Acerbi (2002) where $h$ is additionally assumed to be continuous at 0.

In the next result, we consider the relationship between a distortion riskmetric $\rho_h$ and a convex one dominating $\rho_h$. For this purpose, we introduce the *concave envelope* $h^* : [0,1] \to \mathbb{R}$ of $h \in \mathcal{H}$, defined as

$$h^*(t) = \inf \{g(t) : \ g \in \mathcal{H}, \ g \geqslant h, \ g \text{ is concave on } [0,1]\}. \tag{2.3}$$

One can check that $h^*$ is concave, $h^*(0) = 0$ and $h^*(1) = h(1)$; see Wang et al. (2020) for a simple justification. Theorem 2.3 yields that $\rho_{h^*} : \mathcal{X} \to \mathbb{R}$ is a convex distortion riskmetric if $\mathcal{X} \subseteq \mathcal{L}^1$. We also know that $\rho_{h^*} \geqslant \rho_h$ on their common domain due to Proposition 2.2. The next theorem shows that $\rho_{h^*}$ is actually the smallest law-invariant, convex and continuous-at-infinity functional dominating $\rho_h$; note that it is not obvious whether such a functional exists and whether it is a distortion riskmetric. Below, we say that $\rho_{h^*}$ is finite on $\mathcal{X}$, if the signed Choquet integral $\int X \, \mathrm{d}h^* \circ \mathbb{P}$ is finite for all $X \in \mathcal{X}$.

**Theorem 2.4.** *For a distortion riskmetric $\rho_h : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathcal{L}^1$, if $\rho_{h^*}$ is finite on $\mathcal{X}$, then $\rho_{h^*}$ is the smallest law-invariant, convex and continuous-at-infinity functional dominating $\rho_h$. If $\rho_{h^*}$ is not finite on $\mathcal{X}$, then there is no real-valued law-invariant, convex and continuous-at-infinity functional dominating $\rho_h$.*

Theorem 2.4 implies in particular that $\mathrm{ES}_\alpha$ in (1.2) is the smallest law-invariant and continuous-at-infinity convex functional dominating $\mathrm{VaR}_\alpha$ in (1.1); see Theorem 9 of Kusuoka (2001) and Theorem 4.67 of Föllmer and Schied (2016) for this statement on the set of bounded random variables.

In the next result, we establish a new ES-based representation of convex distortion riskmetrics, which covers the classic ES-based representation of coherent distortion risk measures in Theorem 4.93 of Föllmer and Schied (2016) on $\mathcal{L}^\infty$. As far as we are aware of, the representation (2.4) is new to the literature.

**Theorem 2.5.** *A functional $\rho : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathcal{L}^1$ is a convex distortion riskmetric if and only* *if there exist finite Borel measures $\mu, \nu$ on $[0,1]$ such that*

$$\rho(X) = \int_0^1 \mathrm{ES}_\alpha(X)\,\mathrm{d}\mu(\alpha) + \int_0^1 \mathrm{ES}_\alpha(-X)\,\mathrm{d}\nu(\alpha). \tag{2.4}$$

*Moreover, if $\rho$ is increasing, then we can take $\nu = 0$.*

*Remark* 2.3. In case $\nu$ in (2.4) satisfies $\beta := \int_0^1 \frac{1}{1-\alpha}\,\mathrm{d}\nu(\alpha) < \infty$, using the equality

$$\mathrm{ES}_\alpha(-X) = \frac{1}{1-\alpha}(\alpha \mathrm{ES}_{1-\alpha}(X) - \mathbb{E}[X]), \quad X \in \mathcal{L}^1,$$

we can rewrite (2.4) as

$$\rho(X) = \int_0^1 \mathrm{ES}_\alpha(X)\,\mathrm{d}\hat{\mu}(\alpha) - \beta\mathbb{E}[X], \quad X \in \mathcal{X}, \tag{2.5}$$

where $\hat{\mu}$ is another finite Borel measure on $[0,1]$. Note that the condition $\beta < \infty$ is not automatically satisfied for a general convex distortion riskmetric $\rho$. An example of a convex distortion riskmetric that does not admit the form in (2.5) is $\rho : \mathcal{L}^\infty \to \mathbb{R},\ X \mapsto -F_X^{-1}(0)$. Note that $\rho$ admits the form in (2.4) with $\mu = 0$ and $\nu = \delta_1$, where $\delta_1$ is the point-mass at 1; of course, $\beta = \infty$ in this case.

Finally, we mention the related concept of the *convex level sets* (CxLS) property. A functional $\rho$ has CxLS if the level set $\{F \in \mathcal{M} : \tilde{\rho}(F) = x\}$ of $\tilde{\rho}$ is convex for each $x \in \mathbb{R}$. The CxLS property is a necessary condition for the notions of elicitability, identifiability and backtestability; see Wang and Wei (2020, Section 6) for an explanation. The above three concepts, referring to the quality and validity of risk forecasts, are notably popular in current banking regulation and model risk management. We refer to Gneiting (2011), Fissler and Ziegel (2016) and Acerbi and Szekely (2017) for more discussions on these concepts. Theorem 1 of Wang and Wei (2020) characterizes a signed Choquet integral with CxLS on a convex set $\mathcal{M}$ that contains all three-point distributions, which naturally applies to our distortion riskmetrics on general spaces. In short, up to a constant multiplier, distortion riskmetrics with CxLS only have three forms: the mean, a mixture of left and right $\alpha$-quantiles, $\alpha \in (0,1)$, and a mixture of the essential supremum and the essential infimum.

## 2.4 Continuity of distortion riskmetrics

In this section, we examine continuity of distortion riskmetrics. It is already shown in Remark 2.2 that a distortion riskmetric is Lipschitz-continuous with respect to $\mathcal{L}^\infty$-norm. Namely, for $h \in \mathcal{H}$ and $X, Y \in \mathcal{X}$,

$$|\rho_h(X) - \rho_h(Y)| \leqslant \text{ess-sup}|X - Y| \cdot \mathrm{TV}_h,$$

where $\mathrm{TV}_h$ is the total variation of $h$ on $[0,1]$.

We are then interested in continuity of a distortion riskmetric with respect to convergence in distribution, or equivalently, weak convergence in the set of distributions $\mathcal{M}$. This is closely related to robustness of a risk functional in risk management; see Krätschmer et al. (2014). Before stating the result of such continuity, we write the following relevant definition of $h$-uniform integrability. Given a convex cone $\mathcal{X}$ and $h \in \mathcal{H}$, a set $\mathcal{D} \subseteq \mathcal{X}$ is called *$h$-uniformly integrable* if

$$\lim_{k \downarrow 0} \sup_{X \in \mathcal{D}} \int_0^k |F_X^{-1}(1-t)| \, \mathrm{d}h(t) = 0$$

and

$$\lim_{k \uparrow 1} \sup_{X \in \mathcal{D}} \int_k^1 |F_X^{-1}(1-t)| \, \mathrm{d}h(t) = 0.$$

Note that $h$-uniform integrability reduces to the usual uniform integrability when $h \in \mathcal{H}$ is linear and nonconstant in some neighborhoods of 0 and 1. We give the following result for continuity of distortion riskmetrics with respect to convergence in distribution.

**Theorem 2.6.** *For $h \in \mathcal{H}$ and $X, X_1, X_2, \dots \in \mathcal{X}$, suppose that $X_n \to X$ in distribution as $n \to \infty$ and the set $\{X, X_1, X_2, \dots\}$ is $h$-uniformly integrable. If for all $t \in (0,1)$, either $s \mapsto h(s)$ or $s \mapsto F_X^{-1}(1-s)$ is continuous at $t$, then $\rho_h(X_n) \to \rho_h(X)$ as $n \to \infty$.*

Next, we consider the $\mathcal{L}^p$-continuity of distortion riskmetrics (i.e., continuity with respect to the $\mathcal{L}^p$-norm, defined as $||X||_p = (\mathbb{E}[|X|^p])^{1/p}$, $X \in \mathcal{L}^p$). We give a sufficient condition for a distortion riskmetric to be $\mathcal{L}^p$-continuous without assuming convexity of the functional, as is typically done in the literature.

**Proposition 2.4.** *For $p \in [1, \infty)$ and continuous $h \in \mathcal{H}$, a distortion riskmetric $\rho_h : \mathcal{L}^p \to \mathbb{R}$ is $\mathcal{L}^p$-continuous if $h \in \mathcal{H}_q$ where $q$ is the conjugate of $p$.*

We remark that all convex distortion riskmetrics (i.e., the ones with concave $h$ by Theorem 2.3) on $\mathcal{L}^p$ are $\mathcal{L}^p$-continuous; see Rüschendorf (2013, Corollary 7.10) for the $\mathcal{L}^p$-continuity of the finite-valued convex risk measures on $\mathcal{L}^p$.

## 2.5 Multi-dimensional distortion riskmetrics

In this section, we discuss distortion riskmetrics in a multi-dimensional setting. The importance of multi-dimensional riskmetrics arises in a statistical context, where multi-dimensional forecasting and elicitation of statistical quantities (jointly) has become a popular topic; see Lambert et al. (2008), Fissler and Ziegel (2016) and Frongillo and Kash (2021). Here, multi-dimensionality refers to the range, rather than the domain, of the riskmetrics; in other words, our riskmetrics map $\mathcal{X}$ to $\mathbb{R}^d$ for some $d \geqslant 2$. This formulation is motivated by the statistical applications mentioned above, and in particular, estimating, forecasting, and testing multiple quantities depending on a random object.

In this section, we simply extend the results in Section 2.2 to multi-dimensional distortion riskmetrics. There is essentially nothing new; nevertheless, in view of the importance of multi-dimensional riskmetrics and their applications, we collect some basic results. The distortion riskmetrics of dimension $d \geqslant 2$ are defined as follows.

**Definition 2.2.** A $d$-dimensional distortion riskmetric $\rho_{\mathbf{h}} : \mathcal{X} \to \mathbb{R}^d$ is defined as

$$\rho_{\mathbf{h}}(X) = (\rho_{h_1}(X), \dots, \rho_{h_d}(X)),$$

where $\mathbf{h} = (h_1, \dots, h_d) \in \mathcal{H}^d$. Obviously, each $\rho_{h_i}$ for $i = 1, \dots, d$ is a one-dimensional distortion riskmetric on $\mathcal{X}$.

Properties (a)-(d) in Section 2.2.3 can be equivalently formulated for $d$-dimensional distortion riskmetrics. More precisely, $\rho_{\mathbf{h}} : \mathcal{X} \to \mathbb{R}^d$ with $\mathbf{h} = (h_1, \dots, h_d)$ satisfies some of the properties (a)-(d) in Section 2.2.3 if and only if each one-dimensional distortion riskmetric $\rho_{h_i}$, $i = 1, \dots, d$, satisfies the respective properties. We can now provide the characterization result for multi-dimensional distortion riskmetrics. The same representation on $\mathcal{L}^\infty$ is given by Proposition 5 of Wang and Wei (2020).

**Proposition 2.5.** *A functional $\rho : \mathcal{X} \to \mathbb{R}^d$ is law invariant, comonotonic-additive, continuous at infinity and uniformly sup-continuous if and only if $\rho$ is a $d$-dimensional distortion riskmetric.*

Similarly to Theorem 2.6, the continuity of multi-dimensional distortion riskmetrics with respect to weak convergence is summarized below.

**Proposition 2.6.** *Let $\mathbf{h} = (h_1, \dots, h_d)$ with $h_i \in \mathcal{H}$, $i = 1, \dots, d$. For $X, X_1, X_2, \dots \in \mathcal{X}$, suppose that $X_n \to X$ in distribution as $n \to \infty$ and the set $\{X, X_1, X_2, \dots\}$ is $h_i$-uniformly integrable for all $i = 1, \dots, d$. If for any given $i = 1, \dots, d$ and for all $t \in (0, 1)$, either $s \mapsto h_i(s)$ or $s \mapsto F_X^{-1}(1-s)$ is continuous at $t$, then $\rho_{\mathbf{h}}(X_n) \to \rho_{\mathbf{h}}(X)$ as $n \to \infty$.*

Convexity and concavity cannot be naturally formulated for multi-dimensional functionals due to the lack of complete order in $\mathbb{R}^d$. On the other hand, the CxLS property can be naturally defined for multi-dimensional functionals. Similarly to Section 2.3, a multi-dimensional functional $\rho$ has CxLS if the level set $\{F \in \mathcal{M} : \tilde{\rho}(F) = x\}$ is convex for each $x \in \mathbb{R}^d$. As in the case of dimension one, multi-dimensional CxLS serves as a necessary condition for multi-dimensional elicitability, and hence it is important in the recent study of statistical elicitation.

Unlike the other properties in this section, which do not need new mathematical treatment for multi-dimensional distortion riskmetrics, the multi-dimensional CxLS is highly non-trivial to study or characterize. For instance, one-dimensional distortion riskmetrics with CxLS are characterized by Theorem 1 of Wang and Wei (2020), whereas a full characterization of multi-dimensional distortion riskmetrics with CxLS is a well-known difficult open question; see Fissler and Ziegel (2016) and Kou and Peng (2016). As far as we are aware of, the only existing characterization result on multi-dimensional distortion riskmetrics is given in Theorem 2 of Wang and Wei (2020), which identifies the form of $\rho_h$ such that $(\rho_h, \mathrm{VaR}_\alpha)$ has CxLS; note that $(\rho_h, \mathrm{VaR}_\alpha)$ is a two-dimensional distortion riskmetric.

*Remark* 2.4. Another direction of multi-dimensional generalization of riskmetrics is to consider mappings from $\mathcal{X}^d$ to $\mathbb{R}^m$ where $m$ is a positive integer, usually equal to $d$ or 1. This relates to the study of measures of multivariate risks; see e.g., Embrechts and Puccetti (2006). Our formulation in this section should not be confused with the above one. We stick to the domain $\mathcal{X}$ for the main reason that probability distortion is usually defined and well-understood in dimension one; see the recent work Liu et al. (2020) for a characterization of probability distortion in dimension one.

## 2.6 Proofs of all results

*Proof of Lemma 2.1.* (i) and (ii) can be obtained by combining the results of Lemma 3 in Wang et al. (2020) and Theorems 4 and 6 of Dhaene et al. (2012). We only prove (iii). We first suppose that $h$ is right-continuous. Since $F_X^{-1}$ is continuous on $(0,1)$, we have

$$F_X^{-1}(1-t) = F_X^{-1+}(1-t), \quad \text{for all } t \in [0,1].$$

It then follows from (i) that

$$\int X \,\mathrm{d}h \circ \mathbb{P} = \int_0^1 F_X^{-1+}(1-t)\,\mathrm{d}h(t) = \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t).$$

Then suppose that $h$ is left-continuous. According to (ii), it is straightforward that

$$\int X \,\mathrm{d}h \circ \mathbb{P} = \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t).$$

Then consider a general $h$. Since $h$ is of bounded variation, it has countably many points of discontinuity. Then we can always decompose $h = h_r + h_l$, where $h_r$ and $h_l$ are right-continuous and left-continuous parts of $h$, respectively. From (2.1), it is obvious that

$$\int X \,\mathrm{d}(ah_1 + bh_2) \circ \mathbb{P} = a \int X \,\mathrm{d}h_1 \circ \mathbb{P} + b \int X \,\mathrm{d}h_2 \circ \mathbb{P}$$

for all $h_1, h_2 \in \mathcal{H}$ and $a, b \in \mathbb{R}$. According to the above discussion,

$$\begin{aligned}\int X \,\mathrm{d}h \circ \mathbb{P} &= \int X \,\mathrm{d}h_r \circ \mathbb{P} + \int X \,\mathrm{d}h_l \circ \mathbb{P} \\ &= \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h_r(t) + \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h_l(t) = \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t).\end{aligned}$$

The other equality is similar. □

*Proof of Proposition 2.1.* (i) Recall the quantile representation of the integral $\int X \,\mathrm{d}h \circ \mathbb{P}$,

$$\int X \,\mathrm{d}h \circ \mathbb{P} = \int_0^1 F_X^{-1+}(1-t)\,\mathrm{d}h_r(t) + \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h_l(t). \tag{2.6}$$

We show finiteness of the first term in (2.6) and finiteness of the second term follows similarly. For any $\epsilon \in (0,1)$ such that $h$ is absolutely continuous in $[0,\epsilon) \cup (1-\epsilon, 1]$ and

$$h' \in \mathcal{L}^q((0,\epsilon) \cup (1-\epsilon, 1)),$$

we have $|F_X^{-1+}(1-t)| < \infty$ for all $t \in [\epsilon, 1-\epsilon]$. It follows that

$$\left| \int_\epsilon^{1-\epsilon} F_X^{-1+}(1-t)\,\mathrm{d}h_r(t) \right| < \infty$$

since $h$ is of bounded variation. It then suffices to show that

$$\left| \int_{[0,\epsilon)\cup(1-\epsilon,1]} F_X^{-1+}(1-t)\,\mathrm{d}h_r(t) \right| = \left| \int_{[0,\epsilon)\cup(1-\epsilon,1]} F_X^{-1+}(1-t) h_r'(t)\,\mathrm{d}t \right| < \infty.$$

Since $X \in \mathcal{L}^p$, the right-quantile $F_X^{-1+} \in \mathcal{L}^p([0,1])$. Note that $h_r' \in \mathcal{L}^q((0,\epsilon) \cup (1-\epsilon, 1))$ and $1/p + 1/q = 1$. By Hölder's inequality,

$$\begin{aligned}&\left| \int_{[0,\epsilon)\cup(1-\epsilon,1]} F_X^{-1+}(1-t) h_r'(t)\,\mathrm{d}t \right| \\ &\leqslant \int_{[0,\epsilon)\cup(1-\epsilon,1]} |F_X^{-1+}(1-t)| \cdot |h_r'(t)|\,\mathrm{d}t \\ &\leqslant \left( \int_{[0,\epsilon)\cup(1-\epsilon,1]} |F^{-1+}(1-t)|^p\,\mathrm{d}t \right)^{\frac{1}{p}} \left( \int_{[0,\epsilon)\cup(1-\epsilon,1]} |h_r'(t)|^q\,\mathrm{d}t \right)^{\frac{1}{q}} < \infty.\end{aligned}$$

We then conclude that

$$\left|\int_0^1 F_X^{-1+}(1-t)\,\mathrm{d}h_r(t)\right| < \infty.$$

By similar arguments, $|\int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h_l(t)| < \infty$ holds naturally. Therefore, $\int X\,\mathrm{d}h\circ\mathbb{P}$ is finite.

(ii) Concavity of $h$ implies that $h$ is absolutely continuous on $(0,1)$. Suppose that $h$ is not continuous at 0. Take $X_0 \sim \mathrm{N}(0,1)$ and $X = X_0^{1/p}$. It follows that $F_X^{-1}(1) = \infty$. By Lemma 2.1 (iii),

$$\left|\int X\,\mathrm{d}h\circ\mathbb{P}\right| = \left|\int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t)\right| = \infty,$$

which leads to a contradiction. Therefore, $h$ is continuous at 0. Continuity of $h$ at 1 holds analogously. $h$ is thus absolutely continuous on $[0,1]$. Since $h$ is of bounded variation, we can always use Jordan decomposition $h = h_+ - h_-$, where $h_+$ and $h_-$ are increasing functions. Moreover, $h$ can always be decomposed into $h = h_r + h_l$. It then suffices to prove the property for all increasing and right-continuous $h$.

Since $h$ is concave, we have $h' \in \mathcal{L}^1([0,1])$. Let

$$q' = \sup\{r \geqslant 1 : h' \in \mathcal{L}^r((0,\epsilon)\cup(1-\epsilon,1)) \text{ for some } \epsilon \in (0,1)\}$$

and suppose for the purpose of contradiction that $q' < q$. Note that we have $q'/(q'-1) > p$. Hence, there exists $\delta > 0$ such that

$$q' + \delta < q \quad \text{and} \quad \frac{q'}{q'+\delta-1} > p.$$

Let $q^* = q' + \delta$ and $p^* = q^*/(q^*-1) > p$. Note that $q^*p/p^* = (q'+\delta-1)p < q'$. Construct a random variable $X$ such that

$$\left|F_X^{-1}(1-t)\right| = |h'(t)|^{\frac{q^*}{p^*}},$$

for almost everywhere $t \in [0,1]$. This is always possible due to concavity of $h$, which implies that $h'$ is decreasing and $h'$ has countably many discontinuity points. Since $q^*p/p^* < q'$, we have $h' \in \mathcal{L}^{(q^*p/p^*)}((0,\epsilon)\cup(1-\epsilon,1))$ for some $\epsilon > 0$, and hence $X \in \mathcal{L}^p$. Noting that $h' \notin \mathcal{L}^{q^*}((0,\epsilon)\cup(1-\epsilon,1))$, we have

$$\left|\int_{[0,\epsilon)\cup(1-\epsilon,1]} F_X^{-1}(1-t)h'(t)\,\mathrm{d}t\right| = \int_{[0,\epsilon)\cup(1-\epsilon,1]} |h'(t)|^{\frac{q^*}{p^*}+1}\,\mathrm{d}t = \int_{[0,\epsilon)\cup(1-\epsilon,1]} |h'(t)|^{q^*}\,\mathrm{d}t = \infty,$$

which leads to a contradiction. Therefore, $q' \geqslant q$. □

*Proof of Theorem 2.1.* (i) "$\Rightarrow$": For all $X \in \mathcal{X}$, we define a random variable

$$X_M = X\mathbb{1}_{\{|X|\leqslant M\}} + M\mathbb{1}_{\{X>M\}} - M\mathbb{1}_{\{X<-M\}}, \quad M \geqslant 0.$$

Since $\rho$ is continuous at infinity, we have $\rho(X_M) \to \rho(X)$. Note that $X_M \in \mathcal{L}^\infty$ for any $M \geqslant 0$. It follows from Theorem 1 of Wang et al. (2020) that on $\mathcal{L}^\infty$, the law-invariant, comonotonic-additive and uniformly sup-continuous functional $\rho$ can be represented by a signed Choquet integral

$$\begin{aligned}\rho(X_M) &= \int_{-\infty}^{0} \left(h(\mathbb{P}(X_M \geqslant x)) - h(1)\right) \mathrm{d}x + \int_0^\infty h(\mathbb{P}(X_M \geqslant x))\,\mathrm{d}x \\ &= \int_{-M}^{0} \left(h(\mathbb{P}(X \geqslant x)) - h(1)\right) \mathrm{d}x + \int_0^M h(\mathbb{P}(X \geqslant x))\,\mathrm{d}x,\end{aligned} \tag{2.7}$$

where $h \in \mathcal{H}$. Specifically, $h(t) = \rho(\mathbb{1}_{\{U<t\}}) < \infty$ for $t \in [0,1]$, where $U$ is a uniform random variable on $[0,1]$. Letting $M \to \infty$, we have

$$\rho(X) = \int_{-\infty}^{0} \left(h(\mathbb{P}(X \geqslant x)) - h(1)\right) \mathrm{d}x + \int_0^\infty h(\mathbb{P}(X \geqslant x))\,\mathrm{d}x.$$

(ii) "$\Leftarrow$": Law-invariance is straightforward. Comonotonic-additivity follows from (2.6), since the left- and right-quantiles are well-known to be comonotonic-additive (see Proposition 7.20 of McNeil et al. (2015) for the case of left-quantile). Continuity at infinity holds simply by

$$\begin{aligned}\rho_h(X_M) &= \int_{-\infty}^{0} \left(h(\mathbb{P}(X_M \geqslant x)) - h(1)\right) \mathrm{d}x + \int_0^\infty h(\mathbb{P}(X_M \geqslant x))\,\mathrm{d}x \\ &= \int_{-M}^{0} \left(h(\mathbb{P}(X \geqslant x)) - h(1)\right) \mathrm{d}x + \int_0^M h(\mathbb{P}(X \geqslant x))\,\mathrm{d}x \xrightarrow{M\to\infty} \rho_h(X).\end{aligned}$$

To see the uniform sup-continuity, we take any two random variables $X, Y \in \mathcal{X}$. By representation (2.6), we have

$$\begin{aligned}&|\rho_h(X) - \rho_h(Y)| \\ &\leqslant \left|\int_0^1 \left(F_X^{-1+}(1-t) - F_Y^{-1+}(1-t)\right) \mathrm{d}h_r(t)\right| + \left|\int_0^1 \left(F_X^{-1}(1-t) - F_Y^{-1}(1-t)\right) \mathrm{d}h_l(t)\right| \\ &\leqslant \text{ess-sup}|X-Y| \cdot \mathrm{TV}_h,\end{aligned}$$

where $\mathrm{TV}_h$ is the total variation of the function $h$ on $[0,1]$. $\square$

*Proof of Proposition 2.2.* (i) Sufficiency is straightforward from the definition of distortion riskmetrics. Necessity can be checked by Bernoulli random variables.

(ii) "⇒": We take $X = \mathbb{1}_{\{U \leqslant t_1\}}$ and $Y = \mathbb{1}_{\{U \leqslant t_2\}}$ for all $t_1, t_2 \in [0,1]$ such that $t_1 \leqslant t_2$, where $U \sim \mathrm{U}[0,1]$. Then we have $X \leqslant Y$. Suppose that $\rho_h$ is increasing (resp. decreasing). We have $h(t_1) = \rho_h(X) \leqslant \rho_h(Y) = h(t_2)$ (resp. $h(t_1) = \rho_h(X) \geqslant \rho_h(Y) = h(t_2)$). Thus $h$ is increasing (resp. decreasing).

"⇐": For any random variables $X, Y \in \mathcal{X}$ such that $X \leqslant Y$, we have $\mathbb{P}(X \geqslant x) \leqslant \mathbb{P}(Y \geqslant x)$ for all $x \in \mathbb{R}$. If $h$ is increasing (resp. decreasing), then $h(\mathbb{P}(X \geqslant x)) \leqslant h(\mathbb{P}(Y \geqslant x))$ (resp. $h(\mathbb{P}(X \geqslant x)) \geqslant h(\mathbb{P}(Y \geqslant x))$) for all $x \in \mathbb{R}$. It implies that $\rho_h(X) \leqslant \rho_h(Y)$ (resp. $\rho_h(X) \geqslant \rho_h(Y)$).

(iii) For all $c \in \mathbb{R}$, we first calculate

$$\begin{aligned}\rho_h(c) &= \int_{-\infty}^{0} (h(\mathbb{P}(c \geqslant x)) - h(1))\,\mathrm{d}x + \int_0^\infty h(\mathbb{P}(c \geqslant x))\,\mathrm{d}x \\ &= \int_{0\wedge c}^{0} (-h(1))\,\mathrm{d}x + \int_0^{0\vee c} h(1)\,\mathrm{d}x = ch(1).\end{aligned}$$

Note that any random variable $X \in \mathcal{X}$ and $c$ are comonotonic. By comonotonic-additivity of $\rho_h$, we have $\rho_h(X + c) = \rho_h(X) + \rho_h(c) = \rho_h(X) + ch(1)$.

(iv) For all $\lambda > 0$ and all $X \in \mathcal{X}$,

$$\begin{aligned}\rho_h(\lambda X) &= \int_{-\infty}^{0} (h(\mathbb{P}(\lambda X \geqslant x)) - h(1))\,\mathrm{d}x + \int_0^\infty h(\mathbb{P}(\lambda X \geqslant x))\,\mathrm{d}x \\ &= \int_{-\infty}^{0} (h(\mathbb{P}(X \geqslant \frac{1}{\lambda}x)) - h(1))\,\mathrm{d}x + \int_0^\infty h(\mathbb{P}(X \geqslant \frac{1}{\lambda}x))\,\mathrm{d}x \\ &= \lambda\int_{-\infty}^{0} (h(\mathbb{P}(X \geqslant u)) - h(1))\,\mathrm{d}u + \lambda\int_0^\infty h(\mathbb{P}(X \geqslant u))\,\mathrm{d}u = \lambda\rho_h(X).\end{aligned}$$

(v) This property is already shown in the proof of Lemma 2.1 (ii). □

*Proof of Proposition 2.3.* Since $\rho_h$ is convex on $\mathcal{X}$, we know that it is convex on $\mathcal{L}^\infty$, which implies that $h$ is concave by Theorem 3 of Wang et al. (2020).

Suppose that there exists $X \in \mathcal{X}$ such that $\mathbb{E}[|X|] = \infty$. Note that $\mathbb{E}[|X|] = \infty$ implies either $\mathbb{E}[X_+] = \infty$ or $\mathbb{E}[X_-] = \infty$. If $\mathbb{E}[X_+] = \infty$, then $Y = -X \in \mathcal{X}$ since $\mathcal{X}$ is a linear space, and $\mathbb{E}[Y_-] = \infty$. Similarly, if $\mathbb{E}[X_-] = \infty$, then $\mathbb{E}[Y_+] = \infty$. Therefore, we know that there exist $X, Y \in \mathcal{X}$ such that $\mathbb{E}[X_+] = \mathbb{E}[Y_-] = \infty$.

Take $X \in \mathcal{X}$ with $\mathbb{E}[X_+] = \infty$. Since

$$\rho_h(X) = \int_{-\infty}^{0} (h(\mathbb{P}(X \geqslant x)) - h(1))\,\mathrm{d}x + \int_0^\infty h(\mathbb{P}(X \geqslant x))\,\mathrm{d}x \in \mathbb{R},$$

both $\int_{-\infty}^{0}(h(\mathbb{P}(X \geqslant x))-h(1))\,\mathrm{d}x$ and $\int_{0}^{\infty}h(\mathbb{P}(X \geqslant x))\,\mathrm{d}x$ have to be finite. Since $X$ is unbounded from above, this implies that $h$ is continuous at 0. Similarly, take $Y \in \mathcal{X}$ with $\mathbb{E}[Y_-]=\infty$, and we obtain $h$ is continuous at 1. Further by concavity, $h$ is continuous on $[0,1]$. Using Lemma 2.1, we get

$$\rho_h(X)=\int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t).$$

There exists $\delta>0$ such that $F_X^{-1}(1-\epsilon)>0$ for all $\epsilon \in (0,\delta)$. Moreover,

$$\epsilon\int_0^{\epsilon} F_X^{-1}(1-t)\,\mathrm{d}t=\infty$$

for all $\epsilon \in (0,\delta)$. Let $h'(t)$ be the right-derivative of $h$ at $t \in [0,1)$. Assume that $h'(0)>0$. Since $h$ is concave and continuous, there exists $\epsilon>0$ such that $h'(t)>\epsilon$ for $t\in[0,\epsilon]$. It follows that

$$\int_0^{\epsilon} F_X^{-1}(1-t)\,\mathrm{d}h(t) \geqslant \epsilon\int_0^{\epsilon} F_X^{-1}(1-t)\,\mathrm{d}t=\infty,$$

contradicting the fact that $\rho_h(X)$ is finite. Therefore, $h'(0)\leqslant 0$. Using similar arguments as above for $Y$, we obtain $h'(1)\geqslant 0$ where $h'(1)$ is the left derivative of $h$ at 1. Since $h$ is concave, these two conditions imply that $h=0$ on $[0,1]$, and hence $\rho_h=0$ on $\mathcal{X}$. □

*Proof of Theorem 2.2.* (i) "⇒": Suppose that $X \leqslant_{\mathrm{cx}} Y$. We first consider the case where $h\in\mathcal{H}$ is increasing. For an increasing concave function $h\in\mathcal{H}$, it is well-known (Williamson, 1956, e.g., Theorem 1 of) that there exists some finite Borel measure $\mu$ on $[0,1]$, such that

$$h(t)=\int_0^1 \frac{1}{u}h_u(t)\,\mathrm{d}\mu(u), \quad t\in[0,1], \tag{2.8}$$

where $h_u(t)=t\wedge u$ for $t,u\in[0,1]$ and we use the convention $h_u(t)/u=\mathbb{1}_{\{t>0\}}$ if $u=0$. By the quantile representation of a distortion riskmetric,

$$\rho_{h_u}(X)=\int_0^u F_X^{-1}(1-t)\,\mathrm{d}t=\int_{1-u}^1 F_X^{-1}(u)\,\mathrm{d}u\leqslant\int_{1-u}^1 F_Y^{-1}(u)\,\mathrm{d}u=\rho_{h_u}(Y),$$

where the third inequality holds by Theorem 3.A.5 of Shaked and Shanthikumar (2007). It follows that

$$\rho_h(X)=\int_0^1 \frac{1}{u}\rho_{h_u}(X)\,\mathrm{d}\mu(u)\leqslant\int_0^1 \frac{1}{u}\rho_{h_u}(Y)\,\mathrm{d}\mu(u)=\rho_h(Y).$$

When $h\in\mathcal{H}$ is decreasing, similar to (2.8), we have

$$h(t)=\int_0^1 \frac{1}{1-u}(h_u(t)-t)\,\mathrm{d}\nu(u), \quad t\in[0,1]$$

for some finite Borel measure $\nu$ on $[0,1]$ where the convention is $(h_u(t)-t)/(1-u) = -\mathbb{1}_{\{t=1\}}$ if $u=1$. By definition of $X \leqslant_{\mathrm{cx}} Y$, it implies that $\mathbb{E}[X]=\mathbb{E}[Y]$. It then follows that

$$\rho_h(X) = \int_0^1 \frac{1}{1-u}(\rho_{h_u}(X)-\mathbb{E}[X])\,\mathrm{d}\nu(u) \leqslant \int_0^1 \frac{1}{1-u}(\rho_{h_u}(Y)-\mathbb{E}[Y])\,\mathrm{d}\nu(u) = \rho_h(Y).$$

For any concave function $h$ on $[0,1]$, there always exists $\hat{x}\in[0,1]$, such that $h(\hat{x}) \geqslant h(x)$ for all $x\in[0,1]$. Then $h$ can always be decomposed by $h = h_{\uparrow} + h_{\downarrow}$, where

$$h_{\uparrow}(x) = h(x)\mathbb{1}_{\{0\leqslant x<\hat{x}\}} + h(\hat{x})\mathbb{1}_{\{\hat{x}\leqslant x\leqslant 1\}} \text{ and } h_{\downarrow}(x) = [h(x)-h(\hat{x})]\mathbb{1}_{\{\hat{x}\leqslant x\leqslant 1\}}.$$

Notice that $h_{\uparrow}$ and $h_{\downarrow}$ are increasing and decreasing concave functions, respectively, with

$$h_{\uparrow}(0) = h_{\downarrow}(0) = 0.$$

According to the above arguments, we have

$$\rho_h(X) = \rho_{h_{\uparrow}}(X) + \rho_{h_{\downarrow}}(X) \leqslant \rho_{h_{\uparrow}}(Y) + \rho_{h_{\downarrow}}(Y) = \rho_h(Y).$$

(ii) "$\Leftarrow$": Suppose that $\rho_h(X) \leqslant \rho_h(Y)$ for all concave functions $h\in\mathcal{H}$. For all $t,u\in[0,1]$, choose a concave $h\in\mathcal{H}$ such that $h(t) = h_u(t) = t\wedge u$. Then for all $u\in[0,1]$,

$$\rho_h(X) = \int_{1-u}^1 F_X^{-1}(u)\,\mathrm{d}u \text{ and } \rho_h(Y) = \int_{1-u}^1 F_Y^{-1}(u)\,\mathrm{d}u.$$

It follows that

$$\int_{1-u}^1 F_X^{-1}(u)\,\mathrm{d}u \leqslant \int_{1-u}^1 F_Y^{-1}(u)\,\mathrm{d}u \text{ for all } u\in[0,1],$$

which is equivalent to $X \leqslant_{\mathrm{cx}} Y$ by Theorem 3.A.5 of Shaked and Shanthikumar (2007). □

*Proof of Theorem 2.3.* $(i) \Rightarrow (ii)$ is shown by Theorem 2.2. We proceed in the order $(ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (vi) \Rightarrow (i)$, and the arguments are based on Theorem 3 of Wang et al. (2020).

$(ii) \Rightarrow (iii)$: Take random variables $X, Y, X^c, Y^c \in \mathcal{X}$, such that $X \stackrel{\mathrm{d}}{=} X^c$, $Y \stackrel{\mathrm{d}}{=} Y^c$ and $X^c$ and $Y^c$ are comonotonic. By Theorem 3.5 of Rüschendorf (2013), we have $X+Y \leqslant_{\mathrm{cx}} X^c+Y^c$. It then follows from law-invariance, comonotonic-additivity and convex order consistency of $\rho_h$ that

$$\rho_h(X+Y) \leqslant \rho_h(X^c+Y^c) = \rho_h(X^c)+\rho_h(Y^c) = \rho_h(X)+\rho_h(Y).$$

$(iii) \Rightarrow (iv)$: As $\rho_h$ is positively homogeneous, subadditivity is equivalent to convexity.

$(iv) \Rightarrow (v)$: Directly from the definition of convexity and quasi-convexity.

$(v) \Rightarrow (vi)$: Theorem 3 of Wang et al. (2020) gives that quasi-convexity of $I_h$ on $\mathcal{L}^\infty$ implies that $h$ is concave. Concavity on mixtures follows directly from the concavity of $h$ by the definition of a distortion riskmetric.

$(vi) \Rightarrow (i)$: Theorem 3 of Wang et al. (2020) gives that mixture-concavity of $I_h$ on $\mathcal{L}^\infty$ implies that $h$ is concave. $\square$

*Proof of Theorem 2.4.* Suppose that $\rho : \mathcal{X} \to (-\infty, \infty]$ is a law-invariant, convex and continuous-at-infinity functional dominating $\rho_h$. Using Theorem 5 of Wang et al. (2020), we know that, on $\mathcal{L}^\infty$, $\rho_{h^*}$ is the smallest law-invariant convex functional dominating $\rho_h$. Therefore, $\rho \geqslant \rho_{h^*}$ on $\mathcal{L}^\infty$. If $\rho_{h^*}$ is finite on $\mathcal{X}$, then both $\rho$ and $\rho_{h^*}$ are continuous at infinity on $\mathcal{X}$, and hence $\rho \geqslant \rho_{h^*}$ on $\mathcal{X}$. If $\rho_{h^*}$ is not finite on $\mathcal{X}$, then we know that $\int X \, \mathrm{d}h^* \circ \mathbb{P} = \infty$ (but not $-\infty$ since $\rho_{h^*} \geqslant \rho_h$) for some $X \in \mathcal{X}$. Let

$$X_M = X \mathbb{1}_{\{|X| \leqslant M\}} + M \mathbb{1}_{\{X > M\}} - M \mathbb{1}_{\{X < -M\}}, \quad M \geqslant 0.$$

Using (2.7), $\rho = \rho_{h^*}$ on $\mathcal{L}^\infty$ and $\int X \, \mathrm{d}h^* \circ \mathbb{P} = \infty$, we have, as $M \to \infty$,

$$\rho(X_M) = \rho_{h^*}(X_M) = \int_{-M}^{0} (h^*(\mathbb{P}(X \geqslant x)) - h(1)) \, \mathrm{d}x + \int_0^M h^*(\mathbb{P}(X \geqslant x)) \, \mathrm{d}x \to \infty.$$

The continuity at infinity of $\rho$ implies $\rho(X) = \infty$, and hence $\rho$ cannot be real-valued on $\mathcal{X}$. $\square$

*Proof of Theorem 2.5.* Note that $X \mapsto \mathrm{ES}_\alpha(X)$ and $X \mapsto \mathrm{ES}_\alpha(-X)$ are convex distortion riskmetrics for all $\alpha \in [0,1]$. As a mixture of $X \mapsto \mathrm{ES}_\alpha(X)$ and $X \mapsto \mathrm{ES}_\alpha(-X)$, $\rho$ defined by (2.4) satisfies convexity, comonotonic-additivity, law-invariance, continuity at infinity, and uniform sup-continuity. Hence, $\rho$ is a convex distortion riskmetric. Next we show the "only-if" statement. Denote by $h$ the distortion function of $\rho$, which by Theorem 2.3 is a concave function. Following the same argument in the proof of Theorem 2.2, we can write for some finite Borel measures $\gamma, \nu$ on $[0,1]$,

$$h(t) = \int_0^1 \frac{1}{\alpha} h_\alpha(t) \, \mathrm{d}\gamma(\alpha) + \int_0^1 \frac{1}{1-\alpha} (h_\alpha(t) - t) \, \mathrm{d}\nu(\alpha), \quad t \in [0,1], \tag{2.9}$$

where $h_\alpha(t) = t \wedge \alpha$. Note that $\frac{1}{\alpha} h_\alpha$ is the distortion function of $\mathrm{ES}_{1-\alpha}$. By Proposition 2.2, the distortion function of $X \mapsto \mathrm{ES}_\alpha(-X)$ is given by

$$g_\alpha(t) = \frac{1-t}{1-\alpha} \wedge 1 - 1 = \frac{(\alpha - t) \wedge 0}{1-\alpha} = \frac{1}{1-\alpha} (h_\alpha(t) - t), \quad t \in [0,1].$$

Therefore, (2.9) gives

$$\rho(X) = \int_0^1 \mathrm{ES}_{1-\alpha}(X) \, \mathrm{d}\gamma(\alpha) + \int_0^1 \mathrm{ES}_\alpha(-X) \, \mathrm{d}\nu(\alpha), \qquad X \in \mathcal{X}.$$

Thus (2.4) holds with $\mathrm{d}\mu(\alpha) = \mathrm{d}\gamma(1-\alpha)$. $\square$

*Proof of Theorem 2.6.* Since $h \in \mathcal{H}$ is of bounded variation, it can be decomposed into $h = h_+ - h_-$ where $h_+$ and $h_-$ are increasing functions. It then suffices to prove the result for all increasing function $h$. We denote the distribution function of $X_n$ by $F_n$ for $n \in \mathbb{N}$.

(i) If $h$ is left-continuous and increasing, it induces a Borel measure $\mu$ on $[0,1]$ such that $h(t) = \mu([0,t))$, $t \in [0,1]$. By quantile representation of a distortion riskmetric,

$$\rho_h(X_n) = \int_0^1 F_n^{-1}(1-t)\,\mathrm{d}h(t) \text{ and } \rho_h(X) = \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t).$$

Since $X_n \to X$ in distribution, $F_n^{-1} \to F_X^{-1}$ almost everywhere on $[0,1]$, where $F_X^{-1}$ is continuous. Let

$$A = \{t \in (0,1) : s \mapsto F_X^{-1}(1-s) \text{ is not continuous at } t\}.$$

According to the assumption, $h$ must be continuous on the set $A$, which implies $\mu$ has no point mass on $A$ and $\mu(A) = 0$. It remains to consider the points 0 and 1. Notice that $h$-uniform integrability implies that when $\mu(\{0\}) > 0$, $F_n^{-1}(1) \to F_X^{-1}(1)$ as $n \to \infty$ since $F_n^{-1}(1) = F_X^{-1}(1) = 0$ for all $n \in \mathbb{N}$. Similarly, when $\mu(\{1\}) > 0$, $F_n^{-1}(0) \to F_X^{-1}(0) = 0$ as $n \to \infty$. Therefore, $F_n^{-1} \to F_X^{-1}$ $\mu$-almost surely. In addition, $h$-uniform integrability of $\{X_1, X_2, \dots\}$ is equivalent to uniform integrability of $\{F_1^{-1}, F_2^{-1}, \dots\}$ with respect to the measure $\mu$. It then follows from Vitali's Convergence Theorem (Rudin, 1987, p. 133) that $\rho_h(X_n) \to \rho_h(X)$ as $n \to \infty$.

(ii) If $h$ is right-continuous, we define the Borel measure $\nu$ on $[0,1]$ by $\nu([0,t]) = h(t)$, $t \in [0,1]$. We write the distortion riskmetrics as

$$\rho_h(X_n) = \int_0^1 F_n^{-1+}(1-t)\,\mathrm{d}h(t) \text{ and } \rho_h(X) = \int_0^1 F_X^{-1+}(1-t)\,\mathrm{d}h(t).$$

Note that the set

$$\begin{aligned} B &= \{t \in (0,1) : s \mapsto F_X^{-1+}(1-s) \text{ is not continuous at } t\} \\ &= \{t \in (0,1) : s \mapsto F_X^{-1}(1-s) \text{ is not continuous at } t\}. \end{aligned}$$

This implies $\nu(B) = 0$. By similar argument as (i), we get $F_n^{-1+} \to F_X^{-1+}$ $\nu$-almost surely and $\rho_h(X_n) \to \rho_h(X)$ as $n \to \infty$.

(iii) For a general $h$, we can write $\rho_h$ by (2.6), where $h_r$ and $h_l$ are taken such that the collection of discontinuity points of $h_r$ and $h_l$ coincides with that of $h$. To see that it is always possible, we define countable sets

$$C = \{t \in [0,1] : s \mapsto h(s) \text{ is not continuous at } t\},$$
$$C^+ = \{t \in C : s \mapsto h(s) \text{ is right-continuous at } t\} \text{ and } C^- = C \setminus C^+.$$

Take

$$h_r(x) = \sum_{t\in C^+} [h(t^+) - h(t^-)]\mathbb{1}_{\{x>t\}} + h(x)\mathbb{1}_{\{x\notin C\}} \text{ and } h_l(x) = \sum_{t\in C^-} [h(t^+) - h(t^-)]\mathbb{1}_{\{x\geqslant t\}}$$

for $x \in [0,1]$. Thus $h_r$ and $h_l$ are as desired. It follows that

$$|\rho_h(X_n) - \rho_h(X)| \leqslant |\rho_{h_r}(X_n) - \rho_{h_r}(X)| + |\rho_{h_l}(X_n) - \rho_{h_l}(X)| \to 0$$

as $n \to \infty$. This implies $\rho_h(X_n) \to \rho_h(X)$ as $n \to \infty$ in general. □

*Proof of Proposition 2.4.* Suppose that we have random variables $X_1, X_2, \dots \in \mathcal{L}^p$ such that $X_n \to X$ in $\mathcal{L}^p$ as $n \to \infty$. Let $F_n$ be the distribution function of $X_n$ for $n \in \mathbb{N}$. Since $h \in \mathcal{H}_q$, there exists $\epsilon \in (0,1)$ such that $h' \in \mathcal{L}^q((0,\epsilon) \cup (1-\epsilon, 1))$. Then we have

$$\begin{aligned} |\rho_h(X_n) - \rho_h(X)| \leqslant & \left| \int_{[0,\epsilon)\cup(1-\epsilon,1]} (F_n^{-1}(1-t) - F_X^{-1}(1-t)) \,\mathrm{d}h(t) \right| \\ & + \left| \int_{[\epsilon,1-\epsilon]} (F_n^{-1}(1-t) - F_X^{-1}(1-t)) \,\mathrm{d}h(t) \right|. \end{aligned} \tag{2.10}$$

By Hölder's inequality, the first term of (2.10) satisfies

$$\begin{aligned} & \left| \int_{[0,\epsilon)\cup(1-\epsilon,1]} (F_n^{-1}(1-t) - F_X^{-1}(1-t)) \,\mathrm{d}h(t) \right| \\ & \leqslant \int_{[0,\epsilon)\cup(1-\epsilon,1]} \left|F_n^{-1}(1-t) - F_X^{-1}(1-t)\right| \cdot |h'(t)| \,\mathrm{d}t \\ & \leqslant \left( \int_{[0,\epsilon)\cup(1-\epsilon,1]} \left|F_n^{-1}(1-t) - F_X^{-1}(1-t)\right|^p \mathrm{d}t \right)^{\frac{1}{p}} \left( \int_{[0,\epsilon)\cup(1-\epsilon,1]} |h'(t)|^q \,\mathrm{d}t \right)^{\frac{1}{q}} \xrightarrow{n\to\infty} 0. \end{aligned}$$

It remains to show the second term of (2.10) converges to zero. Note that

$$\begin{aligned} \left| \int_{[\epsilon,1-\epsilon]} (F_n^{-1}(1-t) - F_X^{-1}(1-t)) \,\mathrm{d}h(t) \right| &= \left| \int_0^1 (F_n^{-1}(1-t) - F_X^{-1}(1-t)) \,\mathrm{d}\tilde{h}(t) \right| \\ &= |\rho_{\tilde{h}}(X_n) - \rho_{\tilde{h}}(X)|, \end{aligned}$$

where

$$\tilde{h}(t) = \begin{cases} 0 & t \in [0,\epsilon), \\ h(t) - h(\epsilon) & t \in [\epsilon, 1-\epsilon], \\ h(1-\epsilon) - h(\epsilon) & t \in (1-\epsilon, 1]. \end{cases}$$

Clearly, $\{X, X_1, X_2, \dots\}$ is uniformly $\tilde{h}$-integrable since $\tilde{h}$ stays constant in some neighborhood of 0 and 1. Also, $X_n \to X$ in $\mathcal{L}^p$ implies $X_n \to X$ in distribution and $\tilde{h}$ is continuous due to $h$ being continuous. It then follows from Theorem 2.6 that

$$|\rho_{\tilde{h}}(X_n) - \rho_{\tilde{h}}(X)| \to 0 \quad \text{as } n \to \infty.$$

Therefore, the second term of (2.10) also converges to zero. We conclude that $\rho_h(X_n) \to \rho_h(X)$ as $n \to \infty$, which proves the proposition. □

*Proof of Proposition 2.5.* The proposition follows by applying Theorem 2.2 to each dimension of $\rho$. □

*Proof of Proposition 2.6.* The proposition follows by applying Theorem 2.6 to each dimension of $\rho$. □

# Chapter 3

# Optimizing distortion riskmetrics with distributional uncertainty

## 3.1 Introduction

In this chapter, we focus on *distortion riskmetrics* which is a large class of commonly used measures of risk and variability; see Chapter 2 for the terminology "distortion riskmetrics". Distortion riskmetrics include L-functionals (Huber and Ronchetti, 2009) in statistics, Yaari's dual utilities (Yaari, 1987) in decision theory, distorted premium principles (Wang et al., 1997) in insurance, and spectral risk measures (Acerbi, 2002) in finance; see Chapter 2 for further examples. After a normalization, increasing distortion riskmetrics are *distortion risk measures*, which include, in particular, the two most important risk measures used in current banking and insurance regulation, the Value-at-Risk (VaR) and the Expected Shortfall (ES). Moreover, convex distortion riskmetrics are the building blocks (via taking a supremum) for all convex risk functionals (Liu et al., 2020), including classic risk measures (Artzner et al., 1999; Föllmer and Schied, 2002a) and deviation measures (Rockafellar et al., 2006).

When riskmetrics are evaluated on distributions that are subject to uncertainty, decisions should be taken with respect to the worst (or best) possible values a riskmetric attains over a set of alternative distributions; giving rise to the active subfield of distributionally robust optimization. The set of alternative distributions, the *uncertainty set*, may be characterized by moment constraints (e.g., Popescu, 2007), parameter uncertainty (e.g., Delage and Ye, 2010), probability constraints (e.g., Wiesemann et al., 2014), and distributional distances (e.g., Jiang and Guan, 2016; Esfahani and Kuhn, 2018; Blanchet and Murthy, 2019), amongst others. Distributionally robust optimization problems have been studied under the framework of expected utility (e.g., Popescu, 2007; Chen et al., 2011) and further under shortfall risk measures (e.g., Delage et al., 2022). As an important class of risk measures, distortion risk measures have also been considered as a natural choice of objectives for distributionally robust optimization. Popular distortion risk measures such as VaR and ES are studied extensively in this context; see e.g., Natarajan et al. (2008) and Zhu and Fukushima (2009).

Optimization of convex distortion risk measures, i.e., distortion riskmetrics with an increasing and concave distortion function, is relatively well understood under distributional uncertainty; see Cornilly et al. (2018), Li (2018), and Liu et al. (2020) for some recent work. Nevertheless, many distortion riskmetrics are not convex or monotone. For example, in the Cumulative Prospect Theory of Tversky and Kahneman (1992), the distortion function is typically assumed to be inverse-S-shaped; in financial risk management, the popular risk measure VaR has a non-concave distortion function, and the inter-quantile difference (Wang et al., 2020) has a distortion function that is neither concave nor monotone. Another example is the difference between two distortion risk measures, which is clearly not increasing or convex in general. Optimizing non-convex distortion riskmetrics under distributional uncertainty is difficult and results are available only for special cases; see Li et al. (2018), Cai et al. (2018), Zhu and Shao (2018), Wang et al. (2019), and Bernard et al. (2020), all with an increasing distortion function.

There is, however, a notable common feature in the above mentioned literature when a non-convex distortion risk metric is involved. For numerous special cases, one often obtains an equivalence between the optimization problem with non-convex distortion riskmetric and that with a convex one. Inspired by this observation, the aim of this chapter is to address:

What conditions provide equivalence between a non-convex riskmetric and a convex one in the setting of distributional uncertainty?

An answer to this question is still missing in the literature. In this sense, we offer a novel perspective on distributionally robust optimization problems by converting optimization problems with non-convex objectives to their convex counterpart. Transforming an optimization problem with a non-convex objective to a convex one through approximation and via a direct equivalence has been studied by Zymler et al. (2013) and Cai et al. (2020). Both contributions, however, consider uncertainty sets described by some special forms of constraints. A unifying framework applicable to numerous uncertainty sets and the entire class of distortion riskmetrics is however missing and at the core of this chapter.

The main novelty of our results is three-fold: first, we obtain a unifying result (Theorem 3.1) that allows, under distributional uncertainty, to convert an optimization problem of a non-convex distortion riskmetric to an optimization problem with a convex one. The result covers, to the authors' best knowledge, all known equivalences between optimization problems of non-convex and convex riskmetrics with distributional uncertainty. The proof requires techniques beyond the ones used in the existing literature, as we do not make assumptions such as monotonicity, positiveness, and continuity. Our framework can also be applied to settings with atomic probability space or with uncertainty sets of multi-dimensional distributions. Second, we introduce the concept of *closedness under concentration* as a sufficient condition to establish the equivalence, and it is also a necessary condition on the set of optimizers given that the equivalence holds (Theorem 3.2). We show how the properties of closedness under concentration within a collection of intervals $\mathcal{I}$ and closedness under concentration for all intervals can be verified through direct analysis and provide numerous examples. Third, the classes of distortion riskmetrics and uncertainty formulations considered in this chapter include all special cases studied in the literature; examples are presented in Sections 3.3-3.4. In particular, our class of riskmetrics include all practically used risk measures and variability measures (some via taking a sup), dual utilities with inverse-S-shaped distortion functions of Tversky and Kahneman (1992), and differences between two dual utilities or distortion risk measures. Our uncertainty formulations include both supremum and infimum problems,[1] moment constraints, convex order/risk measure constraints, marginal constraints in risk aggregation with dependence uncertainty (e.g., Embrechts et al., 2015), preference robust optimization (e.g., Armbruster and Delage, 2015; Guo and Xu, 2020), and some one-dimensional and multi-dimensional uncertainty sets induced by Wasserstein metrics.

The generality of our work distinguishes it from the large literature on distributional robust optimization cited above. Our work is of analytical and probabilistic nature, and we focus on theoretical equivalence results which will be also illustrated via numerical implementations. The target problems are formulated in Section 3.2. Sections 3.3 is devoted to our main contribution of the equivalence of the optimization problems with non-convex and convex objectives under distributional uncertainty. We illustrate by many examples the concepts of closedness under conditional

[1]Thus we provide a universal treatment of worst-case and best-case risk values. Calculating best-case risk values allows us to solve economic decision making problems where optimal distributions are chosen to minimize the risk.

expectation and closedness under concentration, and distinguish them in several practical settings. Section 3.4 demonstrates the equivalence results in multi-dimensional settings. In addition to a general multi-dimensional model with a concave loss function, we solve a robust risk aggregation problem with ambiguity on both the marginal distributions and the dependence structure. In Section 3.5, our results are used to solve optimization problems with uncertainty sets defined via moment constraints. In particular, we generalize a few well-known results in the literature on optimization and worst-case values of risk measures. Sections 3.6 and 3.7 contain numerical illustrations of optimizing differences between two distortion riskmetrics, portfolio optimization, and preference robust optimization. Some concluding remarks are put in Section 3.8. Proofs of all results are relegated to Section 7.12.

## 3.2 Distortion riskmetrics with distributional uncertainty

### 3.2.1 Problem formulation

Throughout, we work with an atomless probability space $(\Omega, \mathscr{F}, \mathbb{P})$. For $n \in \mathbb{N}$, $A$ represents a set of actions, $\rho$ is an objective functional, $f : A \times \mathbb{R}^n \to \mathbb{R}$ is a loss function, and $\mathbf{X}$ is an $n$-dimensional random vector with distributional uncertainty. Many problems in distributionally robust optimization have the form

$$\min_{\mathbf{a}\in A} \sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \rho(f(\mathbf{a}, \mathbf{X})), \tag{3.1}$$

where $F_{\mathbf{X}}$ denotes the distribution of $\mathbf{X}$ and $\widetilde{\mathcal{M}}$ is a set of plausible distributions for $\mathbf{X}$. We will first focus on the inner problem

$$\sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \rho(f(\mathbf{a}, \mathbf{X})), \tag{3.2}$$

which we may rewrite as

$$\sup_{F_Y\in\mathcal{M}} \rho(Y), \tag{3.3}$$

where $F_Y$ denotes the distribution of $Y$ and $\mathcal{M}$ is a set of distributions on $\mathbb{R}$. We suppress the reliance on $\mathbf{a}$ as it remains constant in the inner problem (3.2). The supremum in (3.3) is typically referred to as the *worst-case risk measure* in the literature if $\rho$ is monotone.[2] The problem (3.3) can also represent an optimal decision problem, where $\rho$ is an objective to maximize, and a decision maker chooses an optimal distribution from the set $\mathcal{M}$ which is interpreted as an action set instead

[2]A risk measure $\rho : \mathcal{L}^p \to \mathbb{R}$ is *monotone* if $\rho(X) \leqslant \rho(Y)$ for all $X, Y \in \mathcal{L}^p$ with $X \leqslant Y$.

of an uncertainty set (i.e., no uncertainty in this problem). Since the two problems share the same mathematical formulation (3.3), we will navigate through our results mainly with the first interpretation of worst-case risk under uncertainty.

Define the set $\mathcal{H}$ as that in Chapter 2. For $p \in [1, \infty]$ and a distortion function $h \in \mathcal{H}$, a distortion riskmetric $\rho_h : \mathcal{L}^p \to \mathbb{R}$ is defined as in Definition 2.1 of Chapter 2 on the $\mathcal{L}^p$ space. See Proposition 3.5 below for a sufficient condition of the finiteness of $\rho_h$. Note that we allow $h$ to be non-monotone; if $h$ is increasing and $h(1) = 1$, then $\rho_h$ is a distortion risk measure. The distortion riskmetric $\rho_h$ is convex if and only if $h$ is concave; see Wang et al. (2020) for this and other properties of $\rho_h$.

In this chapter we consider the objective functional $\rho$ in (3.1) to be a distortion riskmetric $\rho_h$ for some $h \in \mathcal{H}$, as the class of distortion riskmetrics includes a large class of objective functionals of interest. Note that a general analysis of (3.3) also covers the infimum problem $\inf_{F_Y \in \mathcal{M}} \rho_h(Y)$, since $-\rho_h = \rho_{-h}$ is again a distortion riskmetric. This illustrates an advantage of studying distortion riskmetrics over monotone ones, as our analysis unifies best- and worst-case risk evaluations. Best-case risk measures are also of practical importance. In particular, they may represent risk minimization problems through the second interpretation of (3.3), where $\mathcal{M}$ represents a set of possible actions (see Section 3.3.4 for some examples).

If $\rho_h$ is not convex, or equivalently, $h$ is not concave, problems such as (3.1) and (3.3) are often highly nontrivial. However, the optimization problem of maximizing $\rho_{h^*}(Y)$ over $F_Y \in \mathcal{M}$, where $h^*$ is the smallest concave distortion function dominating $h$, can often be solved either analytically or through numerical methods. Note that $\rho_h$ is mixture concave (i.e., $F_X \mapsto \rho(X)$ is concave) if and only if $h$ is concave by Theorem 3 of Wang et al. (2020). As a consequence, if $f(\mathbf{a}, \mathbf{x})$ is convex in $\mathbf{a}$ (for instance, in portfolio selection, a common choice is $f(\mathbf{a}, \mathbf{x}) = \mathbf{a}^\top \mathbf{x}$), then the optimization (3.1) for $\rho_{h^*}$,

$$\min_{\mathbf{a} \in A} \sup_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_{h^*}(f(\mathbf{a}, \mathbf{X})),$$

has an objective $\rho_{h^*}(f(\mathbf{a}, \mathbf{X}))$ which is convex in $\mathbf{a}$ and concave in $F_{\mathbf{X}}$. This is a standard convex-concave minimax problem in the optimization literature and various computational methods exist (e.g., Korpelevich, 1976; Nemirovski, 2004; Ouyang and Xu, 2021). To utilize this observation for optimizing $\rho_h$, the crucial condition is

$$\sup_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_h(f(\mathbf{a}, \mathbf{X})) = \sup_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_{h^*}(f(\mathbf{a}, \mathbf{X})),$$

that is, with $Y = f(\mathbf{a}, \mathbf{X})$,

$$\sup_{F_Y \in \mathcal{M}} \rho_h(Y) = \sup_{F_Y \in \mathcal{M}} \rho_{h^*}(Y). \tag{3.4}$$

Also note that $\rho_h \leqslant \rho_{h^*}$ always holds, and hence for (3.4), it suffices to study the "$\geqslant$" inequality.

The main contribution of this chapter is a sufficient condition on the uncertainty set $\mathcal{M}$ that guarantees the equivalence (3.4). We will also obtain a necessary condition for (3.4). The equivalence (3.4) makes the optimization problem (3.1) for $\rho_h$ much more tractable in various settings, which will be illustrated through the examples in the following sections.

### 3.2.2 Notation and preliminaries

For $p \geqslant 1$ and $n \in \mathbb{N}$, we denote by $\mathcal{M}_p^n$ the set of all distributions on $\mathbb{R}^n$ with finite $p$-th moment. Let $\mathcal{M}_\infty^n$ be the set of $n$-dimensional distributions of bounded random variables. For $p \in [1, \infty]$, write $\mathcal{M}_p^1 = \mathcal{M}_p$ for simplicity. The set inclusion $\subseteq$ and terms like "increasing" and "decreasing" are in the non-strict sense. Since $h \in \mathcal{H}$ is of bounded variation, its discontinuity points are at most countable and the left- and right-limits exist at each of these points. We write

$$h(t^+) = \begin{cases} \lim_{x \downarrow t} h(x), & t \in [0,1), \\ h(1), & t = 1, \end{cases} \quad \text{and} \quad h(t^-) = \begin{cases} \lim_{x \uparrow t} h(x), & t \in (0,1], \\ h(0), & t = 0, \end{cases}$$

and the upper semicontinuous modification of $h$ is denoted by

$$\hat{h}(t) = h(t^+) \vee h(t^-) \vee h(t), \quad t \in (0,1), \quad \text{with } \hat{h}(0) = 0 \text{ and } \hat{h}(1) = h(1).$$

Note that $\hat{h}(t) = h(t)$ at all continuous points of $h$, and we do not make any modification at the points 0 and 1 even if $h$ has a jump at these points. For $h \in \mathcal{H}$ and $t \in [0,1]$, define its concave and convex envelopes $h^*$ and $h_*$ respectively by

$$h^*(t) = \inf \{ g(t) : \ g \in \mathcal{H}, \ g \geqslant h, \ g \text{ is concave on } [0,1] \},$$

$$h_*(t) = \sup \{ g(t) : \ g \in \mathcal{H}, \ g \leqslant h, \ g \text{ is convex on } [0,1] \}.$$

Both $h^*$ and $h_*$ are continuous functions on $(0,1)$ for all $h \in \mathcal{H}$, and if $h$ is continuous at 0 and 1, then so are $h^*$ and $h_*$ (see Figure 3.4 below for an illustration of $h$ and $h^*$). Denote by $\mathcal{H}^*$ (resp. $\mathcal{H}_*$) the set of concave (resp. convex) functions in $\mathcal{H}$. Note that for all $h \in \mathcal{H}$, we have $h^* \in \mathcal{H}^*$ and $h_* \in \mathcal{H}_*$. As a well-known property of the convex and concave envelopes of a continuous $h$ (e.g., Brighi and Chipot, 1994), $h^*$ (resp. $h_*$) differs from $h$ on a union of disjoint open intervals, and $h^*$ (resp. $h_*$) is linear on these intervals. The functions $h$, $\hat{h}$, $h^*$ and $(\hat{h})^*$ are illustrated in Figure 3.1.

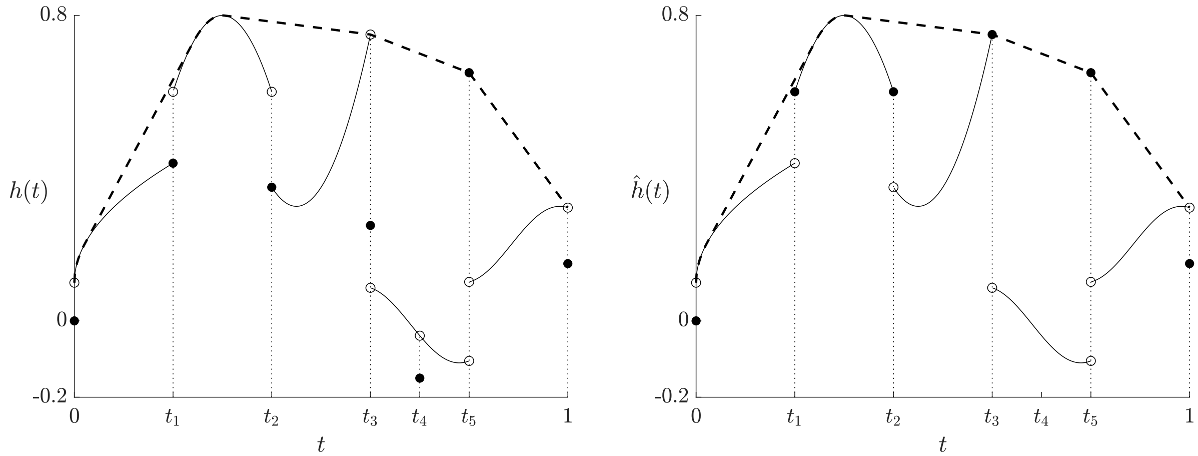


Figure 3.1: An example of $h$ (left) and $\hat{h}$ (right) with the set of discontinuity points $\{t_1, t_2, t_3, t_4, t_5\}$ excluding 0 and 1; the dashed lines represent $h^*$ and $(\hat{h})^*$, which are identical by Proposition 3.1

While in general $\rho_h$ and $\rho_{\hat{h}}$ are different functionals, one has $\rho_h(Y) = \rho_{\hat{h}}(Y)$ for any random variable $Y$ with continuous quantile function; see Lemma 1 of Chapter 2. Moreover, $h^* = (\hat{h})^* \geqslant \hat{h} \geqslant h$ and the four functions are all equal if $h$ is concave. Below, we provide a new result on convex envelopes of distortion functions $h$ that are not necessarily monotone or continuous, which may be of independent interest.

**Proposition 3.1.** *For any $h \in \mathcal{H}$, we have $h^* = (\hat{h})^*$ and the set $\{t \in [0,1] : \hat{h}(t) \neq h^*(t)\}$ is the union of some disjoint open intervals. Moreover, $h^*$ is linear on each of the above intervals.*

In the sequel, we mainly focus on $h^*$, which will be useful when optimizing $\rho_h$ in (3.3). A similar result to Proposition 3.1 holds for $h_*$, useful in the corresponding infimum problem, where the upper semicontinuous modification of $h$ is replaced by the lower semicontinuous one. This follows directly from Proposition 3.1 by setting $g = -h$ which gives $\rho_g = -\rho_h$ and $h_* = -g^*$.

For all distortion functions $h \in \mathcal{H}$, from Proposition 3.1, there exist (countably many) disjoint open intervals on which $\hat{h} \neq h^*$. Using a similar notation to Wang et al. (2019), we define the set

$$\mathcal{I}_h = \{(1-b, 1-a) : \hat{h} \neq h^* \text{ on } (a,b),\ \hat{h}(a) = h^*(a),\ \hat{h}(b) = h^*(b)\}.$$

The set $\mathcal{I}_h$ is straightforward to identify in practice; see Section 3.3.2 for examples of commonly used distortion riskmetrics and their corresponding sets $\mathcal{I}_h$.

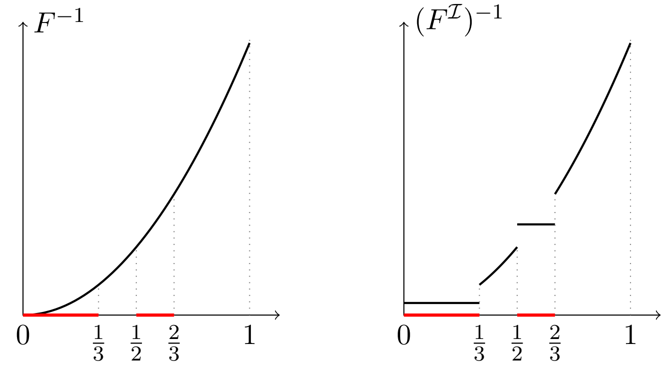


Figure 3.2: Left panel: quantile function of $F$; right panel: quantile function of $F^{\mathcal{I}}$ where $\mathcal{I} = \{(0,1/3),(1/2,2/3)\}$

## 3.3 Equivalence between non-convex and convex riskmetrics

### 3.3.1 Concentration and the main equivalence result

In this section, we introduce the concept of concentration, and use this concept to explain our main equivalence results, Theorems 3.1 and 3.2. For a distribution $F \in \mathcal{M}_1$ and an interval $C \subseteq [0,1]$ (when speaking of an interval in $[0,1]$, we exclude singletons or empty sets), we define the *$C$-concentration of $F$*, denote by $F^C$, as the distribution of the random variable

$$F^{-1}(U)\mathbb{1}_{\{U\notin C\}} + \mathbb{E}[F^{-1}(U)|U\in C]\mathbb{1}_{\{U\in C\}}, \tag{3.5}$$

where $U \sim \mathrm{U}[0,1]$ is a standard uniform random variable. In other words, $F^C$ is obtained by concentrating the probability mass of $F^{-1}(U)$ on $\{U \in C\}$ at its conditional expectation, whereas the rest of the distribution remains unchanged. For $F \in \mathcal{M}_1$ and $0 \leqslant a < b \leqslant 1$, it is clear that the left-quantile function of $F^{(a,b)}$ is given by

$$F^{-1}(t)\mathbb{1}_{\{t\notin(a,b]\}} + \frac{\int_a^b F^{-1}(u)\,\mathrm{d}u}{b-a}\mathbb{1}_{\{t\in(a,b]\}}, \quad t\in[0,1]. \tag{3.6}$$

For a collection $\mathcal{I}$ of (possibly infinitely many) non-overlapping intervals in $[0,1]$, let $F^{\mathcal{I}}$ be the distribution corresponding to the left-quantile function given by the left-continuous version of

$$F^{-1}(t)\mathbb{1}_{\{t\notin\bigcup_{C\in\mathcal{I}}C\}} + \sum_{C\in\mathcal{I}} \frac{\int_C F^{-1}(u)\,\mathrm{d}u}{\lambda(C)}\mathbb{1}_{\{t\in C\}}, \qquad t\in[0,1], \tag{3.7}$$

where $\lambda$ is the Lebesgue measure; see Figure 3.2 for an illustration.

**Definition 3.1.** Let $\mathcal{M}$ be a set of distributions in $\mathcal{M}_1$ and $\mathcal{I}$ be a collection of intervals in $[0,1]$. We say that (a) $\mathcal{M}$ is *closed under concentration within $\mathcal{I}$* if $F^{\mathcal{I}} \in \mathcal{M}$ for all $F \in \mathcal{M}$; (b) $\mathcal{M}$ is *closed under concentration for all intervals* if for all $F \in \mathcal{M}$, we have $F^C \in \mathcal{M}$ for all intervals $C \subseteq [0,1]$; (c) $\mathcal{M}$ is *closed under conditional expectation* if for all $F_X \in \mathcal{M}$, the distribution of any conditional expectation of $X$ is in $\mathcal{M}$.

The relationship between the three properties of closedness in Definition 3.1 is discussed in Propositions 3.2 and 3.3 below. Generally, (c)⇒(b)⇒(a) if $\mathcal{I}$ is finite. Our main equivalence result is summarized in the following theorem.

**Theorem 3.1.** *For $\mathcal{M} \subseteq \mathcal{M}_1$ and $h \in \mathcal{H}$, the following hold.*

*(i) If $h = \hat{h}$, i.e., $h$ is upper semicontinuous on $(0,1)$, and $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$, then*

$$\sup_{F_Y \in \mathcal{M}} \rho_h(Y) = \sup_{F_Y \in \mathcal{M}} \rho_{h^*}(Y). \tag{3.8}$$

*(ii) If $\mathcal{M}$ is closed under concentration for all intervals, then (3.8) holds.*

*(iii) If $h = \hat{h}$, $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$, and the second supremum in (3.8) is attained by some $F \in \mathcal{M}$, then $F^{\mathcal{I}_h}$ attains both suprema.*

Both suprema in (3.8) may be infinite, and this is discussed in Remark 3.5 in Section 3.9.2. The proof of Theorem 3.1 is more technical than similar results in the literature because of the challenges arising from non-monotonicity, non-positivity, and discontinuity of $h$; see Figure 3.1 for a sample of possible complications. In (ii), $h$ does not need to be upper semicontinuous on $(0,1)$ for (3.8) to hold because closedness under concentration for all intervals in (ii) is stronger than the condition in (i).

*Remark* 3.1. For $\mathcal{M} \subseteq \mathcal{M}_1$ and $h \in \mathcal{H}$, if $h = \hat{h}$ and $F^C \in \mathcal{M}$ for all $F \in \mathcal{M}$ and $C \in \mathcal{I}_h$, then the equivalence relation (3.8) also holds. If $\mathcal{I}_h$ is finite, then this condition is generally stronger than closedness under concentration within $\mathcal{I}_h$ in (i).

A natural question from Theorem 3.1 is whether our key condition of closedness under concentration is necessary in some sense for the equivalence (3.8) to hold.[3] It is immediate to notice that adding any distributions $F_Z$ satisfying $\rho_{h^*}(Z) < \sup_{F_Y \in \mathcal{M}} \rho_{h^*}(Y)$ to the set $\mathcal{M}$ does not affect the equivalence, and therefore we turn our attention to the set of maximizers instead of the whole set $\mathcal{M}$. In the next result, we show that closedness under concentration within $\mathcal{I}_h$ of the set of maximizers of (3.3) is necessary for the equivalence (3.8) to hold.

[3] We thank an anonymous referee for raising this question.

**Theorem 3.2.** *For $\mathcal{M} \subseteq \mathcal{M}_1$ and $h \in \mathcal{H}$ such that $h \neq h^*$, suppose that the set $\mathcal{M}_{\text{opt}}$ of all maximizers of* $\max_{F_Y \in \mathcal{M}} \rho_h(Y)$ *is non-empty. If the equivalence (3.8) holds, i.e.,* $\sup_{F_Y \in \mathcal{M}} \rho_h(Y) = \sup_{F_Y \in \mathcal{M}} \rho_{h^*}(Y)$*, then $\mathcal{M}_{\text{opt}}$ is closed under concentration within $\mathcal{I}_h$.*

If the equivalence (3.8) holds, then each $F \in \mathcal{M}_{\text{opt}}$ also maximizes the problem $\sup_{F_Y \in \mathcal{M}} \rho_{h^*}(Y)$. Conversely, if $h = \hat{h}$, then this condition and closedness of $\mathcal{M}_{\text{opt}}$ under concentration within $\mathcal{I}_h$ together are necessary (by Theorem 3.2) and sufficient (by Theorem 3.1) for the equivalence (3.8) to hold. If the maximizer $F$ of the original problem (3.3) is unique, then by Theorem 3.2, $F$ must be equal to $F^{\mathcal{I}_h}$. The equivalence (3.8) does not imply closedness under concentration within $\mathcal{I}_h$ of the uncertainty set $\mathcal{M}$ itself; an example showing this is discussed in Remark 3.2.

### 3.3.2 Some examples of distortion riskmetrics

We provide a few examples of distortion riskmetrics $\rho_h$ commonly used in decision theory and finance, and obtain their corresponding set $\mathcal{I}_h$. The Value-at-Risk (VaR) and the Expected Shortfall (ES) are the most popular risk measures in practice. We introduce them first, followed by an inverse-S-shaped distortion function of Tversky and Kahneman (1992).

**Example 3.1** (VaR and ES)**.** The Value-at-Risk (VaR) is defined in (1.1) on space $\mathcal{L}^0$. Similarly, we define the upper Value-at-Risk ($\mathrm{VaR}^+$) of $Y \in \mathcal{L}^0$ as the right-quantile of $F_Y$:

$$\mathrm{VaR}_\alpha^+(Y) = F_Y^{-1+}(\alpha), \quad \alpha \in [0,1).$$

We also define the Expected Shortfall (ES) as (1.2) on the space $\mathcal{L}^1$. Both $\mathrm{VaR}_\alpha$ and $\mathrm{ES}_\alpha$ belong to the class of distortion riskmetrics. Take $\alpha \in (0,1)$. Let $h(t) = \mathbb{1}_{(1-\alpha,1]}(t)$, $t \in [0,1]$. It follows that $h \in \mathcal{H}$ and $\hat{h}(t) = \mathbb{1}_{[1-\alpha,1]}(t)$, $t \in [0,1]$. In this case, $\rho_h = \mathrm{VaR}_\alpha$. Moreover, $h^*(t) = \frac{t}{1-\alpha} \wedge 1$, $t \in [0,1]$ and $\rho_{h^*} = \mathrm{ES}_\alpha$. Since $h^*$ and $\hat{h}$ differ on $(0, 1-\alpha)$, we have $\mathcal{I}_h = \{(\alpha, 1)\}$.

**Example 3.2** (TK distortion riskmetrics)**.** The following function $h$ is an inverse-S-shaped distortion function (see also Figure 3.4):

$$h(t) = \frac{t^\gamma}{(t^\gamma + (1-t)^\gamma)^{1/\gamma}}, \quad t \in [0,1], \ \gamma \in (0,1). \tag{3.9}$$

Distortion riskmetrics with distortion function (3.9) are commonly used in behavioural economics and finance; see e.g., Tversky and Kahneman (1992). For simplicity, we call such distortion riskmetrics *TK distortion riskmetrics*. Typical values of $\gamma$ are in $[0.5, 0.9]$; see Wu and Gonzalez (1996).

For $h$ in (3.9), it is clear that $h = \hat{h}$ on $[0,1]$ by continuity of $h$. We have $h^* \neq h$ on $(t_0, 1)$, for some $t_0 \in (0,1)$, and $h^*$ is linear on $[t_0, 1]$. Thus, $\mathcal{I}_h = \{(0, 1-t_0)\}$. An example of $h$ in (3.9) and its concave envelope $h^*$ are plotted in Figure 3.3 (left).

For $h_1, h_2 \in \mathcal{H}$, we write $h = h_1 - h_2 \in \mathcal{H}$ and consider the difference between two distortion riskmetrics, that is

$$\rho_h = \rho_{h_1} - \rho_{h_2}. \tag{3.10}$$

Such type of distortion riskmetrics measure the difference or disagreement between two utilities, risk attitudes, or capital requirements. Determining the upper and lower bounds, or the largest absolute values of such measures of disagreement, is of interest in practice but rarely studied in the literature. Note that $h_1 - h_2$ is in general not monotone or concave even when $h_1$ and $h_2$ themselves have the specified properties. Below we show some examples of distortion riskmetrics taking the form of (3.10).

**Example 3.3** (Inter-quantile range and inter-ES range)**.** For $\alpha \in [1/2, 1)$, we take $h_1(t) = \mathbb{1}_{[1-\alpha,1]}(t)$ and $h_2(t) = \mathbb{1}_{(\alpha,1]}(t)$, $t \in [0,1]$. It follows that $h(t) = h_1(t) - h_2(t) = \mathbb{1}_{\{1-\alpha \leqslant t \leqslant \alpha\}}$, $t \in [0,1]$, $\hat{h} = h$, and

$$\rho_h(X) = F_X^{-1+}(\alpha) - F_X^{-1}(1-\alpha), \quad X \in \mathcal{L}^0.$$

Correspondingly, we have $h^*(t) = t/(1-\alpha) \wedge 1 + (\alpha - t)/(1-\alpha) \wedge 0$, $t \in [0,1]$, and

$$\rho_{h^*}(X) = \mathrm{ES}_\alpha(X) + \mathrm{ES}_\alpha(-X), \quad X \in \mathcal{L}^1.$$

This distortion riskmetric $\rho_h$ is called an inter-quantile range and $\rho_{h^*}$ is called an inter-ES range. As the distortion functions $h^*$ and $\hat{h}$ differ on the open intervals $(0, 1-\alpha)$ and $(\alpha, 1)$, we have $\mathcal{I}_h = \{(\alpha, 1), (0, 1-\alpha)\}$. The distortion functions $h$ and $h^*$ are displayed in Figure 3.3 (right).

**Example 3.4** (Difference of two inverse-S-shaped distortion functions)**.** We take $h_1$ and $h_2$ to be the inverse-S-shaped distortion functions in (3.9), with parameters $\gamma_1 = 0.8$ and $\gamma_2 = 0.7$, respectively. By calculation, the function $h = h_1 - h_2$ is convex on $[0, 0.3770]$, concave on $[0.3770, 1]$, and as seen in Figure 3.4 not monotone. The concave envelope $h^*$ is linear on $[0, 0.7578]$ and $h^* = h$ on $[0.7578, 1]$. Thus, we have $\mathcal{I}_h = \{(0.2422, 1)\}$. The graphs of the distortion functions $h_1$, $h_2$, $h$, and $h^*$ are displayed in Figure 3.4.

The functions in $\mathcal{H}$ are a.e. differentiable, and for an absolutely continuous function $h \in \mathcal{H}$, let $h'$ be a (representative) function on $[0,1]$ that is a.e. equal to the derivative of $h$. If $h \in \mathcal{H}$ is

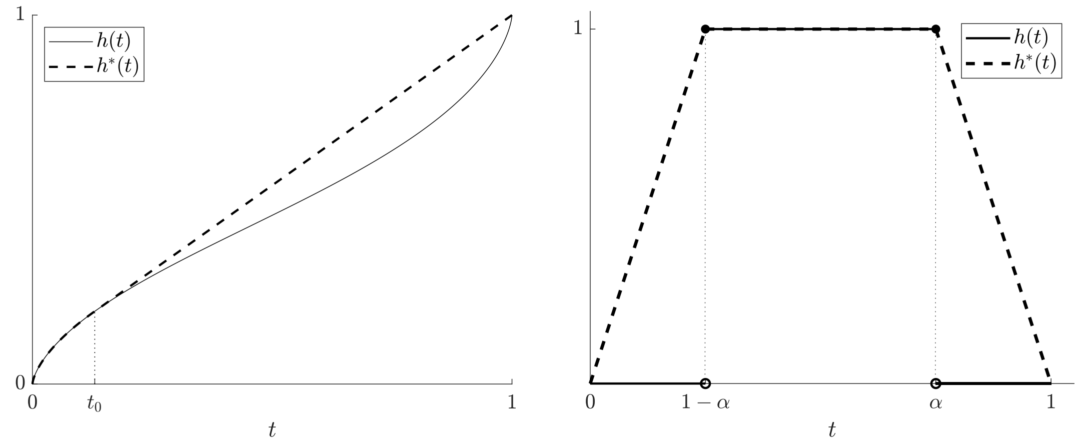


Figure 3.3: Left panel: $h$ and $h^*$ for the TK distortion riskmetric with $\gamma = 0.7$ in Example 3.2; right panel: $h$ and $h^*$ for the inter-quantile range in Example 3.3

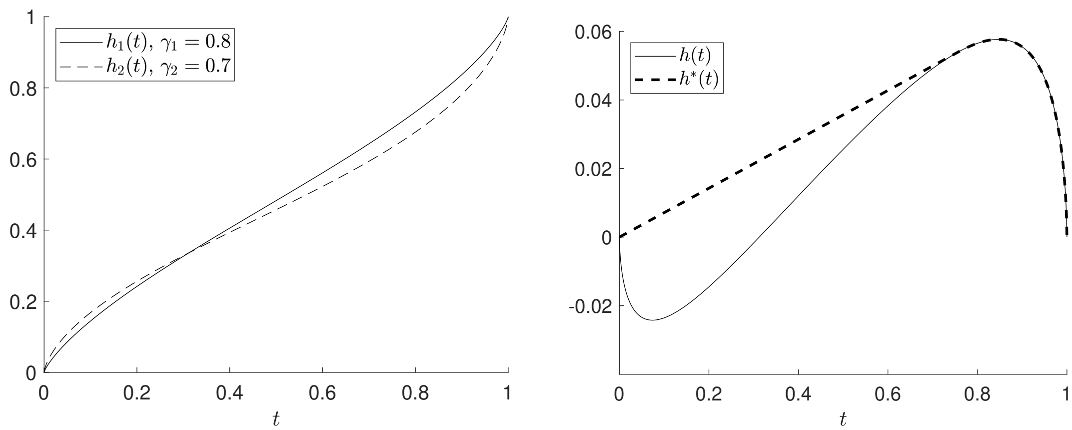


Figure 3.4: Left panel: inverse-S-shaped distortion functions $h_1$ and $h_2$ in Example 3.4; right panel: $h = h_1 - h_2$ and $h^*$ of the same example

left-continuous or $\mathrm{VaR}_t(Y)$ is continuous with respect to $t \in (0, 1)$, the risk measure $\rho_h$ in (1.3) has representation

$$\rho_h(Y) = \int_0^1 \mathrm{VaR}_{1-t}(Y)\, \mathrm{d}h(t), \quad Y \in \mathcal{L}^p; \tag{3.11}$$

see Lemma 1 of Chapter 2. If $h \in \mathcal{H}$ is absolutely continuous it holds

$$\rho_h(Y) = \int_0^1 \mathrm{VaR}_{1-t}(Y) h'(t)\, \mathrm{d}t, \quad Y \in \mathcal{L}^p. \tag{3.12}$$

Another example of a recently introduced distortion riskmetric with concave distortion function may be of independent interest in risk management.

**Example 3.5** (Second-order superquantile)**.** As introduced by Rockafellar and Royset (2018), a second-order superquantile is defined as

$$\mathrm{SSQ}_\alpha(Y) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{ES}_t(Y)\,\mathrm{d}t, \quad \alpha \in (0,1),\ Y \in \mathcal{L}^2.$$

By Theorem 2.4 of Rockafellar and Royset (2018), $\mathrm{SSQ}_\alpha$ is a distortion riskmetric with a concave distortion function $h$ given by

$$h(t) = \begin{cases} \frac{t}{1-\alpha}\left(1+\log\frac{1-\alpha}{t}\right), & 0 \leqslant t < 1-\alpha, \\ 1, & 1-\alpha \leqslant t \leqslant 1. \end{cases}$$

Clearly, $\mathrm{SSQ}_\alpha \geqslant \mathrm{ES}_\alpha$. The difference $\mathrm{SSQ}_\alpha - \mathrm{ES}_\alpha$ between second-order superquantile and ES, which has a similar interpretation as $\mathrm{ES}_\alpha - \mathrm{VaR}_\alpha$, is a distortion riskmetric with a non-concave and non-monotone distortion function $g$, and the set $\mathcal{I}_g$ contains a single interval of the form $(0,\beta)$ for some $\beta \in [\alpha, 1)$.

### 3.3.3 Closedness under concentration for all intervals

In this section, we present some technical results and specific examples about closedness under concentration for all intervals and under conditional expectation. The proposition below clarifies the relationship between closedness under concentration for all intervals and closedness under conditional expectation.

**Proposition 3.2.** *Closedness under conditional expectation implies closedness under concentration for all intervals, but the converse is not true.*

**Example 3.6.** We present 6 classes of sets $\mathcal{M}$ that are closed under conditional expectation, and hence also under concentration for all intervals.

1. (Moment conditions) For $p > 1$, $m \in \mathbb{R}$, and $v > 0$, the set

   $$\mathcal{M}(p,m,v) = \{F_Y \in \mathcal{M}_p : \mathbb{E}[Y] = m,\ \mathbb{E}[|Y-m|^p] \leqslant v^p\}$$

   is closed under conditional expectation by Jensen's inequality. The set $\mathcal{M}(p,m,v)$ corresponds to distributional uncertainty with moment information, and the setting $p = 2$ (mean and variance constraints) is the most commonly studied.

2. (Mean-covariance conditions) For $n \in \mathbb{N}$, $\mathbf{a} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\Sigma \in \mathbb{R}^{n\times n}$ positive semidefinite, let

   $$\mathcal{M}^{\mathrm{mv}}(\mathbf{a},\boldsymbol{\mu},\Sigma) = \{F_{\mathbf{a}^\top \mathbf{X}} \in \mathcal{M}_2 : F_{\mathbf{X}} \in \mathcal{M}_2^n,\ \mathbb{E}[\mathbf{X}] = \boldsymbol{\mu},\ \mathrm{var}(\mathbf{X}) \preceq \Sigma\},$$

where $\mathbf{X} = (X_1, \dots, X_n)$, $\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$, $\mathrm{var}(\mathbf{X})$ is the covariance matrix of $\mathbf{X}$, and $B' \preceq B$ means that the matrix $B - B'$ is positive semidefinite for two positive semidefinite symmetric matrices $B$ and $B'$. With a simple verification in Section 3.9.1, $\mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma) = \mathcal{M}(2, \mathbf{a}^\top \boldsymbol{\mu}, (\mathbf{a}^\top \Sigma \mathbf{a})^{1/2})$.

3. (Convex function conditions) For $n \in \mathbb{N}$, $\mathbf{a} \in \mathbb{R}^n$, $K \subseteq \mathbb{N}$, a collection $\mathbf{f} = (f_k)_{k \in K}$ of convex functions on $\mathbb{R}^n$, and a vector $\mathbf{x} = (x_k)_{k \in K} \in \mathbb{R}^{|K|}$, let

$$\mathcal{M}^{\mathbf{f}}(\mathbf{a}, \mathbf{x}) = \{F_{\mathbf{a}^\top \mathbf{X}} \in \mathcal{M}_1 : \mathbb{E}[f_k(\mathbf{X})] \leqslant x_k \text{ for all } k \in K\}.$$

The set $\mathcal{M}^{\mathbf{f}}$ corresponds to distributional uncertainty with constraints on expected losses or test functions. The set $\mathcal{M}^{\mathbf{f}}$ includes $\mathcal{M}(p, m, v)$ as a special case.

4. (Distortion conditions) For $K \subseteq \mathbb{N}$, a collection $\mathbf{h} = (h_k)_{k \in K} \in (\mathcal{H}^*)^{|K|}$ and a vector $\mathbf{x} = (x_k)_{k \in K} \in \mathbb{R}^{|K|}$, let

$$\mathcal{M}^{\mathbf{h}}(\mathbf{x}) = \{F_Y \in \mathcal{M}_1 : \rho_{h_k}(Y) \leqslant x_k \text{ for all } k \in K\}.$$

The set $\mathcal{M}^{\mathbf{h}}$ corresponds to distributional uncertainty with constraints on preferences modeled by convex dual utilities.

5. (Convex order conditions) For $K \subseteq \mathbb{N}$ and a collection of random variables $\mathbf{Z} = (Z_k)_{k \in K} \in (\mathcal{L}^1)^{|K|}$, let

$$\mathcal{M}^{\mathrm{cx}}(\mathbf{Z}) = \{F_Y \in \mathcal{M}_1 : Y \leqslant_{\mathrm{cx}} Z_k \text{ for all } k \in K\},$$

where $\leqslant_{\mathrm{cx}}$ is the inequality in convex order.[4] Similar to the above two examples, $\mathcal{M}^{\mathrm{cx}}(\mathbf{Z})$ is closed under conditional expectation (cf. Remark 3.6 in Section 3.9.2).

6. (Marginal conditions) For given univariate distributions $F_1, \dots, F_n \in \mathcal{M}_1$, let

$$\mathcal{M}^S(F_1, \dots, F_n) = \{F_{X_1 + \dots + X_n} \in \mathcal{M}_1 : X_i \sim F_i,\ i = 1, \dots, n\}.$$

In other words, $\mathcal{M}^S$ is the set of all possible aggregate risks $X_1 + \dots + X_n$ with given marginal distributions of $X_1, \dots, X_n$; see Embrechts et al. (2015) for some results on $\mathcal{M}^S$. Generally, $\mathcal{M}^S$ is not closed under concentration for all intervals or conditional expectation, since closedness under concentration for all intervals is stronger than joint mixability (Wang and Wang, 2016). In the special case where $F_1 = \dots = F_n = \mathrm{U}[0,1]$, Proposition 1 and Theorem 5 of Mao et al. (2019) imply that $\mathcal{M}^S$ is closed under conditional expectation if and only if $n \geqslant 3$.

[4] Precisely, we write $G \leqslant_{\mathrm{cx}} (\leqslant_{\mathrm{icx}})\, F$ if $\int \phi \,\mathrm{d}G \leqslant \int \phi \,\mathrm{d}F$ for all (increasing) convex functions $\phi$ such that the above two integrals are well defined.

*Remark* 3.2. The uncertainty set $\mathcal{M}(p,m,v)$ of the moment condition in Example 3.6 can be restricted to the set

$$\overline{\mathcal{M}}(p,m,v) = \{F_Y \in \mathcal{M}_p : \mathbb{E}[Y] = m,\ \mathbb{E}[|Y-m|^p] = v^p\},$$

which is the "boundary" of $\mathcal{M}(p,m,v)$. For $\mathcal{M} = \mathcal{M}(p,m,v)$, the suprema on both sides of (3.8) are obtained by some distributions in $\overline{\mathcal{M}}(p,m,v)$; see Theorem 3.5. As a direct consequence, we get

$$\sup_{F_Y \in \overline{\mathcal{M}}(p,m,v)} \rho_{h^*}(Y) = \sup_{F_Y \in \mathcal{M}(p,m,v)} \rho_{h^*}(Y) = \sup_{F_Y \in \mathcal{M}(p,m,v)} \rho_h(Y) = \sup_{F_Y \in \overline{\mathcal{M}}(p,m,v)} \rho_h(Y).$$

Hence, equivalence holds even though $\overline{\mathcal{M}}(p,m,v)$ is not closed under concentration for any interval. By Theorem 3.2, the set of optimizers is closed under concentration within $\mathcal{I}_h$ for each $h \in \mathcal{H}$.

For a distribution $F \in \mathcal{M}_1$ and a collection $\mathcal{I}$ of disjoint intervals in $[0,1]$, we have the following result regarding to the distribution $F^{\mathcal{I}}$.

**Proposition 3.3.** *Let $\mathcal{I}$ be a collection of disjoint intervals in $[0,1]$ and $\mathcal{M}$ be a set of distributions. If $\mathcal{M}$ is closed under concentration for all intervals and $\mathcal{I}$ is finite, or $\mathcal{M}$ is closed under conditional expectation, then $\mathcal{M}$ is closed under concentration within $\mathcal{I}$.*

If $\mathcal{I}$ is infinite, closedness under concentration for all intervals may not be sufficient for closedness under concentration within $\mathcal{I}$; see Remark 3.7 in Section 3.9.2 for a technical explanation. An infinite $\mathcal{I}_h$ does not appear for any distortion riskmetrics in practice.

### 3.3.4 Examples of closedness under concentration within $\mathcal{I}$ but not for all intervals

In practice, it is more tractable to check closedness under concentration within a specific collection of intervals $\mathcal{I}$ than closedness under concentration for all intervals or under conditional expectation. In this section, we show several examples for closedness under concentration within some $\mathcal{I}$.

For distortion functions $h$ such that $\mathcal{I}_h = \{(p,1)\}$ (resp. $\mathcal{I}_h = \{(0,p)\}$) for some $p \in (0,1)$, the result in Theorem 3.1 (i) only requires $\mathcal{M}$ to be closed under concentration within $\{(p,1)\}$ (resp. $\{(0,p)\}$). Such distortion functions include the inverse-S-shaped distortion functions in (3.9), those of $\mathrm{VaR}_p$, and $\mathrm{VaR}_p^+$, and that of the difference between the second-order superquantile and ES in Example 3.5. Below we present some more concrete examples.

**Example 3.7** ($\mathcal{M}$ has two elements)**.** Let $p \in (0,1)$ and $\mathcal{M} = \{\mathrm{U}[0,1],\, p\delta_{p/2} + (1-p)\mathrm{U}[p,1]\}$ where $\delta_{p/2}$ is the point-mass at $p/2$. We can check that $\mathcal{M}$ is closed under concentration within $\{(0,p)\}$ but $\mathcal{M}$ is not closed under concentration for all intervals. Indeed, any set closed under concentration for all intervals and containing $\mathrm{U}[0,1]$ has infinitely many elements. In general, a finite set which contains any non-degenerate distribution is not closed under conditional expectation in an atomless probability space, since there are infinitely many possible distributions for the conditional expectation of a given non-constant random variable. Another similar example that is closed under concentration within $\{(0,p)\}$ is the set of all possible distributions of the sum of several Pareto risks; see Example 5.1 of Wang et al. (2019).

**Example 3.8** (VaR and ES)**.** As we see from Example 3.1, if $\rho_h = \mathrm{VaR}_\alpha^+$ for some $\alpha \in (0,1)$, then $\rho_{h^*}$ is $\mathrm{ES}_\alpha$ and $\mathcal{I}_h = \{(\alpha,1)\}$. Theorem 3.1 (i) implies that if $\mathcal{M}$ is closed under concentration within $\{(\alpha,1)\}$, then

$$\sup_{F_Y \in \mathcal{M}} \mathrm{VaR}_\alpha^+(Y) = \sup_{F_Y \in \mathcal{M}} \mathrm{ES}_\alpha(Y).$$

This observation leads to (with some modifications) the main results in Wang et al. (2015) and Li et al. (2018) on the equivalence between VaR and ES.

**Example 3.9** (TK distortion riskmetric)**.** If we take $h$ to be an inverse-S-shaped distortion function in (3.9), then $\mathcal{I}_h = \{(0,1-t_0)\}$ for some $t_0 \in (0,1)$, and $\rho_h$ is the TK distortion riskmetric. As a direct consequence of Theorem 3.1 (i), if $\mathcal{M}$ is closed under concentration within $\{(0,1-t_0)\}$, then

$$\sup_{F_Y \in \mathcal{M}} \rho_h(Y) = \sup_{F_Y \in \mathcal{M}} \rho_{h^*}(Y).$$

This result implies Theorem 4.11 of Wang et al. (2019) on the robust risk aggregation problem based on dual utilities with inverse-S-shaped distortion functions.

**Example 3.10** (Wasserstein ball, 1-dimensional)**.** Optimization problems under the uncertainty set of a Wasserstein ball are common in literature when quantifying the discrepancy between a benchmark distribution and alternative scenarios; see e.g., Blanchet and Murthy (2019). We discuss the application of the concept of concentration to optimization with Wasserstein distances. For $p \geqslant 1$ and $F, G \in \mathcal{M}_p$, the $p$-*Wasserstein distance* between $F$ and $G$ is defined as

$$W_p(F,G) = \left(\int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right|^p \,\mathrm{d}u\right)^{1/p}.$$

For $\epsilon \geqslant 0$, the uncertainty set of an $\epsilon$-*Wasserstein ball* around a benchmark distribution $\widetilde{G} \in \mathcal{M}_p$ is given by

$$\mathcal{M}(\widetilde{G},\epsilon) = \{F \in \mathcal{M}_p : W_p(F,\widetilde{G}) \leqslant \epsilon\}.$$

Suppose that the benchmark distribution $\widetilde{G}$ has a quantile function that is constant on each element in some collection of disjoint intervals $\widetilde{\mathcal{I}} \subseteq [0,1]$. As shown in Section 3.9.1, $\mathcal{M}(\widetilde{G},\epsilon)$ is closed under concentration within $\mathcal{I}$ for all $\mathcal{I} \subseteq \widetilde{\mathcal{I}}$. Using this closedness property and Theorem 3.1 (i), the equivalence

$$\sup_{F_Y \in \mathcal{M}(\widetilde{G},\epsilon)} \rho_h(Y) = \sup_{F_Y \in \mathcal{M}(\widetilde{G},\epsilon)} \rho_{h^*}(Y) \tag{3.13}$$

holds for all $h \in \mathcal{H}$ such that $\mathcal{I}_h \subseteq \widetilde{\mathcal{I}}$.

*Remark* 3.3. In general, if the quantile function $\widetilde{G}$ in Example 3.10 is not constant on some interval in $\widetilde{\mathcal{I}}$, then $\mathcal{M}(\widetilde{G},\epsilon)$ is not necessarily closed under concentration within $\widetilde{\mathcal{I}}$, and the equivalence (3.13) may not hold. For instance, the worst-case $\mathrm{VaR}_\alpha$ over $\mathcal{M}(\widetilde{G},\epsilon)$ is generally different from the worst-case $\mathrm{ES}_\alpha$ over $\mathcal{M}(\widetilde{G},\epsilon)$ as obtained in Proposition 4 of Liu et al. (2022). We also refer to Bernard et al. (2020) who consider a Wasserstein ball together with moment constraints.

**Example 3.11** (Wasserstein ball, $n$-dimensional)**.** For $n \in \mathbb{N}$, $p \geqslant 1$, $a \geqslant 1$ and $F, G \in \mathcal{M}_p^n$, the $p$-Wasserstein distance on $\mathbb{R}^n$ between $F$ and $G$ is defined as

$$W_{a,p}^n(F,G) = \inf_{\mathbf{X}\sim F,\ \mathbf{Y}\sim G} (\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|_a^p])^{1/p},$$

where $\|\cdot\|_a$ is the $\mathcal{L}^a$-norm on $\mathbb{R}^n$. Similarly to the 1-dimensional case, for $\epsilon \geqslant 0$, an $\epsilon$-Wasserstein ball on $\mathbb{R}^n$ around a benchmark distribution $\widetilde{G} \in \mathcal{M}_p^n$ is defined as

$$\mathcal{M}^n(\widetilde{G},\epsilon) = \{F \in \mathcal{M}_p^n : W_{a,p}^n(F,\widetilde{G}) \leqslant \epsilon\}.$$

In a portfolio selection problem, we consider the worst-case riskmetric of a linear combination of random losses. For $\epsilon \geqslant 0$, $\mathbf{w} \in [0,\infty)^n$, $p > 1$, $a > 1$ and $\mathbf{Z} \in (\mathcal{L}^p)^n$, as shown in Section 3.9.1, the uncertainty set

$$\{F_{\mathbf{w}^\top \mathbf{X}} \in \mathcal{M}_p : F_{\mathbf{X}} \in \mathcal{M}^n(F_{\mathbf{Z}},\epsilon)\}$$

is closed under concentration within $\{(0,t)\}$ for all $t \leqslant p_0$. For a practical example, assume that an investor holds a portfolio of bonds (for simplicity, assume that they have the same maturity). The loss vector $\mathbf{X} \geqslant \mathbf{0}$ from this portfolio at maturity has an estimated benchmark loss distribution $\widetilde{G}$, and the probability of no default from these bonds (i.e., $\mathbf{X} = \mathbf{0}$) is estimated as $p_0 > 0$ (usually quite large). Suppose that the investor uses a distortion riskmetric with an inverse-S-shaped distortion function $h$ given in (3.9) of Example 3.2, and considers a Wasserstein ball around $\widetilde{G}$ with radius $\epsilon$. Note that $\mathcal{I}_h = \{(0,t)\}$ for some $t \in (0,1)$ from Example 3.9. By Theorem 3.1 (i), we obtain an equivalence result on the worst-case riskmetrics for the portfolio with weight vector $\mathbf{w}$,

$$\sup_{F_{\mathbf{X}} \in \mathcal{M}^n(\widetilde{G},\epsilon)} \rho_h(\mathbf{w}^\top \mathbf{X}) = \sup_{F_{\mathbf{X}} \in \mathcal{M}^n(\widetilde{G},\epsilon)} \rho_{h^*}(\mathbf{w}^\top \mathbf{X}),$$

whenever $t \in (0, p_0]$.

**Example 3.12** (Optimal hedging strategy)**.** Suppose that an investor is willing to hedge her random loss $X$ only when it exceeds some certain level $l \in \mathbb{R}$. Mathematically, for a fixed $X \in \mathcal{L}^1$ continuously distributed on $(F_X^{-1}(p_0), F_X^{-1}(1))$ such that $\mathbb{P}(X \leqslant l) = p_0$ for some $p_0 \in (0,1)$ and $l \in \mathbb{R}$, define the set of measurable functions

$$\mathcal{V} = \{V : \mathbb{R} \to \mathbb{R} \mid x \mapsto x - V(x) \text{ is increasing, } V(x) = 0 \text{ for all } x \leqslant l\}$$

representing possible hedging strategies. Let $g : \mathbb{R} \to \mathbb{R}$ be an increasing and convex function. The final payoff obtained by a hedging strategy $V \in \mathcal{V}$ is given by $X - V(X) + g(\mathbb{E}[V(X)])$, where $g(\mathbb{E}[V(X)])$ is a fixed cost of the hedging strategy that depends on the expected value of $V(X)$ calculated by a risk-neutral seller in the market using the same probability measure $\mathbb{P}$. As shown in Section 3.9.1, the action set in this optimization problem,

$$\mathcal{M} = \{F_{X-V(X)+g(\mathbb{E}[V(X)])} \in \mathcal{M}_1 : V \in \mathcal{V}\},$$

is closed under concentration within $\{(p,1)\}$ for all $p \in [p_0, 1)$. On the other hand, it is obvious that $\mathcal{M}$ is not closed under concentration for all intervals or closed under conditional expectation since the quantiles of the distributions in $\mathcal{M}$ are fixed beyond the interval $(p_0, 1)$. The above closedness under concentration property allows us to use Theorem 3.1 to convert the optimal hedging problem for $\rho_h$ with an inverse-S-shaped distortion function $h$ as in (3.9) to a convex version $\rho_{h^*}$.

**Example 3.13** (Risk choice)**.** Suppose that an investor is faced with a random loss $X \in \mathcal{L}^1$. The distortion function $h$ of her riskmetric is inverse-S-shaped with $\mathcal{I}_{-h} = \{(p,1)\}$ for some $p \in (0,1)$. Suppose that $p$ is known to the seller. Since the investor is averse to risk for large losses, the seller may provide her with the option to stick to the initial investment or to convert the upper part of the random loss into a fixed payment to avoid large loss. Specifically, we consider the set $\mathcal{M} = \{F_X, F_X^{(p,1)}\}$ containing two elements, where $\mathbb{P}(X \leqslant u) = p$ for some $u \in \mathbb{R}$. It is clear that $\mathcal{M}$ is closed under concentration within $\{(p,1)\}$ but not closed under conditional expectation. We assume that the costs of the two investment strategies are calculated by expectation and thus are the same. By (i) of Theorem 3.1, it follows that the risk minimization problem satisfies

$$\min_{F_Y \in \mathcal{M}} \rho_h(Y) = \min_{F_Y \in \mathcal{M}} \rho_{h_*}(Y) = \rho_{h_*}(X),$$

where the last equality follows from Theorem 3 of Chapter 2. By (iii) of Theorem 3.1, we further have the minimum of the original problem $\min_{F_Y\in\mathcal{M}}\rho_h(Y)$ is obtained by $F_X^{(p,1)}$; intuitively, the investor will choose to convert the upper part of her loss into a fixed payment.

### 3.3.5 Atomic probability space

The definition of closedness under concentration in Definition 3.1 requires the assumption of an atomless probability space since a uniform random variable is used in the setup. It may be of practical interest in some economic and optimization settings to assume a finite probability space. In this section, we let the sample space be $\Omega_n = \{\omega_1, \dots, \omega_n\}$ for $n \in \mathbb{N}$ and the probability measure $\mathbb{P}_n$ be such that $\mathbb{P}_n(\omega_i) = 1/n$ for all $i = 1, \dots, n$ (such a space is called *adequate* in economics). The possible distributions in such a probability space are supported by at most $n$ points each with probability a multiple of $1/n$, and we denote by $\mathcal{M}_{[n]}$ the set of these distributions.

Define the collection of intervals $\mathcal{I}_n = \{(j/n, k/n] : j, k \in \mathbb{N} \cup \{0\},\ j < k \leqslant n\}$. We say a set of distributions $\mathcal{M} \subseteq \mathcal{M}_{[n]}$ is *closed under grid concentration* within $\mathcal{I} \subseteq \mathcal{I}_n$ if for all $F \in \mathcal{M}$, the distribution of the random variable

$$F^{-1}(U_n)\mathbb{1}_{\{U_n \notin \bigcup_{C\in\mathcal{I}} C\}} + \sum_{C\in\mathcal{I}} \mathbb{E}[F^{-1}(U_n)|U_n \in C]\mathbb{1}_{\{U_n\in C\}}$$

is also in $\mathcal{M}$, where $U_n$ is a random variable such that $U_n(\omega_i) = i/n$ for all $i = 1, \dots, n$. For a distribution $F$ with finite mean and $(a, b] \in \mathcal{I}_n$, it is straightforward that the left-quantile function of $F^{(a,b]}$ is given by (3.6). The following equivalence result holds with additional assumption $\mathcal{I}_h \subseteq \mathcal{I}_n$. The proof can be obtained directly from that of Theorem 3.1.

**Proposition 3.4.** *Let $\mathcal{M} \subseteq \mathcal{M}_{[n]}$ and $h \in \mathcal{H}$. If $h = \hat{h}$, $\mathcal{I}_h \subseteq \mathcal{I}_n$ and $\mathcal{M}$ is closed under grid concentration within $\mathcal{I}_h$, then*

$$\sup_{F_Y\in\mathcal{M}} \rho_h(Y) = \sup_{F_Y\in\mathcal{M}} \rho_{h^*}(Y).$$

We note that the condition $\mathcal{I}_h \subseteq \mathcal{I}_n$ in Proposition 3.4 is satisfied by all distortion functions $h$ which are linear (or constant) on each of $((j-1)/n, j/n]$, $j = 1, \dots, n$. It is common to assume such a distortion function $h$ in an adequate probability space of $n$ states, since any distribution function can only take values in $\{j/n : j = 0, \dots, n\}$.

## 3.4 Multi-dimensional setting

Our main equivalence results in Theorems 3.1 and 3.2 are stated under the context of one-dimensional random variables. In this section, we discuss their generalization to multi-dimensional framework with a few additional steps.

In the multi-dimensional setting, closedness under concentration is not easy to define, as quantile functions are not naturally defined for multivariate distributions. Nevertheless, closedness under conditional expectation can be analogously formulated. For $n \in \mathbb{N}$, we say that $\mathcal{M} \subseteq \mathcal{M}^n$ is *closed under conditional expectation*, if for all $F_{\mathbf{X}} \in \mathcal{M}$, the distribution of any conditional expectation of $\mathbf{X}$ is in $\mathcal{M}$. The following theorem states the multi-dimensional version of our main equivalence result using closedness under conditional expectation.

**Theorem 3.3.** *For $\widetilde{\mathcal{M}} \subseteq \mathcal{M}_1^n$, increasing function $h \in \mathcal{H}$ and $f : A \times \mathbb{R}^n \to \mathbb{R}$ concave in the second argument, if $\widetilde{\mathcal{M}}$ is closed under conditional expectation, then for all $\mathbf{a} \in A$,*

$$\sup_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_h(f(\mathbf{a}, \mathbf{X})) = \sup_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_{h^*}(f(\mathbf{a}, \mathbf{X})). \tag{3.14}$$

*If $h = \hat{h}$ and the second supremum in (3.14) is attained by some $F_{\mathbf{X}} \in \widetilde{\mathcal{M}}$, then $F^{\mathcal{I}_h}_{f(\mathbf{a},\mathbf{X})}$ attains both suprema. Moreover, if $f$ is linear in the second component, then (3.14) holds for all $h \in \mathcal{H}$ (not necessarily monotone).*

*Remark* 3.4. If we assume that $f$ is convex (instead of concave) in the second argument in Theorem 3.3 and keep the other assumptions, then for an increasing $h$,

$$\inf_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_h(f(\mathbf{a}, \mathbf{X})) = \inf_{F_{\mathbf{X}} \in \widetilde{\mathcal{M}}} \rho_{h_*}(f(\mathbf{a}, \mathbf{X})).$$

This statement follows by noting $\rho_{-h} = -\rho_h$. The case of a decreasing $h$ is similar.

Theorem 3.3 is similar to Theorem 3.4 of Cai et al. (2020) which states the equivalence (3.14) for increasing $h$ and a specific set $\widetilde{\mathcal{M}}$ which is a special case in Example 3.14 below. In contrast, our result applies to non-monotone $h$ (with an extra condition on $f$), more general set $\widetilde{\mathcal{M}}$, and also the infimum problem. The setting of a function $f$ linear in the second argument often appears in portfolio selection problems where $f(\mathbf{a}, \mathbf{X}) = \mathbf{a}^\top \mathbf{X}$; see Example 3.11 and Section 3.6.

**Example 3.14.** Similarly to Example 3.6, we give examples of sets of multi-dimensional distributions closed under conditional expectation.

1. (Convex function conditions) For $n \in \mathbb{N}$, a convex set $B \subseteq \mathbb{R}^n$, set $\Psi$ of convex functions on $\mathbb{R}^n$, and a mapping $\pi : \Psi \to \mathbb{R}$, let

$$\widetilde{\mathcal{M}}(B, \Psi, f) = \{F_{\mathbf{X}} \in \mathcal{M}_1^n : \mathbb{P}(\mathbf{X} \in B) = 1,\ \mathbb{E}[\psi(\mathbf{X})] \leqslant \pi(\psi) \text{ for all } \psi \in \Psi\}.$$

It is clear that $\widetilde{\mathcal{M}}(B, \Psi, f)$ is closed under conditional expectation due to Jensen's inequality. The uncertainty set proposed by Delage et al. (2014) and used in Theorem 3.4 of Cai et al. (2020) can be obtained as a special case of this setting by taking $\Psi = \{f_1, \dots, f_n\} \cup \{g_1, \dots, g_n\} \cup \Phi$, where $f_i : (x_1, \dots, x_n) \mapsto x_i$, $g_i : (x_1, \dots, x_n) \mapsto -x_i$ for all $i = 1, \dots, n$, and $\Phi$ is a set of convex functions. The specification for $\pi$ is that $\pi(f_i) = m_i \in \mathbb{R}$, $\pi(g_i) = -m_i$, $\pi(\phi) = 0$ for all $i = 1, \dots, n$, $\phi \in \Phi$.

2. (Distortion conditions) For $n \in \mathbb{N}$, $K \subseteq \mathbb{N}$, $\mathbf{a} = (\mathbf{a}_k)_{k \in K} \in \mathbb{R}^{n \times |K|}$, $\mathbf{h} = (h_k)_{k \in K} \in (\mathcal{H}^*)^{|K|}$ and $\mathbf{x} = (x_k)_{k \in K} \in \mathbb{R}^{|K|}$, the set

$$\widetilde{\mathcal{M}}^{\mathbf{h}}(\mathbf{a}, \mathbf{x}) = \{F_{\mathbf{X}} \in \mathcal{M}_1^n : \rho_{h_k}(\mathbf{a}_k^\top \mathbf{X}) \leqslant x_k \text{ for all } k \in K\}$$

is closed under conditional expectation. In portfolio optimization problems, this setting incorporates distributional uncertainty with constraints on convex distortion risk measures of the total loss. In particular, optimization with the riskmetrics chosen as ES is common in the literature; see e.g., Rockafellar and Uryasev (2002), where ES is called CVaR.

3. (Convex order conditions) For $n \in \mathbb{N}$ and random vectors $\mathbf{Z}_k \in (\mathcal{L}^1)^n$, $k \in K \subseteq \mathbb{N}$, we naturally extend from part 5 of Example 3.6 and obtain that the set

$$\widetilde{\mathcal{M}}^{\mathrm{cx}}(\mathbf{Z}) = \{F_{\mathbf{X}} \in \mathcal{M}_1^n : \mathbf{X} \leqslant_{\mathrm{cx}} \mathbf{Z}_k \text{ for all } k \in K\}$$

is closed under conditional expectation.

Next, we discuss a multi-dimensional problem setting involving concentrations of marginal distributions. For $n \in \mathbb{N}$, we assume that marginal distributions of an $n$-dimensional distribution in $\mathcal{M}_1^n$ are uncertain and are in some sets $\mathcal{F}_1, \dots, \mathcal{F}_n \subseteq \mathcal{M}_1$. For $F_1, \dots, F_n \in \mathcal{M}_1$, define the set

$$\mathcal{D}(F_1, \dots, F_n) = \{\text{cdf of } (X_1, \dots, X_n) : X_i \sim F_i,\ i = 1, \dots, n\},$$

which is the set of all possible joint distributions with specified marginals; see Embrechts et al. (2015). For $\mathbf{a} \in A$, $h \in \mathcal{H}$ and $\mathcal{F}_1, \dots, \mathcal{F}_n \subseteq \mathcal{M}_1$, the worst-case distortion riskmetric can be represented as

$$\sup_{F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n} \sup_{F_{\mathbf{X}} \in \mathcal{D}(F_1, \dots, F_n)} \rho_h(f(\mathbf{a}, \mathbf{X})). \tag{3.15}$$

The outer problem of (3.15) is a robust risk aggregation problem (see Embrechts et al. (2013); Embrechts et al. (2015) and item 6 of Example 3.6), which is typically nontrivial in general when $h$ is not concave. With additional uncertainty of the marginal distributions, (3.15) can be converted to a problem with a convex objective given that $\mathcal{F}_1, \dots, \mathcal{F}_n$ are closed under concentration.

**Theorem 3.4.** *For $\mathcal{F}_1, \dots, \mathcal{F}_n \subseteq \mathcal{M}_1$, increasing $h \in \mathcal{H}$ with $h = \hat{h}$, and $f : A \times \mathbb{R}^n \to \mathbb{R}$ increasing, supermodular and positively homogeneous in the second argument, if $\mathcal{F}_1, \dots, \mathcal{F}_n$ are closed under concentration within $\mathcal{I}_h$, then the following hold.*[5]

*(i) For all $\mathbf{a} \in A$,*

$$\sup_{F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n} \sup_{F_{\mathbf{X}} \in \mathcal{D}(F_1, \dots, F_n)} \rho_h(f(\mathbf{a}, \mathbf{X})) = \sup_{F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n} \sup_{F_{\mathbf{X}} \in \mathcal{D}(F_1, \dots, F_n)} \rho_{h^*}(f(\mathbf{a}, \mathbf{X})). \quad (3.16)$$

*(ii) If the supremum of the right-hand side of (3.16) is attained by some $F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n$ and $F \in \mathcal{D}(F_1, \dots, F_n)$, then for all $\mathbf{a} \in A$, $F_1^{\mathcal{I}_h}, \dots, F_n^{\mathcal{I}_h}$ and a comonotonic random vector $(X_1^{\mathcal{I}_h}, \dots, X_n^{\mathcal{I}_h})$ with $X_i^{\mathcal{I}_h} \sim F_i^{\mathcal{I}_h}$, $i = 1, \dots, n$ attain the suprema on both sides of (3.16).*[6]

Some examples of functions on $\mathbb{R}^n$ that are supermodular and positively homogeneous are given below. These functions are concave due to Theorem 3 of Marinacci and Montrucchio (2008).

**Example 3.15** (Supermodular and positively homogeneous functions)**.** For $n \in \mathbb{N}$, the following functions $f : \mathbb{R}^n \to \mathbb{R}$ are supermodular and positively homogeneous. Write $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.

(i) (Linear function) $f : \mathbf{x} \mapsto \mathbf{a}^\top \mathbf{X}$ for $\mathbf{a} \in \mathbb{R}^n$. The function is increasing for $\mathbf{a} \in \mathbb{R}^n_+$.

(ii) (Geometric mean) $f : \mathbf{x} \mapsto -(\prod_{i=1}^n |x_i|)^{1/n}$ on $\mathbb{R}^n_-$ for odd $n$. The function is also increasing on $\mathbb{R}^n_-$.

(iii) (Negated $p$-norm) $f : \mathbf{x} \mapsto -\|\mathbf{x}\|_p$ for $p \geqslant 1$. The function is increasing on $\mathbb{R}^n_-$.

(iv) (Sum of functions) $f : \mathbf{x} \mapsto \sum_{i=1}^n f_i(x_i)$ for positively homogeneous functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$. The function is increasing if $f_1, \dots, f_n$ are increasing.

[5] For a function $f : \mathbb{R}^n \to \mathbb{R}$, we say $f$ is *supermodular* if $f(\mathbf{x}) + f(\mathbf{y}) \leqslant f(\mathbf{x} \wedge \mathbf{y}) + f(\mathbf{x} \vee \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$; $f$ is *positively homogeneous* if $f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$ for all $\lambda \geqslant 0$ and $\mathbf{x} \in \mathbb{R}^n$.

[6] A random vector $(X_1, \dots, X_n) \in (\mathcal{L}^1)^n$ is called *comonotonic* if there exists a random variable $Z \in \mathcal{X}$ and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ almost surely for all $i = 1, \dots, n$.

## 3.5 One-dimensional uncertainty set with moment constraints

A popular example of an uncertainty set closed under concentration for all intervals is that of distributions with specified moment constraints as in Example 3.6. We investigate this uncertainty set in detail and offer in this section some general results, which generalize several existing results in the literature; none of the results in the literature include non-monotone and non-convex distortion functions. Non-monotone distortion functions create difficulties because of possible complications at their discontinuity points.

For $p > 1$, $m \in \mathbb{R}$ and $v > 0$, we recall the set of interest in Example 3.6:

$$\mathcal{M}(p, m, v) = \{F_Y \in \mathcal{M}_p : \mathbb{E}[Y] = m,\ \mathbb{E}[|Y - m|^p] \leqslant v^p\}.$$

Let $q \in [1, \infty]$ be the Hölder conjugate of $p$, namely $q = (1 - 1/p)^{-1}$, or equivalently, $1/p + 1/q = 1$. For all $h \in \mathcal{H}^*$ or $h \in \mathcal{H}_*$, we denote by

$$\|h' - x\|_q = \left(\int_0^1 |h'(t) - x|^q \,\mathrm{d}t\right)^{1/q},\ q < \infty \quad \text{and} \quad \|h' - x\|_\infty = \max_{t\in[0,1]} |h'(t) - x|, \quad x \in \mathbb{R}. \tag{3.17}$$

We introduce the following quantities:

$$c_{h,q} = \underset{x\in\mathbb{R}}{\arg\min} \|h' - x\|_q \quad \text{and} \quad [h]_q = \min_{x\in\mathbb{R}} \|h' - x\|_q = \|h' - c_{h,q}\|_q.$$

We set $[h]_q = \infty$ if $h$ is not continuous. It is clear that $c_{h,q}$ is unique for $q > 1$. The quantity $[h]_q$ may be interpreted as a $q$-central norm of the function $h$ and $c_{h,q}$ as its $q$-center. Note that for $q = 2$ and $h$ continuous, $[h]_2 = \|h' - h(1)\|_2$ and $c_{h,2} = h(1)$. We also note that the optimization problem is trivial if $[h]_q = 0$, which corresponds to the case that $h' = h(1)\mathbb{1}_{[0,1]}$ and $\rho_h$ is a linear functional, thus a multiple of the expectation. In this case, the supremum and infimum are attained by all random variables whose distributions are in $\mathcal{M}(p, m, v)$, and they are equal to $mh(1)$. Furthermore, for $h \in \mathcal{H}^*$ or $h \in \mathcal{H}_*$, and $q > 1$, we define a function on $[0, 1]$ by

$$\phi_h^q(t) = \frac{|h'(1-t) - c_{h,q}|^q}{h'(1-t) - c_{h,q}} [h]_q^{1-q} \quad \text{if } h'(1-t) - c_{h,q} \neq 0, \quad \text{and } \phi_h^q(t) = 0 \text{ otherwise.}$$

In case $q = 2$, for $t \in [0, 1]$, $\phi_h^2(t) = (h'(1-t) - h(1))\|h' - h(1)\|_2^{-1}$ if $\|h' - h(1)\|_2 > 0$ and 0 otherwise. We summarize our findings in the following theorem.

**Theorem 3.5.** *For any $h \in \mathcal{H}$, $m \in \mathbb{R}$, $v > 0$ and $p > 1$, we have*

$$\sup_{F_Y\in\mathcal{M}(p,m,v)} \rho_h(Y) = mh(1) + v[h^*]_q \quad \textit{and} \quad \inf_{F_Y\in\mathcal{M}(p,m,v)} \rho_h(Y) = mh(1) - v[h_*]_q. \tag{3.18}$$

*Moreover, if $h = \hat{h}$, $0 < [h_*]_q < \infty$ and $0 < [h^*]_q < \infty$, then the supremum and infimum in (3.18) are attained by a random variable $X$ such that $F_X \in \mathcal{M}(p, m, v)$ with its quantile function uniquely specified as a.e. equal to $m + v\phi^q_{h^*}$ and $m - v\phi^q_{h_*}$, respectively.*

The proof of Theorem 3.5 follows from a combination of Lemmas 3.1 and 3.2 in Section 3.10.4 and Theorem 3.1. Note that for $h \in \mathcal{H}^*$ (resp. $h \in \mathcal{H}_*$) and $q > 1$, $\phi^q_h$ is increasing (resp. decreasing) on $[0, 1]$. Hence, $\phi^q_h$ (resp. $-\phi^q_h$) in Theorem 3.5 indeed determines a quantile function.

The following proposition concerns the finiteness of $\rho_h$ on $\mathcal{L}^p$.

**Proposition 3.5.** *For any $h \in \mathcal{H}$ and $p \in [1, \infty]$, $\rho_h$ is finite on $\mathcal{L}^p$ if $[h^*]_q < \infty$ and $[h_*]_q < \infty$.*

As a special case of Proposition 3.5, $\rho_h$ is always finite on $\mathcal{L}^1$ if $h$ is convex or concave with bounded $h'$ because $[h^*]_\infty < \infty$ and $[h_*]_\infty < \infty$.

As a common example of the general result in Theorem 3.5, below we collect our findings for the case of VaR.

**Corollary 3.1.** *For $\alpha \in (0, 1)$, $p > 1$, $m \in \mathbb{R}$ and $v > 0$, we have*

$$\sup_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{VaR}_\alpha(Y) = \max_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{ES}_\alpha(Y) = m + v\alpha \left(\alpha^p(1-\alpha) + (1-\alpha)^p\alpha\right)^{-1/p},$$

*and*

$$\inf_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{VaR}_\alpha(Y) = \min_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{ES}^L_\alpha(Y) = m - v(1-\alpha) \left(\alpha^p(1-\alpha) + (1-\alpha)^p\alpha\right)^{-1/p},$$

*where*

$$\mathrm{ES}^L_\alpha(Y) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_t(Y)\,\mathrm{d}t, \quad Y \in \mathcal{L}^1.$$

We see from Theorem 3.5 that if $h = \hat{h}$, then the supremum and the infimum of $\rho_h(Y)$ over $F_Y \in \mathcal{M}(p, m, v)$ are always attainable. However, in case $h \neq \hat{h}$, the supremum or infimum may no longer be attainable as a maximum or minimum. We illustrate this in Example 3.16 below.

**Example 3.16** (VaR and ES, $p = 2$)**.** Take $\alpha \in (0, 1)$, $p = 2$ and $\rho_h = \mathrm{VaR}_\alpha$, which implies $\rho_{h^*} = \mathrm{ES}_\alpha$. Corollary 3.1 gives $\sup_{F_Y \in \mathcal{M}(2,m,v)} \mathrm{VaR}_\alpha(Y) = \sup_{F_Y \in \mathcal{M}(2,m,v)} \mathrm{ES}_\alpha(Y) = m + v\sqrt{\alpha/(1-\alpha)}$. This is the well-known Cantalli-type formula for ES. By Lemma 3.1, the unique left-quantile function of the random variable $Z$ that attains the supremum of $\mathrm{ES}_\alpha$ is given by $F^{-1}_Z(t) = m + v(\mathbb{1}_{(\alpha,1]}(t)/(1-\alpha) - 1)\sqrt{(1-\alpha)/\alpha}$, $t \in [0, 1]$ a.e. We thus have $\mathrm{VaR}_\alpha(Z) = m - v\sqrt{(1-\alpha)/(\alpha)}$, and hence $Z$ does not attain $\sup_{F_Y \in \mathcal{M}(2,m,v)} \mathrm{VaR}_\alpha(Y)$. It follows by the uniqueness of $F_Z$ that the supremum of

$\mathrm{VaR}_\alpha(Y)$ over $F_Y \in \mathcal{M}(2,m,v)$ cannot be attained. However, the supremum of $\mathrm{VaR}_\alpha^+$ is attained by $Z$ since $\mathrm{VaR}_\alpha^+(Z) = m + v\sqrt{(1-\alpha)/(\alpha)}$.

**Example 3.17** (Difference of two TK distortion riskmetrics)**.** Take $p = 2$ and $h = h_1 - h_2$ to be the difference between two inverse-S-shaped functions in (3.9) with parameters the same as those in Example 3.4 ($\gamma_1 = 0.8$, $\gamma_2 = 0.7$). By Theorem 3.5, the worst-case distortion riskmetrics under the uncertainty set $\mathcal{M}(2,m,v)$ are given by $\sup_{F_Y \in \mathcal{M}(2,m,v)} \rho_h(Y) = \sup_{F_Y \in \mathcal{M}(2,m,v)} \rho_{h^*}(Y) = 0.3345v$, and the unique left-quantile function of the random variable $Z$ attaining both suprema above is given by $F_Z^{-1}(t) = m + 2.9892 \cdot h^{*\prime}(1-t)v$, $t \in [0,1]$ a.e. The worst-case distortion riskmetrics obtained above are independent of the mean $m$ as $h(1) = h_1(1) - h_2(1) = 0$, which is sensible since $\rho_h$ and $\rho_{h^*}$ only incorporate the disagreement between two distortion riskmetrics. Similarly, we can calculate the infimum of $\rho_h(Y)$ over $Y \in \mathcal{M}(2,m,v)$, and thus obtain the largest absolute difference between the two preferences numerically represented by $\rho_{h_1}$ and $\rho_{h_2}$.

## 3.6 Related optimization problems

In this section, we discuss the applications of our main results to some related optimization problems commonly investigated in the literature by including the outer problem of (3.1).

### 3.6.1 Portfolio optimization

Our equivalence results can be applied to robust portfolio optimization problems. For an uncertainty set $\widetilde{\mathcal{M}} \subseteq \mathcal{M}_p^n$ with $p \in [1,\infty]$, let the random vector $\mathbf{X} = (X_1,\dots,X_n) \sim F_{\mathbf{X}} \in \widetilde{\mathcal{M}}$, representing the random losses from $n$ risky assets. For $A \subseteq \mathbb{R}^n$, denote by a vector $\mathbf{a} \in A$ the amounts invested in each of the $n$ risky assets. For a distortion function $h \in \mathcal{H}$ and distortion riskmetric $\rho_h : \mathcal{L}^p \to \mathbb{R}$, we aim to solve the robust portfolio optimization problem given by

$$\min_{\mathbf{a}\in A}\left(\sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \rho_h(\mathbf{a}^\top \mathbf{X}) + \beta(\mathbf{a})\right), \tag{3.19}$$

where $\beta : \mathbb{R}^n \to \mathbb{R}$ is a penalty function of risk concentration. Note that $\beta$ is irrelevant for the inner problem of (3.19). For a general non-concave $h$, there is no known algorithm to solve the inner problem of (3.19). The outer optimization problem is also nontrivial in general. Therefore, we usually cannot obtain closed-form solutions of (3.19) using classical results of optimization problems for non-convex risk measures. However, as a direct consequence of Theorems 3.1 and 3.3, the following proposition converts (3.19) to an equivalent optimization problem with the objective functional $\rho_{h^*}$ being convex and mixture concave, which is usually technically tractable to solve. The proof of Proposition 3.6 follows directly from Theorems 3.1 and 3.3.

**Proposition 3.6.** *Let $h \in \mathcal{H}$, $n \in \mathbb{N}$, $A \subseteq \mathbb{R}^n$, and $\widetilde{\mathcal{M}} \subseteq \mathcal{M}_1^n$.*

*(i) if $h = \hat{h}$ and the set $\{F_{\mathbf{a}^\top \mathbf{X}} \in \mathcal{M}_1 : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\}$ is closed under concentration within $\mathcal{I}_h$ for all $\mathbf{a} \in A$, then*

$$\min_{\mathbf{a}\in A}\left(\sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \rho_h(\mathbf{a}^\top\mathbf{X}) + \beta(\mathbf{a})\right) = \min_{\mathbf{a}\in A}\left(\sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \rho_{h^*}(\mathbf{a}^\top\mathbf{X}) + \beta(\mathbf{a})\right). \tag{3.20}$$

*(ii) if the set $\{F_{\mathbf{a}^\top \mathbf{X}} \in \mathcal{M}_1 : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\}$ is closed under concentration for all intervals for all $\mathbf{a} \in A$, then (3.20) holds.*

*(iii) If $\widetilde{\mathcal{M}}$ is closed under conditional expectation, then (3.20) holds.*

### 3.6.2 Preference robust optimization

We are also able to solve the preference robust optimization problem with distributional uncertainty. For $n \in \mathbb{N}$, an $n$-dimensional action set $A$, a set of plausible distributions $\widetilde{\mathcal{M}} \subseteq \mathcal{M}_1^n$, and a set of possible probability perceptions $\mathcal{G} \subseteq \mathcal{H}$, the problem is formulated as follows:

$$\min_{\mathbf{a}\in A} \sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \sup_{h\in\mathcal{G}} \rho_h(f(\mathbf{a},\mathbf{X})). \tag{3.21}$$

Preference robust optimization refers to the situation when the objective is not completely known, e.g., $h$ is in the set $\mathcal{G}$ but not identified. Therefore, optimization is performed under the worst-case preference in the set $\mathcal{G}$. Also note that the form $\sup_{h\in\mathcal{G}} \rho_h$ includes (but is not limited to) all coherent risk measures via the representation of Kusuoka (2001). See Delage and Li (2018) for the problem of (3.21) without distributional uncertainty (thus, only the minimum and the second supremum), which was further studied by Wang and Xu (2020) for optimization problems of robust spectral risk measures. We have the following result whose proof follows from Theorems 3.1 and 3.3.

**Proposition 3.7.** *Let $\widetilde{\mathcal{M}} \subseteq \mathcal{M}_1^n$ and $A \subseteq \mathbb{R}^n$ with $n \in \mathbb{N}$.*

*(i) If $h = \hat{h}$ and the set $\{F_{f(\mathbf{a},\mathbf{X})} \in \mathcal{M}_1 : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\}$ is closed under concentration within $\mathcal{I}_h$ for all $\mathbf{a} \in A$, then for all $\mathcal{G} \subseteq \mathcal{H}$,*

$$\min_{\mathbf{a}\in A} \sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \sup_{h\in\mathcal{G}} \rho_h(f(\mathbf{a},\mathbf{X})) = \min_{\mathbf{a}\in A} \sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \sup_{h\in\mathcal{G}} \rho_{h^*}(f(\mathbf{a},\mathbf{X})). \tag{3.22}$$

*(ii) If the set $\{F_{f(\mathbf{a},\mathbf{X})} \in \mathcal{M}_1 : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\}$ is closed under concentration for all intervals for all $\mathbf{a} \in A$, then (3.22) holds for all $\mathcal{G} \subseteq \mathcal{H}$.*

*(iii) If $\mathcal{G}$ is a set of increasing functions in $\mathcal{H}$, $f : A \times \mathbb{R}^n \to \mathbb{R}$ is concave in the second component, and $\widetilde{\mathcal{M}}$ is closed under conditional expectation, then (3.22) holds.*

The preference robust optimization problem without distributional uncertainty (i.e., problem (3.21) with only the minimum and the second supremum) is generally difficult to solve when the distortion function $h$ is not concave. However, when the distribution of the random variable is not completely known, we can transfer the original non-convex problem to its convex counterpart using (3.22), provided that the set of plausible distributions is well structured.

## 3.7 Applications and numerical illustrations

Following the discussion in Section 3.6, we provide several applications of our theoretical results to portfolio management for specific sets of plausible distributions. None of the considered optimization problems in this section are convex, and we provide numerical calculations or approximation for the solutions to these optimization problems.[7]

### 3.7.1 Difference of risk measures under moment constraints

We demonstrate a price competition problem as an application of optimizing the difference between two risk measures shown in Example 3.17. Similar to the portfolio management problem discussed in Section 3.6.1, we consider $n$ risky assets with random losses $X_1, \dots, X_n \in \mathcal{L}^2$ that are only known to have a fixed mean and a constrained covariance. That is, we choose the set

$$\widetilde{\mathcal{M}} = \{F_{\mathbf{X}} \in \mathcal{M}_2^n : \mathbb{E}[\mathbf{X}] = \boldsymbol{\mu},\ \mathrm{var}(\mathbf{X}) \preceq \Sigma\},$$

for $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ positive semidefinite. For an $n$-dimensional $\mathbf{a} \in A$, the set of all possible distributions of aggregate portfolio losses

$$\{F_{\mathbf{a}^\top\mathbf{X}} \in \mathcal{M}_2 : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\} = \mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma) = \mathcal{M}\left(2, \mathbf{a}^\top\boldsymbol{\mu}, \left(\mathbf{a}^\top\Sigma\mathbf{a}\right)^{1/2}\right) \tag{3.23}$$

[7]The processors we use are Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60GHz 2.59GHz (2 processors). The numerical results are calculated by MATLAB.

is closed under concentration for all intervals as is shown in Example 3.6. Let $\rho_{h_1} : \mathcal{L}^2 \to \mathbb{R}$ be an investor's own price of the portfolio, while $\rho_{h_2} : \mathcal{L}^2 \to \mathbb{R}$ is her opponent's price of the same portfolio. We choose $h_1$ and $h_2$ to be the inverse-S-shaped distortion functions in (3.9), with parameters the same as those in Example 3.17 ($\gamma_1 = 0.8$ and $\gamma_2 = 0.7$). Write $h = h_1 - h_2$. For an action set $A = \{(a_1, \dots, a_n) \in [0,1]^n : \sum_{i=1}^n a_i = 1\}$, the investor chooses the optimal $\mathbf{a}^* \in A$, such that the worst-case overpricing from her opponent is minimized.

From the calculation in Example 3.17, we get

$$\begin{aligned} D(\Sigma) &:= \min_{\mathbf{a}\in A} \sup_{F_{\mathbf{X}}\in\widetilde{\mathcal{M}}} \left(\rho_{h_1}(\mathbf{a}^\top \mathbf{X}) - \rho_{h_2}(\mathbf{a}^\top \mathbf{X})\right) \\ &= \min_{\mathbf{a}\in A} \sup_{F_Y\in\mathcal{M}^{\mathrm{mv}}(\mathbf{a},\boldsymbol{\mu},\Sigma)} \rho_{h^*}(Y) = 0.3345 \times \min_{\mathbf{a}\in A} \left(\mathbf{a}^\top \Sigma \mathbf{a}\right)^{1/2}. \end{aligned} \tag{3.24}$$

We note that optimizing $\rho_{h_1} - \rho_{h_2}$ is generally nontrivial since the difference between two distortion functions $h_1 - h_2$ is not necessarily monotone, concave, or continuous, even though $h_1$ and $h_2$ themselves may have these properties. The generality of our equivalence result allows us to convert the original problem to the much simpler form (3.24), which can be solved efficiently.[8] Table 3.1 demonstrates the optimal values of $\mathbf{a}^*$ and $D$ for different choices of $\Sigma$.

### 3.7.2 Preference robust portfolio optimization with moment constraints

Next, we discuss an example of preference robust optimization with distributional uncertainty using the results in Sections 3.5. Similarly to Section 3.7.1, we consider the set of plausible aggregate portfolio loss distributions

$$\mathcal{M}^{\mathrm{mv}}(\mathbf{a},\boldsymbol{\mu},\Sigma) = \{F_{\mathbf{a}^\top\mathbf{X}} \in \mathcal{M}_2 : F_{\mathbf{X}} \in \mathcal{M}_2^n,\ \mathbb{E}[\mathbf{X}] = \boldsymbol{\mu},\ \mathrm{var}(\mathbf{X}) \preceq \Sigma\}$$

and the action set $A = \{(a_1, \dots, a_n) \in [0,1]^n : \sum_{i=1}^n a_i = 1\}$ representing the weights the investor assigns to each random loss. The investor considers TK distortion riskmetrics, however, she is not certain about the parameter $\gamma$ of the distortion function $h$. Thus, the investor consider the set of TK distortion riskmetrics with distortion functions in

$$\mathcal{G} = \{h \in \mathcal{H} : h = h^\gamma,\ \gamma \in [0.5, 0.9]\},$$

[8] The convex problem (3.24) is solved numerically by the constrained nonlinear multivariable function "fmincon" with the interior-point method.

Table 3.1: Optimal results in (3.24) for difference between two TK distortion riskmetrics

| $n$ | $\Sigma$ | $\mathbf{a}^*$ | $D$ |
|---|---|---|---|
| 3 | $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ | (0.333, 0.333, 0.333) | 0.193 |
| 3 | $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$ | (0.300, 0.400, 0.300) | 0.150 |
| 3 | $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ | (0.997, 0.002, 0.001) | 0.335 |
| 5 | $\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$ | (0.438, 0.219, 0.146, 0.110, 0.088) | 0.221 |

which is approximately the two-sigma confidence interval of $\gamma$ in Wu and Gonzalez (1996).[9] Therefore, the investor aims to find a optimal portfolio given the uncertainty in the riskmetrics. To penalize deviations from the benchmark parameter $\gamma = 0.71$ (Wu and Gonzalez, 1996), the investor use the term $\mathrm{e}^{c(\gamma-0.71)^2}$ for some $c \geqslant 0$. Since the set $\mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma)$ is closed under concentration for all intervals for all $\mathbf{a} \in A$, Proposition 3.7, (3.23), and Theorem 3.5 lead to

$$\begin{aligned} V(\boldsymbol{\mu}, \Sigma) &:= \min_{\mathbf{a}\in A} \sup_{F_Y \in \mathcal{M}^{\mathrm{mv}}(\mathbf{a},\boldsymbol{\mu},\Sigma)} \sup_{\gamma\in[0.5,0.9]} \left(\rho_{h^\gamma}(Y) - \mathrm{e}^{c(\gamma-0.71)^2}\right) \\ &= \min_{\mathbf{a}\in A} \sup_{F_Y \in \mathcal{M}\left(2,\mathbf{a}^\top\boldsymbol{\mu},\left(\mathbf{a}^\top\Sigma\mathbf{a}\right)^{1/2}\right)} \sup_{\gamma\in[0.5,0.9]} \left(\rho_{(h^\gamma)^*}(Y) - \mathrm{e}^{c(\gamma-0.71)^2}\right) \\ &= \min_{\mathbf{a}\in A} \sup_{\gamma\in[0.5,0.9]} \left(\mathbf{a}^\top\boldsymbol{\mu} + \left(\mathbf{a}^\top\Sigma\mathbf{a}\right)^{1/2} [(h^\gamma)^*]_2 - \mathrm{e}^{c(\gamma-0.71)^2}\right). \end{aligned} \tag{3.25}$$

We calculate the optimal values $V$ for different choices of parameters ($n$, $c$, $\boldsymbol{\mu}$ and $\Sigma$) and report them in Table 3.2, where $\mathbf{a}^*$ and $\hat{\gamma}$ represent the optimal weights and the parameters of the inverse-S-shaped distortion function, respectively. Note that the last optimization problem in

[9]The aggregate least-square estimate of $\gamma$ in Section 5 of Wu and Gonzalez (1996) is 0.71 with standard deviation 0.1.

(3.25) can be calculated numerically.[10]

Table 3.2: Optimal values in (3.25) for TK distortion riskmetrics

| $n$ | $c$ | $\boldsymbol{\mu}$ | $\Sigma$ | $\mathbf{a}^*$ | $\hat{\gamma}$ | $V$ |
|---|---|---|---|---|---|---|
| 3 | 0 | $(1,1,1)$ | $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ | $(0.333,\ 0.333,\ 0.333)$ | 0.610 | 1.41 |
| 3 | 30 | $(2,1,1)$ | $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ | $(0.000,\ 0.500,\ 0.500)$ | 0.676 | 1.29 |
| 3 | 30 | $(1,1,1)$ | $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$ | $(0.300,\ 0.400,\ 0.300)$ | 0.690 | 1.17 |
| 3 | 30 | $(1.2,1,1)$ | $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ | $(0.500,\ 0.331,\ 0.168)$ | 0.630 | 1.57 |
| 5 | 30 | $(1,1,1,1,1)$ | $\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$ | $(0.438,\ 0.219,\ 0.146,\ 0.110,\ 0.088)$ | 0.678 | 1.26 |

### 3.7.3 Portfolio optimization with marginal constraints

A special case of the portfolio optimization problem introduced in Section 3.6.1, which is of interest in robust risk aggregation (see e.g., Blanchet et al., 2020), is to take $\widetilde{\mathcal{M}}$ to be the Fréchet class defined as

$$\widetilde{\mathcal{M}}(F_1, \dots, F_n) = \{F_{\mathbf{X}} \in \mathcal{M}_1^n : X_i \sim F_i,\ i = 1, \dots, n\}, \tag{3.26}$$

for some known marginal distributions $F_1, \dots, F_n \in \mathcal{M}_1$. In this case, although the left-hand side of (3.20) is generally difficult to solve, for $A \subseteq \mathbb{R}_+^n$, the right-hand side of (3.20) can be rewritten

[10] The problem (3.25) is solved numerically by the constrained nonlinear multivariable function "fmincon" with the interior-point method.

using convexity and comonotonicity as

$$\min_{\mathbf{a}\in A}\left(\mathbf{a}^\top(\rho_{h^*}(X_1),\dots,\rho_{h^*}(X_n))+\beta(\mathbf{a})\right), \tag{3.27}$$

where $X_i \sim F_i$, $i=1,\dots,n$. We see that (3.27) is a linear optimization problem with a penalty $\beta$, which often admits closed-form solutions when $\beta$ is properly chosen. For any given $\mathbf{a}\in A$, we define

$$\mathcal{M}(\mathbf{a},F_1,\dots,F_n)=\{F_{\mathbf{a}^\top\mathbf{X}}\in\mathcal{M}_1 : X_i\sim F_i,\ i=1,\dots,n\}. \tag{3.28}$$

The set $\mathcal{M}(\mathbf{a},F_1,\dots,F_n)$ is the weighted version of $\mathcal{M}^S(F_1,\dots,F_n)$ in Example 3.6. Note that $\mathcal{M}(\mathbf{a},F_1,\dots,F_n)$ is generally neither closed under concentration for all intervals nor closed under conditional expectation. However, $\mathcal{M}(\mathbf{a},F_1,\dots,F_n)$ is asymptotically (for large $n$) similar to a set of distributions closed under concentration for all intervals; see Theorem 3.5 of Mao and Wang (2015) for a precise statement in the case of equal weights and identical marginal distributions. Therefore, even though $\mathcal{M}(\mathbf{a},F_1,\dots,F_n)$ is not closed under concentration for all intervals for some $\mathbf{a}\in A$, our result of the problem (3.27) is a good approximation of the original problem for large $n$. Such asymptotic equivalence between worst-case riskmetrics of aggregate risks with equal weights has already been well studied in the literature; see e.g., Theorem 3.3 of Embrechts et al. (2015) for the VaR/ES pair and Theorem 3.5 of Cai et al. (2018) for distortion risk measures.

We conduct numerical calculations to illustrate the equivalence between both sides in (3.20). We choose the action set $A_{a,b}=\{(x_1,\dots,x_n)\in[a,b]^n:\sum_{i=1}^n x_i=1\}$, for $0\leqslant a<1/n<b\leqslant 1$ and the penalty function $\beta$ to be the $\mathcal{L}^2$-norm multiplied by a scaler $c\geqslant 0$, namely $c\|\cdot\|_2$, where the scaler $c$ is a tuning parameter of the $\mathcal{L}^2$ penalty. We first solve the optimization problems separately for the well-known VaR/ES pair at the level of 0.95. Specifically, the two problems are given by

$$V_{\mathrm{VaR}}(a,b,F_1,\dots,F_n)=\min_{\mathbf{a}\in A_{a,b}}\left(\sup_{F_{\mathbf{X}}\in\mathcal{M}(F_1,\dots,F_n)}\mathrm{VaR}_{0.95}(\mathbf{a}^\top\mathbf{X})+c\|\mathbf{a}\|_2\right), \tag{3.29}$$

$$\begin{aligned} V_{\mathrm{ES}}(a,b,F_1,\dots,F_n)&=\min_{\mathbf{a}\in A_{a,b}}\left(\sup_{F_{\mathbf{X}}\in\mathcal{M}(F_1,\dots,F_n)}\mathrm{ES}_{0.95}(\mathbf{a}^\top\mathbf{X})+c\|\mathbf{a}\|_2\right)\\ &=\min_{\mathbf{a}\in A_{a,b}}\left(\mathbf{a}^\top(\mathrm{ES}_{0.95}(F_1),\dots,\mathrm{ES}_{0.95}(F_n))+c\|\mathbf{a}\|_2\right), \end{aligned} \tag{3.30}$$

where the true value of the original inner VaR problem is approximated by the rearrangement algorithm (RA) of Puccetti and Rüschendorf (2012) and Embrechts et al. (2013), whereas the optimal value of the inner ES problem is obtained by simultaneously minimizing the sum of a linear combination of ES and the 2-norm of the vector $\mathbf{a}$, which can be done efficiently.[11] In particular,

[11] The outer problems of (3.29) and (3.30) are solved numerically by the constrained nonlinear multivariable function

if the marginals of the random losses are identical (i.e., $F_1 = \cdots = F_n = F$), the optimal solution is $\mathbf{a}^* = (1/n, \ldots, 1/n)$ and $V_{\mathrm{ES}}(a, b, F_1, \ldots, F_n) = \mathrm{ES}_{0.95}(F) + c/\sqrt{n}$. We consider the following marginal distributions

(i) $F_i$ follows a Pareto distribution with scale parameter 1 and shape parameter $3+(i-1)/(n-1)$ for $i = 1, \ldots, n$;

(ii) $F_i$ is normally distributed with parameters $\mathrm{N}(1, 1+(i-1)/(n-1))$, for $i = 1, \ldots, n$;

(iii) $F_i$ follows an exponential distribution with parameter $1+(i-1)/(n-1)$, for $i = 1, \ldots, n$.

We choose $n$ to be 3, 10, and 20. For comparison, we calculate the value $n\|\Delta\mathbf{a}^*\|_2$, where $\Delta\mathbf{a}^*$ is the difference between the optimal weights of the non-convex problem and the convex problem. In addition, we calculate the absolute differences between the optimal values obtained by the two problems, $\Delta V = V_{\mathrm{ES}} - V_{\mathrm{VaR}} \geqslant 0$, and the percentage differences $\Delta V / V_{\mathrm{VaR}}$. Tables 3.3 and 3.4 show the numerical results that compare both optimization problems with two choices of the action sets $A_{a,b}$. The computation time is reported (in seconds). We observe that the optimal values obtained in the two problems get closer and become approximately the same as $n$ gets larger. As explained before, this is because the set of plausible distributions $\mathcal{M}(F_1, \ldots, F_n)$ is asymptotically equal to a set closed under concentration for all intervals.

Next, we consider a TK distortion riskmetric with parameter $\gamma = 0.7$. Due to the non-concavity of $h$, there are no known ways of directly solving the non-convex optimization problem

$$\min_{\mathbf{a}\in A_{a,b}} \left( \sup_{F_{\mathbf{X}}\in\mathcal{M}(F_1,\ldots,F_n)} \rho_h(\mathbf{a}^\top\mathbf{X}) + c\|\mathbf{a}\|_2 \right). \tag{3.31}$$

We may get an approximation of (3.31) using a lower bound of $\rho_h$ in (3.31) produced with the dependence structure created by the rearrangement algorithm (RA);[12] for simplicity, we denote by $V_h$ this lower bound. On the other hand, by (3.20), the convex counterpart of (3.31) can be written (using Theorem 3.1) as

$$\begin{aligned} V_{h^*}(a, b, F_1, \ldots, F_n) &= \min_{\mathbf{a}\in A_{a,b}} \left( \sup_{F_{\mathbf{X}}\in\mathcal{M}(F_1,\ldots,F_n)} \rho_{h^*}(\mathbf{a}^\top\mathbf{X}) + c\|\mathbf{a}\|_2 \right) \\ &= \min_{\mathbf{a}\in A_{a,b}} \left( \mathbf{a}^\top(\rho_{h^*}(X_1), \ldots, \rho_{h^*}(X_n)) + c\|\mathbf{a}\|_2 \right), \end{aligned} \tag{3.32}$$

"fmincon" with the sequential quadratic programming (SQP) algorithm. The same method is also applied when solving outer problems of (3.31) and (3.32).

[12]Such a dependence structure obviously provides a lower bound for the worst-case value in (3.31). In theory, the result from RA is thus not an optimal dependence structure for (3.31). In our numerical results, this lower bound is very close to an upper bound only for the case of VaR and ES but not for the case of TK distortion riskmetrics.

where $X_i \sim F_i$ for $i = 1, \ldots, n$. We calculate the absolute differences between the optimal values of the convex and non-convex problems $\Delta V = V_{h^*} - V_h \geqslant 0$, and the percentage differences $\Delta V / V_h$. Tables 3.5 and 3.6 compare the numerical results of the two optimization problems with different choices of $A_{a,b}$. We observe that the percentage differences between the RA lower bound $V_h$ for the non-convex problem (3.31) and the minimum value $V_{h^*}$ of the convex problem (3.32) are roughly between 10% to 20%. Note that the RA lower bound is not expected to be very close to the true minimum of (3.31), and hence the differences between the solution of (3.31) and the optimal value in (3.32) are smaller than the observed numbers.

Note that, by transforming an optimization problem with a non-convex objective to a convex one, we significantly reduce the computational time of calculating bounds with negligible errors, as shown in Tables 3.3-3.6.

Table 3.3: Comparison of the numerical results of the two optimization problems (3.29) and (3.30) for $\mathrm{VaR}_{0.95}$ and $\mathrm{ES}_{0.95}$ with $a = 0$ and $b = 1$

| | | $c$ | $V_{\mathrm{VaR}}$ | time | $V_{\mathrm{ES}}$ | time | $n\|\Delta \mathbf{a}^*\|_2$ | $\Delta V$ | $\Delta V / V_{\mathrm{VaR}}$ (%) |
|---|---|---|---|---|---|---|---|---|---|
| (i) Pareto | $n = 3$ | 2.5 | 3.547 | 31.53 | 3.741 | 0.72 | $8.88 \times 10^{-5}$ | 0.194 | 5.48 |
| | $n = 10$ | 3.0 | 3.197 | 153.83 | 3.215 | 1.39 | $9.18 \times 10^{-4}$ | 0.0178 | 0.558 |
| | $n = 20$ | 4.0 | 3.156 | 424.17 | 3.159 | 9.37 | $3.53 \times 10^{-5}$ | $2.68 \times 10^{-3}$ | 0.0850 |
| (ii) Normal | $n = 3$ | 4.0 | 5.766 | 31.19 | 5.785 | 0.18 | $1.39 \times 10^{-3}$ | 0.0186 | 0.323 |
| | $n = 10$ | 2.0 | 4.082 | 97.30 | 4.083 | 0.77 | $1.18 \times 10^{-3}$ | $3.24 \times 10^{-5}$ | $7.93 \times 10^{-4}$ |
| | $n = 20$ | 3.0 | 4.132 | 431.79 | 4.132 | 4.66 | $2.69 \times 10^{-3}$ | $1.88 \times 10^{-5}$ | $4.55 \times 10^{-4}$ |
| (iii) Exp | $n = 3$ | 3.0 | 4.251 | 26.78 | 4.405 | 0.07 | 0.331 | 0.155 | 3.64 |
| | $n = 10$ | 4.0 | 3.892 | 118.23 | 3.893 | 0.50 | $9.74 \times 10^{-4}$ | $2.92 \times 10^{-4}$ | $7.52 \times 10^{-3}$ |
| | $n = 20$ | 7.0 | 4.230 | 543.03 | 4.230 | 3.47 | $3.08 \times 10^{-4}$ | $4.47 \times 10^{-5}$ | $1.06 \times 10^{-3}$ |

## 3.8 Concluding remarks

We introduced the new concept of closedness under concentration, which is, in the context of distributional uncertainty, a sufficient condition to transform an optimization problem with a non-convex distortion riskmetric to its convex counterpart. This concept is genuinely weaker than closedness under conditional expectation, and our main result unifies and improves many existing

Table 3.4: Comparison of the numerical results of the two optimization problems (3.29) and (3.30) for $\mathrm{VaR}_{0.95}$ and $\mathrm{ES}_{0.95}$ with $a = 1/(2n)$ and $b = 2/n$

| | | $c$ | $V_{\mathrm{VaR}}$ | time | $V_{\mathrm{ES}}$ | time | $n\|\Delta\mathbf{a}^*\|_2$ | $\Delta V$ | $\Delta V/V_{\mathrm{VaR}}$ (%) |
|---|---|---|---|---|---|---|---|---|---|
| (i) Pareto | $n = 3$ | 2.5 | 3.546 | 54.59 | 3.741 | 0.19 | $6.58 \times 10^{-4}$ | 0.194 | 5.48 |
| | $n = 10$ | 3.0 | 3.204 | 146.63 | 3.220 | 1.60 | $1.99 \times 10^{-4}$ | 0.0160 | 0.498 |
| | $n = 20$ | 4.0 | 3.162 | 847.13 | 3.163 | 10.08 | $1.69 \times 10^{-3}$ | $2.23 \times 10^{-3}$ | 0.0706 |
| (ii) Normal | $n = 3$ | 4.0 | 5.766 | 57.31 | 5.785 | 0.19 | $1.32 \times 10^{-3}$ | 0.0187 | 0.324 |
| | $n = 10$ | 2.0 | 4.084 | 166.25 | 4.084 | 0.79 | 0 | $2.94 \times 10^{-5}$ | $7.20 \times 10^{-4}$ |
| | $n = 20$ | 3.0 | 4.133 | 691.91 | 4.133 | 5.91 | 0 | $1.99 \times 10^{-5}$ | $4.82 \times 10^{-4}$ |
| (iii) Exp | $n = 3$ | 3.0 | 4.369 | 48.58 | 4.422 | 0.09 | $1.04 \times 10^{-3}$ | 0.0533 | 1.22 |
| | $n = 10$ | 4.0 | 3.916 | 115.18 | 3.916 | 0.50 | $2.54 \times 10^{-5}$ | $1.38 \times 10^{-4}$ | $3.52 \times 10^{-3}$ |
| | $n = 20$ | 7.0 | 4.236 | 665.05 | 4.236 | 3.48 | $2.73 \times 10^{-4}$ | $4.04 \times 10^{-5}$ | $9.54 \times 10^{-4}$ |

results in the literature. Many sets of plausible distributions commonly used in the literature of finance, optimization, and risk management are closed under concentration within some $\mathcal{I}$. Moreover, by focusing on distortion riskmetrics whose distortion functions are not necessarily monotone, concave, or continuous, we are able to solve optimization problems for the class of functionals larger than classical risk measures or deviation measures. In particular, we are able to obtain bounds on differences between two distortion riskmetrics, which represent measures of disagreement between two utilities/risk attitudes. Our result can also be applied to solve the popular problem of optimizing risk measures under moment constraints. In particular, we obtain the worst- and best-case distortion riskmetrics when the underlying random variable has a fixed mean and bounded $p$-th moment.

We demonstrate the applicability of our result by numerically calculating the solution to optimizing the difference between risk measures, preference robust optimization and portfolio optimization under marginal constraints. In all numerical examples, the original non-convex problem is converted or well approximated by a convex one which can be solved efficiently.

Our condition of closedness under concentration within $\mathcal{I}$ in Theorem 3.1 is sufficient but not necessary for the equivalence of optimization problems with non-convex and convex objectives under distributional uncertainty. A necessary condition of the equivalence is closedness under concentration of the set of maximizers in Theorem 3.2. An open question is to find a *necessary*

Table 3.5: Comparison of the numerical results of the two optimization problems (3.31) and (3.32) for TK distortion riskmetrics with $a = 0$ and $b = 1$

| | | $c$ | $V_h$ | time | $V_{h^*}$ | time | $n\|\Delta \mathbf{a}^*\|_2$ | $\Delta V$ | $\Delta V/V_h$ (%) |
|---|---|---|---|---|---|---|---|---|---|
| (i) Pareto | $n = 3$ | 1.0 | 1.076 | 144.75 | 1.185 | 0.23 | 0.488 | 0.109 | 10.2 |
| | $n = 10$ | 2.0 | 1.047 | 220.03 | 1.237 | 1.42 | 0 | 0.190 | 18.1 |
| | $n = 20$ | 4.0 | 1.301 | 826.64 | 1.501 | 8.24 | 0 | 0.200 | 15.4 |
| (ii) Normal | $n = 3$ | 0.5 | 1.240 | 60.76 | 1.493 | 0.16 | 0.0784 | 0.253 | 20.4 |
| | $n = 10$ | 0.5 | 1.141 | 246.31 | 1.363 | 0.72 | 1.28 | 0.222 | 19.4 |
| | $n = 20$ | 0.5 | 1.103 | 1503.35 | 1.316 | 2.80 | 1.78 | 0.213 | 19.3 |
| (iii) Exp | $n = 3$ | 1.0 | 1.305 | 49.79 | 1.427 | 0.23 | 0.360 | 0.122 | 9.32 |
| | $n = 10$ | 2.0 | 1.313 | 198.43 | 1.484 | 1.62 | 0.184 | 0.171 | 13.0 |
| | $n = 20$ | 2.0 | 1.120 | 850.12 | 1.286 | 10.91 | 0.158 | 0.166 | 14.8 |

*and sufficient* condition on the uncertainty set $\mathcal{M}$ itself such that the desired equivalence holds. Pinning down such a condition may facilitate many more applications in decision theory, finance, game theory, and operations research.

## 3.9 Omitted technical details from the chapter

In this section, we present technical details for some examples and as well as some technical remarks omitted from the chapter.

### 3.9.1 Proofs of claims in some Examples

*Proof of the claim in Example 3.6.* We show that $\mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma)$ is equivalent to

$$\{F_S \in \mathcal{M}_2 : \mathbb{E}[S] = \mathbf{a}^\top \boldsymbol{\mu},\ \mathrm{var}(S) \leqslant \mathbf{a}^\top \Sigma \mathbf{a}\} = \mathcal{M}\left(2, \mathbf{a}^\top \boldsymbol{\mu}, \left(\mathbf{a}^\top \Sigma \mathbf{a}\right)^{1/2}\right).$$

For a proof of the equivalence between the sets with fixed mean and covariance matrix, see Popescu (2007). Indeed, it is clear that $\mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma) \subseteq \mathcal{M}(2, \mathbf{a}^\top \boldsymbol{\mu}, (\mathbf{a}^\top \Sigma \mathbf{a})^{1/2})$. On the other hand, for all $F_S \in \mathcal{M}(2, \mathbf{a}^\top \boldsymbol{\mu}, (\mathbf{a}^\top \Sigma \mathbf{a})^{1/2})$, we write $\mathbf{a} = (a_1, \dots, a_n)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, and take $\mathbf{X} = (X_1, \dots, X_n)$ such that $X_i = (S - \mathbf{a}^\top \boldsymbol{\mu})/(na_i) + \mu_i$, for $i = 1, \dots, n$. It follows that $F_S = F_{\mathbf{a}^\top \mathbf{X}} \in \mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma)$. Therefore, we have $\mathcal{M}^{\mathrm{mv}}(\mathbf{a}, \boldsymbol{\mu}, \Sigma) = \mathcal{M}(2, \mathbf{a}^\top \boldsymbol{\mu}, (\mathbf{a}^\top \Sigma \mathbf{a})^{1/2})$. $\square$

Table 3.6: Comparison of the numerical results of the two optimization problems (3.31) and (3.32) for TK distortion riskmetrics with $a = 1/(2n)$ and $b = 2/n$

| | | $c$ | $V_h$ | time | $V_{h^*}$ | time | $n\|\Delta \mathbf{a}^*\|_2$ | $\Delta V$ | $\Delta V/V_h$ (%) |
|---|---|---|---|---|---|---|---|---|---|
| (i) Pareto | $n = 3$ | 1.0 | 1.077 | 73.21 | 1.185 | 0.25 | 0.469 | 0.109 | 10.11 |
| | $n = 10$ | 2.0 | 1.047 | 248.38 | 1.237 | 2.29 | 0.378 | 0.191 | 18.2 |
| | $n = 20$ | 4.0 | 1.301 | 638.24 | 1.501 | 12.21 | 0 | 0.200 | 15.4 |
| (ii) Normal | $n = 3$ | 0.5 | 1.240 | 179.68 | 1.493 | 0.19 | 0.0784 | 0.253 | 20.4 |
| | $n = 10$ | 0.5 | 1.146 | 389.97 | 1.363 | 0.76 | 0.660 | 0.217 | 19.0 |
| | $n = 20$ | 0.5 | 1.103 | 1563.84 | 1.316 | 3.39 | 1.63 | 0.213 | 19.3 |
| (iii) Exp | $n = 3$ | 1.0 | 1.304 | 52.66 | 1.430 | 0.25 | 0.107 | 0.126 | 9.65 |
| | $n = 10$ | 2.0 | 1.312 | 236.15 | 1.485 | 2.27 | 0.214 | 0.172 | 13.1 |
| | $n = 20$ | 2.0 | 1.119 | 879.73 | 1.289 | 10.10 | 0.141 | 0.170 | 15.2 |

*Proof of the claim in Example 3.10.* We will show that $\mathcal{M}(\widetilde{G}, \epsilon)$ is closed under concentration within $\mathcal{I}$ for all $\mathcal{I} \subseteq \widetilde{\mathcal{I}}$. Write $\mathcal{I} = \{C_i : i \in K\}$ for some $K \subseteq \mathbb{N}$. For all $i \in K$ and $F \in \mathcal{M}(\widetilde{G}, \epsilon)$, we have $\widetilde{G}^{-1}(u) = c_i$ for $u \in C_i$ for some $c_i \in \mathbb{R}$. For all $i \in K$, by Jensen's inequality,

$$\frac{1}{\lambda(C_i)} \int_{C_i} \left|F^{-1}(u) - \widetilde{G}^{-1}(u)\right|^p \mathrm{d}u \geqslant \left|\frac{\int_{C_i} F^{-1}(u)\,\mathrm{d}u}{\lambda(C_i)} - c_i\right|^p = \frac{1}{\lambda(C_i)} \int_{C_i} \left|(F^{C_i})^{-1}(u) - \widetilde{G}^{-1}(u)\right|^p \mathrm{d}u.$$

It follows that

$$\begin{aligned}(W_p(F, \widetilde{G}))^p - (W_p(F^{C_i}, \widetilde{G}))^p &= \int_0^1 \left|F^{-1}(u) - \widetilde{G}^{-1}(u)\right|^p \mathrm{d}u - \int_0^1 \left|(F^{C_i})^{-1}(u) - \widetilde{G}^{-1}(u)\right|^p \mathrm{d}u \\ &= \int_{C_i} \left|F^{-1}(u) - \widetilde{G}^{-1}(u)\right|^p \mathrm{d}u - \int_{C_i} \left|(F^{C_i})^{-1}(u) - \widetilde{G}^{-1}(u)\right|^p \mathrm{d}u \geqslant 0,\end{aligned}$$

and thus $W_p(F^{C_i}, \widetilde{G}) \leqslant W_p(F, \widetilde{G}) \leqslant \epsilon$. Moreover, (3.7) and the above argument lead to

$$(W_p(F, \widetilde{G}))^p - (W_p(F^{\mathcal{I}}, \widetilde{G}))^p = \sum_{i \in K} (W_p(F, \widetilde{G}))^p - (W_p(F^{C_i}, \widetilde{G}))^p \geqslant 0.$$

Hence, $W_p(F^{\mathcal{I}}, \widetilde{G}) \leqslant W_p(F, \widetilde{G}) \leqslant \epsilon$. □

*Proof of the claim in Example 3.11.* For $\epsilon \geqslant 0$, $\mathbf{w} \in [0, \infty)^n$, $p > 1$, $a > 1$ and $\mathbf{Z} \in (\mathcal{L}^p)^n$, by Theorem 7 of Mao et al. (2022), the uncertainty set

$$\{F_{\mathbf{w}^\top \mathbf{X}} \in \mathcal{M}_p : F_{\mathbf{X}} \in \mathcal{M}^n(F_{\mathbf{Z}}, \epsilon)\} = \mathcal{M}(F_{\mathbf{w}^\top \mathbf{Z}}, \epsilon \|\mathbf{w}\|_b),$$

where $b$ is the conjugate of $a$ (i.e., $1/a + 1/b = 1$). Suppose that for a benchmark distribution $\widetilde{G} \in \mathcal{M}_p^n$, there exists a random vector $\mathbf{Z} \sim \widetilde{G}$ such that $\mathbf{Z} \geqslant \mathbf{0}$ and $\mathbb{P}(\mathbf{Z} = \mathbf{0}) = p_0$ for some $p_0 \in (0, 1]$. Note that $\mathbb{P}(\mathbf{w}^\top \mathbf{Z} = 0) \geqslant p_0$ and the quantile function of $\mathbf{w}^\top \mathbf{Z}$ is equal to 0 on $(0, p_0]$. It follows from Example 3.10 that the set $\mathcal{M}(F_{\mathbf{w}^\top \mathbf{Z}}, \epsilon \|\mathbf{w}\|_b)$ is closed under concentration within $\{(0, t)\}$ for all $t \leqslant p_0$. □

*Proof of the claim in Example 3.12.* We will show that the set of distributions,

$$\mathcal{M} = \{F_{X - V(X) + g(\mathbb{E}[V(X)])} \in \mathcal{M}_1 : V \in \mathcal{V}\},$$

is closed under concentration within $\{(p, 1)\}$ for all $p \in [p_0, 1)$. For each $V \in \mathcal{V}$ and a standard uniform random variable $U$, we write $a = \mathbb{E}[F^{-1}_{X - V(X)}(U) | U \in (p, 1)]$. Since $F_X^{-1}(p) \geqslant l$, we can take

$$W(x) = V(x)\mathbb{1}_{\{x \leqslant F_X^{-1}(p)\}} + (x - a)\mathbb{1}_{\{x > F_X^{-1}(p)\}}, \quad x \in \mathbb{R}.$$

It follows that $W \in \mathcal{V}$. Noting that $a = \mathbb{E}[X - V(X) | X > F_X^{-1}(p)]$, we have

$$\begin{aligned}
&X - W(X) + g(\mathbb{E}[W(X)]) \\
&= (X - V(X))\mathbb{1}_{\{X \leqslant F_X^{-1}(p)\}} + a\mathbb{1}_{\{X > F_X^{-1}(p)\}} + g\left(\mathbb{E}[V(X)\mathbb{1}_{\{X \leqslant F_X^{-1}(p)\}} + (X - a)\mathbb{1}_{\{X > F_X^{-1}(p)\}}]\right) \\
&= (X - V(X))\mathbb{1}_{\{X \leqslant F_X^{-1}(p)\}} + a\mathbb{1}_{\{X > F_X^{-1}(p)\}} + g(\mathbb{E}[V(X)]),
\end{aligned}$$

which follows the same distribution as $F^{(p,1)}_{X - V(X) + g(\mathbb{E}[V(X)])}$. It follows that $\mathcal{M}$ is closed under concentration within $\{(p, 1)\}$ for all $p \in [p_0, 1)$. □

### 3.9.2 A few additional technical remarks mentioned in the chapter

*Remark* 3.5 (on Theorem 3.1). Using Theorem 3.1, if for some $\mathbf{a} \in A$, the set $\mathcal{M} := \{F_{f(\mathbf{a}, \mathbf{X})} : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\}$ is closed under concentration for all intervals and $\sup\{\rho_{h^*}(f(\mathbf{a}, \mathbf{X})) : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\} = \infty$, then $\sup\{\rho_h(f(\mathbf{a}, \mathbf{X})) : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\} = \infty$. Thus, both objectives in the inner optimization of (3.1) are infinite for this $\mathbf{a}$, which can be excluded from the outer optimization over $A$. Verifying $\sup\{\rho_{h^*}(f(\mathbf{a}, \mathbf{X})) : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\} = \infty$ is easier than verifying $\sup\{\rho_h(f(\mathbf{a}, \mathbf{X})) : F_{\mathbf{X}} \in \widetilde{\mathcal{M}}\} = \infty$ since generally $\rho_h$ is smaller than $\rho_{h^*}$.

*Remark* 3.6 (on Example 3.6). Using Strassen's Theorem (e.g., Theorem 3.A.4 of Shaked and Shanthikumar, 2007), closedness under conditional expectation can equivalently be expressed using convex order. A set $\mathcal{M} \subseteq \mathcal{M}_1$ is closed under conditional expectation if and only if it holds that for $F \in \mathcal{M}$ and $G \leqslant_{\mathrm{cx}} F$, we have $G \in \mathcal{M}$.

*Remark* 3.7 (on Proposition 3.3). In Proposition 3.3, if $\mathcal{M}$ is closed under conditional expectation, $\mathcal{I}$ can be taken as an infinite set. However, $\mathcal{M}$ may not be closed under concentration within an infinite $\mathcal{I}$ if we only assume that $\mathcal{M}$ is closed under concentration for all intervals. Indeed, if we take $\mathcal{M}$ as the set of distributions obtained by some $F \in \mathcal{M}$ with finitely many concentrations, then clearly $\mathcal{M}$ is closed under concentration for all intervals. However, $F^{\mathcal{I}} \notin \mathcal{M}$ when $\mathcal{I}$ is an infinite collection of disjoint intervals. This also serves as a counter-example of the converse statement of Proposition 3.2 since $\mathcal{M}$ is closed under concentration for all intervals but not closed under conditional expectation.

## 3.10 Proofs of all technical results

We present all proofs of technical results in this section. Throughout, we denote the set of discontinuity points of $h$ (excluding 0 and 1) by

$$J_h = \{t \in (0,1) : h(t) \neq h(t^+) \text{ or } h(t) \neq h(t^-)\}. \tag{3.33}$$

Note that $\hat{h}(t)$ can be written as

$$\hat{h}(t) = \begin{cases} h(t^+) \vee h(t^-) \vee h(t), & t \in J_h, \\ h(t), & \text{otherwise.} \end{cases} \tag{3.34}$$

### 3.10.1 Proof of results in Section 3.2

*Proof of Proposition 3.1.* Note that $(\hat{h})^* = h^* = \hat{h} = h$ on 0 and 1. For all $t \in (0,1)$, since $(\hat{h})^*(t) \geqslant \hat{h}(t) \geqslant h(t)$, we have $(\hat{h})^*(t) \geqslant h^*(t)$. On the other hand, we have $h^*(t) \geqslant h(t^+)$ for $t \in (0,1)$. Indeed, if $h^*(t_0) < h(t_0^+)$ for some $t_0 \in (0,1)$, then we have $h^*(t_0+\epsilon) < h(t_0+\epsilon)$ for some $\epsilon > 0$, which leads to a contradiction. Similarly, we have $h^*(t) \geqslant h(t^-)$ for $t \in (0,1)$. Together with $h^* \geqslant h$ on $(0,1)$, we have $h^* \geqslant \hat{h}$ on $(0,1)$, which implies that $h^* \geqslant (\hat{h})^*$ on $(0,1)$. Therefore, $(\hat{h})^* = h^*$ on $[0,1]$.

Next, we assert that the set $\{t \in [0,1] : \hat{h}(t) \neq h^*(t)\}$ is a union of disjoint sets that are not singletons. To show this assertion, assume that the converse is true. There exists $x \in (0,1)$, such that $\hat{h}(x) < h^*(x)$ and $\hat{h}(t) = h^*(t)$ on $t \in (x-\epsilon, x) \cup (x, x+\epsilon)$ for some $0 < \epsilon \leqslant x \wedge (1-x)$. It is clear that $x \in J_h$. Since $h^*$ is continuous on $(x-\epsilon, x+\epsilon)$, we have

$$\hat{h}(x) < h^*(x) = h^*(x^+) = \hat{h}(x^+).$$

This contradicts (3.34). Therefore, the set $\{t \in [0,1] : \hat{h}(t) \neq h^*(t)\}$ is the union of some disjoint intervals, denoted by $\cup_{l \in L} A_l$ for some $L \subseteq \mathbb{N}$. For all $l \in L$, we denote the left and right endpoints of $A_l$ by $a_l$ and $b_l$, respectively, with $a_l < b_l$. Define a function via linear interpolation

$$h^c(t) = \begin{cases} \hat{h}(a_l) + \frac{\hat{h}(b_l) - \hat{h}(a_l)}{b_l - a_l}(t - a_l), & t \in A_l,\ l \in L, \\ \hat{h}(t), & \text{otherwise.} \end{cases}$$

It is clear that $h^c \leqslant h^*$ and $h^c$ is continuous on $(0,1)$. We will prove that $h^c = h^*$ on $\cup_{l \in L} A_l$. Suppose for the purpose of contradiction that $h^c \neq h^*$ on $\cup_{l \in L} A_l$. Since $h^c < h^*$ for some point in $\cup_{l \in L} A_l$, there exists $x_0 \in A_l$ for some $l \in L$ such that $h^c(x_0) < \hat{h}(x_0)$. Thus we can take a point $(x_1, \hat{h}(x_1)) \in (0,1) \times \mathbb{R}$ with $\hat{h}(x_1) > h^c(x_1)$, which has the largest perpendicular distance to the straight line $h^c(t) = \hat{h}(a_l) + \frac{\hat{h}(b_l) - \hat{h}(a_l)}{b_l - a_l}(t - a_l)$, namely,

$$x_1 = \underset{\substack{x \in A_l \\ \hat{h}(x) > h^c(x)}}{\arg\max} \frac{(b_l - a_l)\hat{h}(x) - (\hat{h}(b_l) - \hat{h}(a_l))x - (b_l - a_l)\hat{h}(a_l) + (\hat{h}(b_l) - \hat{h}(a_l))a_l}{\left((\hat{h}(b_l) - \hat{h}(a_l))^2 + (b_l - a_l)^2\right)^{1/2}}.$$

The existence of the maximizer $x_1$ is due to the upper semicontinuity of $\hat{h}$. There exists a function $g$ with $g = h^*$ on $[0,1] \setminus A_l$ and $g(x_1) = \hat{h}(x_1)$, such that $g$ is concave and $\hat{h} \leqslant g \leqslant h^*$ on $[0,1]$. Since $h^* > \hat{h}$ on $A_l$, we have $h^*(x_1) > \hat{h}(x_1) = g(x_1)$. Thus $h^*$ cannot be the concave envelope of $\hat{h}$, which leads to a contradiction. Thus, $h^* = h^c$ on $\cup_{l \in L} A_l$. Since $h^* = \hat{h} = h^c$ on $(0,1) \setminus (\cup_{l \in L} A_l)$, we have $h^* = h^c$. Therefore, $\{t \in [0,1] : \hat{h}(t) \neq h^*(t)\}$ is a union of disjoint open intervals, and $h^*$ is linear on each of the intervals. □

### 3.10.2 Proofs of results in Section 3.3

*Proof of Theorem 3.1.* We will first show that, assuming that $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$, we have

$$\sup_{F_X \in \mathcal{M}} \rho_{\hat{h}}(X) = \sup_{F_X \in \mathcal{M}} \rho_{h^*}(X). \tag{3.35}$$

After proving (3.35), we show the three statements in Theorem 3.1 in the order (i), (ii), and (iii).

For $h \in \mathcal{H}$, suppose that $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$. Take an arbitrary random variable $Y$ with $F_Y \in \mathcal{M}$. Let $G = F_Y^{\mathcal{I}_h}$. For $h \in \mathcal{H}$, write functions $g(t) = 1 - \hat{h}(1-t)$ and $g_*(t) = 1 - h^*(1-t)$ for $t \in [0,1]$. By definition of $\mathcal{I}_h$, $g \neq g_*$ on each set in $\mathcal{I}_h$ and $g = g_*$ on other sets. For any $(a,b) \in \mathcal{I}_h$, we have $G^{-1}(t) = \frac{\int_a^b F_Y^{-1}(u)\,\mathrm{d}u}{b-a}$ for all $t \in (a,b]$ and $G^{-1+}(t) = \frac{\int_a^b F_Y^{-1}(u)\,\mathrm{d}u}{b-a}$

for all $t \in [a, b)$. Using the fact that $g_*$ is linear on $(a, b)$ and $g(t) = g_*(t)$ for $t = a, b$, we have

$$\begin{aligned}\int_{(a,b)} F_Y^{-1}(t)\,\mathrm{d}g_*(t) &= (g_*(b) - g_*(a))\frac{\int_a^b F_Y^{-1}(t)\,\mathrm{d}t}{b-a} \\ &= (g(b) - g(a))\frac{\int_a^b F_Y^{-1}(t)\,\mathrm{d}t}{b-a} \\ &= \int_{(a,b]} G^{-1}(t)\,\mathrm{d}g(t) + G^{-1+}(a)(g(a^+) - g(a)).\end{aligned} \tag{3.36}$$

Define the sets

$$J_+ = \{t \in J_h : \hat{h}(t^+) = \hat{h}(t) \neq \hat{h}(t^-)\}, \quad J_- = \{t \in J_h : \hat{h}(t^+) \neq \hat{h}(t) = \hat{h}(t^-)\},$$
$$\text{and} \quad J_0 = \{t \in J_h : \hat{h}(t^+) \neq \hat{h}(t) \neq \hat{h}(t^-)\}.$$

To better understand these sets, we recall Figure 3.1 (without concave envelopes) as Figure 3.5 to demonstrate an example of a distortion function $h$, the corrresponding $\hat{h}$, the sets $J_h$, $J_+$, $J_-$, and $J_0$, and the sets $\hat{J}$, $\hat{J}_+$, $\hat{J}_-$, $\hat{J}_+^0$, and $\hat{J}_-^0$ (defined in the proof of (i) below).

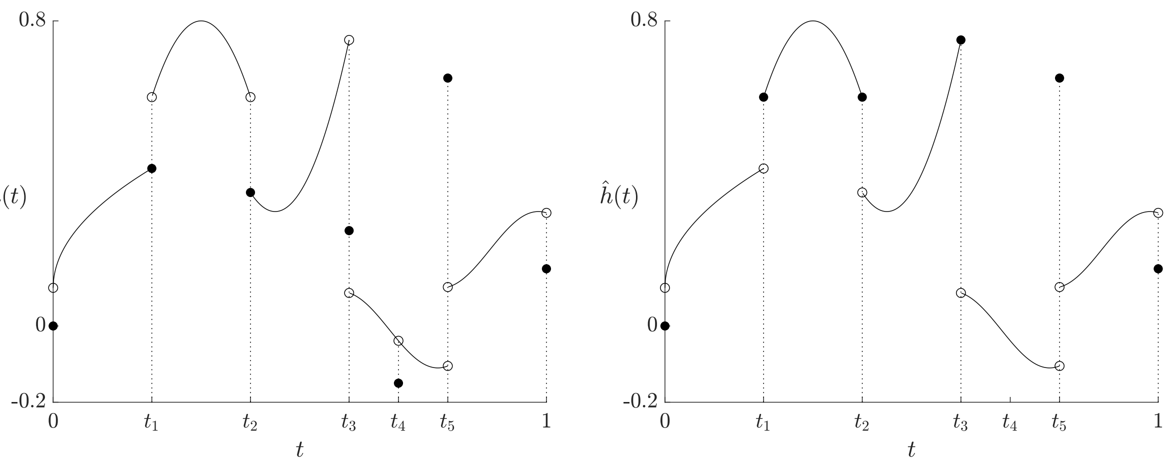


Figure 3.5: An example of $h$ (left) and $\hat{h}$ (right); in this figure, $J_h = \{t_1, t_2, t_3, t_4, t_5\}$, $J_+ = \{t_1\}$, $J_- = \{t_2, t_3\}$, and $J_0 = \{t_5\}$. Moreover, the sets we use in the proof of (i) are $\hat{J} = \{t_1, t_2, t_3, t_4\}$, $\hat{J}_+ = \{t_1, t_4\}$, $\hat{J}_- = \{t_2, t_3\}$, $\hat{J}_+^0 = \{t_4\}$, and $\hat{J}_-^0 = \{t_3\}$

Note that for a random variable $Z_{\mathcal{I}_h} \sim F_Y^{\mathcal{I}_h}$, we have

$$\rho_{\hat{h}}(Z_{\mathcal{I}_h}) = \int_{(0,1]\setminus(J_+\cup J_0)} G^{-1}(t)\,\mathrm{d}g(t) + \sum_{t \in J_+ \cup J_0 \cup \{0\}} G^{-1+}(t)(g(t^+) - g(t)).$$

Hence using (3.36) and (3.7), we get

$$\begin{aligned}
&\rho_{h^*}(Y) - \rho_{\hat{h}}(Z_{\mathcal{I}_h}) \\
&= \int_0^1 F_Y^{-1}(t)\,\mathrm{d}g_*(t) + F_Y^{-1+}(0)(g_*(0^+) - g_*(0)) \\
&\quad - \int_{(0,1]\setminus(J_+\cup J_0)} G^{-1}(t)\,\mathrm{d}g(t) - \sum_{t\in J_+\cup J_0\cup\{0\}} G^{-1+}(t)(g(t^+) - g(t)) \\
&= \sum_{(a,b)\in\mathcal{I}_h} \left( \int_{(a,b)} F_Y^{-1}(t)\,\mathrm{d}g_*(t) - \int_{(a,b]} G^{-1}(t)\,\mathrm{d}g(t) - G^{-1+}(a)(g(a^+) - g(a)) \right) = 0.
\end{aligned} \tag{3.37}$$

Since $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$, we have $F_Y^{\mathcal{I}_h} \in \mathcal{M}$ by definition. Thus we have

$$\rho_{h^*}(Y) = \rho_{\hat{h}}(Z_{\mathcal{I}_h}) \leqslant \sup_{F_X\in\mathcal{M}} \rho_{\hat{h}}(X),$$

which gives our desired equality (3.35) since $\rho_{h^*} = \rho_{(\hat{h})^*} \geqslant \rho_{\hat{h}}$.

Proof of (i): Using $h = \hat{h}$ and (3.35), we have $\sup_{F_X\in\mathcal{M}} \rho_h(X) = \sup_{F_X\in\mathcal{M}} \rho_{h^*}(X)$.

Proof of (ii): We will prove (ii) in two main steps. First, we show that (ii) holds if $\mathcal{I}_h$ is finite and $h$ has finitely many discontinuity points. Next, we discuss general $h$.

***Finite case:*** Here we prove (3.8) under the case where $\mathcal{I}_h$ is finite and $h$ has finitely many discontinuity points (i.e. $J_h$ in (3.33) is a finite set). Suppose that $\mathcal{M}$ is closed under concentration for all intervals, it directly implies that $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$ by Proposition 3.3. Therefore, (3.35) holds for all $h \in \mathcal{H}$. Next, we need to show that $\sup_{F_X\in\mathcal{M}} \rho_h(X) = \sup_{F_X\in\mathcal{M}} \rho_{\hat{h}}(X)$. Define

$$\hat{J} = \{t \in J_h : \hat{h}(t) \neq h(t)\}, \quad \hat{J}_+ = \{t \in \hat{J} : \hat{h}(t) = \hat{h}(t^+)\}, \quad \text{and} \quad \hat{J}_- = \hat{J} \setminus \hat{J}_+.$$

For $n > 0$, write intervals

$$A_s^n = \begin{cases} (1 - s - 1/\sqrt{n}, 1 - s + 1/n), & s \in \hat{J}_-, \\ (1 - s - 1/n, 1 - s + 1/\sqrt{n}), & s \in \hat{J}_+. \end{cases}$$

Let $\mathcal{I}^n = \{A_s^n : s \in \hat{J}\}$. Note that $h \in \mathcal{H}$ has finitely many discontinuity points. Thus the intervals in $\mathcal{I}^n$ are disjoint when $n$ is large enough. For all $F_Y \in \mathcal{M}$ and $Y \sim F_Y$, we define

$$Z_{\mathcal{I}^n} = F_Y^{-1}(U)\mathbb{1}_{\{U\notin\bigcup_{s\in\hat{J}} A_s^n\}} + \sum_{s\in\hat{J}} \mathbb{E}[F_Y^{-1}(U)|U \in A_s^n]\mathbb{1}_{\{U\in A_s^n\}}.$$

It follows that $Z_{\mathcal{I}^n} \sim F_Y^{\mathcal{I}^n}$ and the right-quantile function of $Z_{\mathcal{I}^n}$, denoted by $G_n^{-1+}$, is given by the right-continuous adjusted version of

$$F_Y^{-1+}(t)\mathbb{1}_{\{t\notin\bigcup_{s\in\hat{J}} A_s^n\}} + \sum_{s\in\hat{J}} \frac{\int_{A_s^n} F_Y^{-1}(u)\,\mathrm{d}u}{\lambda(A_s^n)}\mathbb{1}_{\{t\in A_s^n\}}, \quad t \in (0,1).$$

Thus we get

$$\lim_{n\to\infty} G_n^{-1+}(1-t) = \begin{cases} F_Y^{-1}(1-t), & t \in \hat{J}_-, \\ F_Y^{-1+}(1-t), & \text{otherwise.} \end{cases}$$

Similarly, if we denote the left-quantile function of $Z_{\mathcal{I}^n}$ by $G_n^{-1}$, then $G_n^{-1}$ is given by the left-continuous version of

$$F_Y^{-1}(t)\mathbb{1}_{\{t\notin\bigcup_{s\in\hat{J}} A_s^n\}} + \sum_{s\in\hat{J}} \frac{\int_{A_s^n} F_Y^{-1}(u)\,\mathrm{d}u}{\lambda(A_s^n)} \mathbb{1}_{\{t\in A_s^n\}}.$$

It follows that

$$\lim_{n\to\infty} G_n^{-1}(1-t) = \begin{cases} F_Y^{-1+}(1-t), & t \in \hat{J}_+, \\ F_Y^{-1}(1-t), & \text{otherwise.} \end{cases}$$

Define, further, the sets

$$\hat{J}_+^0 = \{t \in \hat{J}_+ : h(t) \neq h(t^-)\} \text{ and } \hat{J}_-^0 = \{t \in \hat{J}_- : h(t) \neq h(t^+)\}.$$

For $u \in [0,1]$, write

$$\begin{aligned}
&h_-(u) = \sum_{t\in\hat{J}_-} (h(t) - h(t^-))\mathbb{1}_{\{u\geqslant t\}}, \quad h_-^0(u) = \sum_{t\in\hat{J}_-^0} (h(t^+) - h(t))\mathbb{1}_{\{u>t\}}, \\
&h_+(u) = \sum_{t\in\hat{J}_+} (h(t^+) - h(t))\mathbb{1}_{\{u>t\}}, \quad h_+^0(u) = \sum_{t\in\hat{J}_+^0} (h(t) - h(t^-))\mathbb{1}_{\{u\geqslant t\}}, \\
&\hat{h}_-(u) = \sum_{t\in\hat{J}_-} (h(t^+) - h(t^-))\mathbb{1}_{\{u>t\}}, \quad \hat{h}_+(u) = \sum_{t\in\hat{J}_+} (h(t^+) - h(t^-))\mathbb{1}_{\{u\geqslant t\}}, \\
&\text{and } h_0(u) = h(u) - h_+(u) - h_-(u) - h_+^0(u) - h_-^0(u) = \hat{h}(u) - \hat{h}_+(u) - \hat{h}_-(u).
\end{aligned}$$

Note that $|Z_{\mathcal{I}^n} - F_Y^{-1}(U)| = 0$ when $U \notin \bigcup_{s\in\hat{J}} A_s^n$ and $0, 1 \in [0,1] \setminus \bigcup_{s\in\hat{J}} A_s^n$. We have $|Z_{\mathcal{I}^n} - F_Y^{-1}(U)| < \infty$. Therefore, by the dominated convergence theorem,

$$\begin{aligned}
&\lim_{n\to\infty} (\rho_{h_-}(Z_{\mathcal{I}^n}) + \rho_{h_-^0}(Z_{\mathcal{I}^n})) \\
&= \lim_{n\to\infty} \int_0^1 G_n^{-1+}(1-u)\,\mathrm{d}h_-(u) + \lim_{n\to\infty} \int_0^1 G_n^{-1}(1-u)\,\mathrm{d}h_-^0(u) \\
&= \sum_{t\in\hat{J}_-} F_Y^{-1}(1-t)(h(t) - h(t^-)) + \sum_{t\in\hat{J}_-^0} F_Y^{-1}(1-t)(h(t^+) - h(t)) \\
&= \sum_{t\in\hat{J}_-\setminus\hat{J}_-^0} F_Y^{-1}(1-t)(h(t) - h(t^-)) + \sum_{t\in\hat{J}_-^0} F_Y^{-1}(1-t)(h(t) - h(t^-) + h(t^+) - h(t)) \\
&= \sum_{t\in\hat{J}_-\setminus\hat{J}_-^0} F_Y^{-1}(1-t)(h(t^+) - h(t^-)) + \sum_{t\in\hat{J}_-^0} F_Y^{-1}(1-t)(h(t^+) - h(t^-)) = \rho_{\hat{h}_-}(Y).
\end{aligned}$$

Similarly, we get $\lim_{n\to\infty}(\rho_{h_+}(Z_{\mathcal{I}^n}) + \rho_{h^0_+}(Z_{\mathcal{I}^n})) = \rho_{\hat{h}_+}(Y)$. On the other hand, it is clear that $\lim_{n\to\infty}\rho_{h_0}(Z_{\mathcal{I}^n}) = \rho_{h_0}(Y)$. Therefore, we have

$$\begin{aligned}\lim_{n\to\infty}\rho_h(Z_{\mathcal{I}^n}) &= \lim_{n\to\infty}(\rho_{h_-}(Z_{\mathcal{I}^n}) + \rho_{h^0_-}(Z_{\mathcal{I}^n}) + \rho_{h_+}(Z_{\mathcal{I}^n}) + \rho_{h^0_+}(Z_{\mathcal{I}^n}) + \rho_{h_0}(Z_{\mathcal{I}^n}))\\ &= \rho_{\hat{h}_-}(Y) + \rho_{\hat{h}_+}(Y) + \rho_{h_0}(Y) = \rho_{\hat{h}}(Y).\end{aligned}$$

Thus we have

$$\rho_{\hat{h}}(Y) = \lim_{n\to\infty}\rho_h(Z_{\mathcal{I}^n}) \leqslant \sup_{F_X\in\mathcal{M}}\rho_h(X). \tag{3.38}$$

Using (3.35) and (3.38), we get

$$\sup_{F_X\in\mathcal{M}}\rho_{h^*}(X) = \sup_{F_X\in\mathcal{M}}\rho_{\hat{h}}(X) \leqslant \sup_{F_X\in\mathcal{M}}\rho_h(X).$$

***General case:*** We prove Theorem 3.1 for all general $h \in \mathcal{H}$ where $\mathcal{I}_h$ or the number of discontinuity points of $h$ is countable.

1. If $\mathcal{I}_h$ is countable, it suffices to prove (3.35). We write $\mathcal{I}_h$ as the collection of $(a_i, b_i)$ for $i \in \mathbb{N}$ and define $\mathcal{I}_1^n = \{(a_i, b_i) : i = 1, \dots, n\}$ for all $n \in \mathbb{N}$. Define the function

$$h_n(t) = \begin{cases} h^*(t), & t \in (1-b_i, 1-a_i),\ i = 1, \dots, n,\\ \hat{h}(t), & \text{otherwise.}\end{cases}$$

It is clear that for all $n \in \mathbb{N}$, the set $\{t \in [0,1] : h_n(t) \neq \hat{h}(t)\}$ is a finite union of disjoint open intervals and $h_n$ is linear on each of the intervals. For all random variables $Y$ with $F_Y \in \mathcal{M}$, let random variable $Z_{\mathcal{I}_1^n} \sim F_Y^{\mathcal{I}_1^n}$. Similar to (3.35), we have

$$\rho_{h_n}(Y) = \rho_{\hat{h}}(Z_{\mathcal{I}_1^n}) \leqslant \sup_{F_X\in\mathcal{M}}\rho_{\hat{h}}(X), \quad \text{for all } n \in \mathbb{N}.$$

Note that $h_n(t) \uparrow h^*(t)$ as $n \to \infty$ for all $t \in (0,1)$. By the monotone convergence theorem, we get $\rho_{h_n}(Y) \to \rho_{h^*}(Y)$ as $n \to \infty$. It follows that

$$\sup_{F_X\in\mathcal{M}}\rho_{\hat{h}}(X) \geqslant \rho_{h_n}(Y) \xrightarrow{n\to\infty} \rho_{h^*}(Y).$$

2. If $h \in \mathcal{H}$ has countably many discontinuity points, it suffices to prove (3.38). There exist series of finite sets $\{\hat{J}^m\}_{m\in\mathbb{N}} \subseteq \hat{J}$, such that $\hat{J}^m \to \hat{J}$ as $m \to \infty$. For all $m \in \mathbb{N}$, write

$$\hat{h}_m(t) = \begin{cases} \hat{h}(t), & t \in \hat{J}^m,\\ h(t), & \text{otherwise,}\end{cases}$$

and define

$$\hat{J}^m_+ = \{t \in \hat{J}^m : \hat{h}_m(t) = \hat{h}_m(t^+)\}, \quad \text{and} \quad \hat{J}^m_- = \hat{J}^m \setminus \hat{J}^m_+.$$

For $n>0$, let $\mathcal{I}_2^{n,m}=\{B_s^{n,m}: i\in \hat{J}^m\}$ with

$$B_s^{n,m}=\begin{cases}(1-s-1/\sqrt{n},1-s+1/n), & s\in\hat{J}_-^m,\\ (1-s-1/n,1-s+1/\sqrt{n}), & s\in\hat{J}_+^m.\end{cases}$$

Following the same argument as (3.38), for all random variable $Y$ with $F_Y\in\mathcal{M}$, we have

$$\sup_{F_X\in\mathcal{M}}\rho_h(X)\geqslant\rho_h(Z_{\mathcal{I}_2^{n,m}})\xrightarrow{n\to\infty}\rho_{\hat{h}_m}(Y),\quad\text{for all } m\in\mathbb{N},$$

where $Z_{\mathcal{I}_2^{n,m}}\sim F_Y^{\mathcal{I}_2^{n,m}}$. Moreover, we have $\hat{h}_m(t)\uparrow\hat{h}(t)$ for all $t\in[0,1]$ as $m\to\infty$. By the monotone convergence theorem, we have $\rho_{\hat{h}_m}(Y)\to\rho_{\hat{h}}(Y)$ as $m\to\infty$. Therefore, we have

$$\sup_{F_X\in\mathcal{M}}\rho_{\hat{h}}(X)\leqslant\sup_{F_X\in\mathcal{M}}\rho_h(X).$$

Proof of (iii): For all $h\in\mathcal{H}$, if $\mathcal{M}$ is closed under concentration within $\mathcal{I}_h$ and $h=\hat{h}$, we have $F_Y^{\mathcal{I}_h}\in\mathcal{M}$ by definition. Since $Z_{\mathcal{I}_h}\sim F_Y^{\mathcal{I}_h}$, (3.37) gives

$$\rho_{h^*}(Y)=\rho_{\hat{h}}(Z_{\mathcal{I}_h})=\rho_h(Z_{\mathcal{I}_h}).$$

Note that $\rho_h\leqslant\rho_{h^*}$ generally. Therefore, if $\max_{F_Y\in\mathcal{M}}\rho_{h^*}(Y)$ is attained by $F_Y$, then so is $\max_{F_Y\in\mathcal{M}}\rho_h(Y)$ by $F_Y^{\mathcal{I}_h}$. Obviously, these two quantities share a common maximizer $F_Y^{\mathcal{I}_h}$ because

$$\rho_{h^*}(Z_{\mathcal{I}_h})\leqslant\max_{F_Y\in\mathcal{M}}\rho_{h^*}(Y)=\max_{F_Y\in\mathcal{M}}\rho_h(Y)=\rho_h(Z_{\mathcal{I}_h})\leqslant\rho_{h^*}(Z_{\mathcal{I}_h}).$$

The proof is complete. □

*Proof of Theorem 3.2.* Suppose for contradiction that $\mathcal{M}_{\text{opt}}$ is not closed under concentration within $\mathcal{I}_h$. There exists $F_Y\in\mathcal{M}_{\text{opt}}$, such that $F_Y^{\mathcal{I}_h}\notin\mathcal{M}_{\text{opt}}$. Define the set $\mathcal{Y}_h=\{(F_Y^{-1}(a),F_Y^{-1}(b)):(a,b)\in\mathcal{I}_h\}$. Since $F_Y^{\mathcal{I}_h}\notin\mathcal{M}_{\text{opt}}$, there exists an interval $(a,b)\in\mathcal{I}_h$, such that $F_Y^{-1}$ is not constant on $(a,b)$. Thus the Lebesgue measure $\lambda((F_Y^{-1}(a),F_Y^{-1}(b)))>0$. Since $h^*>h$ on $(a,b)$,

$$\begin{aligned}\rho_{h^*}(Y)-\rho_h(Y)&=\int_{\mathbb{R}}(h^*(\mathbb{P}(Y>x))-h(\mathbb{P}(Y>x)))\,\mathrm{d}x\\&=\sum_{A\in\mathcal{Y}_h}\int_A(h^*(\mathbb{P}(Y>x))-h(\mathbb{P}(Y>x)))\,\mathrm{d}x>0.\end{aligned}\tag{3.39}$$

On the other hand, we have

$$\rho_{h^*}(Y)\leqslant\sup_{F_X\in\mathcal{M}}\rho_{h^*}(X)=\sup_{F_X\in\mathcal{M}}\rho_h(X)=\rho_h(Y)\leqslant\rho_{h^*}(Y),$$

which leads to a contradiction to (3.39). Therefore, $\mathcal{M}_{\text{opt}}$ is closed under concentration within $\mathcal{I}_h$. □

*Proof of Proposition 3.2.* We first prove that closedness under conditional expectation implies closedness under concentration for all intervals. For all random variables $Y \in \mathcal{L}^1$ and intervals $C \subseteq [0,1]$, define

$$X = F_Y^{-1}(U)\mathbb{1}_{\{U \notin C\}} + \mathbb{E}[F_Y^{-1}(U)|U \in C]\mathbb{1}_{\{U \in C\}},$$

where $U \sim \mathrm{U}[0,1]$. The distribution of $X$ is the concentration $F_Y^C$. For all $\sigma(X)$-measurable random variables $Z$, we have that $Z|\{U \in C\}$ is constant. Hence,

$$\begin{aligned}
\mathbb{E}[XZ] &= \mathbb{E}[ZF_Y^{-1}(U)\mathbb{1}_{\{U \notin C\}} + Z\mathbb{E}[F_Y^{-1}(U)|U \in C]\mathbb{1}_{\{U \in C\}}] \\
&= \mathbb{E}[ZF_Y^{-1}(U)\mathbb{1}_{\{U \notin C\}}] + \mathbb{E}[\mathbb{E}[ZF_Y^{-1}(U)|U \in C]\mathbb{1}_{\{U \in C\}}] \\
&= \mathbb{E}[ZF_Y^{-1}(U)\mathbb{1}_{\{U \notin C\}}] + \mathbb{E}[ZF_Y^{-1}(U)|U \in C]\mathbb{P}(U \in C) \\
&= \mathbb{E}[ZF_Y^{-1}(U)\mathbb{1}_{\{U \notin C\}}] + \mathbb{E}[ZF_Y^{-1}(U)\mathbb{1}_{\{U \in C\}}] = \mathbb{E}[ZF_Y^{-1}(U)].
\end{aligned}$$

It follows that $\mathbb{E}[Y|X] = \mathbb{E}[F_Y^{-1}(U)|X] = X$, $\mathbb{P}$-almost surely. If a set of distributions, $\mathcal{M}$, is closed under conditional expectation and $F_Y \in \mathcal{M}$, then $F_{\mathbb{E}[Y|X]} \in \mathcal{M}$, which implies that $F_Y^C = F_X \in \mathcal{M}$. Thus $\mathcal{M}$ is also closed under concentration for all intervals. $\square$

*Proof of Proposition 3.3.* (i) Suppose that $\mathcal{M}$ is closed under concentration for all intervals and $\mathcal{I}$ is a finite. Using (3.6), we can see that $F^{\mathcal{I}}$ is the resulting distribution obtained by sequentially applying finitely many $C$-concentrations to $F$ over all $C \in \mathcal{I}$. We thus have $F^{\mathcal{I}} \in \mathcal{M}$ for all $F \in \mathcal{M}$.

(ii) Suppose that $\mathcal{M}$ is closed under conditional expectation and $F \in \mathcal{M}$. We define

$$X = F^{-1}(U)\mathbb{1}_{\{U \notin \bigcup_{C \in \mathcal{I}} C\}} + \sum_{C \in \mathcal{I}} \mathbb{E}[F^{-1}(U)|U \in C]\mathbb{1}_{\{U \in C\}},$$

whose left-quantile function is given by (3.7) according to (3.6). Following similar argument to the proof of Proposition 3.2, for all $\sigma(X)$-measurable random variables $Z$, we have

$$\begin{aligned}
\mathbb{E}[XZ] &= \mathbb{E}[ZF^{-1}(U)\mathbb{1}_{\{U \notin \bigcup_{C \in \mathcal{I}} C\}} + \sum_{C \in \mathcal{I}} Z\mathbb{E}[F^{-1}(U)|U \in C]\mathbb{1}_{\{U \in C\}}] \\
&= \mathbb{E}[ZF^{-1}(U)\mathbb{1}_{\{U \notin \bigcup_{C \in \mathcal{I}} C\}}] + \sum_{C \in \mathcal{I}} \mathbb{E}[\mathbb{E}[ZF^{-1}(U)|U \in C]\mathbb{1}_{\{U \in C\}}] \\
&= \mathbb{E}[ZF^{-1}(U)\mathbb{1}_{\{U \notin \bigcup_{C \in \mathcal{I}} C\}}] + \sum_{C \in \mathcal{I}} \mathbb{E}[ZF^{-1}(U)\mathbb{1}_{\{U \in C\}}] = \mathbb{E}[ZF^{-1}(U)].
\end{aligned}$$

Thus $\mathbb{E}[F^{-1}(U)|X] = X$, $\mathbb{P}$-almost surely, which implies that $F^{\mathcal{I}} = F_X \in \mathcal{M}$. $\square$

### 3.10.3 Proofs of results in Section 3.4

*Proof of Theorem 3.3.* To prove the first statement, according to the proof of Theorem 3.1, it suffices to show that for all increasing $h \in \mathcal{H}$, $\mathbf{X} \in (\mathcal{L}^1)^n$ and $\mathscr{G} \subseteq \mathscr{F}$, $\rho_h(\mathbb{E}[f(\mathbf{a},\mathbf{X})|\mathscr{G}]) \leqslant \rho_h(f(\mathbf{a},\mathbb{E}[\mathbf{X}|\mathscr{G}]))$, which holds directly by Jensen's inequality and monotonicity of $\rho_h$. The second statement holds by Theorem 3.1. The last statement follows from $\rho_h(\mathbb{E}[f(\mathbf{a},\mathbf{X})|\mathscr{G}]) = \rho_h(f(\mathbf{a},\mathbb{E}[\mathbf{X}|\mathscr{G}]))$ and using Theorem 3.1. $\square$

*Proof of Theorem 3.4.* (i) For all $\mathbf{X} = (X_1,\dots,X_n) \in (\mathcal{L}^1)^n$, take a comonotonic $\widetilde{\mathbf{X}} = (\widetilde{X}_1,\dots,\widetilde{X}_n) \in (\mathcal{L}^1)^n$ such that $\widetilde{X}_i \overset{\mathrm{d}}{=} X_i$ for all $i = 1,\dots,n$. It follows that $\mathbb{E}[g(\mathbf{X})] \leqslant \mathbb{E}[g(\widetilde{\mathbf{X}})]$ for all supermodular functions $g : \mathbb{R}^n \to \mathbb{R}$ due to Theorem 5 of Tchen (1980). By Proposition 2.2.5 of Simchi-Levi et al. (2005), we have $f(\mathbf{a},\mathbf{X}) \leqslant_{\mathrm{icx}} f(\mathbf{a},\widetilde{\mathbf{X}})$. Moreover, there exists a standard uniform random variable $U$ such that $\widetilde{X}_i = F^{-1}_{\widetilde{X}_i}(U)$ for all $i = 1,\dots,n$ and $f(\mathbf{a},\widetilde{\mathbf{X}}) = F^{-1}_{f(\mathbf{a},\widetilde{\mathbf{X}})}(U)$ almost surely (Denneberg, 1994). Take

$$f(\mathbf{a},\widetilde{\mathbf{X}})^{\mathcal{I}_h} = F^{-1}_{f(\mathbf{a},\widetilde{\mathbf{X}})}(U)\mathbb{1}_{\{U\notin\bigcup_{C\in\mathcal{I}_h} C\}} + \sum_{C\in\mathcal{I}_h} \mathbb{E}[F^{-1}_{f(\mathbf{a},\widetilde{\mathbf{X}})}(U)|U\in C]\mathbb{1}_{\{U\in C\}} \sim F^{\mathcal{I}_h}_{f(\mathbf{a},\widetilde{\mathbf{X}})}.$$

It follows that $f(\mathbf{a},\widetilde{\mathbf{X}})^{\mathcal{I}_h} = \mathbb{E}[f(\mathbf{a},\widetilde{\mathbf{X}})|\mathscr{G}]$, where $\mathscr{G} = \sigma(U\mathbb{1}_{\{U\notin\bigcup_{C\in\mathcal{I}_h} C\}})$. Similarly, $\widetilde{X}_i^{\mathcal{I}_h} = \mathbb{E}[\widetilde{X}_i|\mathscr{G}]$ for all $i = 1,\dots,n$, where

$$\widetilde{X}_i^{\mathcal{I}_h} = F^{-1}_{\widetilde{X}_i}(U)\mathbb{1}_{\{U\notin\bigcup_{C\in\mathcal{I}_h} C\}} + \sum_{C\in\mathcal{I}_h} \mathbb{E}[F^{-1}_{\widetilde{X}_i}(U)|U\in C]\mathbb{1}_{\{U\in C\}} \sim F^{\mathcal{I}_h}_{\widetilde{X}_i}.$$

Since $f$ is supermodular and positively homogeneous, we have by Theorem 3 of Marinacci and Montrucchio (2008) that $f(\mathbf{a},\mathbf{X})$ is concave in $\mathbf{X}$. By Jensen's inequality, we have

$$f(\mathbf{a},\widetilde{\mathbf{X}})^{\mathcal{I}_h} = \mathbb{E}[f(\mathbf{a},\widetilde{\mathbf{X}})|\mathscr{G}] \leqslant f(\mathbf{a},\mathbb{E}[\widetilde{\mathbf{X}}|\mathscr{G}]) = f(\mathbf{a},\widetilde{X}_1^{\mathcal{I}_h},\dots,\widetilde{X}_n^{\mathcal{I}_h}).$$

Thus we have

$$\begin{aligned}\rho_{h^*}(f(\mathbf{a},\mathbf{X})) \leqslant \rho_{h^*}(f(\mathbf{a},\widetilde{\mathbf{X}})) = \rho_h(f(\mathbf{a},\widetilde{\mathbf{X}})^{\mathcal{I}_h}) &\leqslant \rho_h(f(\mathbf{a},\widetilde{X}_1^{\mathcal{I}_h},\dots,\widetilde{X}_n^{\mathcal{I}_h}))\\ &\leqslant \sup_{F_1\in\mathcal{F}_1,\dots,F_n\in\mathcal{F}_n}\ \sup_{F_{\mathbf{Y}}\in\mathcal{D}(F_1,\dots,F_n)} \rho_h(f(\mathbf{a},\mathbf{Y})),\end{aligned}$$

where the first inequality follows from Theorem 4.A.3 of Shaked and Shanthikumar (2007) and Theorem 5 of Chapter 2 and the second equality is by the proof of Theorem 3.1. Combined with the fact that

$$\sup_{F_1\in\mathcal{F}_1,\dots,F_n\in\mathcal{F}_n}\ \sup_{F_{\mathbf{X}}\in\mathcal{D}(F_1,\dots,F_n)} \rho_h(f(\mathbf{a},\mathbf{X})) \leqslant \sup_{F_1\in\mathcal{F}_1,\dots,F_n\in\mathcal{F}_n}\ \sup_{F_{\mathbf{X}}\in\mathcal{D}(F_1,\dots,F_n)} \rho_{h^*}(f(\mathbf{a},\mathbf{X})),$$

we have (3.16) holds.

(ii) Suppose that the supremum of the right-hand side of (3.16) is attained by some $F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n$ and $F_{\mathbf{X}} \in \mathcal{D}(F_1, \dots, F_n)$. For comonotonic $(\widetilde{X}_1, \dots, \widetilde{X}_n)$ such that $\widetilde{X}_i \sim F_i$ for all $i = 1, \dots, n$, using the argument in (i),

$$\rho_{h^*}(f(\mathbf{a}, \mathbf{X})) \leqslant \rho_h(f(\mathbf{a}, \widetilde{X}_1^{\mathcal{I}_h}, \dots, \widetilde{X}_n^{\mathcal{I}_h})),$$

where $(\widetilde{X}_1^{\mathcal{I}_h}, \dots, \widetilde{X}_n^{\mathcal{I}_h})$ is comonotonic and $\widetilde{X}_i^{\mathcal{I}_h} \sim F_i^{\mathcal{I}_h}$ for all $i = 1, \dots, n$. Similarly to the proof of Theorem 3.1 (iii), since $\rho_h \leqslant \rho_{h^*}$, we have the supremum of the left-hand side of (3.16) is attained by $F_1^{\mathcal{I}_h}, \dots, F_n^{\mathcal{I}_h}$ and $(\widetilde{X}_1^{\mathcal{I}_h}, \dots, \widetilde{X}_n^{\mathcal{I}_h})$, which also obtain the supremum of the right-hand side of (3.16) since

$$\begin{aligned}
\rho_{h^*}(f(\mathbf{a}, \widetilde{X}_1^{\mathcal{I}_h}, \dots, \widetilde{X}_n^{\mathcal{I}_h})) &\leqslant \max_{F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n} \max_{F_{\mathbf{X}} \in \mathcal{D}(F_1, \dots, F_n)} \rho_{h^*}(f(\mathbf{a}, \mathbf{X})) \\
&= \max_{F_1 \in \mathcal{F}_1, \dots, F_n \in \mathcal{F}_n} \max_{F_{\mathbf{X}} \in \mathcal{D}(F_1, \dots, F_n)} \rho_h(f(\mathbf{a}, \mathbf{X})) \\
&= \rho_h(f(\mathbf{a}, \widetilde{X}_1^{\mathcal{I}_h}, \dots, \widetilde{X}_n^{\mathcal{I}_h})) \leqslant \rho_{h^*}(f(\mathbf{a}, \widetilde{X}_1^{\mathcal{I}_h}, \dots, \widetilde{X}_n^{\mathcal{I}_h})).
\end{aligned}$$

□

### 3.10.4 Proofs of results in Section 3.5 and related lemmas

In the following, we write $q$ as the Hölder conjugate of $p$. The following lemma closely resembles Theorem 3.4 of Liu et al. (2020) with only an additional statement on the uniqueness of the quantile function of the maximizer.

**Lemma 3.1.** *For $h \in \mathcal{H}^*$, $m \in \mathbb{R}$, $v > 0$ and $p > 1$, we have*

$$\sup_{F_Y \in \mathcal{M}(p,m,v)} \rho_h(Y) = mh(1) + v[h]_q,$$

*If $0 < [h]_q < \infty$, the above supremum is attained by a random variable $X$ such that $F_X \in \mathcal{M}(p, m, v)$ with its quantile function uniquely determined by*

$$\mathrm{VaR}_t(X) = m + v\phi_h^q(t), \quad t \in (0,1) \quad a.e. \tag{3.40}$$

*If $[h]_q = 0$, the above maximum value is attained by any random variable $X$ such that $F_X \in \mathcal{M}(p, m, v)$.*

*Proof.* The only statement that is more than Theorem 3.4 of Liu et al. (2020) is the uniqueness of the quantile function (3.40). Without loss of generality, assume $m = 0$ and $v = 1$. Using the Hölder inequality

$$\sup_{F_Y\in\mathcal{M}(p,0,1)}\int_0^1 h'(t)\mathrm{VaR}_{1-t}(Y)\,\mathrm{d}t = \sup_{F_Y\in\mathcal{M}(p,0,1)}\int_0^1 (h'(t)-c_{h,q})\mathrm{VaR}_{1-t}(Y)\,\mathrm{d}t$$

$$\leqslant \sup_{F_Y\in\mathcal{M}(p,0,1)} \|h'-c_{h,q}\|_q\left(\int_0^1 |\mathrm{VaR}_{1-t}(Y)|^p\,\mathrm{d}t\right)^{1/p} = [h]_q.$$

The maximum is attained by $F_X$ only if the above inequality is an equality, which is equivalent to that the function $t\mapsto|\mathrm{VaR}_{1-t}(X)|^p$ is a multiple of $|h'-c_{h,q}|^q$. Therefore,

$$\mathrm{VaR}_t(X) = \frac{|h'(1-t)-c_{h,q}|^q}{h'(1-t)-c_{h,q}}[h]_q^{1-q} = \phi_h^q(t),\quad t\in(0,1)\ \text{ a.e.}$$

Hence, the quantile function of $X$ is uniquely determined by (3.40). □

**Lemma 3.2.** *For all $h\in\mathcal{H}$ with $h=\hat{h}$, $m\in\mathbb{R}$, $v>0$ and $p>1$, if $[h^*]_q<\infty$, we have*

$$\sup_{F_Y\in\mathcal{M}(p,m,v)}\rho_h(Y) = \sup_{F_Y\in\mathcal{M}(p,m,v)}\rho_{h^*}(Y) = mh(1)+v[h^*]_q,$$

*and the above suprema are simultaneously attained by a random variable $X$ such that $F_X\in\mathcal{M}(p,m,v)$ with*

$$\mathrm{VaR}_t(X) = m+v\phi_{h^*}^q(t),\quad t\in(0,1)\ a.e. \tag{3.41}$$

*Proof.* The statement directly follows from Theorem 3.1 and Lemma 3.1. □

*Proof of Theorem 3.5.* Together with Theorem 3.1, Lemmas 3.1 and 3.2 give the statement in Theorem 3.5 on the supremum. Arguments for the infimum are symmetric. For instance, noting that $(-h)^*=-h_*$, Theorem 3.1 yields

$$\begin{aligned}\inf_{F_Y\in\mathcal{M}(p,m,v)}\rho_h(Y) &= -\sup_{F_Y\in\mathcal{M}(p,m,v)}\rho_{-h}(Y)\\ &= -\sup_{F_Y\in\mathcal{M}(p,m,v)}\rho_{(-h)^*}(Y)\\ &= -\sup_{F_Y\in\mathcal{M}(p,m,v)}\rho_{-h_*}(Y) = \inf_{F_Y\in\mathcal{M}(p,m,v)}\rho_{h_*}(Y).\end{aligned}$$

We omit the detailed arguments for the infimum in Theorem 3.5. □

*Proof of Proposition 3.5.* Note that $\rho_h\leqslant\rho_{h^*}$, which is implied by $h\leqslant h^*$ and (1.3). By Hölder's inequality, for any $Y\in\mathcal{L}^p$, using (3.12), we have

$$\begin{aligned}\int_0^1 h^{*\prime}(t)\mathrm{VaR}_{1-t}(Y)\,\mathrm{d}t &= \int_0^1 (h^{*\prime}(t)-c_{h^*,q})\mathrm{VaR}_{1-t}(Y)\,\mathrm{d}t + c_{h,q}\mathbb{E}[Y]\\ &\leqslant [h^*]_q\|Y\|_p + c_{h^*,q}\mathbb{E}[Y]<\infty.\end{aligned}$$

The other half of the statement is analogous. □

*Proof of Corollary 3.1.* We prove the first half (the suprema). The second half is symmetric to the first half. Theorem 3.5 and Lemma 3.2 give

$$\sup_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{VaR}_\alpha(Y) = \sup_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{ES}_\alpha(Y) = m + v[h^*]_q.$$

By Lemma 3.1, the corresponding random variable $Z$ which attains $\mathrm{ES}_\alpha(Z) = m + v[h^*]_q$ has left-quantile function

$$F_Z^{-1}(t) = m + v\phi_{h^*}^q(t) = m + v\frac{\left|\frac{1}{1-\alpha}\mathbb{1}_{(\alpha,1]}(t) - 1\right|^q}{\frac{1}{1-\alpha}\mathbb{1}_{(\alpha,1]}(t) - 1}[h^*]_q^{1-q}, \quad t \in [0,1] \text{ a.e.}$$

Note that $\phi_{h^*}^q(t)$ only takes two values for $t \geqslant \alpha$ and $t < \alpha$, respectively. Thus $Z$ is a bi-atomic random variable, and using $\mathbb{E}[Z] = m$, we have, for some $k_p > 0$,

$$\mathbb{P}\left(Z = m + \alpha k_p\right) = 1 - \alpha \text{ and } \mathbb{P}\left(Z = m - (1-\alpha)k_p\right) = \alpha.$$

We note that the number $k_p$ can be determined from $\mathbb{E}[|Z - m|^p] = v^p$, that is,

$$k_p = v\left(\alpha^p(1-\alpha) + (1-\alpha)^p\alpha\right)^{-1/p},$$

leading to

$$\sup_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{VaR}_\alpha(Y) = \sup_{F_Y \in \mathcal{M}(p,m,v)} \mathrm{ES}_\alpha(Y) = m + v\alpha\left(\alpha^p(1-\alpha) + (1-\alpha)^p\alpha\right)^{-1/p},$$

and thus the desired equalities in the statement on suprema hold. □

# Chapter 4

# Bayes risk, elicitability, and the Expected Shortfall

## 4.1 Introduction

Mainly through Gneiting (2011), the concept of elicitability has drawn considerable interest within the quantitative risk management literature. The concept is fundamental when comparing different forecasting procedures. We refer to the latter paper for an excellent introduction. In this Introduction we explain our motivation, and more detailed definitions are given in Section 4.2.

Let $\mathcal{X}$ be a linear space of random variables. A set-valued $d$-dimensional functional $\mathcal{S} : \mathcal{X} \to 2^{\mathbb{R}^d}$ is *elicitable* if there exists a measurable function $L : \mathbb{R}^{d+1} \to \mathbb{R}$ (called a loss function) such that

$$\mathcal{S}(X) = \underset{x \in \mathbb{R}^d}{\arg\min}\, \mathbb{E}[L(x, X)], \quad X \in \mathcal{X}. \tag{4.1}$$

In the recent literature, a lot of research has been done on characterizing risk measures that are elicitable for $d = 1$.[1] The case of $d \geqslant 2$ is much more difficult; see Fissler and Ziegel (2016) and Wang and Wei (2020). In sharp contrast to the functional $\mathcal{S}$ in (4.1), much less attention has been paid to functionals $\mathcal{R}$ of the form

$$\mathcal{R}(X) = \min_{x \in \mathbb{R}^d} \mathbb{E}[L(x, X)], \quad X \in \mathcal{X}, \tag{4.2}$$

[1]For characterization of elicitability in dimension $d = 1$, see Ziegel (2016) on coherent risk measures, Bellini and Bignozzi (2015) and Delbaen et al. (2016) on convex risk measures, Kou and Peng (2016) and Wang and Ziegel (2015) on Choquet risk measures, and Liu and Wang (2021) on tail risk measures.

which computes the *Bayes risk* of $X$ for the *Bayes estimator* $\mathcal{S}$. As an important example for $d = 1$, the Value-at-Risk (VaR) at level $\alpha$ is elicited by the loss function $L : (x, y) \mapsto x + \frac{1}{1-\alpha}(y - x)_+$, and the Expected Shortfall (ES) at level $\alpha$ is the corresponding Bayes risk (e.g., Rockafellar and Uryasev, 2002); see (4.4) in Section 4.2.

The functionals in (4.1) and (4.2) are both important in the context of expected loss minimization, and yet only the characterization of (4.1) is extensively studied in the literature of elicitability. In case $d = 1$, the functional $\mathcal{S}$ in (4.1) is a *minimizer* and the functional $\mathcal{R}$ in (4.2) is a *minimum*. The classic interpretation of risk measures, as in Artzner et al. (1999), is the *least amount of capital* needed for a financial loss to be acceptable for the regulator. In other words, an acceptable capital reserve needs to be no smaller than the value of the risk measure. With this interpretation, requiring for a capital reserve to be larger than the minimizer $\mathcal{S}$ does not have a clear financial meaning. On the other hand, the minimum in the Bayes risk $\mathcal{R}$ may be interpreted as a "generalized $L$-distance" from $X$ to the real line,[2] so that the corresponding capital reserve may be interpreted as a penalty for deviating from constancy, thus for bearing risk.[3]

The main focus of this chapter is $(\mathcal{S}, \mathcal{R})$ in (4.1) and (4.2), which we call a Bayes pair. After the formal definition of Bayes pairs and Bayes risk measures in Section 4.2, we derive two main characterization results. In Theorem 4.2 we show that, under a continuity assumption, an ES is the only Bayes risk measure that is either coherent or Choquet, and in Theorem 4.4 we pin down entropic risk measures as the only monetary risk measures which are both elicitable and Bayes.

Currently, ES is the standard risk measure in the Fundamental Review of the Trading Book (BCBS, 2019) in banking. Our characterization of ES as the only coherent Bayes risk measure strengthens the unique role of ES from the perspective of elicitability. This result complements the recent finding of Wang and Zitikis (2021) on an axiomatic characterization of ES from the perspective of portfolio risk aggregation. See also Emmer et al. (2015) and Embrechts et al. (2018) for discussions on comparative advantages of VaR and ES as regulatory risk measures.

[2]For instance, in the simple case $L(x, y) = (x - y)^2$, the Bayes risk $\min_{x\in\mathbb{R}} \mathbb{E}[(X - x)^2] = \mathrm{var}(X)$, which is the squared $\mathcal{L}^2$-distance from $X$ to the real line. If $\mathcal{R}$ is used as a regulatory risk measure, we typically need to adjust the value by the location of $X$ (that is why we call it a "generalized $L$-distance"), e.g., using $L(x, y) = x + \lambda(x - y)^2$ for $\lambda > 0$ would give rise to a mean-variance risk measure. The minimum in (4.2) should not be interpreted as a minimum over economic scenarios; indeed, it is more natural to take a maximum over economic scenarios.

[3]We do acknowledge that elicitability is an important statistical property, especially when comparing competing forecast procedures; see e.g., Fissler and Ziegel (2016). This chapter stresses the important practical difference, even complementarity, between regulatory interpretation (Bayes) and statistical tractability (elicitability).

Our main technical results are closely related to those in Weber (2006), Ziegel (2016) and Wang and Wei (2020) on risk measures with convex level sets, those in Ben-Tal and Teboulle (2007) on optimized certainty equivalents, those in Rockafellar and Uryasev (2013) on risk quadrangles, and those in Frongillo and Kash (2021) on elicitation complexity. More general discussions on elicitability and forecasting risk measures can be found in Davis (2013, 2016) and Nolde and Ziegel (2017). Although we do focus on risk measures, the general theory of elicitability has wide applications outside of finance. For some recent work on interval-valued elicitable functionals, see Fissler et al. (2020) and Brehmer and Gneiting (2021). Elicitability is also closely related to empirical risk minimization; see e.g., Lambert et al. (2008), Steinwart et al. (2014) and Frongillo and Kash (2021) in the context of machine learning.

## 4.2 Bayes pairs and Bayes risk measures

### 4.2.1 Risk measures

In Definition 4.1 below, we slightly generalize the standard definition of scalar risk measures in Artzner et al. (1999) and Föllmer and Schied (2002a) to interval-valued risk measures such as quantiles. In what follows, equalities and inequalities between intervals are understood as holding for both end-points, and so are addition and scalar multiplication. Let $\mathcal{X}$ be a linear space of random variables containing $\mathcal{L}^\infty$, representing the domain of risk measures. Let $I(\mathbb{R})$ be the set of closed real intervals, including $(-\infty, a]$ and $[a, \infty)$, and the interval $[a, a]$ is identified with its element $a$ (hence, $\mathbb{R}$ is treated as a subset of $I(\mathbb{R})$). Below we define some terminology for interval-valued risk measures, different from those in Chapter 1.

**Definition 4.1.** A *risk measure* $\mathcal{S}$ is a mapping from $\mathcal{X}$ to $I(\mathbb{R})$, and it is *scalar* if it maps $\mathcal{X}$ to $\mathbb{R}$. A risk measure $\mathcal{S}$ is *monetary* if it is (i) monotone: $\mathcal{S}(X) \leqslant \mathcal{S}(Y)$ for $X, Y \in \mathcal{X}$ with $X \leqslant Y$, and (ii) translation invariant: $\mathcal{S}(X + c) = \mathcal{S}(X) + c$ for all $X \in \mathcal{X}$ and $c \in \mathbb{R}$. A scalar risk measure $\mathcal{S}$ is *coherent* if it is monetary, (iii) convex: $\mathcal{S}(\lambda X + (1-\lambda)Y) \leqslant \lambda\mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y)$ for all $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$, and (iv) positively homogeneous: $\mathcal{S}(\lambda X) = \lambda\mathcal{S}(X)$ for all $X \in \mathcal{X}$ and $\lambda \in (0, \infty)$. A scalar risk measure $\mathcal{S}$ is *Choquet* if it is monetary and (v) comonotonic-additive: $\mathcal{S}(X + Y) = \mathcal{S}(X) + \mathcal{S}(Y)$ for all comonotonic $X, Y \in \mathcal{X}$.

It was shown by Schmeidler (1986) that comonotonic-additivity characterizes Choquet integrals, and hence we use the name Choquet risk measure. Law-invariant Choquet risk measures are also called *distortion risk measures*; Theorem 1 of Wang et al. (2020) gives a characterization of law-invariant comonotonic-additive functionals. Both coherence and comonotonic-additivity are argued as desirable properties, and a law-invariant risk measure that is both coherent and Choquet is called a spectral risk measure by Acerbi (2002).

We make an important clarification concerning Definition 4.1, which also justifies our focus on risk measures in (4.2). It is well known that with positive homogeneity, convexity is equivalent to subadditivity, which is not easy to financially interpret if $\mathcal{S}$ is interval-valued. In view of this, convexity and coherence are suitable properties for risk measures in (4.2), but it is unclear whether they are suitable for risk measures in (4.1), unless one additionally assumes uniqueness of the optimizer, or some convention (e.g., using the left end-point) is imposed.

The most important examples of risk measures are VaR and ES, widely used in financial regulation. At a probability level $\alpha \in [0,1]$, VaR has two versions, the left- and right-quantiles. Note that we use different notations from (1.1) exclusively in this chapter and define $\mathrm{VaR}_\alpha : \mathcal{X} \to I(\mathbb{R})$ by $\mathrm{VaR}_\alpha(X) = [\mathrm{VaR}^-_\alpha(X), \mathrm{VaR}^+_\alpha(X)]$, where

$$\begin{cases} \mathrm{VaR}^-_\alpha(X) &= \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant \alpha\}; \\ \mathrm{VaR}^+_\alpha(X) &= \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) > \alpha\}. \end{cases} \tag{4.3}$$

By definition, $\mathrm{VaR}^-_0 = -\infty$ and $\mathrm{VaR}^+_1 = \infty$. The interval-valued risk measure $\mathrm{VaR}_\alpha$ is monetary. With sightly different notations from (1.2), the ES (also called CVaR, TVaR and AVaR) at a probability level $\alpha \in [0,1)$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}^-_\beta(X)\,\mathrm{d}\beta, \quad X \in \mathcal{X}.$$

It is well known that an ES is both coherent and Choquet. Rockafellar and Uryasev (2002) obtained the following ES-VaR relation (4.4)

$$\begin{aligned} [\mathrm{VaR}^-_\alpha(X), \mathrm{VaR}^+_\alpha(X)] &= \underset{x\in\mathbb{R}}{\arg\min}\left\{x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+]\right\}; \\ \mathrm{ES}_\alpha(X) &= \min_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+]\right\}. \end{aligned} \tag{4.4}$$

The relation (4.4) will be used repeatedly in this chapter.

### 4.2.2 Bayes pairs and Bayes risk measures

To define the main objects of the chapter, we follow the standard terminology of Bayes estimator and Bayes risk in statistical decision theory. Despite this terminology, our discussion stays purely within the theory of risk measures, and does not require specific knowledge on Bayesian statistics to understand.

**Definition 4.2.** A pair of risk measures $(\mathcal{S}, \mathcal{R}) : \mathcal{X} \to I(\mathbb{R}) \times \mathbb{R}$ is a *Bayes pair* if for some Borel function $L : \mathbb{R}^2 \to \mathbb{R}$, called the *loss function*,

$$\mathcal{S}(X) = \operatorname*{arg\,min}_{x \in \mathbb{R}} \mathbb{E}[L(x, X)] \quad \text{and} \quad \mathcal{R}(X) = \min_{x \in \mathbb{R}} \mathbb{E}[L(x, X)], \quad X \in \mathcal{X}. \tag{4.5}$$

If $\mathcal{S}$ is further translation invariant, then we call $\mathcal{S}$ a *Bayes estimator*, and $\mathcal{R}$ a *Bayes risk measure.*[4]

In a Bayes pair $(\mathcal{S}, \mathcal{R})$, $\mathcal{R}$ is always scalar, whereas $\mathcal{S}$ not necessarily. Therefore, it is appropriate to consider conditions for $\mathcal{R}$, instead of $\mathcal{S}$, to be a coherent risk measure. By (4.4), for $\alpha \in [0, 1)$, the pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ is a Bayes pair, and $\mathrm{ES}_\alpha$ is a coherent Bayes risk measure. Obviously, a Bayes pair is always law invariant.

Let us first explain the important requirement of $\mathcal{S}$ being translation invariant in Definition 4.2. In the next theorem, we show a negative result: if we do not impose any conditions on $\mathcal{S}$, then the interpretation of Bayes estimator and Bayes risk is lost.

**Theorem 4.1.** *A risk measure $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ satisfies (4.5) for some $\mathcal{S} : \mathcal{X} \to 2^{\mathbb{R}}$ and loss function $L$ if and only if there exists a set $\mathcal{A}$ of real Borel functions such that*

$$\mathcal{R}(X) = \min_{\ell \in \mathcal{A}} \mathbb{E}[\ell(X)], \quad X \in \mathcal{X}. \tag{4.6}$$

*Proof.* The $\Rightarrow$ implication follows directly by setting $\mathcal{A} = \{y \mapsto L(x, y) : x \in \mathbb{R}\}$. Next we show the $\Leftarrow$ implication. Let $\phi$ be a one-to-one mapping from $\mathbb{R}$ to the set of real Borel functions on $\mathbb{R}$ (since both sets have the same cardinality), and define $L(x, y) = \phi(x)(y)$ for $x, y \in \mathbb{R}$. Let

$$\mathcal{S}(X) = \phi^{-1}\left(\operatorname*{arg\,min}_{\ell \in \mathcal{A}} \mathbb{E}[\ell(X)]\right) = \operatorname*{arg\,min}_{x \in \mathbb{R}} \mathbb{E}[L(x, X)]. \tag{4.7}$$

Hence, (4.5) holds. $\square$

The negative result in Theorem 4.1 is very simple, but it is important for the motivation behind the concept of Bayes risk measures as in Definition 4.2. If no property is imposed on $\mathcal{S}$, then we can directly define $\mathcal{R}$ by (4.6) without introducing $\mathcal{S}$. However, this would be problematic because the Bayes estimator $\mathcal{S}$ is not interpretable as there is nothing to estimate. A similar problem appears

[4]$\mathcal{S}$ is also called a Bayes act; see e.g., Grünwald and Dawid (2004). When we say that a risk measure is Bayes in this chapter, we mean that it is a Bayes risk measure (instead of a Bayes estimator).

in Frongillo and Kash (2021) when they define elicitation complexity.[5] In the context of Bayes estimation, both $\mathcal{S}$ and $\mathcal{R}$ have a concrete meaning: $\mathcal{S}$ is the estimated parameter and $\mathcal{R}$ is the risk of this estimation. Therefore, directly defining a risk measure $\mathcal{R}$ by (4.6) cannot be called a Bayes risk measure because it is not the Bayes risk of any interpretable parameter. For this reason, we impose translation invariance on $\mathcal{S}$, which means that the parameter of interest of the unknown financial loss is additive under location shift. This is similar to the consideration of Artzner et al. (1999), where location shift is interpreted as capital injection. Other types of regularization on $\mathcal{S}$ may also be considered, among which translation invariance seems both natural and easy to work with. See Examples 4.2 and 4.3 in Section 4.3 for instances of $\mathcal{R}$ in (4.5) where translation invariance of $\mathcal{S}$ is not assumed.

For a Bayes pair $(\mathcal{S}, \mathcal{R})$ with loss function $L$, by defining

$$\mathcal{R}' : X \mapsto \lambda \mathcal{R}(X) + (1-\lambda)\mathbb{E}[f(X)] \tag{4.8}$$

for any real function $f$ and $\lambda \in (0, 1]$, the pair $(\mathcal{S}, \mathcal{R}')$ is also a Bayes pair with loss function $L' : (x, y) \mapsto \lambda L(x, y) + (1-\lambda)f(y)$. Hence, some conditions on $\mathcal{R}$ also need to be imposed to obtain an economically meaningful risk measure. For this, we have plenty of candidates in the literature, notably in the theories of coherent and Choquet risk measures.

Some advantages of Bayes risk measures follow from the definition and results in this chapter, and we briefly summarize them below. Bayes risk measures are (i) convenient to optimize due to their form as a minimizer to a linear mapping on distributions;[6] (ii) concave in mixtures and thus correctly measuring randomness (see Section 4.5); (iii) relatively easy to evaluate forecasts due to their second-order elicitability (Corollary 4.1); (iv) relatively easy to compute due to their low elicitation complexity (Frongillo and Kash, 2021), which is at most 2.[7]

[5] Frongillo and Kash (2021) argue that, through a one-to-one mapping from $\mathbb{R}$ to the set of real Borel functions on $\mathbb{R}$ like $\phi$ in the proof of Theorem 4.1, one arrives at a counter-intuitive statement that all functionals have elicitation complexity 1. Hence, some regularity requirements are needed.

[6] A typical optimization problem is to minimize $\mathcal{R}(f(a, Y))$ over $a \in A$ where $A$ is a set of actions, $Y$ is a random vector, and $f : A \times \mathbb{R}^d \to \mathbb{R}$ is a function; this includes the classic problem of portfolio selection with risk measures. If $\mathcal{R}$ is a Bayes risk measure with loss function $L$, then via the relationship $\min_a \mathcal{R}(f(a, Y)) = \min_x \min_a \mathbb{E}[L(x, f(a, Y))]$, the above optimization problem can be solved by first minimizing an expected loss over $a \in A$, which is well studied. See Rockafellar and Uryasev (2013) and the references therein for optimizing risk measures of the form (4.2).

[7] Roughly speaking, the elicitation complexity of a functional $\mathcal{R}$ is the lowest dimension of $\mathcal{R}'$ such that (i) $\mathcal{R}'$ is elicitable; (ii) $\mathcal{R}$ is determined by $\mathcal{R}'$; (iii) $\mathcal{R}'$ satisfies some regularity conditions. We omit a detailed definition in this chapter since some heavy preparation is needed for a proper definition of elicitation complexity. The interested reader is referred to Frongillo and Kash (2021).

### 4.2.3 Examples of Bayes pairs

We present some common examples of Bayes pairs. Except for the ES/$\mathbb{E}$-mixture, none of the other Bayes risk measures in Example 4.1 are coherent risk measures; this gives a hint on the unique role of ES/$\mathbb{E}$-mixtures among coherent Bayes risk measures.

**Example 4.1.** In all examples below, $\mathcal{S}$ in a Bayes pair $(\mathcal{S},\mathcal{R})$ is translation invariant, and hence $\mathcal{R}$ is a Bayes risk measure in Definition 4.2.

(i) $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha^\lambda)$: As we have seen from (4.4), for $\alpha \in [0,1)$, $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ is a Bayes pair with loss function $L:(x,y)\mapsto x+\frac{1}{1-\alpha}(y-x)_+$, and $\mathrm{ES}_\alpha$ is a coherent Bayes risk measure. Moreover, using (4.8), the convex combination of $\mathrm{ES}_\alpha$ and $\mathbb{E}$, called an ES/$\mathbb{E}$-*mixture* and denoted by $\mathrm{ES}_\alpha^\lambda$, i.e.,

$$\mathrm{ES}_\alpha^\lambda = \lambda \mathrm{ES}_\alpha + (1-\lambda)\mathbb{E}, \quad \lambda \in [0,1],\ \alpha \in (0,1), \tag{4.9}$$

is a coherent Bayes risk measure with loss function $L:(x,y)\mapsto x+\lambda(y-x)+\frac{1-\lambda}{1-\alpha}(y-x)_+$. Note that $\lambda=0$ corresponds to the mean, and $\lambda=1$ corresponds to $\mathrm{ES}_\alpha$.

(ii) $(\mathrm{ER}_\gamma, \mathrm{ER}_\gamma)$: An entropic risk measure (ER) is defined as

$$\mathrm{ER}_\gamma(X) = \frac{1}{\gamma}\log \mathbb{E}[e^{\gamma X}], \quad X \in \mathcal{L}^\infty,$$

for $\gamma \in (0,\infty)$, with the limiting case $\mathrm{ER}_0 = \mathbb{E}$.[8] The entropic risk measure $\mathrm{ER}_\gamma$ is known to be convex but not coherent. Next we see that $\mathrm{ER}_\gamma$ is both Bayes and elicitable for the same loss function $L:(x,y)\mapsto x+(e^{\gamma(y-x)}-1)/\gamma$. Indeed, by defining

$$\mathcal{R}(X) := \min_{x\in\mathbb{R}} \mathbb{E}[L(x,X)] = \min_{x\in\mathbb{R}} \left\{ x + \frac{1}{\gamma}\mathbb{E}[e^{\gamma(X-x)}-1] \right\}, \tag{4.10}$$

one can verify that the minimizer of (4.10) is $\mathcal{S}(X) = \frac{1}{\gamma}\log\mathbb{E}[e^{\gamma X}] = \mathrm{ER}_\gamma(X)$. Substituting it into (4.10), we have

$$\mathcal{R}(X) = \frac{1}{\gamma}\log\mathbb{E}[e^{\gamma X}] + \frac{\mathbb{E}[e^{\gamma X}]}{\mathbb{E}[\gamma e^{\gamma X}]} - \frac{1}{\gamma} = \frac{1}{\gamma}\log\mathbb{E}[e^{\gamma X}] = \mathrm{ER}_\gamma(X).$$

(iii) $(\mathbb{E}, \sigma^2)$: The variance

$$\sigma^2(X) := \mathbb{E}[(X-\mathbb{E}[X])^2] = \min_{x\in\mathbb{R}} \mathbb{E}[(X-x)^2], \quad X \in \mathcal{L}^2,$$

is a Bayes risk measure with loss function $L:(x,y)\mapsto(y-x)^2$. The corresponding Bayes estimator is the mean $\mathbb{E}$.

[8] The domain of $\mathrm{ER}_\gamma$ can be enlarged to include random variables with finite exponential moments, such as normal random variables.

(iv) $(\mathbb{E}, \mathbb{E} + \lambda\sigma^2)$: The mean-variance functional, for $\lambda > 0$,

$$\mathbb{E}[X] + \lambda\sigma^2(X), \quad X \in \mathcal{L}^2,$$

is a Bayes risk measure with loss function $L : (x, y) \mapsto y + \lambda(y - x)^2$. The Bayes estimator corresponding to the mean-variance functional is also the mean $\mathbb{E}$.

(v) $(\mathrm{VaR}_{1/2}, \mathrm{MD})$: The mean-median deviation,

$$\mathrm{MD}(X) := \min_{x \in \mathbb{R}} \mathbb{E}[|X - x|], \quad X \in \mathcal{L}^1,$$

is a Bayes risk measure with loss function $L : (x, y) \mapsto |y - x|$, and the corresponding Bayes estimator is the median (interval) $\mathrm{VaR}_{1/2}(X)$. The mean-median deviation is a signed Choquet integral with distortion function $h(t) = \min\{t, 1 - t\}$, $t \in [0, 1]$; see Wang et al. (2020).

(vi) $(\mathrm{ex}_\alpha, \mathrm{var}_\alpha)$: The variantile (e.g., Wang and Wei, 2020),

$$\mathrm{var}_\alpha(X) := \min_{x \in \mathbb{R}} \left\{ \alpha\, \mathbb{E}[(X - x)_+^2] + (1 - \alpha)\, \mathbb{E}[(X - x)_-^2] \right\}, \quad X \in \mathcal{L}^2,$$

where $\alpha \in (0, 1)$, is a Bayes risk measure with loss function

$$L : (x, y) \mapsto \alpha(y - x)_+^2 + (1 - \alpha)(y - x)_-^2.$$

The Bayes estimator corresponding to the variantile $\mathrm{var}_\alpha$ is the *expectile* at the level $\alpha$, denoted by $\mathrm{ex}_\alpha$; see Bellini et al. (2014) and Ziegel (2016).

## 4.3 Characterizing ES as a Bayes risk measure

We will present below our first main result on the characterization of Bayes risk measures. Recall that the ES/$\mathbb{E}$-mixtures in (4.9) of Example 4.1 are coherent and Choquet Bayes risk measures. Theorem 4.2 below further shows that they are the only possible class of Bayes risk measures which are either coherent or Choquet. This result is illustrated by the Venn diagram in Figure 4.1.

Below, lower semicontinuity is defined with respect to almost sure convergence. This form of lower semicontinuity is used to formulate the *prudence axiom* of Wang and Zitikis (2021), and the interpretation is that a consistent statistical approximation of the true risk should not underestimate the risk measure.

**Theorem 4.2.** *Suppose that $\mathcal{L}^\infty \subseteq \mathcal{X} \subseteq \mathcal{L}^1$. For a risk measure $\mathcal{R} : \mathcal{X} \to \mathbb{R}$, the following are equivalent:*

*(i)* $\mathcal{R}$ *is a coherent Bayes risk measure;*

*(ii)* $\mathcal{R}$ *is a Choquet Bayes risk measure;*

*(iii)* $\mathcal{R} = \mathrm{ES}_\alpha^\lambda$ *for some* $\alpha \in (0,1)$ *and* $\lambda \in [0,1]$.

*If* $\mathcal{R}$ *further satisfies lower semicontinuity, then* $\mathcal{R} = \mathrm{ES}_\alpha$ *for some* $\alpha \in (0,1)$.

*Proof.* The full proof is presented in Section 4.8.1, and we give a sketch of the main steps here. The implication (iii)⇒(ii) is obvious. The implication (ii)⇒(i) is implied by Proposition 4.3 in Section 4.5. Below are the main steps for the most important implication (i)⇒(iii). Assume $(\mathcal{S}, \mathcal{R})$ is a Bayes pair in which $\mathcal{S}$ is translation invariant and $\mathcal{R}$ is coherent.

We first show in Lemma 4.1, that, using the fact that $\mathcal{S}$ and $\mathcal{R}$ are both translation invariant, we can choose a loss function for $(\mathcal{S}, \mathcal{R})$ in the form $(x, y) \mapsto x + v(y - x)$ for some real function $v$. Thus, we have

$$\mathcal{R}(X) = \inf_{c \in \mathbb{R}} \{c + \mathbb{E}[v(X - c)]\}, \quad X \in \mathcal{X}. \tag{4.11}$$

Using the monotonicity of $\mathcal{R}$, we proceed to show in Lemma 4.2 that such $v$ can be replaced by an increasing function $\tilde{v}$ without changing $\mathcal{R}$. Next, using the convexity of $\mathcal{R}$, we show in Lemma 4.3 that $\tilde{v}$ can be replaced by an increasing convex function $\hat{v}$. Using the positive homogeneity of $\mathcal{R}$, in Lemma 4.5 we show that $\hat{v}$ can be replaced by the piece-wise linear function $\bar{v}(x) = \lambda x + (\gamma - \lambda) x_+$ for some $\gamma \geqslant 1$ and $\lambda \in [0,1]$. Finally, with the above loss function, we derive $\mathcal{R} = \mathrm{ES}_\alpha^{1-\lambda}$ where $\alpha = (\gamma - 1)/(\gamma - \lambda) \in (0,1)$.

For the last statement of the theorem, the lower semicontinuity of $\mathrm{ES}_\alpha^{1-\lambda}$ implies $\lambda = 0$ since $\mathrm{ES}_\alpha$ is lower semicontinuous and $\mathbb{E}$ is not, as implied by Theorem 1 of Wang and Zitikis (2021). □

*Remark* 4.1. As we see in the proof of Theorem 4.2, a key step is to show that a translation-invariant Bayes risk measure $\mathcal{R}$ has the form (4.11) in Lemma 4.1. Risk measures directly defined via the form (4.11) have appeared in the literature, and we make two notable connections.

1. The *optimized certainty equivalent* (OCE) of Ben-Tal and Teboulle (2007) has the form (4.11) where $v$ is increasing, convex, and satisfying $v(0) = 0$ and $v'(0+) \geqslant 1$; here it is adapted to our convention that a positive value of $X$ represents a loss. Theorem 3.1 of Ben-Tal and Teboulle (2007), which is closely related to Theorem 4.2, states that, assuming that $v$ is real-valued, increasing, convex, $v(0) = 0$, $v'(0+) > 0$, and $v(x) > x$ for all $x \neq 0$, the only coherent risk

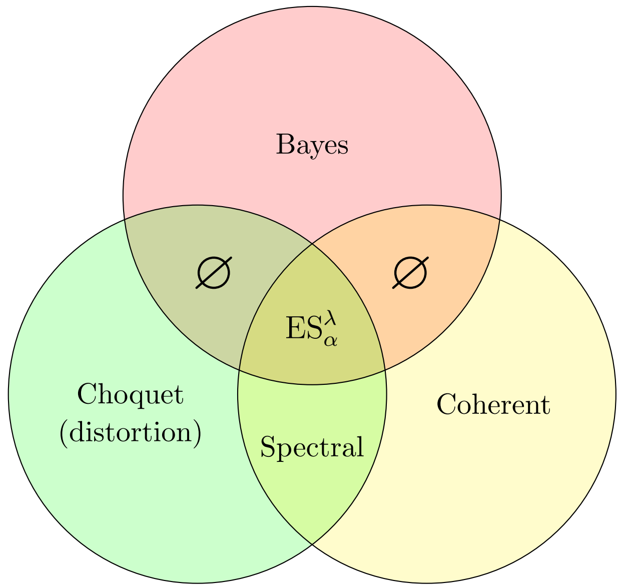


Figure 4.1: A Venn diagram for three classes of law-invariant risk measures

measure in the OCE class is generated by $v(x) = \lambda x + (\gamma - \lambda)x_+$ for some $\infty > \gamma > 1 > \lambda \geqslant 0$, which is an ES/$\mathbb{E}$-mixture, similar to Lemma 4.5 except for the boundary cases of $\gamma = \infty$, $\gamma = 1$ and $\lambda = 1$. Different from Ben-Tal and Teboulle (2007), all our assumptions are made on $\mathcal{R}$ and not on the form of $v$ in (4.11).

2. The form (4.11) also appears in the expectation quadrangle of Rockafellar and Uryasev (2013), where $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ also serves as an important example. Our choice of notation, especially $\mathcal{S}$ and $\mathcal{R}$, is consistent with the notation of Rockafellar and Uryasev (2013). Nevertheless, our interpretation of the Bayes pair and our focus on characterization are different from their framework. See also the recent paper Chong et al. (2021) where (4.11) appears as an optimized objective in the context of capital allocation.

*Remark* 4.2. Using Lemmas 4.2 and 4.3 in Section 4.8.1, we also obtain the forms of monetary and convex Bayes risk measures. A risk measure $\mathcal{R}$ is a monetary (resp. monetary and convex) Bayes risk measure if and only if (4.11) holds for some increasing (resp. increasing convex) function $v : \mathbb{R} \to \mathbb{R}$.

Below we further present two examples of a coherent risk measure $\mathcal{R}$ satisfying (4.5), but the corresponding minimizer $\mathcal{S}$ is not translation invariant (indeed, not interpretable). By Theorem

4.1, these risk measures have the form (4.6), for a set of loss functions $\mathcal{A}$,

$$\mathcal{R}(X) = \min_{\ell \in \mathcal{A}} \mathbb{E}[\ell(X)], \quad X \in \mathcal{X}.$$

The examples also show that the assumption of translation invariance on $\mathcal{S}$ in Definition 4.2 is essential for the characterization of the Bayes risk measures in Theorem 4.2.

**Example 4.2.** The first example is a convex combination of ES at different levels. Define a risk measure $\mathcal{R} = \frac{1}{2}\mathrm{ES}_\alpha + \frac{1}{2}\mathrm{ES}_\beta$ for some distinct numbers $\alpha, \beta \in (0,1)$. Clearly, $\mathcal{R}$ is a coherent risk measure. By (4.4), for $X \in \mathcal{L}^1$,

$$\begin{aligned}\mathcal{R}(X) &= \frac{1}{2}\min_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+]\right\} + \frac{1}{2}\min_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\beta}\mathbb{E}[(X-x)_+]\right\} \\ &= \frac{1}{2}\min_{x_1,x_2\in\mathbb{R}}\left\{x_1 + x_2 + \frac{1}{1-\alpha}\mathbb{E}[(X-x_1)_+] + \frac{1}{1-\beta}\mathbb{E}[(X-x_2)_+]\right\}.\end{aligned}$$

By Theorem 4.1, $\mathcal{R}$ satisfies (4.5) for some $\mathcal{S}$. Since $\mathcal{R}$ is not an ES/$\mathbb{E}$-mixture, by Theorem 4.2, $\mathcal{R}$ is not a Bayes risk measure. This implies that any minimizer $\mathcal{S}$ satisfying (4.5) is not translation invariant. We can also see from this example that $\mathcal{S}(X)$ should be a one-to-one function of the minimizer $(x_1, x_2)$ above, which is difficult to interpret in a financial context (one-to-one mappings from $\mathbb{R}^2$ to $\mathbb{R}$ are usually quite strange).

**Example 4.3.** The second example is the coherent entropic risk measure introduced by Föllmer and Knispel (2011), defined, for some $c > 0$, as

$$\mathcal{R}(X) = \min_{\gamma>0}\left\{\frac{1}{\gamma}\log\mathbb{E}[e^{\gamma X}] + \frac{c}{\gamma}\right\}, \qquad X \in \mathcal{L}^\infty.$$

Föllmer and Knispel (2011) showed that $\mathcal{R}$ is a coherent risk measure; it satisfies (4.5) by Theorem 4.1. Since $\mathcal{R}$ is not an ES/$\mathbb{E}$-mixture, by Theorem 4.2, $\mathcal{R}$ is not a Bayes risk measure.

Before ending this section, we show that the Bayes pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha^\lambda)$ can be characterized if $\mathcal{R}$ is coherent or Choquet. This result slightly generalizes Theorem 4.2 which only gives the form of $\mathcal{R}$ but not that of $\mathcal{S}$. A proof of Proposition 4.1 is put in Section 4.8.2.

**Proposition 4.1.** *For a Bayes pair $(\mathcal{S}, \mathcal{R})$ with loss function $L$, the following are equivalent:*

*(i) $\mathcal{S}(0) = 0$ and $\mathcal{R}$ is a coherent Bayes risk measure;*

*(ii) $\mathcal{S}(0) = 0$ and $\mathcal{R}$ is a Choquet Bayes risk measure;*

*(iii) $(\mathcal{S}, \mathcal{R}) = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha^\lambda)$ for some $\alpha \in (0,1)$ and $\lambda \in [0,1)$;*

(iv) *the loss function can be chosen as* $L : (x, y) \mapsto x + (1 - \lambda)(y - x) + \frac{\lambda}{1-\alpha}(y - x)_+$ *for some* $\alpha \in (0, 1)$ *and* $\lambda \in (0, 1]$.

For a given $\alpha \in (0, 1)$, Wang and Wei (2020, Theorem 6.9) showed that an ES/$\mathbb{E}$-mixture is the only coherent Choquet risk measure $\rho$ such that $(\rho, \mathrm{VaR}_\alpha)$ is elicitable. This result does not imply, and is not implied by, Theorem 4.2 and Proposition 4.1, although the similarity is visible.

## 4.4 Elicitability of Bayes risk measures

In this section we study the connection between Bayes pairs and elicitability. Recall that the functional $\mathcal{S}$ is elicitable if there exists a loss function $L : \mathbb{R}^{d+1} \to \mathbb{R}$ such that

$$\mathcal{S}(X) = \operatorname*{arg\,min}_{y \in \mathbb{R}^d} \mathbb{E}[L(y, X)], \quad X \in \mathcal{X}. \tag{4.12}$$

The first observation is that a Bayes pair $(\mathcal{S}, \mathcal{R})$ is always elicitable. This was essentially shown in Theorem 1 of Frongillo and Kash (2021) where $\mathcal{S}$ takes scalar values. We present a similar proof which is adapted to our slightly different definitions.

**Theorem 4.3.** *Any Bayes pair* $(\mathcal{S}, \mathcal{R})$ *with loss function* $L$ *is elicitable by*

$$L^*(x, y, z) = \int_0^y h(t)\,\mathrm{d}t + h(y)(L(x, z) - y), \quad (x, y) \in D, \ z \in \mathbb{R},$$

*where* $h$ *is any positive and strictly decreasing function on* $\mathbb{R}$ *and* $D$ *is the range of* $(\mathcal{S}, \mathcal{R})$.

*Proof.* We need to show that $L^*$ elicits $(\mathcal{S}, \mathcal{R})$; that is, for $X \in \mathcal{X}$,

$$(\mathcal{S}, \mathcal{R})(X) = \operatorname*{arg\,min}_{(x,y) \in D} \mathbb{E}[L^*(x, y, X)] = \operatorname*{arg\,min}_{(x,y) \in D} \left\{ \int_0^y h(t)\,\mathrm{d}t + h(y)(\mathbb{E}[L(x, X)] - y) \right\}. \tag{4.13}$$

First, for a fixed $(y, X)$, the minimizers $x^*$ to (4.13) are the same as the minimizers of $\mathbb{E}[L(x, X)]$. Therefore, we know that the set of minimizers $x^*$ are precisely $\mathcal{S}(X)$, and $\mathbb{E}[L(x^*, X)] = \mathcal{R}(X)$. Next, we need to find the minimizers for

$$\operatorname*{arg\,min}_{y \in \mathbb{R}} \left\{ \int_0^y h(t)\,\mathrm{d}t + h(y)(\mathcal{R}(X) - y) \right\},$$

which gives $y^* = \mathcal{R}(X)$ since $h$ is a strictly decreasing function. ☐

*Remark* 4.3. In Theorem 4.3, the loss function which elicits $(\mathcal{S}, \mathcal{R})$ is not unique. For instance, if $\mathcal{S}(X)$ is itself elicited by a loss function $L'$, then $(x, y, z) \mapsto L^*(x, y, z) + L'(x, z)$ also elicits $(\mathcal{S}, \mathcal{R})$.

Following the terminology of Emmer et al. (2015) and Fissler and Ziegel (2016), a functional $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is *second-order elicitable* if it is a component of a 2-dimensional elicitable functional, and it is *conditionally elicitable* on another functional $\mathcal{S}$ if for each $r \in \mathbb{R}$ and some loss function $L_r$, we have

$$\mathcal{R}(X) = \underset{y \in \mathbb{R}^d}{\arg\min}\, \mathbb{E}[L_r(y, X)], \quad X \in \mathcal{X} \text{ with } \mathcal{S}(X) = r.$$

Theorem 4.3 immediately yields that any Bayes risk measure $\mathcal{R}$ is second-order elicitable as a component of the Bayes pair $(\mathcal{S}, \mathcal{R})$ and conditionally elicitable on $\mathcal{S}$ via $L_r : (y, z) \mapsto L^*(r, y, z)$.

**Corollary 4.1.** *Any Bayes risk measure is second-order elicitable and conditionally elicitable.*

*Remark* 4.4. Another direct consequence of Theorem 4.3 is that any Bayes risk measure has elicitation complexity of at most 2 (implied by second-order elicitability), and hence they are relatively simple to estimate via empirical risk minimization; see Frongillo and Kash (2021) for a precise definition and related discussions.

## 4.5 Other properties of Bayes risk measures

The following two results of a Bayes risk measure do not require monotonicity. Let $\mathcal{M}$ be the set of distributions of the elements in $\mathcal{X}$. For any scalar law-invariant risk measure $\mathcal{R}$, we write $\widehat{\mathcal{R}} : F \mapsto \mathcal{R}(X)$ where $X \sim F \in \mathcal{M}$. Thus, $\widehat{\mathcal{R}}$ represents the risk measure $\mathcal{R}$ treated as a mapping from $\mathcal{M}$ to $\mathbb{R}$. We say that $\mathcal{R}$ has *convex level sets (CxLS)* if the set $\{F \in \mathcal{M} : \widehat{\mathcal{R}}(F) = r\}$ is convex for each $r \in \mathbb{R}$. Mixture concavity represents that using a mixture of models (i.e., introducing a stochastic factor) increases randomness, and it is a desirable property for both risk and deviation measures. Moreover, for Choquet risk measures, mixture concavity is equivalent to coherence (Theorem 3 of Wang et al. (2020)). The CxLS property is a necessary condition for elicitability (Osband, 1985) and has been widely studied in the risk measure literature (e.g., Weber, 2006; Ziegel, 2016; Delbaen et al., 2016; Wang and Wei, 2020). The following two properties are useful in the proofs of Theorems 4.2 and 4.4. Moreover, Proposition 4.2 directly inspires the study in Section 6.6.

**Proposition 4.2.** *A Bayes risk measure is necessarily mixture concave, and a Bayes estimator necessarily has CxLS.*[9]

[9] Mixture concavity of risk measures is as defined in Chapter 2.

*Proof.* By definition, $\widehat{\mathcal{R}} : F \mapsto \inf_{x\in\mathbb{R}}\{\int_{\mathbb{R}} L(x,y)\,\mathrm{d}F(y)\}$ is the infimum of linear functions on $\mathcal{M}$, and hence concave. Thus, $\mathcal{R}$ is mixture concave. The second statement is due to the fact that any Bayes estimator is elicitable, and it is well known that elicitable functionals have CxLS (e.g., Ziegel, 2016). $\square$

**Proposition 4.3.** *A Choquet Bayes risk measure is necessarily coherent.*

*Proof.* Note that a monetary risk measure is uniformly continuous with respect to the $L^\infty$-norm. Using Theorem 1 of Wang et al. (2020), a law-invariant, uniformly $L^\infty$-continuous and comonotonic-additive functional admits a representation as a Choquet integral. Theorem 3 of Wang et al. (2020) further implies that mixture concavity is equivalent to convexity. Therefore, as Choquet risk measures are automatically positively homogeneous, $\mathcal{R}$ is coherent. $\square$

## 4.6 Elicitable Bayes risk measures

Any Bayes risk measure is mixture concave (Proposition 4.2), and any elicitable risk measure, such as the Bayes estimator, has CxLS. We wonder what is the intersection of the two classes of risk measures. This question is not only driven by mathematical curiosity, but also has interesting connections with some classical results in decision theory.

As we have seen above, the mean is both elicitable (with loss function $L(x,y) = (y-x)^2$) and Bayes (with loss function $L(x,y) = x + (y-x)_+$). Moreover, the entropic risk measure ER in Example 4.1 is mixture concave and has CxLS, since it is both elicitable and Bayes. The next result shows that ER is the only risk measure that is mixture concave and has CxLS under the following continuity assumption (recall that $\widehat{\mathcal{R}}(F) = \mathcal{R}(X)$ where $X \sim F$)

(C) For any $x < y$, the mapping $\alpha \mapsto \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y)$ on $[0,1]$ is continuous at $\alpha = 0$, where $\delta_z$ is the point-mass at $z \in \mathbb{R}$.

Clearly, continuity (C) is weaker than continuity from above.

**Theorem 4.4.** *Let $\mathcal{R}$ be a law-invariant monetary risk measure on $\mathcal{X} = \mathcal{L}^\infty$ satisfying continuity (C) with $\mathcal{R}(0) = 0$. Then $\mathcal{R}$ is mixture concave and has CxLS if and only if it is an entropic risk measure.*

The proof of Theorem 4.4 is technical and put in Section 4.8.3. Below we illustrate some intuition of this result by connecting mixture concavity and CxLS to the notions of betweenness (Chew, 1983) and associativity (Grant et al., 2000) in decision theory.[10] We say that $\mathcal{R}$ satisfies *associativity* if for any $F, G, H \in \mathcal{M}$ and $\lambda \in (0,1)$,

$$\widehat{\mathcal{R}}(F) = \widehat{\mathcal{R}}(G) \implies \widehat{\mathcal{R}}(\lambda F + (1-\lambda)H) = \widehat{\mathcal{R}}(\lambda G + (1-\lambda)H). \tag{4.14}$$

Lemma 2 of Grant et al. (2000) shows that associativity holds under the assumption of a suitable continuity condition, mixture concavity and betweenness. The betweenness property is slightly stronger than the CxLS property, but they are equivalent under some mild assumption (e.g., Lemma 14 of Steinwart et al., 2014). If $\widehat{\mathcal{R}}$ satisfies associativity, then by the de Finetti-Kolmogorov-Nagumo Theorem (see e.g., Cifarelli and Regazzini, 1996), $\mathcal{R}$ is a certainty equivalent, that is, there exists a continuous and strictly increasing function $u : \mathbb{R} \to \mathbb{R}$ such that $\mathcal{R}(X) = u^{-1}(\mathbb{E}[u(X)])$, $X \in \mathcal{X}$. Finally, by translation-invariance of $\mathcal{R}$, one can conclude that $u(x) = e^{cx}$ for $c > 0$ or $u(x) = x$, $x \in \mathbb{R}$. As a consequence, $\mathcal{R}$ must be an entropic risk measure. The main gap in the above informal argument is to verify the conclusion of Lemma 2 of Grant et al. (2000) under CxLS and (C), which is a complicated mathematical task although intuitively clear. In Section 4.8.3, we provide a full proof without using the results of Grant et al. (2000).

*Remark* 4.5. Continuity (C) is not satisfied by the essential supremum $X \mapsto \mathrm{VaR}_1^-(X)$, which is mixture concave and has CxLS. In our proof of Theorem 4.4, the continuity condition (C) is essential and we were not able to relax it. Nevertheless, we conjecture that by including $\mathrm{VaR}_1^- = \mathrm{ER}_\infty$ as an extended member of the ER family, one may remove or weaken (C) in Theorem 4.4.

A consequence of Theorem 4.4 is that a risk measure $\mathcal{R}$ with the form

$$\mathcal{R}(X) = \inf\{x \in \mathbb{R} : \mathbb{E}[g(X - x)] \leqslant z\} \tag{4.15}$$

for a strictly increasing $g$ cannot be mixture concave unless it is an entropic risk measure. In particular, this implies that expectiles defined in Example 4.1 (vi) are not mixture concave. This fact is shown by Bellini et al. (2018), and it is (surprisingly) not easy to directly verify.

**Corollary 4.2.** *Let $\mathcal{R}$ be defined by (4.15) for some increasing function $g$ and constant $z$ satisfying $g(-t) < z < g(t)$ for all $t > 0$. Then $\mathcal{R}$ is mixture concave if and only if it is an entropic risk measure.*

[10] We thank an anonymous referee for brining up this connection.

*Proof.* It is clear that the risk measure $\mathcal{R}$ defined by (4.15) is monetary. Note that the condition $g(-t) < z < g(t)$ for all $t > 0$ implies $\mathcal{R}(0) = 0$. Lemma 4.7 guarantees (C) from the above condition on $g$. The rest follows by applying Theorem 4.4. □

Finally, we obtain a characterization of entropic risk measures as the intersection of Bayes estimators and Bayes risk measures. Moreover, as we see in Example 4.1 (ii), an entropic risk measure is a Bayes estimator and a Bayes risk measure with the same loss function.

**Corollary 4.3.** *A monetary risk measure $\mathcal{R}$ with $\mathcal{R}(0) = 0$ is elicitable and Bayes if and only if it is an entropic risk measure.*

*Proof.* Note that an elicitable risk measure satisfies CxLS and a Bayes risk measure satisfies mixture concavity. Using Theorem 4.4, it suffices to verify that a Bayes risk measure $\mathcal{R}(X) = \min_x \mathbb{E}[L(x, X)]$ satisfies continuity (C). That is, for any $x < y$, the function

$$\alpha \mapsto H(\alpha) := \min_{s\in\mathbb{R}}\{(1-\alpha)L(s,x) + \alpha L(s,y)\}$$

is continuous at $\alpha = 0$. Note that there exists $s_0$ such that $L(s_0, x) = x$. Then

$$x \leqslant \liminf_{\alpha\downarrow 0} H(\alpha) \leqslant \limsup_{\alpha\downarrow 0} H(\alpha) \leqslant \lim_{\alpha\downarrow 0}\{(1-\alpha)L(s_0,x) + \alpha L(s_0,y)\} = L(s_0,x) = x.$$

Hence, we have $H(\alpha)$ is continuous at $\alpha = 0$, which gives the desired condition (C). □

## 4.7 Concluding remarks

In this chapter, we introduce the concepts of Bayes pairs and Bayes risk measures, and offer some characterization results. In particular, Theorem 4.2 yields a new characterization of ES in the context of statistical inference and optimization, complementing the ES characterization of Wang and Zitikis (2021) based on portfolio risk aggregation.

It is known that entropic risk measures are the only dynamically consistent law-invariant risk measures (Kupper and Schachermayer, 2009), and they are also the only intersection of the class of optimized certainty equivalents (OCE) and the class of shortfall risk measures (Ben-Tal and Teboulle, 2007; Föllmer and Schied, 2016). Theorem 4.4 further shows that, under a continuity assumption, the entropic risk measures are the only monetary risk measures satisfying mixture concavity (a property of the OCE) and CxLS (a property of the shortfall risk measures).

Bayes risk measures are closely related to elicitability, and they are second-order elicitable (Theorem 4.3). There are several open questions on the theory of Bayes pairs and Bayes risk measures which will be explored in the future; we discuss a few of them here.

The first question is regarding the special role of Bayes pairs among elicitable two-dimensional functionals. Almost all examples of elicitable two-dimensional functionals $(\mathcal{S}, \mathcal{R})$ in the literature are one-to-one transforms of either a Bayes pair, such as those in Example 4.1, or a pair whose components are both elicitable, such as $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta)$, $(\mathrm{ex}_\alpha, \mathrm{ex}_\beta)$, or the modal interval (see Brehmer and Gneiting, 2021). We wonder under what conditions an elicitable two-dimensional functional has to be obtained from a Bayes pair.

Next, we focus on the Bayes risk measure $\mathcal{R}$. We say that a risk measure is genuinely 2-elicitable if it is second-order elicitable but not elicitable. Since coherent risk measures are not elicitable except for the expectiles (Ziegel, 2016), it is natural to study the class of genuinely 2-elicitable coherent risk measures. A non-elicitable Bayes risk measure is genuinely 2-elicitable (see Corollary 4.1), but the converse is not true; it is unclear what special role Bayes risk measures play among genuinely 2-elicitable risk measures.

There are at least two very different ways to construct a genuinely 2-elicitable coherent risk measure. The first is to combine two elicitable risk measures, such as a mixture of two different expectiles, $(1-\lambda)\mathrm{ex}_\alpha + \lambda \mathrm{ex}_\beta$, and the second is to use a Bayes risk measure, such as an ES/$\mathbb{E}$-mixture. We conjecture that a coherent Choquet risk measure is genuinely 2-elicitable if and only if it is an ES/$\mathbb{E}$-mixture (except for the mean). We also wonder under what conditions, a genuinely 2-elicitable coherent risk measure has to be a mixture of two expectiles.

Finally, if we replace translation invariance of $\mathcal{S}$ in Definition 4.2 with another property, the characterization in Theorem 4.2 may fail to hold, as we see in Examples 4.2 and 4.3. A full characterization of coherent $\mathcal{R}$ without translation invariance of $\mathcal{S}$ is open at the moment. A similar question arises in a setting where $\mathcal{S}$ is allowed to be multi-dimensional; these questions are planned for future research.

## 4.8 Lemmas and proofs of several results

### 4.8.1 Lemmas in the proof of Theorem 4.2

Below we derive a few lemmas which lead to the proof of Theorem 4.2, implication (i)⇒(iii). The other implications are already shown in the proof sketch in Section 4.3. In all lemmas below, $\mathcal{X}$ is a linear space satisfying $\mathcal{L}^\infty \subseteq \mathcal{X} \subseteq \mathcal{L}^1$.

**Lemma 4.1.** *Suppose that $(\mathcal{S}, \mathcal{R})$ is a Bayes pair with loss function $L$, and $\mathcal{S}$ and $\mathcal{R}$ are translation invariant. The function $(x, y) \mapsto x + L(0, y - x)$ is also a loss function for $(\mathcal{S}, \mathcal{R})$. As consequence, there exists a real function $v$ such that*

$$\mathcal{R}(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X - x)]\},\ X \in \mathcal{X}.$$

*Proof.* Define $L^*(x, y) = x + L(0, y - x),\ (x, y) \in \mathbb{R}^2$, and

$$\mathcal{R}^*(X) = \min_{x\in\mathbb{R}} \mathbb{E}[L^*(x, X)], \quad X \in \mathcal{X}.$$

We aim to show $\mathcal{R}^* = \mathcal{R}$. For a random variable $X \in \mathcal{X}$, denote by $\mathcal{S}^*$ the left end-point of $\mathcal{S}$, that is, $\mathcal{S}^*(X) = \min\{\arg\min_{x\in\mathbb{R}} \mathbb{E}[L(x, X)]\}$, $X \in \mathcal{X}$. By translation invariance of $\mathcal{S}$, we have $\mathcal{S}^*(X+c) = \mathcal{S}^*(X)+c$ for $c \in \mathbb{R}$. Then by translation invariance of $\mathcal{R}$ we have $\mathcal{R}(X) = c+\mathcal{R}(X-c)$, that is,

$$\min_{x\in\mathbb{R}} \mathbb{E}[L(x, X)] = c + \min_{x\in\mathbb{R}} \mathbb{E}[L(x, X - c)].$$

As a consequence,

$$\mathbb{E}[L(\mathcal{S}^*(X), X)] = c + \mathbb{E}[L(\mathcal{S}^*(X - c), X - c)] = c + \mathbb{E}[L(\mathcal{S}^*(X) - c, X - c)], \tag{4.16}$$

where the last equality follows from $\mathcal{S}^*(X + c) = \mathcal{S}^*(X) + c$. By setting $x^* = \mathcal{S}^*(X)$, we have

$$\begin{aligned}
\mathcal{R}^*(X) &= \min_{x\in\mathbb{R}} \mathbb{E}[L^*(x, X)] = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[L(0, X - x)]\} \\
&\leqslant x^* + \mathbb{E}[L(0, X - x^*)] = x^* + \mathbb{E}[L(\mathcal{S}^*(X) - x^*, X - x^*)] \\
&= x^* + \mathbb{E}[L(\mathcal{S}^*(X - x^*), X - x^*)] = x^* + \min_{x\in\mathbb{R}} \mathbb{E}[L(x, X - x^*)] \\
&= x^* + \mathcal{R}(X - x^*) = \mathcal{R}(X),
\end{aligned}$$

where the last equality is due to the translation invariance of $\mathcal{R}$. Thus we have $\mathcal{R}^* \leqslant \mathcal{R}$. In order to show $\mathcal{R}^* \geqslant \mathcal{R}$, take $y^* \in \arg\min_{x\in\mathbb{R}}\{x + \mathbb{E}[L(0, X - x)]\}$. By (4.16), we have

$$\begin{aligned}
\mathcal{R}^*(X) &= y^* + \mathbb{E}[L(0, X - y^*)] \geqslant y^* + \min_{x\in\mathbb{R}} \mathbb{E}[L(x, X - y^*)] \\
&= y^* + \mathbb{E}[L(\mathcal{S}^*(X - y^*), X - y^*)] = \mathbb{E}[L(\mathcal{S}^*(X), X)] = \mathcal{R}(X).
\end{aligned}$$

Hence, we have $\mathcal{R} = \mathcal{R}^*$. Taking $v(y) = L(0, y)$ gives the last statement. $\square$

Using Lemma 4.1, we can write

$$\mathcal{R}(X) = \min_{x \in \mathbb{R}} \{x + \mathbb{E}[v(X - x)]\}, \quad \text{where} \ \ v(y) = L(0, y). \tag{4.17}$$

In the following lemmas, we allow $v$ to take the value $\infty$, and obtain results that are slightly more general than required, i.e., we will also include $\mathrm{ES}_1$ which is the essential supremum. We define the increasing version of $v$ as

$$\tilde{v}(x) = \inf_{y \geqslant x} v(y), \quad x \in \mathbb{R}.$$

Note that $\mathcal{R}$ in (4.17) is real-valued, and

$$\mathcal{R}(x) = \inf_{c \in \mathbb{R}} \{c + v(x - c)\} \leqslant \inf_{y \geqslant x} v(y).$$

Hence, $\tilde{v}(x) > -\infty$ for all $x \in \mathbb{R}$. The finiteness of $\mathcal{R}$ also implies that $v$ is not always $\infty$ on $\mathbb{R}$.

**Lemma 4.2.** *Suppose that $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ in (4.17) is monotone. Then*

$$\mathcal{R}(X) = \inf_{c \in \mathbb{R}} \{c + \mathbb{E}[\tilde{v}(X - c)]\}, \quad X \in \mathcal{X}.$$

*Proof.* Let us denote by $\widetilde{\mathcal{R}}(X) = \inf_{c \in \mathbb{R}} \{c + \mathbb{E}[\tilde{v}(X - c)]\}$, $X \in \mathcal{X}$. Obviously, $\mathcal{R}(X) \geqslant \widetilde{\mathcal{R}}(X)$, $X \in \mathcal{X}$. Below we show $\mathcal{R} \leqslant \widetilde{\mathcal{R}}$. Take $\epsilon > 0$ and $c \in \mathbb{R}$. By definition of $\tilde{v}$, for any $x \in \mathbb{R}$, there exists $y \geqslant x$ such that $v(y) \leqslant \tilde{v}(x) + \epsilon$, and such $y$ admits an increasing (hence measurable) selection. As a consequence, there exists $Y \in \mathcal{X}$ such that $Y \geqslant X$ and $v(Y - c) \leqslant \tilde{v}(X - c) + \epsilon$. This implies $c + \mathbb{E}[v(Y - c)] \leqslant c + \mathbb{E}[\tilde{v}(X - c)] + \epsilon$. By monotonicity of $\mathcal{R}$ and $Y \geqslant X$, we further have

$$\mathcal{R}(X) \leqslant \mathcal{R}(Y) \leqslant c + \mathbb{E}[v(Y - c)] \leqslant c + \mathbb{E}[\tilde{v}(X - c)] + \epsilon.$$

Taking an infimum of the above inequality over $c \in \mathbb{R}$ and $\epsilon > 0$ yields $\mathcal{R}(X) \leqslant \widetilde{\mathcal{R}}(X)$. $\square$

Next, for an increasing function $v$, we define the largest convex function dominated by $v$ as

$$\hat{v}(x) = \sup\{g(x) : g \leqslant v \text{ on } \mathbb{R}, \ g \text{ is convex}\}, \quad x \in \mathbb{R}.$$

By definition, $\hat{v}$ is convex. To state the following lemma, we define

$$\mathcal{U} = \big\{v : \ v \text{ is increasing and convex}, 1 \in \operatorname{int} \partial v(\mathbb{R})\big\}, \tag{4.18}$$

where $\partial v(\mathbb{R}) = \operatorname{cx}\{v'_-(x), v'_+(x), \ x \in \mathbb{R}\}$, $\operatorname{int} A$ is the interior of a set $A$, and $\operatorname{cx}(A)$ is the convex hull of $A$. Here we define the right derivative $v'_+(x) = \infty$ if $v(y) = \infty$ for any $y > x$.

**Lemma 4.3.** *Suppose that $\mathcal{R}:\mathcal{X}\to\mathbb{R}$ in (4.17) is monetary and convex, and $v$ is an increasing function. Then*

$$\mathcal{R}(X)=\inf_{c\in\mathbb{R}}\{c+\mathbb{E}[\hat{v}(X-c)]\},\quad X\in\mathcal{X}.$$

*Specifically, if $\hat{v}\notin\mathcal{U}$, then either $\mathcal{R}(X)\equiv-\infty$ or $\mathcal{R}(X)=\mathbb{E}[X]-\hat{v}^*(1)$, where $\hat{v}^*(x)=\sup_y\{xy-\hat{v}(y)\}$ is the conjugate function of $\hat{v}$.*

*Proof.* Let us denote by

$$\mathcal{R}'(X)=\inf_{c\in\mathbb{R}}\{c+\mathbb{E}[\hat{v}(X-c)]\},\quad X\in\mathcal{X}.$$

Obviously, $\mathcal{R}\geqslant\mathcal{R}'$. Below we show $\mathcal{R}\leqslant\mathcal{R}'$. Take $\epsilon>0$ and $c\in\mathbb{R}$. Note that a law-invariant convex risk measure is monotonic with respect to convex order (e.g., Proposition 3.2 of Mao and Wang, 2020). For $X\in\mathcal{X}$, we assert that there exists $Y\in\mathcal{X}$ be such that

$$X\prec_{\mathrm{cx}}Y\ \text{ and }\ \mathbb{E}[v(Y-c)]\leqslant\mathbb{E}[\hat{v}(X-c)]+\epsilon.\tag{4.19}$$

To show this assertion, we use Theorem 4.1 of Mao et al. (2018), which gives

$$\mathbb{E}[\hat{v}(X)]=\lim_{n\to\infty}\frac{1}{n}\inf\left\{\mathbb{E}\left[v\left(X_1\right)\right]+\cdots+\mathbb{E}\left[v\left(X_n\right)\right]:X_1+\cdots+X_n=nX\right\}.$$

Then for any $\epsilon>0$, there exist $n\in\mathbb{N}$ and $X_1,\dots,X_n$ such that $X_1+\cdots+X_n=nX$ and

$$\frac{1}{n}\big(\mathbb{E}\left[v\left(X_1\right)\right]+\cdots+\mathbb{E}\left[v\left(X_n\right)\right]\big)\leqslant\mathbb{E}[\hat{v}(X)]+\epsilon.$$

Denote by $F_i$ the distribution of $X_i$, $i=1,\dots,n$ and take a random variable $Y$ such that its distribution is $H=\sum_{i=1}^n F_i/n$. We then have

$$\mathbb{E}[v(Y)]=\int_{\mathbb{R}}v(y)\,\mathrm{d}H(y)=\frac{1}{n}\sum_{i=1}^n\mathbb{E}[v(X_i)]\leqslant\mathbb{E}[\hat{v}(X)]+\epsilon.$$

For any convex function $\ell$, we have $\mathbb{E}[\ell(Y)]=\frac{1}{n}\sum_{i=1}^n\mathbb{E}[\ell(X_i)]\geqslant\mathbb{E}[\ell(X)]$, where the inequality follows convexity. This implies $X\prec_{\mathrm{cx}}Y$, and hence (4.19) holds. This implies $c+\mathbb{E}[v(Y-c)]\leqslant c+\mathbb{E}[\hat{v}(X-c)]+\epsilon$. By monotonicity of $\mathcal{R}$ with respect to convex order and $X\prec_{\mathrm{cx}}Y$, we further have

$$\mathcal{R}(X)\leqslant\mathcal{R}(Y)\leqslant c+\mathbb{E}[v(Y-c)]\leqslant c+\mathbb{E}[\hat{v}(X-c)]+\epsilon.$$

Taking an infimum of the above inequality over $c\in\mathbb{R}$ and $\epsilon>0$ yields $\mathcal{R}(X)\leqslant\mathcal{R}'(X)$. Therefore, $\mathcal{R}=\mathcal{R}'$. The last statement is shown in the discussion below Definition 3.1 of Wu et al. (2020). □

By Lemma 4.3, in order to avoid the trivial cases of $\mathcal{R}$, we only need to consider $v\in\mathcal{U}$.

**Lemma 4.4.** *Let $v$ be a function such that $v(x) \geqslant \overline{v}(x) := \lambda x + (\gamma - \lambda)x_+$, $x \in \mathbb{R}$, and*

$$\min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X - x)]\} = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[\overline{v}(X - x)]\}, \quad X \in \mathcal{X}.$$

*Then $v = \overline{v}$.*

*Proof.* We show the result by contradiction. Suppose that $v(x_0) > \overline{v}(x_0)$ for some $x_0 \in \mathbb{R}$. Define a random variable $X$ such that $1 - \mathbb{P}(X = 0) = \mathbb{P}(X = x_0) = p$, where $p \in (0,1)$ satisfies $1 - p > \alpha := 1 - (1-\lambda)/\gamma$. Then we have $\mathrm{VaR}_\alpha(X) = 0$ and thus

$$\mathcal{R}(X) = \min_x\{x + \mathbb{E}[\overline{v}(X - x)]\} = 0 + \mathbb{E}[\overline{v}(X - 0)] = \overline{v}(x_0).$$

Note that 0 is the unique minimizer of the above minimization problem, which implies that

$$x + \mathbb{E}[v(X - x)] \geqslant x + \mathbb{E}[\overline{v}(X - x)] > \mathcal{R}(X), \quad x \neq 0. \tag{4.20}$$

For $x = 0$, note that $\mathbb{E}[v(X)] = (1-p)v(0) + pv(x_0) > p\overline{v}(x_0) = \mathcal{R}(X)$. This combined with (4.20) yields a contradiction to the fact that $\min_{x\in\mathbb{R}} x + \mathbb{E}[v(X - x)]$ can be attained. Hence, $v = \overline{v}$. □

Theorem 3.1 of Ben-Tal and Teboulle (2007) showed that an OCE in (4.17) is positively homogeneous if and only if $v(x) = \lambda x + (\gamma - \lambda)x_+$ for some $\gamma, \lambda$. The following lemma gives a similar result under slightly different conditions. It states that $\mathcal{R}$ defined by (4.17) with $v \in \mathcal{U}$ is positively homogeneous if and only if $v$ can be replaced by $\bar{v}$. For completeness, we give a self-contained proof that is different from Ben-Tal and Teboulle (2007).

**Lemma 4.5.** *For $v \in \mathcal{U}$, suppose that $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ in (4.17) is positively homogeneous and $v$ is an increasing convex function. Then there exist $\gamma \in [1,\infty]$ and $\lambda \in [0,1]$, such that*

$$\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[\bar{v}(X - c)]\}, \quad X \in \mathcal{X}, \tag{4.21}$$

*where $\bar{v}(x) = \lambda x + (\gamma - \lambda)x_+$ for all $x \neq 0$.*

*Proof.* Since $\mathcal{R}$ is positively homogeneous, we have $\mathcal{R}(0) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v(-c)]\} = 0$. Note that one minimizer of the above infimum is $-\zeta_v$ given by $\zeta_v := \inf\{x : v'_-(x) \geqslant 1\} \in \mathbb{R}$, where $v'_-$ is the left-derivative of $v$. One can easily verify that $1 \in [v'_-(\zeta_v), v'_+(\zeta_v)]$ and $v(\zeta_v) = \zeta_v$. Define $u(x) := v(x + \zeta_v) - \zeta_v$, $x \in \mathbb{R}$. We have $1 \in [u'_-(0), u'_+(0)]$, and thus, $u \in \mathcal{U}$ with $u(0) = 0$. It is obvious that $\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[u(X - c)]\}$, $X \in \mathcal{X}$. Hence, without loss of generality, we assume that $v \in \mathcal{U}$ and $v(0) = 0$.

For $\lambda > 0$, denote by $v_\lambda(x) = v(\lambda x)/\lambda$ and $\bar{v}(x) = \inf_{\lambda>0} v_\lambda(x),\ x \in \mathbb{R}$. It is clear that $\bar{v}$ is convex, increasing and positively homogeneous. Since $v$ is convex, we have, for $0 < \lambda \leqslant \gamma$,

$$v(\lambda x) \leqslant \frac{\lambda}{\gamma} v(\gamma x) + \left(1 - \frac{\lambda}{\gamma}\right) v(0) = \frac{\lambda}{\gamma} v(\gamma x).$$

As a consequence, $v_\lambda(x) \leqslant v_\gamma(x)$. Thus we know that $v_\lambda(x)$ is increasing in $\lambda$. Hence, $\bar{v}(x) = \lim_{\lambda\downarrow 0} v_\lambda(x)$. Note that if $v(x) = \infty$ for any $x > 0$, then we have $\bar{v}(x) = \infty$ for $x > 0$; if however $v(x) < \infty$ for some $x > 0$, then for any $x \in \mathbb{R}$, there exists $\lambda > 0$ such that $v_\lambda(x) < \infty$. For $X \in \mathcal{X}$ with an upper bound and $c \in \mathbb{R}$, the Monotone Convergence Theorem gives

$$\mathbb{E}[\bar{v}(X-c)] = \mathbb{E}\left[\lim_{\lambda\downarrow 0} v_\lambda(X-c)\right] = \lim_{\lambda\downarrow 0} \mathbb{E}\left[v_\lambda(X-c)\right] = \inf_{\lambda>0} \mathbb{E}\left[v_\lambda(X-c)\right].$$

By definition, for $\lambda > 0$ and $X \in \mathcal{X}$,

$$\mathcal{R}(\lambda X) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v(\lambda X - c)]\} = \inf_{c\in\mathbb{R}}\{\lambda c + \mathbb{E}[v(\lambda(X-c))]\} = \lambda \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v_\lambda(X-c)]\}.$$

Hence, positive homogeneity of $\mathcal{R}$ implies

$$\mathcal{R}(X) = \frac{\mathcal{R}(\lambda X)}{\lambda} = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v_\lambda(X-c)]\}.$$

Taking an infimum over $\lambda > 0$ yields that for any $X \in \mathcal{X}$ with an upper bound

$$\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\left\{c + \inf_{\lambda>0} \mathbb{E}[v_\lambda(X-c)]\right\} = \inf_{c\in\mathbb{R}}\left\{c + \mathbb{E}[\bar{v}(X-c)]\right\},$$

thus showing that (4.21) holds for $X \in \mathcal{X}$ with an upper bound. By Lemma 4.4, we have $v = \bar{v}$, and thus, (4.21) holds for all $X \in \mathcal{X}$. Positive homogeneity and monotonicity of $\bar{v}$ imply that

$$\bar{v}(x) = \gamma x_+ - \lambda x_- = \lambda x + (\gamma - \lambda)x_+, \quad x \neq 0,$$

for some $\gamma \in [0,\infty]$ and $\lambda \in [0,\infty)$. Using Lemma 4.3, we further know that either $\mathcal{R}(X) = \mathbb{E}[X] + c$ for some constant $c$ or $\bar{v} \in \mathcal{U}$. If $\mathcal{R}(X) = \mathbb{E}[X] + c$, then $c = 0$ due to positive homogeneity of $\mathcal{R}$. In this case, $\bar{v}$ can be chosen as $\bar{v}(x) = x$, corresponding to $\gamma = \lambda = 1$. If $\bar{v} \in \mathcal{U}$, then $\operatorname{int} \partial v(\mathbb{R}) = (\lambda, \gamma)$, which implies $\lambda < 1 < \gamma$. □

*Proof of Theorem 4.2.* It remains to prove the last step in the implication (i)⇒(iii). Combining Lemmas 4.1-4.5, we know that (4.17) holds, and $v$ can be chosen as $v(x) = \lambda x + (\gamma - \lambda)x_+$ for some $\gamma \in [1,\infty]$ and $\lambda \in [0,1]$. If $\lambda < 1$, write $\alpha = (\gamma - 1)/(\gamma - \lambda) \in (0,1]$. Using (4.4), including the case $\alpha = 1$, we have

$$\begin{aligned}\mathcal{R}(X) &= \inf_{c\in\mathbb{R}}\{c + \lambda\mathbb{E}[(X-c)] + (\gamma - \lambda)\mathbb{E}[(X-c)_+]\} \\ &= (1-\lambda)\mathrm{ES}_\alpha(X) + \lambda\mathbb{E}[X].\end{aligned}$$

If $\lambda = 1$, then $\mathcal{R}(X) = \inf_{c\in\mathbb{R}} \{(\gamma - 1)\mathbb{E}[(X - c)_+]\} + \mathbb{E}[X] = \mathbb{E}[X]$. In either case, $\mathcal{R} = \mathrm{ES}_\alpha^{1-\lambda} = (1-\lambda)\mathrm{ES}_\alpha + \lambda\mathbb{E}$. If, moreover, $v$ is real-valued (it is in the definition of the loss function for a Bayes pair), then we have $\gamma < \infty$ and thus, $\alpha < 1$. $\square$

### 4.8.2 Proof of Proposition 4.1

*Proof of Proposition 4.1.* Note that the implications (iii)$\Leftrightarrow$(iv)$\Rightarrow$(ii) are obvious, and the implication (ii)$\Rightarrow$(i) is implied by Proposition 4.3 in Section 4.5. We next show the implication (i)$\Rightarrow$(iii). By Lemma 4.1, there exists a function $v : \mathbb{R} \to \mathbb{R}$ such that

$$\mathcal{S}(X) = \arg\min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X - x)]\}, \quad \mathcal{R}(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X - x)]\}.$$

On the other hand, by Theorem 4.2, $\mathcal{R}(X) = \mathrm{ES}_\alpha^\lambda(X)$ for some $\lambda \in [0, 1]$ and $\alpha \in (0, 1)$. That is, there exists $\overline{v}(x) = \lambda' x + (\gamma - \lambda')x_+$ with $\lambda' = 1 - \lambda \in [0, 1]$ and $\gamma = (1 - \alpha\lambda')/(1 - \alpha) \in [1, \infty)$ such that

$$\mathcal{R}(X) = \mathrm{ES}_\alpha^\lambda(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[\overline{v}(X - x)]\}.$$

Denote by $\hat{v}$ the largest increasing convex function dominated by $v$. That is,

$$\hat{v}(x) = \sup\{g(x) : g \leqslant \tilde{v} \text{ on } \mathbb{R},\ g \text{ is convex}\}, \quad x \in \mathbb{R}, \tag{4.22}$$

with $\tilde{v}(x) = \inf_{y\geqslant x} v(y)$. We then show the result by considering the following two cases.

(i) If $\hat{v}(0) = 0$, then by the proofs of Lemmas 4.2 to 4.5, we have $\overline{v} \leqslant v$, and hence $v = \overline{v}$ by Lemma 4.4. By $\mathcal{S}(0) = 0$, which excludes $\lambda = 1$, we have $\mathcal{S}(X) = \mathrm{VaR}_\alpha(X)$.

(ii) If $\hat{v}(0) > 0$, then by the proof of Lemma 4.5, there exists $c \in \mathbb{R}$ such that $\hat{v}(c) = c$. Define $v^*(x) = v(x + c) - c$, $x \in \mathbb{R}$ and the corresponding $\hat{v^*}$ of $v^*$ by (4.22). One can verify that $\hat{v^*}(x) = \hat{v}(x + c) - c$, which implies $\hat{v^*}(0) = 0$, and

$$\mathcal{R}(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[v^*(X - x)]\}.$$

Similar to Case (i), we have $v^* \geqslant \overline{v}$ and thus, $v^* = \overline{v}$ by Lemma 4.4. Since $\mathcal{S}(0) = 0$ excludes $\lambda = 1$, it follows that $\mathcal{S}(X) = \mathrm{VaR}_\alpha(X) - c$. By $\mathcal{S}(0) = 0$ again, we have $c = 0$. This completes the proof. $\square$

### 4.8.3 Proof of Theorem 4.4

We present two lemmas used in the proof of Theorem 4.4. The first lemma uses a weaker continuity (C') than (C) in Theorem 4.4. This result is similar to Theorem 3.1 of Weber (2006) which uses a different condition of $\psi$-weak lower semi-continuity to replace mixture concavity.

(C') There exists $x \leqslant 0$ such that for any $y > 0$, $\widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \leqslant 0$ for small enough $\alpha > 0$.

**Lemma 4.6.** *Let $\mathcal{R}$ be a monetary risk measure on $\mathcal{X} = \mathcal{L}^\infty$ satisfying continuity (C') with $\mathcal{R}(0) = 0$. If $\mathcal{R}$ is mixture concave and has CxLS, then there exist $z \in \mathbb{R}$ and an increasing and left-continuous $g$ such that*

$$\mathcal{R}(X) = \inf\{x \in \mathbb{R} : \mathbb{E}[g(X - x)] \leqslant z\}, \quad X \in \mathcal{X}. \tag{4.23}$$

*Proof.* Take $x \leqslant 0$ in assumption (C') and any fixed constant $y > 0$, and let $z \in (0, 1)$ be such that $[0, z] = \{\alpha \in [0, 1] : \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \leqslant 0\}$; the interval is closed since $\alpha \mapsto \mathcal{R}((1-\alpha)\delta_x + \alpha\delta_y)$ is concave. We define the function $g$ as

$$g(t) = \begin{cases} \frac{z-\overline{\alpha}(t)}{1-\overline{\alpha}(t)}, & t \leqslant 0, \\ \frac{z}{\overline{\alpha}(t)}, & t > 0, \end{cases} \tag{4.24}$$

where

$$\overline{\alpha}(t) = \begin{cases} \sup\{\alpha \in [0, 1] : \widehat{\mathcal{R}}((1-\alpha)\delta_t + \alpha\delta_y) \leqslant 0\}, & t \leqslant 0, \\ \sup\{\alpha \in [0, 1] : \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_t) \leqslant 0\}, & t > 0. \end{cases} \tag{4.25}$$

This is the same construction as in Eq. (3.5) and (3.7) of Weber (2006). Since for any $t \leqslant 0$, $\alpha \mapsto \widehat{\mathcal{R}}((1-\alpha)\delta_t + \alpha\delta_y)$ is increasing concave in $\alpha \in [0, 1]$ and $\widehat{\mathcal{R}}(\delta_y) = y > 0$, we have $\overline{\alpha}(t) < 1$ for $t \leqslant 0$. By (C'), we have $\overline{\alpha}(t) > 0$ for $t > 0$. Hence, $g$ is well defined. By monotonicity of $\mathcal{R}$, one can verify that $g$ is increasing and satisfies $g(0) = z$. Next we show (4.23) and the left-continuity of $g$.

1. Denote by $\mathcal{S}$ the convex hull of $\{\delta_{x_i},\ i = 1, \dots, n\}$, that is,

   $$\mathcal{S} = \left\{\sum_{i=1}^n \alpha_i\delta_{x_i} : \alpha_i \geqslant 0,\ i = 1, \dots, n,\ \sum_{i=1}^n \alpha_i = 1\right\},$$

   where $x_1 = x$, $x_2 = y$ are those fixed above. By mixture concavity of $\mathcal{R}$, one can verify that $(\alpha_1, \dots, \alpha_n) \mapsto \widehat{\mathcal{R}}(\sum_{i=1}^n \alpha_i\delta_{x_i})$ is a concave function, and thus is lower-semicontinuous by Theorem 10.2 of Rockafellar (1970). It follows that $\mathcal{N} := \{(\alpha_1, \dots, \alpha_n) : \widehat{\mathcal{R}}(\sum_{i=1}^n \alpha_i\delta_{x_i}) \leqslant 0\}$ is closed sets in the Euclidean topology. By CxLS of $\mathcal{R}$, we have $\mathcal{N}$ and $\mathcal{S}\setminus\mathcal{N}$ are both convex sets. Using similar arguments of Weber (2006), one can verify that (4.23) holds for random variable $X$ which has distribution in $\mathcal{S}$.

2. We next show $g$ is left-continuous. Since $g$ is increasing, it suffices to show $\lim_{s\uparrow t} g(s) = g(t)$ for $t \in \mathbb{R}$, which is equivalent to $\lim_{s\uparrow t} \overline{\alpha}(s) \leqslant \overline{\alpha}(t)$ as $\overline{\alpha}$ is decreasing.

   (a) For $t > 0$, if $\overline{\alpha}(t) = 1$, then $\lim_{s\uparrow t} \overline{\alpha}(s) \leqslant \overline{\alpha}(t)$ holds trivially as $\overline{\alpha}(s) \leqslant 1$ for any $s$.

   (b) For $t > 0$ with $\overline{\alpha}(t) < 1$, by definition of $\overline{\alpha}(t)$, we have for any $\epsilon \in (0, 1 - \overline{\alpha}(t))$, $\widehat{\mathcal{R}}((1 - \overline{\alpha}(t) - \epsilon)\delta_x + (\overline{\alpha}(t) + \epsilon)\delta_t) > 0$. Since $\mathcal{R}$ is monetary, $\widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_t)$ is continuous in $t \in \mathbb{R}$, and hence there exists $s_0 < t$ such that $\widehat{\mathcal{R}}((1-\overline{\alpha}(t)-\epsilon)\delta_x + (\overline{\alpha}(t)+\epsilon)\delta_{s_0}) > 0$. That is, $\overline{\alpha}(s_0) \leqslant \overline{\alpha}(t) + \epsilon$. Therefore, $\lim_{s\uparrow t} \overline{\alpha}(s) \leqslant \overline{\alpha}(s_0) \leqslant \overline{\alpha}(t) + \epsilon$. As $\epsilon$ is arbitrary, we have $\lim_{s\uparrow t} \overline{\alpha}(s) \leqslant \overline{\alpha}(t)$.

   (c) For $t \leqslant 0$, we have $\overline{\alpha}(t) < 1$. Similar arguments as in (b) yield $\lim_{s\uparrow t} \overline{\alpha}(s) \leqslant \overline{\alpha}(t)$.

3. Next we show that (4.23) holds for any $X \in \mathcal{X}$. Since $\mathcal{R}$ is monetary, there exist $X_n$, $n \in \mathbb{N}$, each taking values in a finite set, such that $X_n \uparrow X$ and $\lim_{n\to\infty} \mathcal{R}(X_n) = \mathcal{R}(X)$. Since $g$ is left-continuous, we have $g(X_n - x) \uparrow g(X - x)$ and by the Monotone Convergence Theorem, we obtain $\lim_{n\to\infty} \mathbb{E}[g(X_n - x)] = \mathbb{E}[g(X - x)]$ for any $x \in \mathbb{R}$. This implies

$$\lim_{n\to\infty} \inf\{x \in \mathbb{R} : \mathbb{E}[g(X_n - x)] \leqslant z\} = \inf\{x \in \mathbb{R} : \mathbb{E}[g(X - x)] \leqslant z\}.$$

   It then follows from $\lim_{n\to\infty} \mathcal{R}(X_n) = \mathcal{R}(X)$ and $\mathcal{R}(X_n) = \inf\{x \in \mathbb{R} : \mathbb{E}[g(X_n - x)] \leqslant z\}$ that (4.23) holds for any $X \in \mathcal{X}$. □

**Lemma 4.7.** *Let $\mathcal{R}$ be defined by (4.23) for an increasing function $g$ and $z \in \mathbb{R}$ satisfying $\mathcal{R}(0) = 0$. Then $\mathcal{R}$ satisfies (C) if and only if $g(t) < z$ for all $t < 0$. Moreover, if $\mathcal{R}$ is mixture concave and satisfies (C), then $g$ in (4.23) can be chosen continuous and strictly increasing on either $(-\infty, 0)$ or $(0, \infty)$, and $\mathcal{R}(X)$ is the unique solution $x$ to the equation $\mathbb{E}[g(X - x)] = z$.*

*Proof.* 1. We first show that $\mathcal{R}$ satisfies (C) if and only if $g(t) < z$ for $t < 0$. To see the "if" statement, by translation invariance, it suffices to show that $\widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y)$ is continuous at $\alpha = 0$ for $x \leqslant 0 < y$. For this, we will verify $\lim_{\alpha\downarrow 0} \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \leqslant x$. For any $\epsilon \in (0, y - x)$, we have

$$\lim_{\alpha\downarrow 0}(1-\alpha)g(x - x - \epsilon) + \alpha g(y - x - \epsilon) = g(-\epsilon) < z.$$

Hence, there exists $\alpha_0 \in (0,1)$ such that $(1-\alpha_0)g(x-x-\epsilon)+\alpha_0 g(y-x-\epsilon) < z$, implying $\widehat{\mathcal{R}}((1-\alpha_0)\delta_x+\alpha_0\delta_y) \leqslant x+\epsilon$. By monotonicity of $\mathcal{R}$, we have $\lim_{\alpha\downarrow 0}\widehat{\mathcal{R}}((1-\alpha)\delta_x+\alpha\delta_y) \leqslant \widehat{\mathcal{R}}((1-\alpha_0)\delta_x+\alpha_0\delta_y) \leqslant x+\epsilon$. As $\epsilon$ is arbitrary, we have $\lim_{\alpha\downarrow 0}\widehat{\mathcal{R}}((1-\alpha)\delta_x+\alpha\delta_y) \leqslant x$.

To see the "only if" statement, suppose that there exists $\epsilon > 0$ such that $g(t) = z$ for $z \in (-\epsilon, 0)$. By $\mathcal{R}(0) = 0$, we have $z < g(t)$ for any $t > 0$. It follows that $\widehat{\mathcal{R}}((1-\alpha)\delta_0+\alpha\delta_{\epsilon/2}) \geqslant \epsilon/2$ for any $\alpha > 0$. This contradicts (C).

2. In what follows, we take $g$ from (4.24) in the proof of Lemma 4.6. We will show that $g$ is continuous and strictly increasing on either $(-\infty,0)$ or $(0,\infty)$ by contradiction. We have seen from Step 1 above that (C) and $\mathcal{R}(0) = 0$ together imply that $g$ is strictly increasing at 0. Suppose that there exist $a < b < 0 < c < d$ such that $[a,b] = \{x : g(x) = g(a)\}$ and $[c,d] = \{x : g(x) = g(c)\}$. As $g(a) < z < g(c)$, there exists $\alpha_0 \in [0,1)$ such that $(1-\alpha_0)g(a)+\alpha_0 g(c) = g(0)$, and hence
$$(1-\alpha)g(a)+\alpha g(c) > z \quad \text{for any } \alpha > \alpha_0. \tag{4.26}$$
Since $g(t) < z$ for $t < 0$, there exist $\alpha_1 > \alpha_0$ and $\epsilon \in (0, \min\{(b-a)/2, (d-c)/2\})$ such that
$$\frac{1-\alpha_1}{2}g(a)+\frac{\alpha_1}{2}g(c)+\frac{1}{2}g(-\epsilon) < z. \tag{4.27}$$
Define the distribution $F_{x,y} = (1-\alpha_1)\delta_x+\alpha_1\delta_y$ with $x \in [a+2\epsilon, b]$ and $y \in [c+2\epsilon, d]$. Then by (4.26), we have $\widehat{\mathcal{R}}(F_{x,y}) > 2\epsilon$. By letting $G = \frac{1}{2}F_{x,y}+\frac{1}{2}\delta_0$, (4.27) implies $\widehat{\mathcal{R}}(G) \leqslant \epsilon < \widehat{\mathcal{R}}(F_{x,y})/2+\widehat{\mathcal{R}}(\delta_0)/2$, yielding a contradiction. Hence, $g$ is strictly increasing on either $(-\infty,0)$ or $(0,\infty)$.

3. Using results in Steps 1 and 2, for any $X \in \mathcal{X}$, we have $\mathbb{E}[g(X-x)]$ is strictly decreasing for $x$ in the range of $X$, and thus, the equation $\mathbb{E}[g(X-x)] = z$ has a unique solution $x$.

4. Finally we show that $g$ is continuous. Using (C), $\overline{\alpha}$ defined by (4.25) satisfies that $\overline{\alpha}(t)$ is the unique solution to the equation in $\alpha$, $\widehat{\mathcal{R}}((1-\alpha)\delta_t+\alpha\delta_y) = 0$, if $t \leqslant 0$, and it is the unique solution $\alpha$ to $\widehat{\mathcal{R}}((1-\alpha)\delta_x+\alpha\delta_t) = 0$ if $t > 0$. Since $\mathcal{R}$ is monetary and hence $L^\infty$-continuous, we have that $\widehat{\mathcal{R}}((1-\alpha)\delta_{x_1}+\alpha\delta_{x_2})$ is continuous in $(x_1,x_2) \in \mathbb{R}^2$. Hence, $\overline{\alpha}$ is continuous and thus, $g$ is continuous. □

*Proof of Theorem 4.4.* The "if" statement is argued in (ii) of Example 4.1, where we see that the entropic risk measure is both a Bayes risk measure and a Bayes estimator. Hence, it is mixture concave and has CxLS. To prove the "only if" statement, first note that by using Lemmas 4.6 and

4.7, we have (4.15) holds with $g$ and $z$ satisfying that $g$ is continuous, and strictly increasing on either $(-\infty, 0)$ or $(0, \infty)$, and the equation $\mathbb{E}[g(X-x)] = z$ always has a unique solution. Further, by Lemma 4.7 and $\mathcal{R}(0) = 0$, we have $g(-t) < z < g(t)$ for all $t > 0$. We then employ the following steps to show that $\mathcal{R}$ must be an entropic risk measure.

1. Note that a monotone function $g$ has derivatives almost everywhere. Let $t$ be a point such that $g(t) < z$ and $g$ is differentiable at $t$. Take arbitrary $x, y > 0$. Since $g(t) < z < \min\{g(x), g(y)\}$, for each $\epsilon \in (0, \min\{x, y\})$, there exist unique $\lambda_1(\epsilon) \in (0,1)$ and $\lambda_2(\epsilon) \in (0,1)$ such that

$$\lambda_1(\epsilon) g(x-\epsilon) + (1-\lambda_1(\epsilon)) g(t) = z \quad \text{and} \quad \lambda_2(\epsilon) g(y-\epsilon) + (1-\lambda_2(\epsilon)) g(t) = z. \tag{4.28}$$

   As $g$ is increasing, $\lambda_i(\epsilon)$ is decreasing in $\epsilon$, $i = 1, 2$. Let $\lambda_i^0 = \lim_{\epsilon \downarrow 0} \lambda_i(\epsilon) \in (0,1)$ for $i = 1, 2$, and we have

$$\lambda_1^0 g(x) + (1-\lambda_1^0) g(t) = z \quad \text{and} \quad \lambda_2^0 g(y) + (1-\lambda_2^0) g(t) = z. \tag{4.29}$$

2. Let a random variable $X$ be given by $\mathbb{P}(X = x) = \lambda_1^0$ and $\mathbb{P}(X = t) = 1 - \lambda_1^0$, and $Y_\epsilon$ be given by $\mathbb{P}(Y_\epsilon = y + \epsilon) = \lambda_2(\epsilon)$ and $\mathbb{P}(Y_\epsilon = t + 2\epsilon) = 1 - \lambda_2(\epsilon)$ for $\epsilon > 0$.

3. Since $g$ is strictly increasing at either $t$ or $x$, the first equation of (4.29) gives the inequality $\lambda_1^0 g(x+\delta) + (1-\lambda_1^0) g(t+\delta) > z$ for any $\delta > 0$. This implies $\mathcal{R}(X) = 0$. Similarly, we have $\mathcal{R}(Y_\epsilon) = 2\epsilon$.

4. Let $Z$ have a distribution with is a mixture of the distributions of $X$ and $Y_\epsilon$ with weight $1/2$ each. Using mixture concavity, we have $\mathcal{R}(Z) \geqslant \epsilon$, meaning that $\mathbb{E}[g(Z-\epsilon)] \geqslant z$. Hence, we have

$$\lambda_1^0 g(x-\epsilon) + (1-\lambda_1^0) g(t-\epsilon) + \lambda_2(\epsilon) g(y) + (1-\lambda_2(\epsilon)) g(t+\epsilon) \geqslant 2z.$$

   Subtracting the second equality in (4.28) and the first equality in (4.29) from the above equation, we get

$$\begin{aligned} &\lambda_1^0 \left(g(x-\epsilon) - g(x)\right) + \lambda_2(\epsilon) \left(g(y) - g(y-\epsilon)\right) \\ &+ (1-\lambda_1^0) \left(g(t-\epsilon) - g(t)\right) + (1-\lambda_2(\epsilon)) \left(g(t+\epsilon) - g(t)\right) \geqslant 0. \end{aligned}$$

   Divide the above equation by $\epsilon$, and letting $\epsilon \downarrow 0$, we obtain

$$\lambda_2^0 \liminf_{\epsilon \downarrow 0} \frac{g(y) - g(y-\epsilon)}{\epsilon} - \lambda_1^0 \limsup_{\epsilon \downarrow 0} \frac{g(x) - g(x-\epsilon)}{\epsilon} \geqslant (\lambda_2^0 - \lambda_1^0) g'(t), \tag{4.30}$$

where we use $\lambda_2^0 = \lim_{\epsilon\downarrow 0} \lambda_2(\epsilon)$. Since the positions of $(x,\, \lambda_1^0)$ and $(y,\, \lambda_2^0)$ are symmetric, we also have

$$\lambda_1^0 \liminf_{\epsilon\downarrow 0} \frac{g(x)-g(x-\epsilon)}{\epsilon} - \lambda_2^0 \limsup_{\epsilon\downarrow 0} \frac{g(y)-g(y-\epsilon)}{\epsilon} \geqslant (\lambda_1^0 - \lambda_2^0) g'(t), \tag{4.31}$$

Combining (4.30) and (4.31), we conclude that the two inequalities in (4.30) and (4.31) are both equalities, and because $\lambda_1^0, \lambda_2^0 > 0$, we have

$$\liminf_{\epsilon\downarrow 0} \frac{g(x)-g(x-\epsilon)}{\epsilon} = \limsup_{\epsilon\downarrow 0} \frac{g(x)-g(x-\epsilon)}{\epsilon}.$$

That is, $g$ has a left-derivative at $x$ and $y$. Similarly, we can show that it also has a right-derivative at $x$ and $y$, so that

$$\lambda_2^0 g'(y) - \lambda_1^0 g'(x) = (\lambda_2^0 - \lambda_1^0) g'(t), \tag{4.32}$$

for all $x, y > 0$. By (4.29), we can write $\lambda_1^0 = \frac{z-g(t)}{g(x)-g(t)}$ and $\lambda_2^0 = \frac{z-g(t)}{g(y)-g(t)}$. Substituting them into (4.32) yields

$$(g'(x) - g'(t))(g(y) - g(t)) = (g'(y) - g'(t))(g(x) - g(t)). \tag{4.33}$$

5. By fixing $t$ and $y$ and noting that $g(y) > g(t)$, (4.33) can be rewritten as

$$g'(x) - bg(x) = d \tag{4.34}$$

for some constants $b$ and $d$. Solving (4.34), we obtain that, on $(0, \infty)$, either $g$ is linear or $g(x) = ae^{bx} + c$ for some constants $a, b, c$. Similarly, by fixing $x$ and $y$, we have that, almost everywhere on $(-\infty, 0)$, either $g$ is linear or $g(x) = a'e^{b'x} + c'$ for some constants $a', b', c'$; continuity of $g$ now implies that the above for holds on $(-\infty, 0)$.

6. From the previous step, $g$ indeed has a positive derivative at any point $t < 0$. Hence, (4.33) holds for all $x, y > 0$ and $t < 0$. Note that (4.33) and the continuity of $g$ imply that $g'$ is continuous at 0. The forms of $g$ on $(-\infty, 0)$ and on $(0, \infty)$ have three parameters each (the linear case corresponds to the limit of $b \to 0$ after normalization). We obtain three equations from $g'(x)(g(y) - c') = g'(y)(g(x) - c')$ (obtained by letting $t \to -\infty$) and the continuity of $g$ and $g'$ at 0, and these three equations give $a = a'$, $b = b'$ and $c = c'$. Hence, we conclude that either $g$ is linear or $g(x) = ae^{bx} + c$ on $\mathbb{R}$.

7. If $g$ is linear, then $\mathcal{R} = \mathrm{ER}_0 = \mathbb{E}$. If $g$ is not linear, then $ab > 0$ since $\mathcal{R}$ is monotone. Moreover, (4.15) implies

$$\mathcal{R}(X) = \frac{1}{b} \log \mathbb{E}[e^{bX}], \quad X \in \mathcal{X}.$$

Since log is a concave function, mixture concavity does not hold if $b < 0$ (in this case, $\mathcal{R}$ is mixture convex). Hence, $b > 0$, and $\mathcal{R} = \mathrm{ER}_b$. $\square$

# Chapter 5

# Risk measures induced by efficient insurance contracts

## 5.1 Introduction

Optimal insurance and reinsurance design problems have been a prevalent topic for both researchers and practitioners in insurance for decades, since the seminal work of Arrow (1963) showing that deductible insurance is optimal for a risk-averse insured when the insurer is risk neutral. As natural extensions, Raviv (1979) studied conditions for optimality of deductible insurances when the insured and the insurer are both risk averse. Schlesinger (1981) examined the optimal choice of a risk-averse insured given that the insurance is of deductible type.

Previous studies on optimal (re)insurance design problems have shown considerations from several different perspectives. The majority of the studies focus on optimization under specific classes of optimization criteria quantifying the risk of decision makers; see e.g., Gollier and Schlesinger (1996) and Schlesinger (1997) for criteria preserving second-order stochastic dominance; Cai and Tan (2007), Cai et al. (2008) and Bernard and Tian (2009) for Value-at-Risk (VaR) and the Expected Shortfall (ES, also called CTE or TVaR in the above literature); Cui et al. (2013) for distortion risk measures or dual utilities (Yaari, 1987); and Braun and Muermann (2004) for regret-theoretical expected utilities. For more recent developments on optimal insurance with risk measures, we refer to Cai and Chi (2020) and the references therein. Moreover, optimal (re)insurance contract design problems are studied under a variety of constraints and formulations. We refer to studies on efficient insurance contracts with background risk (e.g., Gollier, 1996; Dana and Scarsini, 2007) and limited liability (e.g., Cummins and Mahul, 2004; Hofmann et al., 2019). More recently, Lo et al. (2021) analyzed the set of universally marketable indemnities with risk measures preserving convex orders.

Most of the previous literature aims to derive optimal forms of ceded loss functions under various scenarios and constraints. To the best of our knowledge, there is no relevant research on (re)insurance contract design problems focusing on identifying risk measures adopted by the insured and the insurer. Therefore, we study optimal insurance contract design problems through a distinctive perspective if compared to previous literature. Namely, the main goal of the present chapter is to answer the following (converse) question: In order for efficient contracts to be some sets of contracts commonly seen in insurance practice (e.g., of deductible form), which risk measures should the insurer and the insured use as their objectives? Specifically, we characterize different classes of risk measures adopted by the insured and the insurer given different sets of ceded loss functions that are Pareto optimal.

The risk measure ES has been widely applied in the contexts of financial regulation, risk management, and insurance. In particular, ES is the standard measure for market risk in the Fundamental Review of the Trading Book (FRTB) of BCBS (2016, 2019). In the insurance regulation framework of Solvency II, the risk measure Value-at-Risk (VaR) is the dominating risk measure. There is a growing academic literature on various problems using ES in actuarial science (where ES is often called TVaR). Most of these studies motivate the use of ES as a coherent risk measure (Artzner et al., 1999) and its advantages over VaR. Recently, Wang and Zitikis (2021) proposed the axiom called "no reward for concentration" (NRC) which, together with a few other standard axioms, characterizes ES.[1] The main objective of Wang and Zitikis (2021) is to separate ES from other coherent risk measures via the axiom of NRC, thus answering the question of why one uses ES instead of other risk measures from an axiomatic point of view. The interpretation and implication of the NRC axiom in financial regulation have been extensively discussed in Wang and Zitikis (2021) in the context of FRTB; see also an alternative formulation for axiomatizing ES in Han et al. (2021).

Given the big volume of research with ES in actuarial science, it is of great interest to understand whether ES plays a special role in insurance. The NRC axiom of Wang and Zitikis (2021) does not apply in the insurance context since it is interpreted as a requirement of portfolio risk

[1]The NRC axiom for a risk measure $\rho$ means that there exists a regulatory stress event $A$ such that $\rho(X+Y) = \rho(X)+\rho(Y)$ whenever $X$ and $Y$ both have the tail event $A$, meaning that $X$ satisfies $X(\omega) \geqslant X(\omega')$ for almost surely all $\omega \in A$ and $\omega' \in A^c$, and so does $Y$.

assessment. To understand the special role of ES in insurance, new insights that are specific to insurance design are therefore needed.

We work mainly within the framework of convex risk measures of Föllmer and Schied (2002b), which is a flexible and popular class of risk measures in risk management. As the main contribution of this chapter, we show that the set of efficient ceded loss functions of deductible form corresponds to the family of mixtures of ES and the mean (Theorem 5.2). If we further impose lower semicontinuity as in Wang and Zitikis (2021), then we arrive at the family of ES (Lemma 5.3). Our work also extends Chapter 4, which characterizes the mixture of the mean and ES, called an ES/$\mathbb{E}$-mixture, as the only coherent Bayes risk measure from the perspective of statistical inference. In addition, if the set of efficient ceded loss functions is the set of all slowly growing (1-Lipschitz) functions, then the corresponding risk measures are precisely the convex distortion risk measures (Theorem 5.1). Mathematically, our results are based on connecting various risk measures with different additivity forms over the ceded losses and the retained losses.

For illustrative purposes, we take the perspective of an insurance design problem between an insurer and an insured. Our technical results can certainly be applied in the reinsurance setting as well, where risk measures are often encountered.

The rest of the chapter is organized as follows. Section 5.2 contains some preliminaries on insurance losses and risk measures. Section 5.3 sets up the formulation of the insurance contract design problem and states economic assumptions. Section 5.4 contains our main characterization results of the risk measures used by the insured and the insurer given different Pareto-optimal sets of ceded loss functions. The results make natural connections between some common sets of ceded loss functions and common classes of risk measures in insurance practice. We also discuss economic implications of these results on the design of insurance menus by the insurer. Section 5.6 contains proofs of the main results accompanied with relevant technical lemmas.

## 5.2 Preliminaries on risk measures

Let $\mathcal{X}$ be the set of all bounded random variables, and let $\mathcal{X}_+$ be the set of all non-negative random variables in $\mathcal{X}$ representing insurable losses. Let $\mathcal{I}$ be a class of non-negative functions on $[0, \infty)$ which represent possible insurance ceded loss functions. For an insurable loss random variable $X \in \mathcal{X}_+$ and a contract $f \in \mathcal{I}$, $f(X)$ represents the payment to the insured, and $X - f(X)$ represents the retained loss of the insured. Losses are usually quantified by risk measures which are mappings from $\mathcal{X}$ to the set of real numbers, representing riskiness. For $X \in \mathcal{X}$, a *distortion risk measure* is defined as

$$\rho(X) = \int_0^\infty h(\mathbb{P}(X > x))\,\mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}(X > x)) - 1)\,\mathrm{d}x,$$

where $h : [0,1] \to [0,1]$ is an increasing function with $h(0) = 0$ and $h(1) = 1$, and $h$ is called the *distortion function* of $\rho$. Distortion risk measures are always monetary, positively homogeneous, and law invariant, and they are coherent if and only if their distortion functions are concave; see e.g., Yaari (1987) and Wang et al. (1997). For the application of distortion risk measures to insurance premium principle calculation, see Wang et al. (1997). For $X \in \mathcal{X}$ and $p \in (0,1]$, the *left-ES* risk measure (see e.g., Embrechts et al., 2015) is defined by

$$\mathrm{ES}_p^-(X) = \frac{1}{p}\int_0^p \mathrm{VaR}_t(X)\,\mathrm{d}t.$$

## 5.3 Optimal insurance contract design

In this section, we explain the optimal insurance design problem. For the economic setting, we make the following assumptions:

(A) The insured and the insurer may hold different attitudes towards risk. The insured adopts the risk measure $\rho : \mathcal{X} \to \mathbb{R}$ while the insurer uses the risk measure $\psi : \mathcal{X} \to \mathbb{R}$. The insured and the insurer do not observe the risk measure of their counterparty.

(B) The premium functional is specified as $\pi : \mathcal{I} \to \mathbb{R}$, which usually does not take negative values. For insurance loss $X \in \mathcal{X}_+$, note that $X - f(X) + \pi(f)$ is the total risk (i.e., total loss random variable) of the insured, and $f(X) - \pi(f)$ is the total risk of the insurer. Thus, the risk values of the insurance loss to the insured and the insurer are $\rho(X - f(X) + \pi(f))$ and $\psi(f(X) - \pi(f))$, respectively.

(C) The insured and the insurer agree on an insurance contract $f \in \mathcal{I}$ that is *Pareto optimal* defined next.

**Definition 5.1.** For $X \in \mathcal{X}_+$, $\pi : \mathcal{I} \to \mathbb{R}$, and $\rho, \psi : \mathcal{X} \to \mathbb{R}$, an insurance contract $f \in \mathcal{I}$ is called Pareto optimal if there is no $g \in \mathcal{I}$, such that

$$\rho(X - f(X) + \pi(f)) \geqslant \rho(X - g(X) + \pi(g))$$

and

$$\psi(f(X)-\pi(f)) \geqslant \psi(g(X)-\pi(g)),$$

with at least one of the two inequalities strict. Pareto optimality is also known as (Pareto) *efficiency*.

A Pareto optimization problem is closely related to the minimization of a convex combination of the objective functionals of all parties, which can be seen in, e.g., Gerber (1974), Barrieu and Scandolo (2008), Cai et al. (2017) and Embrechts et al. (2018). For $X \in \mathcal{X}_+$, $\pi : \mathcal{I} \to \mathbb{R}$, and $\rho, \psi : \mathcal{X} \to \mathbb{R}$, we define the set of minimizers of the sum of the two objectives for the insured and the insurer as

$$\mathcal{I}^X_{\rho,\psi} = \underset{g\in\mathcal{I}}{\arg\min}\{\rho(X-g(X)+\pi(g)) + \psi(g(X)-\pi(g))\}.$$

If we further assume that $\rho$ and $\psi$ are translation invariant, then we have

$$\mathcal{I}^X_{\rho,\psi} = \underset{g\in\mathcal{I}}{\arg\min}\{\rho(X-g(X)) + \psi(g(X))\}. \tag{5.1}$$

In this case, the set $\mathcal{I}^X_{\rho,\psi}$ is independent of the choice of the premium functional $\pi$. Below we give a characterization of the Pareto-optimal problem in our context as the minimization of the total insurance value of the insured and the insurer.

**Proposition 5.1.** *For two translation-invariant risk measures $\rho, \psi : \mathcal{X} \to \mathbb{R}$ and $X \in \mathcal{X}_+$, the following are equivalent:*

*(i) an insurance contract $f \in \mathcal{I}$ is Pareto optimal for all $\pi : \mathcal{I} \to \mathbb{R}_+$;*

*(ii) an insurance contract $f \in \mathcal{I}$ is Pareto optimal for $\pi : h \mapsto \psi(h(X))$;*[2]

*(iii) $f \in \mathcal{I}^X_{\rho,\psi}$.*

Proofs of all results in this chapter are in Section 5.6.

In a similar spirit to Proposition 5.1, a characterization of Pareto optimality in the context of risk sharing problems can be found in Embrechts et al. (2018). Proposition 5.1 ensures that if the objectives $\rho$ and $\psi$ for the two parties are translation invariant, then by (5.1), a Pareto-optimal insurance contract can typically be obtained by solving the following minimization problem:

$$\min_{g\in\mathcal{I}} \left\{\rho(X-g(X)) + \psi(g(X))\right\}. \tag{5.2}$$

[2]This means $\phi(h) = \psi(h(X))$ for all $h \in \mathcal{I}$.

A minimizer of (5.2) may not be unique in many situations. Hence, the set $\mathcal{I}^X_{\rho,\psi}$ of efficient ceded loss functions is not a singleton in general. In the literature on optimal insurance design problems, there are many common sets of ceded loss functions. Some notable refinements include:

1. The set $\mathcal{I}_0$ of all non-negative functions $f$ on $[0,\infty)$ satisfying $f(x) \leqslant x$ for $x \geqslant 0$. This property means that the payment cannot exceed the total loss incurred, and it is a common feature of almost all insurance contracts in practice. In particular, $f(0) = 0$, and thus there is no insurance payment if there is no loss incurred.

2. The set $\mathcal{I}_1$ of all increasing functions in $\mathcal{I}_0$. This property means that larger incurred losses lead to higher payments to the insured.

3. The set $\mathcal{I}_2 = \{f \in \mathcal{I}_1 : f(y) - f(x) \leqslant y - x \text{ for all } y \geqslant x \geqslant 0\}$, which is the set of all *slowly growing* increasing functions in $\mathcal{I}_1$. The slowly growing property is commonly assumed to avoid the problem of ex-post moral hazard (Huberman et al. (1983)) via the concept of comonotonicity; see Proposition 5.2 below.

4. The set $\mathcal{I}^d_1 = \{f \in \mathcal{I}_1 : f(x) \leqslant (x-d)_+ \text{ for all } x \geqslant 0\}$. Ceded loss functions within this set does not exceed the direct deductible form. Note that

   $$\mathcal{I}^d_1 = \{f \in \mathcal{I}_1 : f(d) = 0,\ x - f(x) \geqslant d \text{ for all } x > d\}.$$

   Thus this class includes contract functions with deductible $d \geqslant 0$. Also, we require that the retained loss of the insured should be at least at the deductible level $d$, given that the random loss exceeds the deductible level. In particular, we have $\mathcal{I}^0_1 = \mathcal{I}_1$.

Among the above sets, we have

$$\mathcal{I}_2 \subset \mathcal{I}_1 \subset \mathcal{I}_0 \quad \text{and} \quad \mathcal{I}^d_1 \subset \mathcal{I}_1 \subset \mathcal{I}_0.$$

Throughout, $\subset$ represents non-strict set inclusion. Contracts of deductible forms within the set $\mathcal{I}^d_1$ are commonly seen in the insurance market. We next give some examples.

**Example 5.1** (Deductible insurance with coinsurance)**.** Consider the following ceded loss function:

$$f(x) = \alpha(x-d)_+, \quad x \geqslant 0,$$

which presents an insurance contract with deductible $d \geqslant 0$ and coinsurance parameter $\alpha \in [0,1]$. We have $f \in \mathcal{I}^d_1$ since $f$ is bounded from above by $(x-d)_+$. See Figure 5.1 (left-hand panel).

**Example 5.2** (Deductible insurance with policy limit)**.** The following ceded loss function

$$f(x) = (x-d)_+ \wedge u, \quad x \geqslant 0,$$

is also in the set $\mathcal{I}_1^d$. It represents an insurance contract truncated at deductible $d \geqslant 0$ and censored at the policy upper limit $u \geqslant 0$. The function is plotted in Figure 5.1 (right-hand panel).

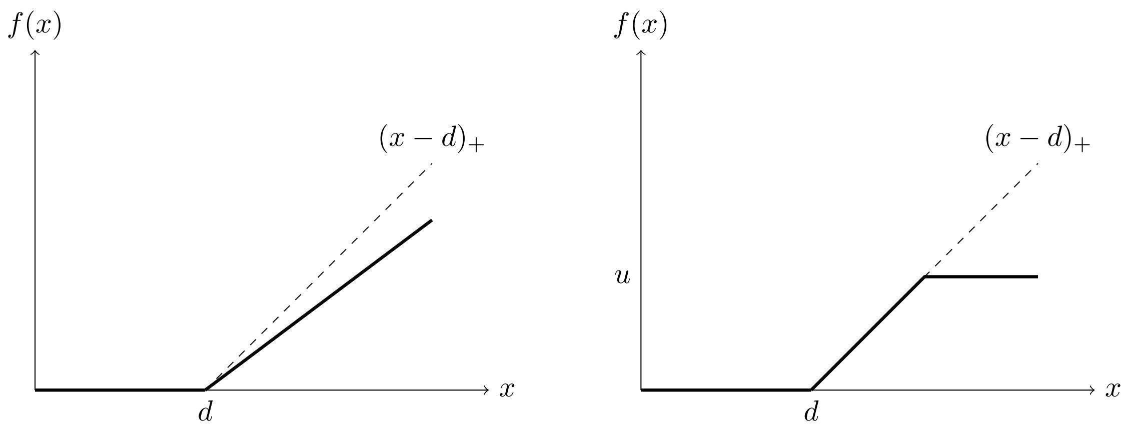


Figure 5.1: Solid lines represent the ceded loss functions of deductible insurance with coinsurance (left-hand panel) and deductible insurance with policy limit (right-hand panel); dashed lines represent ceded loss function with direct deductible

We focus on the above three subsets due to their prominence in real-world insurance contracts. Other subsets of $\mathcal{I}_1$, such as classes of convex functions, piece-wise linear functions, or functions with the Vajda condition, have also been studied in the literature, but they correspond to different practical considerations; see e.g., Vajda (1962), Cai et al. (2008), Chi and Weng (2013) and Chen (2021).

## 5.4 Risk measures implied by Pareto-optimal contracts

### 5.4.1 Main characterization results

In this section, we characterize measures $\rho$ and $\psi$ for the insured and the insurer in the optimal insurance design problem with different Pareto-optimal sets of ceded loss functions.

We first collect some dependence concepts that will be helpful to distinguish different properties of risk measures in our main results. A risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is said to be *comonotonic-additive* if

$\rho(X+Y)=\rho(X)+\rho(Y)$ for all comonotonic $(X,Y)\in\mathcal{X}^2$. Following similar definitions as those of Wang and Zitikis (2021), for an event $A\in\mathcal{F}$ with $0<\mathbb{P}(A)<1$, we call $A$ a *tail event* of a random variable $X\in\mathcal{X}$ if $X(\omega)\geqslant X(\omega')$ for almost surely all $\omega\in A$ and $\omega'\in A^c$. A tail event $A$ is called a *p-tail event* if $\mathbb{P}(A)=1-p$. We say that a random vector $(X_1,\dots,X_n)\in\mathcal{X}^n$ is *p-concentrated* if there exists a common $p$-tail event of $X_1,\dots,X_n$. For fixed $d\geqslant 0$ and $p\in[0,1]$, define the sets

$$\mathcal{X}_p^d=\{X\in\mathcal{X}_+ : p=\mathbb{P}(X\leqslant d)\}$$

and

$$\mathcal{X}_p=\{X\in\mathcal{X} : p=\mathbb{P}(X\leqslant d) \text{ for some } d\geqslant 0\}\supset\bigcup_{d\geqslant 0}\mathcal{X}_p^d.$$

We note that $\mathcal{X}_p\supset\mathcal{X}_p^d$ and $\mathcal{X}_p$ contains random variables that may take negative values and may be discrete. The following proposition connects the dependence structure of $(f(X),X-f(X))$ with the function $f\in\mathcal{I}_1$.

**Proposition 5.2.** *The following statements hold.*

*(i)* $(f(X),X-f(X))$ *is comonotonic for all* $f\in\mathcal{I}_2$ *and* $X\in\mathcal{X}_+$.

*(ii) For fixed* $d>0$ *and* $p\in[0,1)$, $(f(X),X-f(X))$ *is p-concentrated for all* $f\in\mathcal{I}_1^d$ *and* $X\in\mathcal{X}_p^d$.

Following the terminology in Chapter 4, for $\lambda\in\mathbb{R}$ and $p\in(0,1)$, we say that the linear combination

$$\mathrm{ES}_p^\lambda(X)=\lambda\mathrm{ES}_p(X)+(1-\lambda)\mathbb{E}[X],\quad X\in\mathcal{X}$$

of the mean and $\mathrm{ES}_p$ is an ES/$\mathbb{E}$-*mixture*. Note that we allow $\lambda<0$ in the definition of $\mathrm{ES}_p^\lambda$, so the ES/$\mathbb{E}$-mixture is not necessarily a monotone risk measure. Define the sets

$$\mathcal{I}_{\rho,\psi}=\bigcap_{X\in\mathcal{X}_+}\mathcal{I}_{\rho,\psi}^X \quad\text{and}\quad \mathcal{I}_{\rho,\psi}^{p,d}=\bigcap_{X\in\mathcal{X}_p^d}\mathcal{I}_{\rho,\psi}^X,$$

which are the intersections of all Pareto optimal contract sets with respect to all models of random losses in $\mathcal{X}_+$ and $\mathcal{X}_p^d$, respectively. Different choices of $\mathcal{I}_{\rho,\psi}$ pin down different forms of $\rho$ and $\psi$, as we will show below. Obviously, we shall arrive at a narrower class of risk measures as the set of efficient contracts enlarges.

**Theorem 5.1.** *Suppose that* $\rho$ *and* $\psi$ *are law-invariant convex risk measures. Then:*

*(i)* $\mathcal{I}_{\rho,\psi}=\mathcal{I}_2$ *if and only if* $\rho=\psi$ *and* $\rho$ *is a convex distortion risk measure on* $\mathcal{X}$*;*

*(ii)* $\mathcal{I}_{\rho,\psi} = \mathcal{I}_0$ *if and only if* $\rho = \psi = \mathbb{E}$ *on* $\mathcal{X}$*.*

Our next result, Theorem 5.2, establishes a relationship between deductible contracts and ES, and it is the most sophisticated result of the present chapter. The proofs of Theorems 5.1 and 5.2 are technical and rely on additional lemmas, which are presented in Section 5.6 together with proofs of the theorems.

**Theorem 5.2.** *Suppose that* $\rho$ *and* $\psi$ *are law-invariant convex risk measures with* $\rho(0) = \psi(0) = 0$*. For any fixed* $d \geqslant 0$ *and* $p \in [0, 1)$*, we have* $\mathcal{I}^{p,d}_{\rho,\psi} \supset \mathcal{I}^d_1$ *if and only if* $\rho = \psi = \mathrm{ES}^{\lambda}_p$ *on* $\mathcal{X}_p$ *for some* $\lambda \geqslant 0$*.*

We note that, given that the ceded loss functions in the set $\mathcal{I}^d_1$ are Pareto optimal for all insurance losses in the set $\mathcal{X}^d_p$, in Theorem 5.2 we can identify the risk measure adopted by the insured and the insurer as an ES/$\mathbb{E}$-mixture on a larger space of random losses $\mathcal{X}_p$, which does not depend on the deductible level $d$. In particular, the set $\mathcal{X}_p$ includes all random variables with continuous distributions on bounded supports.

Theorems 5.1 and 5.2 reveal profound connections between common sets of ceded loss functions and common classes of risk measures, as shown in Table 5.1.

| Sets of ceded loss functions | | Classes of risk measures |
|---|---|---|
| all 1-Lipschitz ceded loss functions | $\Longleftrightarrow$ | distortion risk measures |
| all non-negative ceded loss functions | $\Longleftrightarrow$ | the mean |
| ceded loss functions with deductible form | $\Longleftrightarrow$ | an ES/$\mathbb{E}$-mixture |

Table 5.1: Connections between sets of ceded loss functions and classes of risk measures

As one of the most important economic interpretations of the above results, we show that if the set of Pareto-optimal contracts between the insured and the insurer contains the set $\mathcal{I}^d_1$, then the risk measures of the two parties have to be an ES/$\mathbb{E}$-mixture. Furthermore, if the ES/$\mathbb{E}$-mixture in Theorem 5.2 satisfies lower semicontinuity with respect to almost sure convergence, then it has to be an ES; see Lemma 5.3.

If we remove some conditions from the convex risk measures $\rho$ in Theorems 5.1 and 5.2, then we arrive at larger classes of risk measures. For instance, without monotonicity in statement (i) of Theorem 5.1, we expect to arrive at the distortion riskmetrics in Chapter 2.

### 5.4.2 Designing insurance menus

In this section, we discuss economic implications of our characterization results of risk measures. We assume that the risk measures $\rho$ and $\psi$ for the insured and the insurer are coherent throughout this section.

Apart from the link between the common sets of ceded loss functions and the popular classes of risk measures, it is also interesting that all the three sets of Pareto-optimal contracts in Theorems 5.1 and 5.2 lead to the fact that the two risk measures $\rho$ and $\psi$ of the insured and the insurer are the same. In fact, when the risk measures $\rho$ and $\psi$ are coherent, a set of Pareto-optimal contracts with identical risk measures of the two parties is large enough to include all efficient contracts where the insurer is more optimistic than the insured, which can be seen from the next proposition. In this sense, the Pareto-optimal set that we obtain with identical risk measures is the union of Pareto-optimal sets with general risk measures $\rho \geqslant \psi$.

**Proposition 5.3.** *We have $\mathcal{I}^X_{\rho,\psi} \subset \mathcal{I}^X_{\psi,\psi}$ for all $X \in \mathcal{X}_+$ and all coherent risk measures $\rho$ and $\psi$ such that $\rho \geqslant \psi$.*

The relation $\rho \geqslant \psi$ in Proposition 5.3 indicates that the insured is more pessimistic, or more risk averse, than the insurer in the sense of Pratt (1964). Indeed, the certainty equivalent of any random loss $X$ under the preference described by the coherent risk measure $\rho$ is the risk measure $\rho(X)$ itself. Therefore, we compare risk aversion of the insured and the insurer through a direct comparison of magnitudes between coherent risk measures $\rho$ and $\psi$.

In practice, the insurer with the risk measure $\psi$ does not know the risk measure $\rho$ of the insured. Thus it is necessary for the insurer to provide a menu of contracts that is large enough to include all possible efficient contracts that might be chosen by the insured who is more pessimistic than the insurer. Specifically, we consider the following process for the design of insurance menus.

1. An insurer adopts the coherent risk measure $\psi$ as her own risk attitude.

2. The insurer does not have exact information about the risk attitudes of her customers. In other words, the insurer does not know the coherent risk measure $\rho$ held by any insured. However, in order to achieve the deal, the insured should be more pessimistic than the insurer (i.e. $\rho \geqslant \psi$).

3. Due to incomplete information, the insurer provides a menu of contracts $\mathcal{I}_{\psi,\psi}^X = \bigcup_{\rho\geqslant\psi} \mathcal{I}_{\rho,\psi}^X$ for a random loss $X \in \mathcal{X}_+$. The set $\mathcal{I}_{\psi,\psi}^X$ is large enough so that Pareto optimality can be obtained for any insured that is more pessimistic than the insurer. The deal can be achieved as long as we have $\rho \geqslant \psi$ since both parties benefit from the final deal.

4. If the insurer aims to design a "universal" menu of contracts so that Pareto optimality can be achieved for a bundle of random losses, the menu is then obtained by taking intersections of $\mathcal{I}_{\psi,\psi}^X$ with respect to a set of random losses. In this case, the insurer must choose specific classes of risk measures $\psi$, provided that the "universal" menu of contracts contains some common sets of contracts in the insurance market. Specifically, Table 5.2 illustrates our characterization results.

| Pareto-optimal menu | | Insurer's risk measure $\psi$ |
|---|---|---|
| $\mathcal{I}_{\psi,\psi} = \mathcal{I}_2$ | $\Longleftrightarrow$ | $\psi$ is a distortion risk measure |
| $\mathcal{I}_{\psi,\psi} = \mathcal{I}_0$ | $\Longleftrightarrow$ | $\psi = \mathbb{E}$ |
| $\mathcal{I}_{\psi,\psi}^{p,d} \supset \mathcal{I}_1^d$ | $\Longleftrightarrow$ | $\psi = \mathrm{ES}_p^\lambda$ |

Table 5.2: Connections between Pareto-optimal sets of contracts and the insurer's risk measures

## 5.5 Concluding remarks

In this chapter, the optimal insurance design problem is considered in the sense of Pareto optimality. Unlike existing studies, we solved a characterization problem of the risk measures of the insured and the insurer given the form of the Pareto-optimal contracts, and thus this chapter is in an opposite direction to the vast majority of the literature on optimal insurance. As our main finding, we are able to link the ES family, the most popular convex risk measures, to the set of ceded loss functions with a deductible form, commonly seen in insurance practice. It is not our intention to assert that ES dominates other convex risk measures in the insurance market, since there are so many other factors that need to be taken into account. Nevertheless, given the large volume of research based on ES in insurance and actuarial science, we hope that the present chapter brings in additional insights on why ES is a natural risk measure to use by the insurer when evaluating risks in the insurance market.

We note that our characterization results can be extended to the multi-player case with multiple insurers. This naturally links our study to the characterization of risk measures in risk sharing problems. Another potential application that can be further developed through our characterization results is that insurance companies may wish to evaluate risk attitudes of their customers based on contracts chosen from provided menus. This research direction requires more experimental studies as well as theoretical justifications. As yet another future direction, viewing the insured and the insurer as two economic agents in a competitive game, characterization problems may be explored via game theoretic approaches.

## 5.6 Proofs of main results and related technical lemmas

In this section, we present proofs of our main results as well as several related lemmas. As we will see, the results are technical and require highly sophisticated analysis.

### 5.6.1 Technical lemmas

We first collect technical lemmas that are related to, or are needed for proving, Theorems 5.1 and 5.2. We note in this regard that some parts of the proofs of the main results needed characterizations without assuming translation invariance. Hence, our next lemma characterizes risk measures $\rho$ and $\psi$ without this assumption and is restricted to the space $\mathcal{X}_0^0$. The lemma was used in the proof of Theorem 5.2.

**Lemma 5.1.** *Suppose that risk measures $\rho$ and $\psi$ are law invariant, monotone, convex and uniformly continuous with respect to $L^\infty$-norm. Then we have the following two characterization results.*

*(i) The inclusion*

$$\bigcap_{X\in\mathcal{X}_0^0} \arg\min_{g\in\mathcal{I}}\{\rho(X-g(X))+\psi(g(X))\} \supset \mathcal{I}_2$$

*holds if and only if*

$$\rho(X)=\psi(X)=\int_0^\infty h(\mathbb{P}(X\geqslant x))\,\mathrm{d}x \tag{5.3}$$

*for all $X\in\mathcal{X}_0^0$, where $h:[0,1]\to[0,\infty)$ is an increasing concave function with $h(0)=0$.*[3]

[3] Functionals of form (5.3) belong to the family of distortion riskmetrics in Chapter 2 with increasing distortion functions.

*(ii) The inclusion*

$$\bigcap_{X \in \mathcal{X}_0^0} \underset{g \in \mathcal{I}}{\arg\min}\{\rho(X - g(X)) + \psi(g(X))\} \supset \mathcal{I}_1$$

*holds if and only if $\rho = \psi = \lambda\mathbb{E}$ on $\mathcal{X}_0^0$ for some $\lambda \geqslant 0$.*

*Proof.* (i) For convenience of the proof for Theorem 5.1, we prove the result on space $\mathcal{X}_+$, and the proof of (i) holds by directly changing $\mathcal{X}_+$ to $\mathcal{X}_0^0$.

"$\Rightarrow$": Suppose that $\mathcal{I}_{\rho,\psi} \supset \mathcal{I}_2$. Let $h_0(x) = 0$ and $h_1(x) = x$, $x \geqslant 0$, the constant zero function and the identity, respectively. Since $h_0, h_1 \in \mathcal{I}_2$, we have

$$\rho(X) = \psi(X) = \min_{g \in \mathcal{I}} \{\rho(X - g(X)) + \psi(g(X))\}, \quad X \in \mathcal{X}_+.$$

Hence $\rho = \psi$ on $\mathcal{X}_+$ and $\rho(X - f(X)) + \rho(f(X)) = \rho(X)$ for all $f \in \mathcal{I}_2$ and $X \in \mathcal{X}_+$.

By Proposition 4.5 of Denneberg (1994), for any comonotonic $(Y, Z) \in \mathcal{X}_+^2$ with $Y + Z = X$, there exists $f \in \mathcal{I}_2$ such that $Y = f(X)$ and $Z = X - f(X)$. Since $X$ is arbitrary, we therefore have the equation $\rho(Y) + \rho(Z) = \rho(Y + Z)$ for all comonotonic $(Y, Z) \in \mathcal{X}_+^2$. This shows that $\rho$ is comonotonic-additive on $\mathcal{X}_+$. Thus (5.3) holds by Theorems 1 and 3 of Wang et al. (2020).

"$\Leftarrow$": Suppose that $\rho$ and $\psi$ satisfy (5.3) on $\mathcal{X}_+$. For all $f \in \mathcal{I}_2$ and $X \in \mathcal{X}_+$, we have by Proposition 5.2 that $(f(X), X - f(X))$ is comonotonic. By comonotonic-additivity of $\rho$, we have $\rho(X - f(X)) + \rho(f(X)) = \rho(X)$. Furthermore, due to subadditivity of $\rho$, we have $f \in \mathcal{I}_{\rho,\rho}$. It follows that $\mathcal{I}_2 \subset \mathcal{I}_{\rho,\rho}$.

(ii) The "if" part is straightforward by linearity of the mean. We prove the "only if" part. Since $\mathcal{I}_1 \supset \mathcal{I}_2$, by (i), we have $\rho(X) = \psi(X) = \int_0^\infty h(\mathbb{P}(X \geqslant x))\,\mathrm{d}x$ for all $X \in \mathcal{X}_0^0$. By Theorem 2.5 in Chapter 2, there is a finite Borel measure $\mu$ on $[0,1]$ such that $\rho(X) = \int_0^1 \mathrm{ES}_\alpha(X)\,\mu(\mathrm{d}\alpha)$ for $X \in \mathcal{X}_0^0$. For all $0 < \alpha \leqslant 1$, there exists differentiable $f \in \mathcal{I}_1$ such that $f'(x) \leqslant 1$ for all $x \in [0, \mathrm{VaR}_\alpha(X))$ and $f'(x) > 1$ for all $x \in [\mathrm{VaR}_\alpha(X), \infty)$. Thus $x \mapsto x - f(x)$ is increasing on $[0, \mathrm{VaR}_\alpha(X))$ and decreasing on $[\mathrm{VaR}_\alpha(X), \infty)$ in strict sense. According to Lemma A.3 and Lemma A.7 of Wang and Zitikis (2021), we have a $p$-tail event $A$ of $X$ and $f(X)$ with

$$\{X > \mathrm{VaR}_\alpha(X)\} \subset A \subset \{X \geqslant \mathrm{VaR}_\alpha(X)\}$$

such that

$$\mathrm{ES}_\alpha(X) = \mathbb{E}[X|A] \;\text{ and }\; \mathrm{ES}_\alpha(f(X)) = \mathbb{E}[f(X)|A].$$

On the other hand, for a $p$-tail event $B$ of $X - f(X)$ satisfying

$$\{X - f(X) > \mathrm{VaR}_\alpha(X - f(X))\} \subset B \subset \{X - f(X) \geqslant \mathrm{VaR}_\alpha(X - f(X))\},$$

we have

$$\mathrm{ES}_\alpha(X - f(X)) = \mathbb{E}[X - f(X)|B] > \mathbb{E}[X - f(X)|A].$$

Thus we have

$$\mathrm{ES}_\alpha(f(X)) + \mathrm{ES}_\alpha(X - f(X)) > \mathbb{E}[f(X)|A] + \mathbb{E}[X - f(X)|A] = \mathbb{E}[X|A] = \mathrm{ES}_\alpha(X)$$

and so

$$\begin{aligned}\rho(f(X)) + \rho(X - f(X)) &= \int_0^1 \mathrm{ES}_\alpha(f(X)) + \mathrm{ES}_\alpha(X - f(X))\,\mu(\mathrm{d}\alpha)\\ &> \int_0^1 \mathrm{ES}_\alpha(X)\,\mu(\mathrm{d}\alpha) = \rho(X),\end{aligned}$$

which leads to a contradiction. Hence, $\mu((0,1]) = 0$ and $\rho(X) = \psi(X) = \lambda\mathbb{E}[X]$ for some $\lambda \geqslant 0$ and for all $X \in \mathcal{X}_0^0$. □

The next lemma characterizes an ES/$\mathbb{E}$-mixture. The lemma implies that a law-invariant convex risk measure dominated by an ES/$\mathbb{E}$-mixture must be the ES/$\mathbb{E}$-mixture itself provided that it coincides with the ES/$\mathbb{E}$-mixture somewhere. We used the lemma when proving Theorem 5.2.

**Lemma 5.2.** *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant convex risk measure. Fix $d \geqslant 0$ and $p \in (0,1)$. We have $\rho(X) = \rho((X-d)_+) + \rho(X \wedge d)$ and $\rho(X \vee d) = \lambda\mathrm{ES}_p(X) + (1-\lambda)d$ for all $X \in \mathcal{X}_p^d$ with $\lambda \in \mathbb{R}$ if and only if $\rho(X) = \lambda\mathrm{ES}_p(X) + (1-\lambda)\mathrm{ES}_p^-(X)$ for all $X \in \mathcal{X}_p^d$ with $\lambda \geqslant 1-p$.*

*Proof.* The "if" part follows immediately from the definitions of $\mathrm{ES}_p$ and $\mathrm{ES}_p^-$. Hence, we prove the "only if" part.

Since $\rho$ is a law-invariant convex risk measure, for all $X \in \mathcal{X}_p^d$ we write

$$\rho(X) = \sup_{Z \in \mathcal{Q}}\{\mathbb{E}[ZX] + V(Z)\},$$

where $\mathcal{Q}$ is a set of Radon-Nikodym derivatives and $V$ is a mapping from $\mathcal{Q}$ to $[-\infty, 0]$ (see e.g., Jouini et al., 2006). We first show that $Z \leqslant \lambda/(1-p)$ for all $Z \in \mathcal{Q}$. Assume for the sake of contradiction that $\mathbb{P}(Z' > \lambda/(1-p)) > 0$ for some $Z' \in \mathcal{Q}$. Take $A \subset \{Z' > \lambda/(1-p)\}$ and $Y = \mathbb{1}_A(d+1)\gamma + \mathbb{1}_B(d+1)$ for $\gamma > 1$, where $\mathbb{P}(A \cup B) = 1-p$ and $A \cap B = \emptyset$. It is clear that

$Y \in \mathcal{X}_p^d$. We have

$$\begin{aligned}\sup_{Z\in\mathcal{Q}}\{\mathbb{E}[ZY]+V(Z)\} &\geqslant \mathbb{E}[Z'(\mathbb{1}_A(d+1)\gamma+\mathbb{1}_B(d+1))]+V(Z')\\ &\geqslant (d+1)\gamma\mathbb{E}[Z'\mathbb{1}_A]+V(Z')\\ &= \mathbb{E}[Z'|A]\mathbb{E}[\mathbb{1}_A(d+1)\gamma]+V(Z').\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\lambda\mathrm{ES}_p(Y)+(1-\lambda)d &= \lambda\mathrm{ES}_p(\mathbb{1}_A(d+1)\gamma+\mathbb{1}_B(d+1))+(1-\lambda)d\\ &= \frac{\lambda}{1-p}\mathbb{E}[\mathbb{1}_A(d+1)\gamma+\mathbb{1}_B(d+1)]+(1-\lambda)d.\end{aligned}$$

Since $\mathbb{E}[Z'|A] > \lambda/(1-p)$, we have

$$\lim_{\gamma\to\infty}(\mathbb{E}[Z'|A]\mathbb{E}[\mathbb{1}_A(d+1)\gamma]+V(Z')) > \lim_{\gamma\to\infty}(\lambda\mathrm{ES}_p(Y)+(1-\lambda)d),$$

which contradicts the assumption that $\rho(X) \leqslant \lambda\mathrm{ES}_p(X)+(1-\lambda)d$ for all $X \in \mathcal{X}_p^d$. Therefore, we have $Z \leqslant \lambda/(1-p)$ for all $Z \in \mathcal{Q}$. On the other hand, since $\mathbb{E}[Z]=1$, we have $\lambda/(1-p) \geqslant 1$ and thus $\lambda \geqslant 1-p$.

We next show that $\rho(X) = \lambda\mathrm{ES}_p(X)+(1-\lambda)\mathrm{ES}_p^-(X)$ for all $X \in \mathcal{X}_p^d$. Note that $\{X>d\}$ is a common $p$-tail event of $X$ and $X\vee d$. We have $\mathrm{ES}_p(X)=\mathrm{ES}_p(X\vee d)$ and

$$d = \frac{1}{p}\mathbb{E}[(X\vee d)\mathbb{1}_{\{X\leqslant d\}}] = \mathrm{ES}_p^-(X\vee d).$$

It follows that

$$\begin{aligned}\sup_{Z\in\mathcal{Q}}\{\mathbb{E}[Z(X\vee d)]+V(Z)\} &= \rho(X\vee d)\\ &= \lambda\mathrm{ES}_p(X)+(1-\lambda)d = \lambda\mathrm{ES}_p(X\vee d)+(1-\lambda)\mathrm{ES}_p^-(X\vee d).\end{aligned}$$

For $X_1, X_2, \ldots \in \mathcal{X}_p^d$ and $X_n \downarrow X$, since $Z$ is non-negative and bounded from above by $1/(1-p)$, the dominated convergence theorem implies

$$\lim_{n\to\infty}\sup_{Z\in\mathcal{Q}}\{\mathbb{E}[ZX_n]+V(Z)\} = \sup_{Z\in\mathcal{Q}}\{\mathbb{E}[ZX]+V(Z)\},$$

which means that $\rho$ is continuous from above. Hence,

$$\rho(X) = \max_{Z\in\mathcal{Q}}\{\mathbb{E}[ZX]+V(Z)\}$$

for all $X \in \mathcal{X}_p^d$; see e.g., Corollary 4.35 of Föllmer and Schied (2016). It follows that there exists $Z_0 \in \mathcal{Q}$ such that

$$\mathbb{E}[Z_0(X\vee d)]+V(Z_0) = \frac{\lambda}{1-p}\mathbb{E}[(X\vee d)\mathbb{1}_{\{X>d\}}]+\frac{1-\lambda}{p}\mathbb{E}[(X\vee d)\mathbb{1}_{\{X\leqslant d\}}]. \tag{5.4}$$

We claim that $Z_0 = \lambda \mathbb{1}_{\{X>d\}}/(1-p)+(1-\lambda)\mathbb{1}_{\{X\leqslant d\}}/p$. Indeed, assume for the sake of contradiction that $Z_0 \neq \lambda \mathbb{1}_{\{X>d\}}/(1-p)+(1-\lambda)\mathbb{1}_{\{X\leqslant d\}}/p$. Since

$$\mathbb{E}[Z_0] = 1 = \mathbb{E}\left[\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} + \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right],$$

we have

$$\mathbb{P}\left(\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_+ > 0\right) > 0$$

and

$$\mathbb{P}\left(\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_- > 0\right) > 0.$$

Note that $\lambda/(1-p) \geqslant 1 \geqslant (1-\lambda)/p$. Hence,

$$\left\{\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_+ > 0\right\} \subset \{X \leqslant d\}.$$

We also note that

$$\left\{\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_- > 0\right\} \cap \{X > d\} \neq \emptyset.$$

Otherwise, we must have $Z_0 = \lambda/(1-p)$ and

$$\mathbb{P}\left(\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_- > 0\right) = 0,$$

which leads to contradiction. These considerations imply that

$$\begin{aligned}
&\mathbb{E}\left[\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)(X \vee d)\right] \\
&= \mathbb{E}\left[\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_+ (X \vee d)\right] \\
&\quad - \mathbb{E}\left[\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_- (X \vee d)\right] \\
&< d\left(\mathbb{E}\left[\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_+\right] - \mathbb{E}\left[\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X\leqslant d\}}\right)_-\right]\right) \\
&= 0,
\end{aligned}$$

which contradicts equation (5.4). Therefore, we must have $Z_0 = \lambda \mathbb{1}_{\{X>d\}}/(1-p)+(1-\lambda)\mathbb{1}_{\{X\leqslant d\}}/p$. Hence, $Z_0 = \lambda \mathbb{1}_{\{X>d\}}/(1-p)+(1-\lambda)\mathbb{1}_{\{X\leqslant d\}}/p \in \mathcal{Q}$ and $V(Z_0) = 0$. It follows that

$$\sup_{Z\in\mathcal{Q}}\{\mathbb{E}[ZX] + V(Z)\} \geqslant \mathbb{E}\left[\frac{\lambda}{1-p}X\mathbb{1}_{\{X>d\}} + \frac{1-\lambda}{p}X\mathbb{1}_{\{X\leqslant d\}}\right] = \lambda \mathrm{ES}_p(X) + (1-\lambda)\mathrm{ES}_p^-(X).$$

On the other hand, we have

$$\rho(X) \leqslant \lambda \mathrm{ES}_p(X) + (1-\lambda)d = \gamma \mathrm{ES}_p(X) + (1-\gamma)\mathrm{ES}_p^-(X),$$

for some $1-p \leqslant \lambda \leqslant \gamma \leqslant 1$ since $\mathrm{ES}_p^-(X) \leqslant d \leqslant \mathrm{ES}_p(X)$. Hence, there exists $\lambda \leqslant \lambda' \leqslant \gamma$, such that $\rho(X) = \lambda'\mathrm{ES}_p(X) + (1-\lambda')\mathrm{ES}_p^-(X)$.

Take $X_m = X\mathbb{1}_{\{X\leqslant d\}} + (X+m)\mathbb{1}_{\{X>d\}}$ for $m>0$. We have $X_m \in \mathcal{X}_p^d$. For some $\lambda_m \in [\lambda, 1]$,

$$\rho(X_m) = \lambda_m\mathrm{ES}_p(X_m) + (1-\lambda_m)\mathrm{ES}_p^-(X_m) = \lambda_m\mathrm{ES}_p(X) + \lambda_m m + (1-\lambda_m)\mathrm{ES}_p^-(X). \tag{5.5}$$

Since $\rho(X_m \vee d) = \lambda\mathrm{ES}_p(X) + \lambda m + (1-\lambda)d$, this implies that there exists $m>0$ such that $\lambda_m = \lambda$. Indeed, otherwise we can take $m \to \infty$ and have a contradiction to $\rho(X_m) \leqslant \rho(X_m \vee d)$ by monotonicity of $\rho$. On the other hand, for $m$ such that $\lambda_m = \lambda$, we have

$$\begin{aligned}\rho(X_m) = \rho(X_m \vee d) - d + \rho(X_m \wedge d) &= \lambda\mathrm{ES}_p(X) + \lambda m - \lambda d + \rho(X \wedge d)\\ &= \rho(X) + \lambda m = \lambda'\mathrm{ES}_p(X) + (1-\lambda')\mathrm{ES}_p^-(X) + \lambda m.\end{aligned} \tag{5.6}$$

Equations (5.5) and (5.6), together with $\lambda_m = \lambda$, yield that $\lambda' = \lambda$ for all $X \in \mathcal{X}_p^d$. This completes the proof. $\square$

Finally, we give a lemma on properties of ES/$\mathbb{E}$-mixtures that can precisely pin down the family of ES within the class of ES/$\mathbb{E}$-mixtures obtained in Theorem 5.2.

**Lemma 5.3.** *For an* ES/$\mathbb{E}$*-mixture* $\rho = \lambda\mathrm{ES}_p + (1-\lambda)\mathbb{E}$*, we have the following statements:*

*(i) $\rho$ is lower semicontinuous with respect to almost sure convergence if and only if $\lambda \geqslant 1$;*

*(ii) $\rho$ is convex if and only if $\lambda \geqslant 0$;*

*(iii) $\rho$ is monotone if and only if $\lambda \in [1-1/p, 1]$.*

*In particular, $\rho$ is monotone and lower semicontinuous with respect to almost sure convergence if and only if it is* $\mathrm{ES}_p$*.*

*Proof.* (i) Suppose that $\lambda < 1$. Let $X_k = -k\mathbb{1}_{\{U<1/k\}}$, where $U \sim \mathrm{U}[0,1]$. Clearly, $X_k \to 0$ almost surely as $k \to \infty$, $\mathbb{E}[X_k] = -1$, and $\mathrm{ES}_p(X_k) = 0$ for $k > 1/p$. Therefore,

$$\liminf_{k\to\infty} \left((1-\lambda)\mathbb{E}[X_k] + \lambda\mathrm{ES}_p(X_k)\right) = -(1-\lambda) < 0 = \rho(0),$$

contradicting lower semicontinuity.

(ii) We note that $\rho$ is a signed Choquet integral of Wang et al. (2020) with the (not necessarily increasing) distortion function

$$h(t) = \lambda\left(\frac{t}{1-p} \wedge 1\right) + (1-\lambda)t, \quad t \in [0,1].$$

By Theorem 3 of Wang et al. (2020), $\rho$ is convex if and only if $h$ is concave. It is straightforward to verify that $h$ is concave if and only if $\lambda \geqslant 0$.

(iii) By Lemma 1 (i) of Wang et al. (2020), $\rho$ is monotone if and only if $h$ is increasing. Clearly, $\lambda > 1$ implies that $h$ is strictly decreasing on $(1-p, 1]$. For $\lambda \leqslant 1$, increasing monotonicity of $h$ is equivalent to

$$\frac{\lambda}{1-p} + 1 - \lambda \geqslant 0 \iff \lambda \geqslant 1 - \frac{1}{p}.$$

Hence, $\rho$ is monotone if and only if $\lambda \in [1 - 1/p, 1]$. □

### 5.6.2 Proofs of all results

*Proof of Proposition 5.1.* "(i)⇒(ii)": This is straightforward by taking $\pi : h \mapsto \psi(h(X))$.

"(ii)⇒(iii)": Suppose that $f \in \mathcal{I}$ is Pareto optimal for $\pi : h \mapsto \psi(h(X))$. Assume for the sake of contradiction that $f \notin \mathcal{I}^X_{\rho,\psi}$. It follows that there exists $g \in \mathcal{I}$, such that

$$\rho(X - g(X)) + \psi(g(X)) < \rho(X - f(X)) + \psi(f(X)).$$

By translation invariance of $\rho$ and $\psi$, we have

$$\begin{aligned}\rho(X - g(X) + \pi(g)) &= \rho(X - g(X)) + \psi(g(X)) \\ &< \rho(X - f(X)) + \psi(f(X)) = \rho(X - f(X) + \pi(f))\end{aligned}$$

and

$$\psi(g(X) - \pi(g)) = \psi(g(X) - \psi(g(X))) = 0 = \psi(f(X) - \pi(f)),$$

which leads to a contradiction to Pareto optimality of $f$. Therefore, $f \in \mathcal{I}^X_{\rho,\psi}$.

"(iii)⇒(i)": Suppose that $f \in \mathcal{I}^X_{\rho,\psi}$. Assume for the sake of contradiction that $f$ is not Pareto optimal for some $\pi : \mathcal{I} \to \mathbb{R}$. It follows that there exists $g \in \mathcal{I}$ such that

$$\rho(X - g(X) + \pi(g)) \leqslant \rho(X - f(X) + \pi(f))$$

and

$$\psi(g(X) - \pi(g)) \leqslant \psi(f(X) - \pi(f)),$$

with at least one of the above two inequalities being strict. Hence,

$$\rho(X - g(X) + \pi(g)) + \psi(g(X) - \pi(g)) < \rho(X - f(X) + \pi(f)) + \psi(f(X) - \pi(f)),$$

which contradicts the fact that $f \in \mathcal{I}^X_{\rho,\psi}$. Therefore, the function $f$ is Pareto optimal for all $\pi : \mathcal{I} \to \mathbb{R}$. □

*Proof of Proposition 5.2.* (i) Suppose that $f \in \mathcal{I}_2$. Define the function $g$ by $g(x) = x - f(x)$ for $x \in [0, \infty)$. For all $X \in \mathcal{X}_+$, we have $X - f(X) = g(X)$. Since $f \in \mathcal{I}_2$, the function $g$ is increasing and $(f(X), g(X))$ is comonotonic.

(ii) Suppose that $f \in \mathcal{I}_1^d$ for $d > 0$. For all $X \in \mathcal{X}_p^d$, the set $\{X > d\}$ is a common tail event of $f(X)$ and $X - f(X)$ by the definitions of the tail event and the set $\mathcal{I}_1^d$. Also note that $\mathbb{P}(X > d) = 1 - p$. Therefore, $(f(X), X - f(X))$ is $p$-concentrated. $\square$

Here we present the proof of Theorem 5.2 first because it is useful for the proof of Theorem 5.1.

*Proof of Theorem 5.2.* "$\Leftarrow$": For all $f \in \mathcal{I}_1^d$, note that $(f(X), X - f(X))$ is $p$-concentrated for all $X \in \mathcal{X}_p^d$ by Proposition 5.2. By $p$-additivity of $\mathrm{ES}_p$ (see Wang and Zitikis, 2021), we have $\mathrm{ES}_p(X - f(X)) + \mathrm{ES}_p(f(X)) = \mathrm{ES}_p(X)$ and thus $f \in \mathcal{I}^{d,p}_{\mathrm{ES}_p,\mathrm{ES}_p}$. Hence $\mathcal{I}^{d,p}_{\mathrm{ES}_p,\mathrm{ES}_p} \supset \mathcal{I}_1^d$.

"$\Rightarrow$": It suffices to show that $\rho = \psi = \mathrm{ES}_p^\lambda$ on $\mathcal{X}_p^d$ for some $\lambda \geqslant 0$, and that $\rho = \psi = \mathrm{ES}_p^\lambda$ on $\mathcal{X}_p$ holds due to translation invariance of $\rho$ and $\psi$. Write $h_d(x) = (x - d)_+$, $x \geqslant 0$, for all $d \geqslant 0$ and recall that $h_0(x) = 0$, $x \geqslant 0$. Since $h_0, h_d \in \mathcal{I}_1^d$, we have

$$\rho(X) = \rho(X \wedge d) + \psi((X - d)_+) = \min_{g \in \mathcal{I}} \{\rho(X - g(X)) + \psi(g(X))\} \tag{5.7}$$

for all $X \in \mathcal{X}_p^d$.

We first prove the case when $d = p = 0$. We know from Lemma 5.1 that $\rho(X) = \psi(X) = \lambda\mathbb{E}[X]$ for some $\lambda \geqslant 0$ and for all $X \in \mathcal{X}_0^0$. Since $\rho$ is translation invariant and $X + c \in \mathcal{X}_0^0$ for all $X \in \mathcal{X}_0^0$ and $c \geqslant 0$, we have

$$\lambda\mathbb{E}[X] + c = \rho(X) + c = \rho(X + c) = \lambda\mathbb{E}[X + c] = \lambda\mathbb{E}[X] + \lambda c.$$

It follows that $\lambda = 1$.

We now prove the case when $d = 0$ and $p \in (0, 1)$. We know from statement (5.7) that $\rho = \psi$ on $X_p^0$. For all $X \in \mathcal{X}_0^0$, we define $\phi(X) = \rho(X\mathbb{1}_A)$ by taking an event $A$ independent of $X$ with $\mathbb{P}(A) = 1 - p$ (a specific choice of $A$ does not matter since $\rho$ is law invariant). It is clear that $\phi$ is law invariant, monotone, convex and uniformly continuous with respect to $L^\infty$-norm. Note that for all $X \in \mathcal{X}_0^0$ and all events $B$ and $C$ independent of $X$ with $\mathbb{P}(B) = \mathbb{P}(C) = 1 - p$, we have $X\mathbb{1}_B \overset{\mathrm{d}}{=} X\mathbb{1}_C$. Hence, $\phi(X) = \rho(X\mathbb{1}_B) = \rho(X\mathbb{1}_C)$ and thus $\phi$ is well defined. Since $X\mathbb{1}_A \in \mathcal{X}_p^0$ and

$\mathcal{I}_{\rho,\psi}^{0,p} \supset \mathcal{I}_1$, we have

$$\begin{aligned}\phi(f(X)) + \phi(X - f(X)) &= \rho(f(X)\mathbb{1}_A) + \rho((X - f(X))\mathbb{1}_A) \\ &= \rho(f(X\mathbb{1}_A)) + \rho(X\mathbb{1}_A - f(X\mathbb{1}_A)) = \rho(X\mathbb{1}_A) = \phi(X)\end{aligned}$$

for all $f \in \mathcal{I}_1$ and $X \in \mathcal{X}_0^0$. It follows from Lemma 5.1 that $\phi(X) = \lambda\mathbb{E}[X]$ for some $\lambda \geqslant 0$ and for all $X \in \mathcal{X}_0^0$. For all $X \in \mathcal{X}_p^0$, we take any random variable $Y$ such that $Y \overset{\mathrm{d}}{=} X|X > 0$. We have $Y \in \mathcal{X}_0^0$ and $X\mathbb{1}_{\{X>0\}} \overset{\mathrm{d}}{=} Y\mathbb{1}_A$. Thus

$$\rho(X\mathbb{1}_{\{X>0\}}) = \rho(Y\mathbb{1}_A) = \lambda\mathbb{E}[Y] = \lambda\mathrm{ES}_p(X).$$

It follows that

$$\rho(X) = \psi(X) = \rho(X\mathbb{1}_{\{X>0\}}) + \rho(X\mathbb{1}_{\{X=0\}}) = \lambda\mathrm{ES}_p(X) \tag{5.8}$$

for all $X \in \mathcal{X}_p^0$. Note that for all $X \in \mathcal{X}_p^0$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\mathbb{E}[X\mathbb{1}_{X>0}] = \frac{1}{1-p}\mathbb{E}[X].$$

Hence, we have

$$\rho(X) = \lambda'\mathrm{ES}_p(X) + (1 - \lambda')\mathbb{E}[X]$$

for all $X \in \mathcal{X}_p^0$, where $\lambda' = (\lambda - 1 + p)/p$. By equation (5.8) and Lemma 5.2, we have $\lambda \geqslant 1 - p$ and thus $\lambda' \geqslant 0$.

Next, we prove the case when $d > 0$ and $p = 0$. For all $X \in \mathcal{X}_0^0$, we have $X + d \in \mathcal{X}_0^d$. We obtain from $\mathcal{I}_{\rho,\psi}^{d,0} \supset \mathcal{I}_1^d$ that

$$\rho(X + d - f(X + d)) + \psi(f(X + d)) = \rho(X + d) \tag{5.9}$$

for all $f \in \mathcal{I}_1^d$. Take those $f$ that are of the form $f(x) = g(x - d)$ for any $g \in \mathcal{I}_1$ and all $x \geqslant d$. Noting that $\rho$ is translation invariant, we have

$$\rho(X - g(X)) + \psi(g(X)) = \rho(X) \tag{5.10}$$

for all $g \in \mathcal{I}_1$. Hence, $\rho(X) = \psi(X) = \lambda\mathbb{E}[X] = \lambda\mathrm{ES}_0(X)$ for some $\lambda \geqslant 0$ and for all $X \in \mathcal{X}_0^0$ by Lemma 5.1. Since $\rho$ is translation invariant, we have $\lambda = 1$.

We finally prove the case when $d > 0$ and $p \in (0, 1)$. For all $X \in \mathcal{X}_p^0$, we have $X + d \in \mathcal{X}_p^d$. Following similar arguments as those we used to derive equations (5.9) and (5.10), we obtain

$$\rho(X - g(X)) + \psi(g(X)) = \rho(X)$$

for all $g \in \mathcal{I}_1$. Hence, $\rho(X) = \psi(X) = \lambda \mathrm{ES}_p(X)$ for some $\lambda \geqslant 1-p$ and for all $X \in \mathcal{X}_p^0$ by equation (5.8). For all $X \in \mathcal{X}_p^d$, we have $(X-d)_+ \in \mathcal{X}_p^0$. Therefore,

$$\rho((X-d)_+) = \psi((X-d)_+) = \lambda \mathrm{ES}_p[(X-d)_+] = \lambda(\mathrm{ES}_p(X) - d).$$

Hence, $\rho(X \vee d) = \rho((X-d)_+ + d) = \lambda \mathrm{ES}_p(X) + (1-\lambda)d$ and $\psi(X \vee d) = \lambda \mathrm{ES}_p(X) + (1-\lambda)d$. By Lemma 5.2, we have $\rho(X) = \psi(X) = \lambda \mathrm{ES}_p(X) + (1-\lambda)\mathrm{ES}_p^-(X)$ for all $X \in \mathcal{X}_p^d$. Since

$$(1-p)\mathrm{ES}_p(X) + p\mathrm{ES}_p^-(X) = \mathbb{E}[X],$$

we have $\rho(X) = \psi(X) = \gamma \mathrm{ES}_p(X) + (1-\gamma)\mathbb{E}[X]$, where $\gamma = 1 - (1-\lambda)/p \geqslant 0$. □

*Proof of Theorem 5.1.* Let $h_0(x) = 0$ and $h_1(x) = x$, $x \geqslant 0$, the constant zero function and the identity, respectively.

(i) "$\Rightarrow$": Suppose that $\mathcal{I}_{\rho,\psi} = \mathcal{I}_2$. By the proof of Lemma 5.1 (i) and translation invariance of $\rho$ and $\psi$, we have $\rho = \psi$ on $\mathcal{X}$ and $\rho$ is comonotonic-additive on $\mathcal{X}$. Moreover, we know that $\rho$ is uniformly continuous with respect to $L^\infty$-norm since $\rho$ is monetary, and $\rho$ is law invariant. Hence, $\rho$ is a convex distortion risk measure on $\mathcal{X}$ (see e.g., Kusuoka, 2001).

"$\Leftarrow$": Suppose that $\rho = \psi$ is a convex distortion risk measure on $\mathcal{X}$. We will prove that $\mathcal{I}_{\rho,\rho} = \mathcal{I}_2$. Since $\rho$ is a convex distortion risk measure, it is also coherent by e.g., Corollary 2.1 of Chapter 2; see Acerbi (2002). Following the same logic as the proof of Lemma 5.1 (i), we have $\mathcal{I}_2 \subset \mathcal{I}_{\rho,\rho}$.

We next prove that $\mathcal{I}_{\rho,\rho} \subset \mathcal{I}_2$. For each $f \notin \mathcal{I}_2$, we will show that there exists $X \in \mathcal{X}_+$ such that $\rho(X - f(X)) + \rho(f(X)) > \rho(X)$. Indeed, there exists $0 \leqslant x < y$, such that $|f(y) - f(x)| > y - x$. It is clear that $f(x) \neq f(y)$. Since $\rho$ is a coherent distortion risk measure, there exists a Borel measure $\mu$ on $[0,1]$ such that $\rho = \int_0^1 \mathrm{ES}_t \,\mathrm{d}\mu(t)$ on $\mathcal{X}$. Take $X = x\mathbb{1}_A + y\mathbb{1}_{A^c}$ where $\mathbb{P}(A) = 1/2$. If $f(x) < f(y)$, then

$$\mathrm{ES}_t(X) = \begin{cases} \frac{(1-2t)x+y}{2-2t}, & 0 \leqslant t \leqslant 1/2, \\ y, & 1/2 < t < 1, \end{cases}$$

$$\mathrm{ES}_t(f(X)) = \begin{cases} \frac{(1-2t)f(x)+f(y)}{2-2t}, & 0 \leqslant t \leqslant 1/2, \\ f(y), & 1/2 < t < 1, \end{cases}$$

$$\mathrm{ES}_t(X - f(X)) = \begin{cases} \frac{x-f(x)+(1-2t)(y-f(y))}{2-2t}, & 0 \leqslant t \leqslant 1/2, \\ x - f(x), & 1/2 < t < 1. \end{cases}$$

Hence,

$$\begin{aligned}
&\frac{\rho(X-f(X))+\rho(f(X))-\rho(X)}{y-x}\\
&=\int_0^{1/2}\frac{t}{1-t}\left(\frac{f(y)-f(x)}{y-x}-1\right)\mathrm{d}\mu(t)+\int_{1/2}^{1}\frac{f(y)-f(x)}{y-x}-1\,\mathrm{d}\mu(t)>0.
\end{aligned}$$

Similarly, if $f(x)>f(y)$, then we have

$$\mathrm{ES}_t(X)=\begin{cases}\frac{(1-2t)x+y}{2-2t}, & 0\leqslant t\leqslant 1/2,\\ y, & 1/2<t<1,\end{cases}$$

$$\mathrm{ES}_t(f(X))=\begin{cases}\frac{f(x)+(1-2t)f(y)}{2-2t}, & 0\leqslant t\leqslant 1/2,\\ f(x), & 1/2<t<1,\end{cases}$$

$$\mathrm{ES}_t(X-f(X))=\begin{cases}\frac{(1-2t)(x-f(x))+y-f(y)}{2-2t}, & 0\leqslant t\leqslant 1/2,\\ y-f(y), & 1/2<t<1,\end{cases}$$

and thus

$$\begin{aligned}
&\rho(X-f(X))+\rho(f(X))-\rho(X)\\
&=\int_0^{1/2}\frac{t}{1-t}\,(f(x)-f(y))\,\mathrm{d}\mu(t)+\int_{1/2}^{1}f(x)-f(y)\,\mathrm{d}\mu(t)>0.
\end{aligned}$$

Therefore, $\mathcal{I}_{\rho,\rho}\subset\mathcal{I}_2$ and thus $\mathcal{I}_{\rho,\rho}=\mathcal{I}_2$.

(ii) The "if" part is straightforward by linearity of the mean. Hence, we prove the "only if" part. Similar to (i), since $h_0,h_1\in\mathcal{I}_0$, we have by translation invariance of $\rho$ and $\psi$ that $\rho=\psi$ on $\mathcal{X}$. Since $\mathcal{I}_1\subset\mathcal{I}_0$ and $\mathcal{X}_0^0\subset\mathcal{X}_+$, we know from Theorem 5.2 that $\rho(X)=\mathbb{E}[X]$ for all $X\in\mathcal{X}_0^0$. Since $X\in\mathcal{X}$ is bounded, we take $c>0$ such that $X+c\in\mathcal{X}_0^0$. It follows that $\rho(X+c)=\mathbb{E}[X+c]$. Hence, translation invariance of $\rho$ implies $\rho(X)=\mathbb{E}[X]$. □

*Proof of Proposition 5.3.* Take any $X\in\mathcal{X}_+$ and coherent risk measures $\rho,\psi:\mathcal{X}\to\mathbb{R}$ with $\rho\geqslant\psi$. For all $f\notin\mathcal{I}_{\psi,\psi}^X$, we have

$$\rho(X-f(X))+\psi(f(X))\geqslant\psi(X-f(X))+\psi(f(X))>\psi(X),$$

where the last inequality is due to subadditivity of $\psi$. With $h_1(x)=x$, $x\geqslant 0$, which belongs to $\mathcal{I}$, we have $\rho(X-h_1(X))+\psi(h_1(X))=\psi(X)$ and thus

$$\min_{g\in\mathcal{I}}\{\rho(X-g(X))+\psi(g(X))\}\leqslant\psi(X).$$

It follows that $f\notin\mathcal{I}_{\rho,\psi}^X$ and therefore $\mathcal{I}_{\rho,\psi}^X\subset\mathcal{I}_{\psi,\psi}^X$. □

# Chapter 6

# Cash-subadditive risk measures without quasi-convexity

## 6.1 Introduction

The quantification of market risk for pricing, portfolio selection, and risk management purposes has long been a point of interest to researchers and practitioners in finance. Measures of risk have been widely adopted to assess the riskiness of financial positions and determine capital reserves. Value-at-risk (VaR) has been one of the most commonly adopted risk measures in industry but is criticized due to its fundamental deficiencies; for instance, it does not account for "tail risk" and it lacks for subadditivity or convexity; see e.g., Daníelsson et al. (2001) and McNeil et al. (2015). In light of this, the notion of coherent risk measures that satisfy a set of reasonable axioms (monotonicity, cash additivity, subadditivity and positive homogeneity) was introduced by Artzner et al. (1999) and extensively treated by Delbaen (2002). Convex risk measures were introduced by Frittelli and Rosazza Gianin (2002) and Föllmer and Schied (2002a) with convexity replacing subadditivity and positive homogeneity. There have been many other developments in the past two decades on various directions; see Föllmer and Schied (2016) and the references therein.

A common feature of all above risk measures is that the axiom of cash additivity (also called cash invariance or translation invariance) is employed. The cash additivity axiom has been challenged, in particular, by El Karoui and Ravanelli (2009), in a relevant context. The main motivation for cash additivity is that the random losses should be discounted by a constant numéraire. Therefore, cash additivity fails as soon as there is any form of uncertainty about interest rates. For this reason, El Karoui and Ravanelli (2009) replaced cash additivity by cash subadditivity and provided a representation result for convex cash-subadditive risk measures. In this context, Cerreia-Vioglio et al. (2011) argued that quasi-convexity rather than convexity is the appropriate mathematical translation of the statement "diversification should not increase the risk" and introduced the notion of quasi-convex cash-subadditivie risk measurses. Farkas et al. (2014) studied general risk measures to model defaultable contingent claims and discussed their relationship with cash-subadditive risk measures. For other related work on cash subadditivity and quasi-convexity, see Frittelli and Maggis (2011), Cont et al. (2013), Drapeau and Kupper (2013) and Frittelli et al. (2014). In decision theory, the economic counterpart of quasi-convexity of risk measures is quasi-concavity of utility functions, which is classically associated to uncertainty aversion in the economics of uncertainty; see, e.g., Schmeidler (1989), Cerreia-Vioglio et al. (2011) and Mastrogiacomo and Rosazza Gianin (2015).

The main aim of this chapter is a thorough understanding of cash-subadditive risk measures when quasi-convexity, or the stronger property of convexity, is absent. This class of risk measures is very broad and, with proper normalization, it contains a wide majority of risk measure or preference functional considered in the literature. By relaxing cash additivity, both the theory of risk measures and that of expected utility and rank-dependent utility (Quiggin, 1982) can be included within the same framework. For instance, the mapping $X \mapsto -\mathbb{E}[u(-X)]/m$ for any increasing utility function $u$ with derivative bounded above by $m > 0$ belongs to this class (recall that the constant $m$ does not matter when modeling utility preferences); the same holds true if $\mathbb{E}$ is replaced by a non-additive and normalized Choquet integral (Yaari, 1987; Schmeidler, 1989). Here, the a loss function $\ell$ may not be convex or concave; see the recent discussions and examples in Müller et al. (2017) and Castagnoli et al. (2022) for non-convex and non-concave loss and utility functions.

Cash-additive risk measures without convexity have been actively studied in the recent literature. In particular, a few representation results were obtained by Mao and Wang (2020), Jia et al. (2021) and Castagnoli et al. (2022). As a common feature, such risk measures can be represented as the infimum over a collection of convex and cash-additive risk measures (see Table 6.1 below),[1] in contrast to the classic theory of convex risk measures where representations are typically based on a supremum. In a similar fashion, one of our main results states that a general cash-subadditive risk measure can be represented as the lower envelope of a family of quasi-convex cash-subadditive

[1]The risk measures studied by Jia et al. (2021) and Castagnoli et al. (2022) have similar representations; their differences are studied by Moresco and Righi (2022).

risk measures.

In addition to a representation of general cash-subadditive risk measures, we will also give implicit and explicit representations of cash-subadditive risk measures with additional properties including quasi-star-shapedness, normalization and SSD-consistency (that is, consistency with second-order stochastic dominance). In particular, similarly to the argument that convexity does not fit well with cash-subadditive risk measures, star-shapedness introduced by Castagnoli et al. (2022) is no longer a natural property beyond the framework of cash-additive risk measures. In this sense, we introduce the property of quasi-star-shapedness induced naturally from quasi-convexity, and obtain a representation result of cash-subadditive risk measures that are normalized and quasi-star-shaped. It turns out that the representation result also holds true if we change normalization to a weaker version which we call quasi-normalization. We examine a few other problems studied by Mao and Wang (2020), now under a general framework of cash subadditivity. Apart from the major differences, it also turns out that some of results obtained by Mao and Wang (2020) hold under the extended framework of cash subadditivity. We summarize some related results in the literature and compare them with our results in Table 6.1.

| | a (...) risk measure | is an infimum of (...) risk measures |
|---|---|---|
| Mao and Wang (2020) | CA, SSD-consistent | CA, convex, law-invariant |
| Jia et al. (2021) | CA | CA, convex |
| Castagnoli et al. (2022) | CA, star-shaped, normalized | CA, convex, normalized |
| Theorem 6.4 | CS, SSD-consistent | CS, quasi-convex, law-invariant |
| Theorem 6.2 | CS | CS, quasi-convex |
| Theorem 6.3 | CS, quasi-star-shaped, normalized | CS, quasi-convex, normalized |

Table 6.1: Representation results related to this chapter, where monotonicity is always assumed; definitions of the properties are in Sections 6.2 and 6.3. CA stands for cash additivity and CS stands for cash subadditivity.

The new property of quasi-star-shapedness has a sound decision-theoretic foundation. Translating it into the setting of Anscombe and Aumann (1963), it means that the decision maker always prefers to replace part of an uncertain (random) payoff with an equally favourable certain (non-random) payoff. This property is a weaker requirement than the uncertainty aversion axiom studied by Maccheroni et al. (2006), which corresponds to quasi-convexity in our setting.

The rest of the chapter is organized as follows. In Section 6.2, some preliminaries on risk measures are collected and the definition of cash-subadditive risk measures is given. Two new properties, quasi-star-shapedness and quasi-normalization, are introduced in Section 6.3 and we provide a few related results. In particular, we obtain a new formula (Theorem 6.1) for $\Lambda$VaR introduced by Frittelli et al. (2014), which is an example of quasi-star-shaped, quasi-normalized and cash-subadditive risk measures. In Section 6.4, representation results for general cash-subadditive risk measures are established. Section 6.5 contains representation results and other technical results on cash-subadditive risk measures with further properties including quasi-star-shapedness and SSD-consistency. Section 6.6 concludes the chapter, and Section 6.7 contains some further technical results and discussions that are not directly used in the main text. As the main message of this chapter, most of the existing results on non-convex cash-additive risk measures have a nice parallel version on non-quasi-convex cash-subadditive risk measures, although they often require more sophisticated analysis to establish.

## 6.2 Cash-subadditive risk measures

Fix an atomless probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{M}_f$ denote the set of finitely additive probabilities on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to $P$, and let $\mathcal{M}$ represent the subset of $\mathcal{M}_f$ consisting of all its countably additive elements, i.e., probability measures. Let $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, P)$ be the set of all essentially bounded random variables on $(\Omega, \mathcal{F}, P)$, where $P$-a.s. equal random variables are treated as identical.[2] Let a random variable $X \in \mathcal{X}$ represent the random loss faced by financial institutions in a fixed period of time.

To clarify, a functional $\rho : \mathcal{X} \to \mathbb{R}$ is called a *risk measure* if it is monotone. Monotonicity is self-explanatory and common in the literature of risk management, e.g., Artzner et al. (1999). It means that if the loss increases for almost all scenarios $\omega \in \Omega$, then the capital requirement in order for the financial position to be acceptable should increase as well. As one of the fundamental properties to define monetary risk measures, cash additivity (also called cash invariance or translation invariance),

[2]As such, equalities and inequalities should be understood in a $P$-a.s. sense.

defined in Chapter 1, intuitively guarantees the risk measure $\rho(X)$ to be the capital required added to the financial position $X$. Cash additivity is a nice and simplifying mathematical property, but the class of cash-additive risk measures is too restricted to include some common functionals such as the expectation of a convex loss function. From the viewpoint of financial practice, the assumption of cash additivity of a risk measure may fail when uncertainty of interested rates is taken into account. In this sense, we consider the more general class of risk measures $\rho$, as argued by El Karoui and Ravanelli (2009), satisfying

*Cash subadditivity*: $\rho(X+m) \leqslant \rho(X)+m$ for all $X \in \mathcal{X}$ and $m \geqslant 0$.[3]

The assumption of cash subadditivity allows non-linear increase of the capital requirement as cash is added to the financial position but the increase should not exceed linear growth. Cash-subadditive risk measures are often studied in the literature together with convexity, or more generally, with quasi-convexity; see El Karoui and Ravanelli (2009), Cerreia-Vioglio et al. (2011) and Frittelli et al. (2014).

As the main objective of this chapter is to study cash-subadditive risk measures without quasi-convexity, we first note that the lack of quasi-convexity arises in many economically relevant contexts, such as aggregation of risk measures, non-convex utility functions, and risk mitigation. Castagnoli et al. (2022) argued with examples that many operations on a collection of convex risk measures lead to a non-convex one; the same applies in the context of quasi-convexity. Other than those built from operations, we provide a few simple examples of cash-subadditive risk measures in the literature, which are not cash-additive or quasi-convex.

**Example 6.1** (Expected insured loss)**.** Suppose that an insurance contract pays $f(X)$ for an insurable loss $X$ (often non-negative), where $f$ is an increasing function on $\mathbb{R}$ that is 1-Lipschitz and $f(x)=0$ for $x \leqslant 0$.[4] A typical example is $f(x)=(x-d)_+ \wedge \ell$ for some $\ell > d > 0$, which represents an insurance contract with deductible $d$ and limit $\ell$. The expected losses to the policy holder and to the insurer are given by, respectively,

$$\rho_{\mathrm{ph}}(X)=\mathbb{E}[X-f(X)] \quad \text{and} \quad \rho_{\mathrm{in}}(X)=\mathbb{E}[f(X)].$$

It is straightforward to check that $\rho_{\mathrm{ph}}$ and $\rho_{\mathrm{in}}$ are both monotone and cash subadditive, but generally neither cash additive nor quasi-convex. In particular, $\rho_{\mathrm{in}}$ (resp. $\rho_{\mathrm{ph}}$) is concave if $f$ is concave

[3] An equivalent definition of cash subadditivity is $\rho(X+m) \geqslant \rho(X)+m$ for all $X \in \mathcal{X}$ and $m \leqslant 0$.

[4] A function $f: \mathbb{R} \to \mathbb{R}$ is called 1-*Lipschitz* if $|f(x)-f(y)| \leqslant |x-y|$ for all $x, y \in \mathbb{R}$.

(resp. convex). For a related example in finance, take $f : x \mapsto x_+$ and fix a probability measure $Q$ representing a pricing measure in a financial market. The put option premium on the insolvency of a firm with future asset value $-X$ is defined as $\mathbb{E}_Q[X_+]$, which is convex and cash subadditive but not cash additive; see Jarrow (2002) and El Karoui and Ravanelli (2009) for a connection between the put option premium and risk measures.

The risk measure *Value-at-Risk* (VaR) is given by, for $t \in (0, 1]$,

$$\mathrm{VaR}_t(X) = \inf\{x \in \mathbb{R} \mid P(X \leqslant x) \geqslant t\}, \quad X \in \mathcal{X}. \tag{6.1}$$

Note that $\mathrm{VaR}_1(X) = \text{ess-sup}(X)$. VaR is one of the most popular risk measures used in the banking industry; see McNeil et al. (2015). The next example is a generalization of VaR introduced by Frittelli et al. (2014) without cash additivity.

**Example 6.2** ($\Lambda$-Value-at-Risk)**.** The risk measure $\Lambda$-Value-at-Risk is defined as, for some function $\Lambda : \mathbb{R} \to [0, 1]$ that is not constantly 0,

$$\Lambda\mathrm{VaR}(X) = \inf\{x \in \mathbb{R} : P(X \leqslant x) \geqslant \Lambda(x)\}, \quad X \in \mathcal{X}. \tag{6.2}$$

In particular, if $\Lambda$ is a constant $t \in (0, 1)$, then $\Lambda\mathrm{VaR} = \mathrm{VaR}_t$. The function $\Lambda$ is often chosen as a decreasing function to avoid pathological cases; see the discussion in Bellini and Peri (2021). Here, we will assume that $\Lambda$ is a decreasing function. Since for $c \geqslant 0$, $\Lambda\mathrm{VaR}(X + c) = \Lambda^*\mathrm{VaR}(X) + c$ where $\Lambda^*(t) = \Lambda(t + c) \leqslant \Lambda(t)$ for $t \in \mathbb{R}$, we obtain $\Lambda\mathrm{VaR}(X + c) \leqslant \Lambda\mathrm{VaR}(X) + c$, and therefore $\Lambda\mathrm{VaR}$ is cash subadditive; we can check that it is not cash additive in general. Moreover, $\Lambda\mathrm{VaR}$ is generally not quasi-convex either, as the following argument illustrates. For any decreasing $\Lambda : \mathbb{R} \to (0, 1/3]$ and a standard normal random variable $X$, we have $\Lambda\mathrm{VaR}(X) = \Lambda\mathrm{VaR}(-X) \leqslant z_{1/3} < 0$, where $z_{1/3}$ is the 1/3-quantile of the standard normal distribution. Hence, $\Lambda\mathrm{VaR}(0) = 0 > \max\{\Lambda\mathrm{VaR}(X), \Lambda\mathrm{VaR}(-X)\}$ violating quasi-convexity.

**Example 6.3** (Certainty equivalent with discount factor ambiguity)**.** Consider the following $\alpha$-maxmin expected utility ($\alpha$-MEU, see e.g., Ghirardato et al., 2004; Marinacci, 2002) with a profit-loss adjustment:

$$\alpha \min_{Q_1 \in \mathcal{Q}_1} \mathbb{E}_{Q_1}[\mathrm{e}^{\gamma X}] + (1 - \alpha) \max_{Q_2 \in \mathcal{Q}_2} \mathbb{E}_{Q_2}[\mathrm{e}^{\gamma X}], \quad X \in \mathcal{X},\ \alpha \in [0, 1],\ \gamma > 0$$

with the loss function $x \mapsto \mathrm{e}^{\gamma x}$, where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are two nonempty, weak*-compact and convex sets of finitely additive functions. Its certainty equivalent with ambiguity of a discount factor is given by

$$\rho(X)=\sup_{\lambda\in I}\left\{\frac{1}{\gamma}\log\left(\alpha\min_{Q_1\in\mathcal{Q}_1}\mathbb{E}_{Q_1}[\mathrm{e}^{\gamma\lambda X}]+(1-\alpha)\max_{Q_2\in\mathcal{Q}_2}\mathbb{E}_{Q_2}[\mathrm{e}^{\gamma\lambda X}]\right)\right\},\quad X\in\mathcal{X}$$

with the ambiguity set $I\subseteq[0,1]$. Cash subadditivity induced by ambiguity of the discount factor was discussed in El Karoui and Ravanelli (2009) with stochastic discount factors. We can check that $\rho$ is a cash-subadditive risk measure, and it is generally not cash additive. Moreover, because of the presence of both minimum and maximum in $\alpha$-MEU, quasi-convexity does not hold for $\rho$.

**Example 6.4** (Risk measures based on an eligible risky asset)**.** Take an acceptance set $\mathcal{A}\subseteq\mathcal{X}$ and a reference asset $S=(S_0,S_T)\in\mathcal{X}^2$, where $S_T$ is a nonzero and positive terminal payoff. Define the mapping $\rho_{\mathcal{A},S}$ as in Farkas et al. (2014) by

$$\rho_{\mathcal{A},S}(X)=\inf\left\{m\in\mathbb{R}:X-\frac{m}{S_0}S_T\in\mathcal{A}\right\},\quad X\in\mathcal{X}.$$

They showed that $\rho_{\mathcal{A},S}$ is *S-additive* and monotone.[5] Cash subadditivity of $\rho_{\mathcal{A},S}$ is equivalent to

$$\rho_{\mathcal{A},S}\left(X+\lambda\right)\leqslant\rho_{\mathcal{A},S}\left(X+\frac{\lambda}{S_0}S_T\right)\text{ for all }\lambda\geqslant0\text{ and }X\in\mathcal{X}.$$

Hence, under the assumption of $P\left(S_T<S_0\right)=0$ (e.g., $S$ generates a non-negative random interest rate), the risk measure $\rho_{\mathcal{A},S}$ is cash subadditive. As shown by Farkas et al. (2014), $\rho_{\mathcal{A},S}$ is quasi-convex if and only if $\mathcal{A}$ is convex; thus, such risk measures are not quasi-convex in general.

Some other relevant properties for a risk measure $\rho$ are collected below, which will be used throughout the chapter; we refer to Föllmer and Schied (2016) for a comprehensive treatment of properties of risk measures.

*Normalization*: $\rho(t)=t$ for all $t\in\mathbb{R}$.

*Law invariance*: $\rho(X)=\rho(Y)$ for all $X,Y\in\mathcal{X}$ with $X\overset{\mathrm{d}}{=}Y$.

Monetary, convex and positively homogeneous risk measures are called coherent by Artzner et al. (1999).[6]

Next, we define two most important notions of stochastic dominance in decision theory, the first-order stochastic dominance (FSD) and the second-order stochastic dominance (SSD). Given two random variables $X,Y\in\mathcal{X}$, we denote by $X\succeq_1 Y$, if $\mathbb{E}[f(X)]\geqslant\mathbb{E}[f(Y)]$ for all increasing

[5] The functional $\rho$ is said to be $S$-additive if $\rho\left(X+\lambda S_T\right)=\rho(X)+\lambda S_0$ holds for all $\lambda\in\mathbb{R}$ and $X\in\mathcal{X}$.

[6] The functional $\rho$ is said to be positively homogeneous if $\rho(\lambda X)=\lambda\rho(X)$ for all $X\in\mathcal{X}$ and $\lambda\geqslant0$.

functions $f : \mathbb{R} \to \mathbb{R}$, and denote by $X \succeq_2 Y$, if $\mathbb{E}[f(X)] \geqslant \mathbb{E}[f(Y)]$ for all increasing convex functions $f : \mathbb{R} \to \mathbb{R}$. Consistency with respect to FSD or SSD is defined as monotonicity in these partial orders.

*FSD-consistency*: $\rho(X) \geqslant \rho(Y)$ for all $X, Y \in \mathcal{X}$ whenever $X \succeq_1 Y$.

*SSD-consistency*: $\rho(X) \geqslant \rho(Y)$ for all $X, Y \in \mathcal{X}$ whenever $X \succeq_2 Y$.

It is well known that either FSD-consistency or SSD-consistency implies law invariance. For monetary risk measures, SSD-consistency is characterized by Theorem 3.1 of Mao and Wang (2020).

Finally, the notion of comonotonicity is useful for some results in this chapter. A random vector $(X_1, \ldots, X_n) \in \mathcal{X}^n$ is called *comonotonic* if there exists a random variable $Z \in \mathcal{X}$ and increasing functions $f_1, \ldots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ almost surely for all $i = 1, \ldots, n$.

## 6.3 Quasi-star-shapedness, quasi-normalization, and Lambda VaR

In this section, we discuss two new properties that are specific to cash-subadditive risk measures without quasi-convexity, and they will be used in the representation results in Section 6.5.1.

### 6.3.1 Quasi-star-shapedness and quasi-normalization

In the context of cash-additive risk measures, Castagnoli et al. (2022) studied a weaker property than convexity:

*Star-shapedness*: $\rho(\lambda X) \leqslant \lambda \rho(X) + (1-\lambda)\rho(0)$ for all $X \in \mathcal{X}$ and $\lambda \in [0, 1]$,

and formulated star-shapedness via $\rho(\lambda X) \leqslant \lambda \rho(X)$ for $\lambda \in [0, 1]$ with the extra normalization $\rho(0) = 0$. Star-shapedness is discussed in Artzner et al. (1999) and it has a natural economic motivation that additional liquidity risk may arise if a position is multiplied by a factor larger than 1. In case $\rho(0) \neq 0$, it is more natural to define star-shapedness via our formulation, which means convexity at 0 (has also been called "positive superhomogeneity" for obvious mathematical reasons), thus weaker than convexity. In the context of the cash-additive risk measures, we introduce the corresponding property for cash-subadditive risk measures:

*Quasi-star-shapedness*: $\rho(\lambda X + (1-\lambda)t) \leqslant \max\{\rho(X), \rho(t)\}$ for all $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\lambda \in [0, 1]$.

Since quasi-star-shapedness is new to the literature of risk measures, it may need some explanation. As explained by Castagnoli et al. (2022), star-shapedness reflects the consideration of liquidity risk, in a way similar to (but weaker than) convexity which reflects the consideration of diversification. For cash-additive risk measures, star-shapedness is equivalent to $\rho(\lambda X+(1-\lambda)t) \leqslant \lambda\rho(X)+(1-\lambda)\rho(t)$ for all $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\lambda \in [0,1]$; indeed, it means that $\rho$ has convexity at each constant. This reformulation of star-shapedness implies our quasi-star-shapedness, which means that $\rho$ has quasi-convexity at each constant. Obviously, quasi-star-shapedness is weaker than quasi-convexity.

Quasi-star-shapedness has a sound decision-theoretic interpretation, which we explain in Proposition 6.1 below. For a risk measure $\rho : \mathcal{X} \to \mathbb{R}$, the preference associated with $\rho$ is a binary relation $\succeq$ on $\mathcal{X}$ defined by, for all $X, Y \in \mathcal{X}$, $X \succeq Y \iff \rho(X) \leqslant \rho(Y)$. The equivalence relation of $\succeq$ is denoted by $\simeq$. In other words, $\succeq$ represents the preference of an agent favouring less risk evaluated via $\rho$.

**Proposition 6.1.** *An $L^\infty$-continuous risk measure $\rho : \mathcal{X} \to \mathbb{R}$ satisfies quasi-star-shapedness if and only if its associated preference $\succeq$ satisfies, for $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\lambda \in [0,1]$,*

$$X \simeq t \implies \lambda X + (1-\lambda)t \succeq X. \tag{6.3}$$

*Proof.* By definition of $\succeq$, (6.3) is equivalent to

$$\rho(X) = \rho(t) \implies \rho(\lambda X + (1-\lambda)t) \leqslant \rho(X), \tag{6.4}$$

which is clearly implied by quasi-star-shapedness. Hence, "only-if" statement holds true. To show the "if" statement, take arbitrary $X \in \mathcal{X}$ and $t \in \mathbb{R}$. If $\rho(X) \leqslant \rho(t)$, then we take $s \geqslant 0$ such that $\rho(X+s) = \rho(t)$. Such $s$ exists since $s \mapsto \rho(X+s)$ is continuous and $X+s \geqslant t$ for $s$ large enough. Using monotonicity of $\rho$ and (6.4), we have

$$\rho(\lambda X + (1-\lambda)t) \leqslant \rho(\lambda(X+s) + (1-\lambda)t) \leqslant \rho(t) = \max\{\rho(X), \rho(t)\}$$

for each $\lambda \in [0,1]$. If $\rho(X) > \rho(t)$, then we take $s \geqslant 0$ such that $\rho(X) = \rho(t+s)$. Such $s$ exists since $s \mapsto \rho(t+s)$ is continuous and $t+s \geqslant X$ for $s$ large enough. Using monotonicity of $\rho$ and (6.4), we have

$$\rho(\lambda X + (1-\lambda)t) \leqslant \rho(\lambda X + (1-\lambda)(t+s)) \leqslant \rho(X) = \max\{\rho(X), \rho(t)\}$$

for each $\lambda \in [0,1]$. Hence, quasi-star-shapedness holds. $\square$

*Remark* 6.1. Proposition 6.1 requires $L^\infty$-continuity, which is a weak property satisfied by essentially all risk measures in the literature. We can check from definition that all cash-subadditive risk measures are $L^\infty$-continuous. The result in Proposition 6.1 holds true with the same proof if $L^\infty$-continuity is replaced by the property of *solvability* in decision theory: For each $X \in \mathcal{X}$, there exists $t \in \mathbb{R}$ such that $\rho(X) = \rho(t)$.

Proposition 6.1 gives the following decision-theoretic interpretation of quasi-star-shapedness. Suppose that the preference $\succeq$ of an agent satisfies (6.3). If a random loss $X$ is seen as equally favourable as a constant loss $t$, then $\lambda X + (1-\lambda)t$ is favourable compared to $X$. That is, a combination of random $X$ and constant $t$ reduces the riskiness of $X$. In contrast, quasi-convexity requires the above relation to hold for random $Y$ in replace of constant $t$. Indeed, in the setting of Anscombe and Aumann (1963) where $X$ and $Y$ represent acts with uncertainty (thus, they are not necessarily $\mathbb{R}$-valued), the property, for $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$,

$$X \simeq Y \implies \lambda X + (1-\lambda)Y \succeq X, \tag{6.5}$$

is the uncertainty aversion axiom of Maccheroni et al. (2006), and it corresponds to quasi-convexity of the risk measure $\rho$ in our setting. It is clear that (6.3) is weaker than (6.5) as the riskiness of $X$ is only reduced when combined with an equally favourable constant loss, instead of an arbitrary equally favourable loss $Y$.

The difference between quasi-star-shapedness and quasi-convexity, or between (6.3) and (6.5), can also be explained via considerations for the dependence between pooled risks. For a law-invariant $\rho$ and two losses $X$ and $Y$ with fixed distributions, the dependence structure of $X$ and $Y$ affects $\rho(\lambda X + (1-\lambda)Y)$ but not $\rho(X)$ or $\rho(Y)$, and hence quasi-convexity imposes inequalities over all dependence structures. Such an issue does not appear for $\lambda X + (1-\lambda)t$ as dependence is irrelevant between a random variable $X$ and a constant $t$. Hence, relaxing quasi-convexity to quasi-star-shapedness gives rise to more flexibility on preferences over dependence. In particular, under quasi-convexity, comonotonicity is the worst-case dependence in risk aggregation; see Lemmas 6.3 and 6.4. This is not the case for quasi-star-shapedness, since $\mathrm{VaR}_t$ for $t \in (0,1)$ is quasi-star-shaped but it does not take comonotonicity as the worst-case dependence.

Next, we discuss the issue of normalization. The risk measures in Examples 6.1 and 6.2 are not necessarily normalized. In general, cash-subadditive risk measures may not have the range of the entire real line. Hence, normalization may also need to be weakened in our setting of cash subadditive risk measures, which we define as follows.

*Quasi-normalization*: $\rho(t) = t$ for all $t \in D_\rho$, where $D_\rho = \{\rho(X) \mid X \in \mathcal{X}\}$ is the range of $\rho$.

The risk measure $X \mapsto \mathbb{E}[\min\{X, d\}]$ in Example 6.1 satisfies quasi-normalization with range $(-\infty, d]$, and $\Lambda\mathrm{VaR}$ in Example 6.2 satisfies quasi-normalization with range $(-\infty, z]$ where $z = \inf\{x \in \mathbb{R} : \Lambda(x) = 0\}$ with the convention $\inf \emptyset = \infty$.

### 6.3.2 A new representation of Lambda VaR

The next result gives quasi-star-shapedness of $\Lambda\mathrm{VaR}$, complementing the fact observed by Castagnoli et al. (2022) that VaR is star-shaped. We also obtain, as a by-product, an alternative representation of $\Lambda\mathrm{VaR}$. In what follows, set $\mathrm{VaR}_0(X) = -\infty$ for any $X \in \mathcal{X}$, which follows from plugging $t = 0$ in (6.1).

**Theorem 6.1.** *Let $\Lambda : \mathbb{R} \to [0,1]$ be a decreasing function that is not constantly $0$. The risk measure $\Lambda\mathrm{VaR}$ in (6.2) has the representation*

$$\Lambda\mathrm{VaR}(X) = \inf_{x \in \mathbb{R}} \left\{ \mathrm{VaR}_{\Lambda(x)}(X) \vee x \right\} = \sup_{x \in \mathbb{R}} \left\{ \mathrm{VaR}_{\Lambda(x)}(X) \wedge x \right\}, \qquad X \in \mathcal{X}, \tag{6.6}$$

*and moreover, $\Lambda\mathrm{VaR}$ is quasi-star-shaped.*

*Proof.* Note that for $X \in \mathcal{X}$, $x \in \mathbb{R}$ and $t \in [0,1]$, $P(X \leqslant x) \geqslant t$ if and only if $\mathrm{VaR}_t(X) \leqslant x$. Moreover, since $\Lambda$ is decreasing, the set $\{x \in \mathbb{R} : \mathrm{VaR}_{\Lambda(x)}(X) \leqslant x\}$ is an interval with right end-point $\infty$. By definition, for $X \in \mathcal{X}$,

$$\begin{aligned}
\Lambda\mathrm{VaR}(X) &= \inf\{x \in \mathbb{R} : P(X \leqslant x) \geqslant \Lambda(x)\} \\
&= \inf\{x \in \mathbb{R} : \mathrm{VaR}_{\Lambda(x)}(X) \leqslant x\} \\
&= \inf\{\mathrm{VaR}_{\Lambda(x)}(X) \vee x : \mathrm{VaR}_{\Lambda(x)}(X) \leqslant x\} \geqslant \inf_{x \in \mathbb{R}} \left\{ \mathrm{VaR}_{\Lambda(x)}(X) \vee x \right\}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\Lambda\mathrm{VaR}(X) &= \inf\{x \in \mathbb{R} : \mathrm{VaR}_{\Lambda(x)}(X) \leqslant x\} \\
&= \sup\{x \in \mathbb{R} : \mathrm{VaR}_{\Lambda(x)}(X) > x\} \\
&= \sup\{\mathrm{VaR}_{\Lambda(x)}(X) \wedge x : \mathrm{VaR}_{\Lambda(x)}(X) > x\} \leqslant \sup_{x \in \mathbb{R}} \left\{ \mathrm{VaR}_{\Lambda(x)}(X) \wedge x \right\}.
\end{aligned}$$

Since $\mathrm{VaR}_{\Lambda(x)}(X) \wedge x \leqslant \mathrm{VaR}_{\Lambda(y)}(X) \vee y$ for any $x, y \in \mathbb{R}$, we have

$$\Lambda\mathrm{VaR}(X) \leqslant \sup_{x \in \mathbb{R}} \left\{ \mathrm{VaR}_{\Lambda(x)}(X) \wedge x \right\} \leqslant \inf_{x \in \mathbb{R}} \left\{ \mathrm{VaR}_{\Lambda(x)}(X) \vee x \right\} \leqslant \Lambda\mathrm{VaR}(X),$$

thus showing (6.6). Next, verify that the mapping $X \mapsto \mathrm{VaR}_\alpha(X) \vee x$ is quasi-star-shaped for all $\alpha \in [0,1]$ and $x \in \mathbb{R}$. Note that if $\alpha = 0$ then it is trivial. If $\alpha > 0$, then $\mathrm{VaR}_\alpha(X) \vee x = \mathrm{VaR}_\alpha(X \vee x)$. For all $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\lambda \in [0,1]$,

$$\begin{aligned}\mathrm{VaR}_\alpha(x \vee (\lambda X + (1-\lambda)t)) &\leqslant \mathrm{VaR}_\alpha(\lambda(x \vee X) + (1-\lambda)(x \vee t)) \\ &= \lambda \mathrm{VaR}_\alpha(x \vee X) + (1-\lambda)\mathrm{VaR}_\alpha(x \vee t) \\ &\leqslant \max\{\mathrm{VaR}_\alpha(x \vee X), \mathrm{VaR}_\alpha(x \vee t)\}. \qquad (6.7)\end{aligned}$$

Finally, we need to use Lemma 6.1 below, which states that the infimum of quasi-normalized, quasi-star-shaped and cash subadditive risk measures is quasi-star-shaped. Since $X \mapsto \mathrm{VaR}_{\Lambda(x)}(X) \vee x$ is quasi-normalized, quasi-star-shaped and cash subadditive for all $x \in \mathbb{R}$, $\Lambda\mathrm{VaR}$ is quasi-star-shaped by Lemma 6.1. $\square$

*Remark* 6.2. We note that $\Lambda\mathrm{VaR}$ is generally not star-shaped. For instance, take $\Lambda : x \mapsto \mathbb{1}_{\{x \leqslant 1\}}$. For this choice, we have $\Lambda\mathrm{VaR}(x) = x \wedge 1$ for $x \in \mathbb{R}$. It follows that $\Lambda\mathrm{VaR}(1) = 1 > 1/2 = \Lambda\mathrm{VaR}(2)/2 + \Lambda\mathrm{VaR}(0)/2$, and hence $\Lambda\mathrm{VaR}$ is not star-shaped. Indeed, any $\Lambda$ with $\inf\{x \in \mathbb{R} : \Lambda(x) = 0\} = 1$ suffices for this example. Note that each $X \mapsto \mathrm{VaR}_\alpha(X) \vee x$ in the representation (6.6) is star-shaped (including $\alpha = 0$); see (6.7). Therefore, the infimum of quasi-normalized, star-shaped and cash-subadditive risk measures is not necessarily star-shaped, in sharp contrast to the corresponding result on quasi-star-shaped risk measures in Lemma 6.1. This example shows that quasi-star-shapedness is more natural than, and genuinely different from, star-shapedness in the context of cash-subadditive risk measures.

Theorem 6.1 can be applied to solve portfolio optimization problems with $\Lambda\mathrm{VaR}$ constraints. Let $\Lambda : \mathbb{R} \to [0,1]$ be a decreasing function which is not constantly 0. In a portfolio optimization problem, one often maximizes an objective, e.g., an expected utility or an expected return, under the constraint that a risk measure does not exceed a certain level $z$ (and often together with a budget constraint). For $X \in \mathcal{X}$, by Theorem 6.1, we have

$$\Lambda\mathrm{VaR}(X) \leqslant z \iff \inf_{x \in \mathbb{R}} \{\mathrm{VaR}_{\Lambda(x)}(X) \vee x\} \leqslant z \iff \inf_{x \leqslant z} \mathrm{VaR}_{\Lambda(x)}(X) \leqslant z \iff \mathrm{VaR}_{\Lambda(z)}(X) \leqslant z.$$

Therefore, optimization under a $\Lambda\mathrm{VaR}$ constraint below a constant level $z$ is equivalent to that under a $\mathrm{VaR}_{\Lambda(z)}$ constraint below the same level $z$, which has been well studied in the risk management literature; see e.g., Basak and Shapiro (2001) and Basak et al. (2006).

### 6.3.3 A few useful technical results

The next lemma shows that quasi-normalization and quasi-star-shapedness are preserved under a minimum operation, a fact used in the proof of Theorem 6.1.

**Lemma 6.1.** *The infimum of quasi-normalized, quasi-star-shaped, and cash-subadditive risk measures (assuming it is real-valued) is again quasi-normalized, quasi-star-shaped and cash-subadditive.*

*Proof.* Let $\mathcal{C}$ be a class of quasi-normalized, quasi-star-shaped and cash-subadditive risk measures, and denote by $\rho = \inf_{\psi \in \mathcal{C}} \psi$. It is obvious that $\rho$ is cash subadditive and monotone. It remains to show that $\rho$ is quasi-normalized and quasi-star-shaped. Denote by $d = \inf D_\rho$, $u = \sup D_\rho$, $d_\psi = \inf D_\psi$ and $u_\psi = \sup D_\psi$ for $\psi \in \mathcal{C}$. For any $X \in \mathcal{X}$ and $\psi \in \mathcal{C}$, if $u < d_\psi$, then $\rho(X) \leqslant u < \psi(X)$. Hence, we can write

$$\rho(X) = \inf_{\psi \in \mathcal{C}'} \psi(X) \text{ where } \mathcal{C}' = \{\psi \in \mathcal{C} \mid u \geqslant d_\psi\}.$$

Note that $d \leqslant d_\psi \leqslant u \leqslant u_\psi$ for each $\psi \in \mathcal{C}'$. Moreover, for any $\psi \in \mathcal{C} \cup \{\rho\}$, $\psi(t) = t \wedge u_\psi$ for all $t \geqslant d_\psi$ and $\psi(t) = t \vee d_\psi$ for all $t \leqslant u_\psi$ by monotonicity and quasi-normalization of $\psi$.

We first show that $\rho$ is quasi-normalized. Take a constant $t \in (d, u]$. Since $t \leqslant u \leqslant u_\psi$, we have $\psi(t) \geqslant t$ for all $\psi \in \mathcal{C}'$. Hence, $\rho(t) = \inf_{\psi \in \mathcal{C}'} \psi(t) \geqslant t$. Moreover, since $t > d$ and $d = \inf_{\psi \in \mathcal{C}'} d_\psi$, there exists $\psi \in \mathcal{C}'$ such that $d_\psi < t$. This gives $\psi(t) \leqslant t$. Hence, $\rho(t) = \inf_{\psi \in \mathcal{C}'} \psi(t) \leqslant t$. Thus, we obtain $\rho(t) = t$ for $t \in (d, u]$. It remains to verify $\rho(d) = d$ if $d > -\infty$. This follows from the fact that a cash-subadditive risk measure is $L^\infty$-continuous. Therefore, $\rho(t) = t$ for $t \in D_\rho$, and thus $\rho$ is quasi-normalized.

Next, we show that $\rho$ is quasi-star-shaped. For $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\lambda \in [0, 1]$, quasi-star-shapendess of $\psi \in \mathcal{C}'$ gives

$$\rho(\lambda X + (1 - \lambda)t) = \inf_{\psi \in \mathcal{C}'} \psi(\lambda X + (1 - \lambda)t) \leqslant \inf_{\psi \in \mathcal{C}'} \max\{\psi(X), \psi(t)\}. \tag{6.8}$$

If $t \geqslant u$, then $\rho(t) = u$ and

$$\rho(\lambda X + (1 - \lambda)t) \leqslant u = \rho(t).$$

If $t < u$, then $\psi(t) = t \vee d_\psi$ for $\psi \in \mathcal{C}'$ and $\rho(t) = t \vee d$. It follows that

$$\begin{aligned} \inf_{\psi \in \mathcal{C}'} \max\{\psi(X), \psi(t)\} &= \inf_{\psi \in \mathcal{C}'} \max\{\psi(X), t, d_\psi\} \\ &= \inf_{\psi \in \mathcal{C}'} \max\{\psi(X), t\} = \max\left\{\inf_{\psi \in \mathcal{C}'} \psi(X), t\right\} \leqslant \max\{\rho(X), \rho(t)\}. \end{aligned}$$

Using (6.8) and combining both cases, we obtain $\rho(\lambda X + (1-\lambda)t) \leqslant \max\{\rho(X), \rho(t)\}$ for all $\lambda \in [0,1]$, $X \in \mathcal{X}$ and $t \in \mathbb{R}$ and thus $\rho$ is quasi-star-shaped. $\square$

Finally, we show that in the classic setting of cash-additive risk measures, we do not need to distinguish between each of normalization, star-shapedness and covexity and their quasi-versions. This result further illustrates that quasi-star-shapedness is a natural property to consider for cash-subadditive risk measures.

**Proposition 6.2.** *For cash-additive risk measures,*

*(i) normalization is equivalent to quasi-normalization;*

*(ii) star-shapedness is equivalent to quasi-star-shapedness;*

*(iii) convexity is equivalent to quasi-convexity.*

*In contrast, for cash-subadditive risk measures, none of the above equivalence holds true.*

*Proof.* The statements on normalization are straightforward. Those on convexity are well known and can be checked with acceptance sets; see e.g., Proposition 2.1 and Example 2.2 of Cerreia-Vioglio et al. (2011). We only show the statements on star-shapedness.

(a) For cash-subadditive risk measures, the fact that these star-shapedness and quasi-star-shapedness are not necessarily equivalent is illustrated in Remark 6.2.

(b) Suppose that a cash-additive risk measure $\rho$ is star-shaped. Cash additivity and star-shapedness yield that, for all $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\lambda \in [0,1]$,

$$\rho(\lambda X + (1-\lambda)t) = \rho(\lambda X) + (1-\lambda)t \leqslant \lambda\rho(X) + (1-\lambda)\rho(t) \leqslant \max\{\rho(X), \rho(t)\},$$

which implies that $\rho$ is quasi-star-shaped.

(c) Suppose that a cash-additive risk measure $\rho$ is quasi-star-shaped. Let $\tilde{\rho} = \rho - \rho(0)$, and hence $\tilde{\rho}$ is normalized. The acceptance set of $\tilde{\rho}$ is given by

$$\mathcal{A}_{\tilde{\rho}} = \{X \in \mathcal{X} : \tilde{\rho}(X) \leqslant 0\}.$$

Note that $0 \in \mathcal{A}_{\tilde{\rho}}$ and $\tilde{\rho}$ is quasi-star-shaped. Therefore, for any $X \in \mathcal{A}_{\tilde{\rho}}$ and $\lambda \in [0,1]$, we have $\tilde{\rho}(\lambda X) \leqslant \max\{\tilde{\rho}(X), \rho(0)\} \leqslant 0$. Hence, $\lambda X \in \mathcal{A}_{\tilde{\rho}}$, and thus the set $\mathcal{A}_{\tilde{\rho}}$ is star-shaped. By Proposition 2 of Castagnoli et al. (2022), we know that $\tilde{\rho}$ is star-shaped. In turn, this implies that $\rho$ is star-shaped. $\square$

## 6.4 Representation results on cash-subadditive risk measures

In this section, we present a representation result, Theorem 6.2, of general cash-subadditive risk measures, which illustrates that a cash-subadditive risk measure is the lower envelope of a family of quasi-convex cash-subadditive risk measures. Quasi-convexity of risk measures has been extensively studied for instance in Cerreia-Vioglio et al. (2011) and Drapeau and Kupper (2013), and it is argued by many authors as the minimal property that a risk measure needs to satisfy to properly reflect diversification effects.

**Theorem 6.2.** *For a functional $\rho : \mathcal{X} \to \mathbb{R}$, the following statements are equivalent.*

*(i) $\rho$ is a cash-subadditive risk measure.*

*(ii) There exists a set $\mathcal{C}$ of quasi-convex cash-subadditive risk measures such that*

$$\rho(X) = \min_{\psi \in \mathcal{C}} \psi(X), \quad \textit{for all } X \in \mathcal{X}. \tag{6.9}$$

In order to prove Theorem 6.2, we need the following lemma, which will also be useful for a few other results.

**Lemma 6.2.** *If $\rho : \mathcal{X} \to \mathbb{R}$ is a risk measure, then $\rho(X) = \min_{Z \in \mathcal{X}} \rho_Z(X)$ for all $X \in \mathcal{X}$, where*

$$\rho_Z(X) = \inf\{\rho(Z + m) \mid m \in \mathbb{R},\ Z + m \geqslant X\}, \quad X, Z \in \mathcal{X}.$$

*Proof.* For all $X, Z \in \mathcal{X}$, by the definition of $\rho_Z$, we have

$$\rho_Z(X) = \rho(Z + \text{ess-sup}(X - Z)).$$

Note that $\rho$ is monotone, $Z + \text{ess-sup}(X - Z) \geqslant X$, and $\rho_X(X) = \rho(X)$. We have $\min_{Z \in \mathcal{X}} \rho_Z(X) = \rho(X)$. □

*Proof of Theorem 6.2.* "(ii) ⇒ (i)" is obvious. We now prove "(i) ⇒ (ii)". Assume that $\rho$ is a cash-subadditive risk measure. By Lemma 6.2, we have $\rho(X) = \min_{Z \in \mathcal{X}} \rho_Z(X)$ for all $X \in \mathcal{X}$, where

$$\rho_Z(X) = \inf\{\rho(Z + m) \mid m \in \mathbb{R}, Z + m \geqslant X\} = \rho(Z + \text{ess-sup}(X - Z)), \quad X, Z \in \mathcal{X}.$$

It is clear that $\rho_Z$ is monotonic. We show that $\rho_Z$ is cash subadditive. Indeed, for all $m \geqslant 0$ and $X \in \mathcal{X}$, we have

$$\begin{aligned}\rho_Z(X+m) = \rho(Z + \text{ess-sup}(X+m-Z)) &= \rho(Z + \text{ess-sup}(X-Z) + m) \\ &\leqslant \rho(Z + \text{ess-sup}(X-Z)) + m = \rho_Z(X) + m.\end{aligned}$$

Next, we show that $\rho_Z$ is quasi-convex. To this end, we need to show that, for all $\alpha \in \mathbb{R}$, $X_1, X_2 \in \mathcal{X}$ and $\lambda \in [0,1]$,

$$\rho_Z(X_i) \leqslant \alpha,\ i=1,2 \implies \rho_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \alpha.$$

Assume that $\rho_Z(X_i) \leqslant \alpha$ for $i = 1,2$. For all $\varepsilon > 0$ and $i = 1,2$, there exists some $m_i \in \mathbb{R}$ such that $Z + m_i \geqslant X_i$ and $\rho(Z+m_i) \leqslant \rho_Z(X_i) + \varepsilon \leqslant \alpha + \varepsilon$. Thus we have

$$\lambda X_1 + (1-\lambda)X_2 \leqslant Z + \lambda m_1 + (1-\lambda)m_2.$$

It then follows that

$$\rho_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \rho(Z + \lambda m_1 + (1-\lambda)m_2) \leqslant \rho(Z + \max\{m_1, m_2\}) \leqslant \alpha + \varepsilon.$$

The arbitrariness of $\varepsilon$ implies that $\rho_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \alpha$. Therefore, $\rho_Z$ is quasi-convex. Finally, $\{\rho_Z \mid Z \in \mathcal{X}\}$ is a desired family of quasi-convex cash-subadditive risk measures. □

*Remark* 6.3. Theorem 6.2 is more general than the result of Jia et al. (2021) for monetary risk measures, which says that any monetary risk measure can be written as the infimum of some convex risk measures. Indeed, the proof of Theorem 6.2 does not depend on, but leads to Theorem 3.1 of Jia et al. (2021) as a special case; see Proposition 6.7 in Section 6.7 for more details on this statement.

As far as we are aware of, Theorem 6.2 is the first characterization result of cash-subadditive risk measures that are not necessarily quasi-convex. We also note that, by straightforward argument, an equivalent statement of Theorem 6.2 (ii) is

$$\rho(X) = \min\{\psi(X) \mid \psi \text{ is a quasi-convex cash-subadditive risk measure},\ \psi \geqslant \rho\}, \quad X \in \mathcal{X}.$$

*Remark* 6.4. Using the same argument as for Theorem 6.2, a similar result holds for risk measures without cash subadditivity; that is, a functional $\rho : \mathcal{X} \to \mathbb{R}$ is a risk measure if and only if

$$\rho(X) = \min\{\psi(X) \mid \psi \text{ is a quasi-convex risk measure},\ \psi \geqslant \rho\}, \quad X \in \mathcal{X}.$$

**Example 6.5** ($\Lambda$VaR)**.** For a decreasing function $\Lambda : \mathbb{R} \to [0,1]$ that is not constantly 0, by Theorem 6.1, the $\Lambda$VaR in (6.2) admits the representation $\Lambda\mathrm{VaR}(X) = \inf_{x\in\mathbb{R}}\{\mathrm{VaR}_{\Lambda(x)}(X) \vee x\}$, $X \in \mathcal{X}$. Since VaR commutes with continuous increasing transforms, we have

$$\Lambda\mathrm{VaR}(X) = \inf_{x\in\mathbb{R}} \mathrm{VaR}_{\Lambda(x)}(X \vee x). \tag{6.10}$$

Let $\mathcal{C}_t$ be the set of coherent risk measures dominating $\mathrm{VaR}_t$ for $t \in (0,1)$. By Proposition 5.2 of Artzner et al. (1999), $\mathrm{VaR}_t$ has the representation

$$\mathrm{VaR}_t(X) = \min_{\rho\in\mathcal{C}_t} \rho(X), \qquad X \in \mathcal{X}. \tag{6.11}$$

For $x \in \mathbb{R}$, denote by $\tau_x : X \mapsto \tau(X \vee x)$ for $\tau \in \mathcal{C}_{\Lambda(x)}$ and by $\mathcal{C}_{\Lambda,x} = \{\tau_x : \tau \in \mathcal{C}_{\Lambda(x)}\}$. Using (6.10) and (6.11), we get the representation (6.9) for $\Lambda$VaR as

$$\Lambda\mathrm{VaR}(X) = \min\left\{\rho(X) \;\middle|\; \rho \in \bigcup_{x\in\mathbb{R}} \mathcal{C}_{\Lambda,x}\right\}, \qquad X \in \mathcal{X}. \tag{6.12}$$

We check that $\tau_x \in \mathcal{C}_{\Lambda,x}$ is cash subadditive and quasi-convex for any $x \in \mathbb{R}$. Indeed, $\tau_x$ is convex since it is the composition of a convex risk measure $\tau$ and a convex transform $y \mapsto y \vee x$. To see that it is cash subadditive, it suffices to note that $\tau_x(X + c) \leqslant \tau(X \vee x + c) = \tau_x(X) + c$ for $c \geqslant 0$.

A special case of $\Lambda$VaR is the two-level $\Lambda$VaR in Example 7 of Bellini and Peri (2021), which is the simplest form of $\Lambda$VaR different from VaR; we give a more explicit formula for this case. Fix $0 < \alpha < \beta < 1$ and $z \in \mathbb{R}$. Define $\Lambda' : x \mapsto \beta\mathbb{1}_{\{x\leqslant z\}} + \alpha\mathbb{1}_{\{x>z\}}$. The corresponding risk measure is given by $\Lambda'\mathrm{VaR}(X) = \min\{\mathrm{VaR}_\beta(X), \mathrm{VaR}_\alpha(X \vee z)\}$, $X \in \mathcal{X}$. Write $\mathcal{C}_{t,x} = \{\tau_x : \tau \in \mathcal{C}_t\}$ for $x \in \mathbb{R}$ and $t \in (0,1)$. By (6.12),

$$\Lambda'\mathrm{VaR}(X) = \min\{\rho(X) \mid \rho \in \mathcal{C}_\beta \cup \mathcal{C}_{\alpha,z}\}, \quad X \in \mathcal{X}.$$

Next, we look at a more explicit representation of cash-subadditive risk measures. An existing result of Cerreia-Vioglio et al. (2011) states that a quasi-convex cash-subadditive risk measure can be represented by the supremum of a family of functions $(t,Q) \mapsto R(t,Q)$ that are upper semi-continuous, quasi-concave, increasing and 1-Lipschitz in its first argument $t$. Combining Theorem 6.2 and Theorem 3.1 of Cerreia-Vioglio et al. (2011), we obtain a representation of a general cash-subadditive risk measure based on the above functions $R$.

**Proposition 6.3.** *A functional $\rho : \mathcal{X} \to \mathbb{R}$ is a cash-subadditive risk measure if and only if there exists a set $\mathcal{R}$ of upper semi-continuous, quasi-concave, increasing and 1-Lipschitz in the first argument functions $R : \mathbb{R} \times \mathcal{M}_f \to \mathbb{R}$ such that*

$$\rho(X) = \min_{R\in\mathcal{R}} \max_{Q\in\mathcal{M}_f} R\left(\mathbb{E}_Q[X], Q\right), \quad \textit{for all } X \in \mathcal{X}.$$

Proposition 6.3 has a similar form to the minimax representation of star-shaped risk measures in Proposition 8 of Castagnoli et al. (2022).

## 6.5 Cash-subadditive risk measures with further properties

### 6.5.1 Normalized and quasi-star-shaped cash-subadditive risk measures

In this section, we present the representation result, Theorem 6.3, of cash-subadditive risk measures that are normalized and quasi-star-shaped and other relevant technical results. Before showing Theorem 6.3, we need the representation result below of quasi-normalized, qusi-star-shaped and cash-subadditive risk measures, which is in similar sense with Lemma 6.2 but based on a more sophisticated construction with techniques different from the literature. In what follows, the convention is $\sup\emptyset = -\infty$ so that all quantities are well defined.

**Proposition 6.4.** *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a quasi-normalized, quasi-star-shaped and cash-subadditive risk measure. For $Z \in \mathcal{X}$ and $t \in \mathbb{R}$, define*

$$m_Z(t) = \sup\{m \in \mathbb{R} \mid \rho(Z+m) = t\} \quad \textit{and} \quad \mathcal{A}_Z^t = \bigcup_{\lambda\in[0,1]} \{X \in \mathcal{X} \mid X \leqslant \lambda(Z + m_Z(t)) + (1-\lambda)t\}.$$

*We have $\rho(X) = \min_{Z\in\mathcal{X}} \tilde{\rho}_Z(X)$ for $X \in \mathcal{X}$, where $\tilde{\rho}_Z(X) = \inf\{t \in \mathbb{R} \mid X \in \mathcal{A}_Z^t\}$.*

*Proof.* Since a cash-subadditive risk measure is $L^\infty$-continuous, for each $Z \in \mathcal{X}$, the range of the function $m \mapsto \rho(Z+m)$ on $\mathbb{R}$ is an interval of $\mathbb{R}$. Moreover, recall the definition of $D_\rho = \{\rho(X) \mid X \in \mathcal{X}\}$, since $\rho$ is quasi-normalized, the function $m \mapsto \rho(m)$ on $\mathbb{R}$ takes all possible values in $D_\rho$, which is an interval on $\mathbb{R}$, and by monotonicity, so does $m \mapsto \rho(Z+m)$. Hence, $\rho(Z + m_Z(t)) = t$ for all $t \in D_\rho$. For $X, Z \in \mathcal{X}$, we can write

$$\begin{aligned}\tilde{\rho}_Z(X) &= \inf\{t \in \mathbb{R} \mid X \in \mathcal{A}_Z^t\}\\ &= \inf\{t \in \mathbb{R} \mid X \leqslant \lambda(Z + m_Z(t)) + (1-\lambda)t \text{ for some } \lambda \in [0,1]\}\\ &= \inf_{\lambda\in[0,1]} \inf\{t \in \mathbb{R} \mid X \leqslant \lambda(Z + m_Z(t)) + (1-\lambda)t\}.\end{aligned}$$

It is straightforward that $\tilde{\rho}_Z(X) \in D_\rho$. For $X, Z \in \mathcal{X}$ and $t \in D_\rho$, if $\tilde{\rho}_Z(X) < t$, then $X \leqslant \lambda(Z + m_Z(t)) + (1-\lambda)t$ for some $\lambda \in [0,1]$. By monotonicity, quasi-normalization and quasi-star-shapedness of $\rho$, we have

$$\rho(X) \leqslant \rho(\lambda(Z + m_Z(t)) + (1-\lambda)t) \leqslant \max\{\rho(Z + m_Z(t)), t\} = t.$$

Thus we have $\rho(X) \leqslant \inf_{Z \in \mathcal{X}} \tilde{\rho}_Z(X)$. On the other hand,

$$\begin{aligned}\tilde{\rho}_Z(X) &\leqslant \inf\{t \in \mathbb{R} \mid X \leqslant Z + m_Z(t)\} \\ &= \inf\{t \in \mathbb{R} \mid m_Z(t) = \text{ess-sup}(X - Z)\} = \rho(Z + \text{ess-sup}(X - Z)),\end{aligned}$$

which is $\rho_Z(X)$ in Lemma 6.2. Using Lemma 6.2, we have

$$\rho(X) = \min_{Z \in \mathcal{X}} \rho_Z(X) \geqslant \inf_{Z \in \mathcal{X}} \tilde{\rho}_Z(X) \geqslant \rho(X). \tag{6.13}$$

Moreover, attainability of the infimum is guaranteed by $\rho(X) = \rho_X(X) \geqslant \tilde{\rho}_X(X) \geqslant \rho(X)$. Therefore, $\rho(X) = \min_{Z \in \mathcal{X}} \tilde{\rho}_Z(X)$ holds. $\square$

The representation in Proposition 6.4 is closely linked to that in Lemma 6.2 through (6.13).

*Remark* 6.5. Although arising from completely different considerations, the risk measure $\tilde{\rho}_Z$ in Proposition 6.4 has a similar form to an acceptability index of Cherny and Madan (2009). For more recent results on acceptability indices, see e.g., Righi (2021).

The following representation result concerns cash-subadditive risk measures that are normalized and quasi-star-shaped. We show that a normalized, quasi-star-shaped and cash-subadditive risk measure can be represented by the lower envelope of a family of ones that are normalized, quasi-convex, and cash subadditive.

**Theorem 6.3.** *For a functional $\rho : \mathcal{X} \to \mathbb{R}$, the following statements are equivalent.*

*(i) $\rho$ is a normalized, quasi-star-shaped and cash-subadditive risk measure.*

*(ii) There exists a family $\mathcal{C}$ of normalized, quasi-convex and cash-subadditive risk measures such that*

$$\rho(X) = \min_{\psi \in \mathcal{C}} \psi(X), \quad \textit{for all } X \in \mathcal{X}. \tag{6.14}$$

*Proof.* "(ii) $\Rightarrow$ (i)": Assume that there exists a family $\mathcal{C}$ of normalized, quasi-convex and cash-subadditive risk measures such that $\rho = \min_{\psi \in \mathcal{C}} \psi$. Monotonicity, normalization and cash subadditivity of $\rho$ are straightforward. Quasi-star-shapedness follows from Lemma 6.1.

"(i) $\Rightarrow$ (ii)": Assume that $\rho$ is a normalized, quasi-star-shaped and cash-subadditive risk measure. Using Proposition 6.4, it suffices to show that $\tilde{\rho}_Z(X)$ defined via

$$\tilde{\rho}_Z(X) = \inf_{\lambda \in [0,1]} \inf\{t \in \mathbb{R} \mid X \leqslant \lambda(Z + m_Z(t)) + (1 - \lambda)t\}$$

for each $Z \in \mathcal{X}$ is a normalized, quasi-convex and cash-subadditive risk measure.

We first verify that each $\tilde{\rho}_Z$ is normalized. For all $s \in \mathbb{R}$, by taking $\lambda = 0$, we have $\tilde{\rho}_Z(s) \leqslant \inf\{t \in \mathbb{R} \mid s \leqslant t\} = s$. On the other hand, for all $t \in \mathbb{R}$ and $\lambda \in [0,1]$ such that $s \leqslant \lambda(Z + m_Z(t)) + (1-\lambda)t$, by normalization, monotonicity and quasi-star-shapedness of $\rho$, we have

$$s = \rho(s) \leqslant \rho(\lambda(Z + m_Z(t)) + (1-\lambda)t) \leqslant \max\{\rho(Z + m_Z(t)), t\} = t.$$

Hence we obtain $\tilde{\rho}_Z(s) \geqslant s$, and further $\tilde{\rho}_Z(s) = s$.

Next, we show that each $\tilde{\rho}_Z$ is quasi-convex. We first note that $\mathcal{A}_Z^t$ is a convex set for each $t \in \mathbb{R}$, which follows from the fact that $\mathcal{A}_Z^t$ is the set of all $X \in \mathcal{X}$ dominated by the segment $\{\lambda(Z+m_Z(t))+(1-\lambda)t \mid \lambda \in [0,1]\}$. Take $t \in \mathbb{R}$. For $X_1, X_2$ satisfying $\tilde{\rho}_Z(X_1) \leqslant \tilde{\rho}_Z(X_2) \leqslant t$, for any $s > t$, we have $X_1, X_2 \in \mathcal{A}_Z^s$. Convexity of $\mathcal{A}_Z^s$ implies, for each $\lambda \in [0,1]$, $\lambda X_1 + (1-\lambda)X_2 \in \mathcal{A}_Z^s$, and it further gives $\tilde{\rho}_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant s$. Since $s > t$ is arbitrary, we have $\tilde{\rho}_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant t$. This gives quasi-convexity of $\tilde{\rho}_Z$.

Finally, we prove that $\tilde{\rho}_Z$ is cash subadditive for all $Z \in \mathcal{X}$. Since $\rho$ is cash subadditive, for all $Z \in \mathcal{X}$, $t \in \mathbb{R}$ and $c \geqslant 0$,

$$\begin{aligned} m_Z(t+c) &= \sup\{m + c \in \mathbb{R} \mid \rho(Z + m + c) = t + c\} \\ &\geqslant \sup\{m \in \mathbb{R} \mid \rho(Z + m) = t\} + c = m_Z(t) + c. \end{aligned}$$

For all $c \geqslant 0$ and $X \in \mathcal{X}$, we have

$$\begin{aligned} \tilde{\rho}_Z(X + c) &= \inf_{\lambda \in [0,1]} \inf\{t \in \mathbb{R} \mid X + c \leqslant \lambda(Z + m_Z(t)) + (1-\lambda)t\} \\ &= \inf_{\lambda \in [0,1]} \inf\{t + c \in \mathbb{R} \mid X + c \leqslant \lambda(Z + m_Z(t+c)) + (1-\lambda)(t+c)\} \\ &= \inf_{\lambda \in [0,1]} \inf\{t \in \mathbb{R} \mid X \leqslant \lambda(Z + m_Z(t+c) - (t+c)) + t\} + c \\ &\leqslant \inf_{\lambda \in [0,1]} \inf\{t \in \mathbb{R} \mid X \leqslant \lambda(Z + m_Z(t) - t) + t\} + c = \tilde{\rho}_Z(X) + c. \end{aligned}$$

In summary, $\{\tilde{\rho}_Z \mid Z \in \mathcal{X}\}$ is a desired family of normalized, quasi-convex and cash-subadditive risk measures. $\square$

The proof of Theorem 6.3 is based on a delicate construction of the dominating risk measures, different from those used for Theorem 6.2. Normalization in both (i) and (ii) of Theorem 6.3 is important and cannot be removed, but it can be replaced by quasi-normalization. The modified version of Theorem 6.3 using quasi-normalization follows from combining Proposition 6.4 and Lemma 6.1.

Theorem 6.3 can be seen as a parallel result, although obtained via different techniques, to the representation result of Castagnoli et al. (2022), which uses star-shapedness, convexity, and cash additivity instead of quasi-star-shapedness, quasi-convexity and cash subadditivity. It is clear that (ii) of Theorem 6.3 is equivalent to the following alternative formulation

$$\rho(X) = \min\left\{ \psi(X) \;\middle|\; \begin{array}{l} \psi \text{ is a normalized, quasi-convex and} \\ \text{cash-subadditive risk measure, } \psi \geqslant \rho \end{array} \right\}, \quad X \in \mathcal{X}. \tag{6.15}$$

### 6.5.2 SSD-consistent cash-subadditive risk measures

In this section, we present below the representation result of SSD-consistent cash-subadditive risk measures. For this, we define the *Expected Shortfall* (ES) at level $t \in [0,1]$ as

$$\mathrm{ES}_t(X) = \frac{1}{1-t}\int_t^1 \mathrm{VaR}_\alpha(X)\,\mathrm{d}\alpha,\ t \in [0,1) \ \text{ and } \ \mathrm{ES}_1(X) = \text{ess-sup}(X), \quad X \in \mathcal{X}.$$

As a coherent alternative to VaR, ES is the most important risk measure in current banking regulation; see Wang and Zitikis (2021) for its role in the Basel Accords and an axiomatization. It is well known that the class of ES characterizes SSD via

$$X \succeq_2 Y \iff \mathrm{ES}_t(X) \geqslant \mathrm{ES}_t(Y) \text{ for all } t \in [0,1].$$

Mao and Wang (2020) investigated SSD-consistent monetary risk measures and provided four equivalent conditions of SSD-consistency; see their Theorem 2.1. The result can also be extended to cash-subadditive risk measures, which is shown in the following lemma.

**Lemma 6.3.** *Let $\rho$ be a cash-subadditive risk measure on $\mathcal{X}$. The following are equivalent.*

*(i)* $\rho$ *is SSD-consistent.*

*(ii)* $\rho(X) \geqslant \rho(Y)$ *for all* $X, Y \in \mathcal{X}$ *with* $X \succeq_2 Y$ *and* $\mathbb{E}[X] = \mathbb{E}[Y]$.

*(iii)* $\rho(X) \geqslant \rho(Y)$ *for all* $X, Y \in \mathcal{X}$ *with* $\mathbb{E}[(X-K)_+] \geqslant \mathbb{E}[(Y-K)_+]$ *for all* $K \in \mathbb{R}$.

*(iv)* $\rho(X) \geqslant \rho(Y)$ *for all* $X, Y \in \mathcal{X}$ *with* $Y = \mathbb{E}[X \mid Y]$.

*(v)* $\rho(X^c + Y^c) \geqslant \rho(X+Y)$ *for all* $X, Y, X^c, Y^c \in \mathcal{X}$ *such that* $(X^c, Y^c)$ *is comonotonic,* $X \overset{\mathrm{d}}{=} X^c$, *and* $Y \overset{\mathrm{d}}{=} Y^c$.

*Moreover, any of these properties imply that* $\rho$ *is law invariant.*

*Proof.* The equivalence among (i)-(iv) is easy to verify from classic properties of SSD by the same logic of the proof of Theorem 2.1 in Mao and Wang (2020). The equivalence between (i) and (v) for continuous functions follows from Theorem 2 of Wang and Wei (2020). Hence, we only need to show that cash-subadditive risk measures also satisfy $\|\cdot\|_\infty$-continuity; namely, $\lim_{n\to\infty} \rho(X_n) = \rho(X)$ for any sequence $X_n \in \mathcal{X}$ satisfying ess-sup$(|X_n - X|) \to 0$ as $n \to \infty$. Clearly, for all $X, Y \in \mathcal{X}$, $X \leqslant Y + \|X - Y\|$. By monotonicity and cash subadditivity of $\rho$, we have $\rho(X) - \rho(Y) \leqslant \|X - Y\|$. Switching the roles of $X$ and $Y$ yields the assertion. □

The following lemma is needed in the proof of Theorem 6.4, which was obtained by Cerreia-Vioglio et al. (2011) with the additional assumption of continuity from above. We include a self-contained proof of Lemma 6.4.

**Lemma 6.4.** *If $\rho : \mathcal{X} \to \mathbb{R}$ is a quasi-convex cash-subadditive risk measure, then $\rho$ is law invariant if and only if $\rho$ is SSD-consistent.*

*Proof.* It is obvious that SSD-consistency implies law invariance. We will only show the "only if" statement. By Lemma 6.3, it suffices to show that $\rho(X) \leqslant \rho(Y)$ for $X \preceq_2 Y$ satisfying $\mathbb{E}[X] = \mathbb{E}[Y]$. By Proposition 3.6 of Mao and Wang (2015), there exists a sequence of $\mathbf{Y}^k = (Y_1^k, \ldots, Y_{n_k}^k)$, $k \in \mathbb{N}$, such that each $Y_j^k \stackrel{\mathrm{d}}{=} Y$, $n_k \to \infty$ as $k \to \infty$, and

$$\frac{1}{n_k} \sum_{j=1}^{n_k} Y_j^k \to X \quad \text{in } L^\infty.$$

Note that a cash-subadditive risk measure is $L^\infty$-continuous. Quasi-convexity and $L^\infty$-continuity lead to

$$\rho(Y) \geqslant \rho\left(\frac{1}{n_k} \sum_{j=1}^{n_k} Y_j^k\right) \to \rho(X),$$

and thus $\rho$ is SSD-consistent. □

In the following theorem, we establish a representation for an SSD-consistent cash-subadditive risk measure as the lower envelope of some family of law-invariant, quasi-convex and cash-subadditive risk measures.

**Theorem 6.4.** *For a functional $\rho : \mathcal{X} \to \mathbb{R}$, the following statements are equivalent.*

*(i) $\rho$ is an SSD-consistent cash-subadditive risk measure.*

(ii) *There exists a family $\mathcal{C}$ of law-invariant, quasi-convex and cash-subadditive risk measures such that*

$$\rho(X) = \min_{\psi\in\mathcal{C}} \psi(X), \quad \text{for all } X \in \mathcal{X}.$$

*Proof.* "(ii) $\Rightarrow$ (i)" is implied by Lemma 6.4 and the fact that cash subadditivity and SSD-consistency are preserved under the infimum operation. We will show "(i) $\Rightarrow$ (ii)".

Suppose that $\rho$ is an SSD-consistent cash-subadditive risk measure. For all $X \in \mathcal{X}$ and $Z \in \mathcal{X}$, define the risk measure

$$\psi_Z(X) = \inf\{\rho(Z+m) \mid m \in \mathbb{R},\ Z+m \succeq_2 X\}.$$

It is straightforward to check that $\rho(X) = \min_{Z\in\mathcal{X}} \psi_Z(X)$ and

$$\begin{aligned}\psi_Z(X) &= \inf\{\rho(Z+m) \mid m \in \mathbb{R},\ \mathrm{ES}_t(Z)+m \geqslant \mathrm{ES}_t(X), \text{ for all } t \in [0,1]\} \\ &= \rho\left(Z + \sup_{t\in[0,1]} (\mathrm{ES}_t(X) - \mathrm{ES}_t(Z))\right).\end{aligned}$$

It is clear that $\psi_Z$ is monotone, cash subadditive and law invariant. We prove that $\psi_Z$ is quasi-convex with similar manner to the proof of Theorem 6.2. Assume that $\psi_Z(X_i) \leqslant \alpha$ for $i = 1, 2$. For all $\varepsilon > 0$ and $i = 1, 2$, there exists some $m_i \in \mathbb{R}$ such that $Z+m_i \succeq_2 X_i$ and $\rho(Z+m_i) \leqslant \psi_Z(X_i)+\varepsilon \leqslant \alpha+\varepsilon$. We have $Z + \lambda m_1 + (1-\lambda)m_2 \succeq_2 \lambda X_1 + (1-\lambda)X_2$ for all $\lambda \in [0, 1]$. This can be obtained by observing that

$$\mathrm{ES}_t(\lambda X_1 + (1-\lambda)X_2) \leqslant \lambda \mathrm{ES}_t(X_1) + (1-\lambda)\mathrm{ES}_t(X_2) \leqslant \mathrm{ES}_t(Z + \lambda m_1 + (1-\lambda)m_2),$$

for all $t \in [0, 1]$, due to convexity and cash additivity of $\mathrm{ES}_t$. It follows that

$$\psi_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \rho(Z + \lambda m_1 + (1-\lambda)m_2) \leqslant \rho(Z + \max\{m_1, m_2\}) \leqslant \alpha + \varepsilon.$$

The arbitrariness of $\varepsilon$ implies that $\psi_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \alpha$. Therefore, $\psi_Z$ is quasi-convex. We conclude that $\{\psi_Z \mid Z \in \mathcal{X}\}$ is a desired family of law-invariant, quasi-convex and cash-subadditive risk measures. □

Theorem 6.4 can be seen as a parallel result to Theorem 3.3 of Mao and Wang (2020) which showed that any SSD-consistent monetary risk measure is the lower envelope of law-invariant and convex monetary risk measures. Similarly to (6.15), we can reformulate (ii) of Theorem 6.4 as

$$\rho(X) = \min\left\{\psi(X) \;\middle|\; \begin{array}{l}\psi \text{ is a law-invariant, quasi-convex and} \\ \text{cash-subadditive risk measure, } \psi \geqslant \rho\end{array}\right\}, \quad X \in \mathcal{X}.$$

A representation result in a similar spirit to Proposition 6.3 for SSD-consistent cash-subadditive risk measures follows directly from Theorem 5.1 of Cerreia-Vioglio et al. (2011) and Theorem 6.4.

**Proposition 6.5.** *A functional $\rho : \mathcal{X} \to \mathbb{R}$ is an SSD-consistent cash-subadditive risk measure if and only if there exists a set $\mathcal{R}$ of upper semi-continuous, quasi-concave, increasing and 1-Lipschitz in the first component functions $R : \mathbb{R} \times \mathcal{M} \to \mathbb{R}$ such that*

$$\rho(X) = \min_{R\in\mathcal{R}} \max_{Q\in\mathcal{M}_f} R\left(\int_0^1 \mathrm{VaR}_t(X)\mathrm{VaR}_t\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}t,\ Q\right), \quad \textit{for all } X \in \mathcal{X}.$$

### 6.5.3 Risk sharing with SSD-consistent cash-subadditive risk measures

Risk sharing problems for law-invariant and cash-additive risk measures are extensively studied in the literature; see e.g., Barrieu and El Karoui (2005), Jouini et al. (2008) and Filipović and Svindland (2008) for convex risk measures and Embrechts et al. (2018, 2020), Liebrich (2021) and Liu et al. (2022) for some classes of non-convex risk measures. Moreover, Mao and Wang (2020) discussed the risk sharing problem for consistent risk measures, i.e., risk measures that are SSD-consistent and cash additive. Cash additivity is assumed in the above results. We present a simple result in this section on cash-subadditive and SSD-consistent risk measures.

For a given random loss $X \in \mathcal{X}$, the set of its possible allocations is defined as $\mathbb{A}_n(X) = \{(X_1, \dots, X_n) \in \mathcal{X}^n : X_1 + \cdots + X_n = X\}$. The *inf-convolution* of risk measures $\rho_1, \dots, \rho_n : \mathcal{X} \to \mathbb{R}$ is defined as

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \inf\left\{\sum_{i=1}^{n} \rho_i(X_i) \;\middle|\; (X_1, \dots, X_n) \in \mathbb{A}_n(X)\right\}, \quad X \in \mathcal{X}. \tag{6.16}$$

The solution $(X_1^*, \dots, X_n^*) \in \mathbb{A}_n(X)$ to the problem (6.16), if exists, is called *optimal*.

**Proposition 6.6.** *Let $\rho_1, \dots, \rho_n : \mathcal{X} \to \mathbb{R}$ be SSD-consistent cash-subadditive risk measures.*

*(i) Define $\mathbb{A}_n^c(X) = \{(X_1, \dots, X_n) \in \mathbb{A}_n(X) \mid (X_1, \dots, X_n)$ is comonotonic$\}$. We have*

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \inf\left\{\sum_{i=1}^{n} \rho_i(X_i) \;\middle|\; (X_1, \dots, X_n) \in \mathbb{A}_n^c(X)\right\}, \quad X \in \mathcal{X}.$$

*(ii) $\square_{i=1}^{n} \rho_i$ is an SSD-consistent cash-subadditive risk measure.*

*Proof.* (i) This follows similarly from the proof of Theorem 4.1 of Mao and Wang (2020).

(ii) It is straightforward that $\square_{i=1}^{n} \rho_i$ is a cash-subadditive risk measure. To prove that $\square_{i=1}^{n} \rho_i$ is SSD-consistent, by Lemma 6.3, we show that $\square_{i=1}^{n} \rho_i(X) \geqslant \square_{i=1}^{n} \rho_i(Y)$ for all $X, Y \in \mathcal{X}$ such that $Y = \mathbb{E}[X \mid Y]$. Indeed, we take $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ and $Y_i = \mathbb{E}[X_i \mid Y]$ for all $i = 1, \dots, n$ and have $(Y_1, \dots, Y_n) \in \mathbb{A}_n(Y)$. Since $\rho_1, \dots, \rho_n$ are SSD-consistent and cash subadditive, we have by Lemma 6.3 that $\rho_i(X_i) \geqslant \rho_i(Y_i)$ for all $i = 1, \dots, n$. It follows that

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \inf\left\{ \sum_{i=1}^{n} \rho_i(X_i) \;\middle|\; (X_1, \dots, X_n) \in \mathbb{A}_n(X) \right\}$$
$$\geqslant \inf\left\{ \sum_{i=1}^{n} \rho_i(Y_i) \;\middle|\; (Y_1, \dots, Y_n) \in \mathbb{A}_n(Y) \right\} = \mathop{\square}_{i=1}^{n} \rho_i(Y). \qquad \square$$

*Remark* 6.6. Different from Theorem 4.1 of Mao and Wang (2020), we note that the inf-convolution of SSD-consistent cash-subadditive risk measures does not need to be finite. An example is given by $\rho : X \mapsto \min\{\mathbb{E}[X], 0\}$. It is straightforward to verify that $\rho$ is an SSD-consistent cash-subadditive risk measure. The inf-convolution of $\rho$ and $\rho$ at 0 is $\inf_{X \in \mathcal{X}}\{\min\{\mathbb{E}[X], 0\} + \min\{-\mathbb{E}[X], 0\}\} = -\infty$.

## 6.6 Conclusion

We provide a systemic study of cash-subadditive risk measures, which were traditionally studied together with convexity (El Karoui and Ravanelli, 2009) or quasi-convexity (Cerreia-Vioglio et al., 2011). Different from the literature, our study focuses on cash-subadditive risk measures without quasi-convexity, which include many natural examples as discussed in the chapter. As our major technical contributions, a general cash-subadditive risk measure is shown to be representable by the lower envelope of a family of quasi-convex cash-subadditive risk measures (Theorem 6.2). The notions of quasi-star-shapedness and quasi-normalization were introduced as analogues of star-shapedness and normalization studied by Castagnoli et al. (2022). It turns out that quasi-star-shapedness and quasi-normalization fit naturally in the setting of cash subadditivity, leading to a new representation result (Theorem 6.3). A representation result of SSD-consistent cash-subadditive risk measures was also obtained (Theorem 6.4). Furthermore, we obtain several results on the risk measure ΛVaR proposed by Frittelli et al. (2014), including a new representation result (Theorem 6.1). In particular, the class of ΛVaR serves as a natural example of quasi-star-shaped, quasi-normalized and cash-subadditive risk measures, which are not star-shaped, normalized, or cash additive.

Risk measures without cash additivity have received increasing attention in the recent literature due to their technical generality and intimate connection to decision analysis, risk transforms, portfolio optimization, and stochastic interest rates; many references and examples were mentioned in the introduction and throughout the chapter. Results in this chapter serve as a building block for future studies on cash subadditivity and the new properties of quasi-star-shapedness and quasi-normalization, for which many questions and applications remain to be explored.

## 6.7 Additional results and technical discussions

This section includes a few additional technical results, examples and discussions of the representation results of cash-subadditive risk measures, which are not used in the main text of the chapter. Some of them may be of independent interest.

### 6.7.1 Connection to a representation of monetary risk measures

The proposition below illustrates how Lemma 6.2 can be used to show that any monetary risk measure is the minimum of some convex risk measures (Jia et al., 2021, Theorem 3.1).

**Proposition 6.7.** *A functional $\rho : \mathcal{X} \to \mathbb{R}$ is a monetary risk measure if and only if*

$$\rho(X) = \min_{Z \in \mathcal{A}_\rho} \text{ess-sup}(X - Z), \quad \textit{for all } X \in \mathcal{X},$$

*where $\mathcal{A}_\rho$ is the acceptance set of $\rho$ given by $\mathcal{A}_\rho = \{Z \in \mathcal{X} \mid \rho(Z) \leqslant 0\}$.*

*Proof.* The "if" part is straightforward. We prove the "only if" part. For all $X \in \mathcal{X}$, by Lemma 6.2 and cash additivity of $\rho$, we have

$$\rho(X) = \min_{Z \in \mathcal{X}} \rho(Z + \text{ess-sup}(X - Z)) = \min_{Z \in \mathcal{X}} \{\rho(Z) + \text{ess-sup}(X - Z)\}.$$

By taking $Z_0 = X - \rho(X)$, we have $\rho(Z_0) + \text{ess-sup}(X - Z_0) = \rho(X)$, where the minimum is obtained. Define $\mathcal{A}_\rho^0 = \{Z \in \mathcal{X} \mid \rho(Z) = 0\}$. We have $Z_0 \in \mathcal{A}_\rho^0$ and thus

$$\rho(X) = \min_{Z \in \mathcal{A}_\rho^0} (\rho(Z) + \text{ess-sup}(X - Z)) = \min_{Z \in \mathcal{A}_\rho^0} \text{ess-sup}(X - Z) \geqslant \min_{Z \in \mathcal{A}_\rho} \text{ess-sup}(X - Z).$$

On the other hand, since $\rho(Z) \leqslant 0$ for all $Z \in \mathcal{A}_\rho$, we have

$$\rho(X) = \min_{Z \in \mathcal{A}_\rho} \{\rho(Z) + \text{ess-sup}(X - Z)\} \leqslant \min_{Z \in \mathcal{A}_\rho} \text{ess-sup}(X - Z).$$

Therefore, we have $\rho(X) = \min_{Z \in \mathcal{A}_\rho} \text{ess-sup}(X - Z)$. □

### 6.7.2 Comonotonic quasi-convexity

Since law-invariant, quasi-convex and cash-subsdditive risk measures are SSD-consistent (Lemma 6.4), a general law-invariant cash-subadditive risk measure (such as VaR in Section 6.2) does not admit a representation via the lower envelope of a family of law-invariant, quasi-convex and cash-subsdditive risk measures. One remaining question is whether a law-invariant cash-subadditive risk measure can be represented as the infimum of a set of law-invariant cash-subadditive risk measures with some other properties. For such a representation, we need comonotonic quasi-convexity.

*Comonotonic quasi-convexity*: $\rho(\lambda X + (1-\lambda)Y) \leqslant \max\{\rho(X), \rho(Y)\}$ for all comonotonic $(X, Y) \in \mathcal{X}^2$ and $\lambda \in [0, 1]$.

The property of comonotonic quasi-convexity appeared in various contexts; e.g., Xia (2013), Tian and Long (2015) and Li and Wang (2019). Before showing the representation result, we first give the following equivalence result demonstrating the relations among several properties of $\rho$, similarly to Lemma 6.4.

**Lemma 6.5.** *If $\rho : \mathcal{X} \to \mathbb{R}$ is a cash-subadditive risk measure, then $\rho$ is law invariant and quasi-convex if and only if $\rho$ is SSD-consistent and comonotonic quasi-convex.*

*Proof.* The "only if" part follows directly from Lemma 6.4. We prove the "if" part. Suppose that $\rho$ is SSD-consistent and comonotonic quasi-convex. It is clear that $\rho$ is law invariant by taking $X \stackrel{\mathrm{d}}{=} Y$ and observing $X \succeq_2 Y$ and $Y \succeq_2 X$. For all $X, Y \in \mathcal{X}$, take $X^c, Y^c \in \mathcal{X}$ such that $(X^c, Y^c)$ is comonotonic, $X^c \stackrel{\mathrm{d}}{=} X$, and $Y^c \stackrel{\mathrm{d}}{=} Y$. It follows that $\lambda X^c + (1-\lambda)Y^c \succeq_2 \lambda X + (1-\lambda)Y$ for all $\lambda \in [0, 1]$ (see e.g., Rüschendorf, 2013, Theorem 3.5). Hence, we have

$$\rho(\lambda X + (1-\lambda)Y) \leqslant \rho(\lambda X^c + (1-\lambda)Y^c) \leqslant \max\{\rho(X^c), \rho(Y^c)\} = \max\{\rho(X), \rho(Y)\},$$

which indicates that $\rho$ is quasi-convex. □

With the extra requirement of comonotonic quasi-convexity, we obtain a unifying umbrella for the representation of cash-subadditive risk measures with various properties. This result is parallel to the result of Jia et al. (2021) on monetary risk measures, where comonotonic convexity (Song and Yan, 2006, 2009) is in place of our comonotonic quasi-convexity.

**Proposition 6.8.** *For a functional $\rho : \mathcal{X} \to \mathbb{R}$, we have the following statements.*

*(i) $\rho$ is a cash-subadditive risk measure if and only if it is the lower envelope of a family of comonotonic quasi-convex and cash-subadditive risk measures.*

*(ii) $\rho$ is a law-invariant cash-subadditive risk measure if and only if it is the lower envelope of a family of law-invariant, comonotonic quasi-convex and cash-subadditive risk measures.*

*The equivalence (ii) holds true if "law-invariant" is replaced by "normalized and quasi-star-shaped" or "SSD-consistent".*

*Proof.* Note that each of law invariance, normalization, quasi-star-shapedness, SSD-consistency and cash subadditivity is preserved under taking an infimum, and hence the "if" parts in all statements are obvious. Since comonotonic quasi-convexity is weaker than quasi-convexity, the representations ("only if") in Theorems 6.2, 6.3 and 6.4 hold true by replacing quasi-convexity with comonotonic quasi-convexity. This, together with Lemma 6.4, gives the "only if" parts except for the case of law-invariant cash-subadditive risk measures in (ii). Below we show this part.

Assume $\rho$ is a law-invariant cash-subadditive risk measure. According to Proposition 6.10 below, for all $X \in \mathcal{X}$, we have $\rho(X) = \min_{Z\in\mathcal{X}} \phi_Z(X)$ in which

$$\phi_Z(X) = \rho\left(Z + \sup_{t\in(0,1)} (\mathrm{VaR}_t(X) - \mathrm{VaR}_t(Z))\right).$$

It is clear that $\phi_Z$ is monotone, cash subadditive and law invariant. We prove that $\phi_Z$ is comonotonic quasi-convex by the similar way to Theorems 6.2 and 6.4. Assume that $(X_1, X_2) \in \mathcal{X}^2$ is comonotonic and $\phi_Z(X_i) \leqslant \alpha$ for $i = 1, 2$. For all $\varepsilon > 0$ and $i = 1, 2$, there exists some $m_i \in \mathbb{R}$ such that $Z + m_i \succeq_1 X_i$ and $\rho(Z + m_i) \leqslant \phi_Z(X_i) + \varepsilon \leqslant \alpha + \varepsilon$. For all $\lambda \in [0, 1]$, comonotonic additivity of $\mathrm{VaR}_t$ yields that

$$\mathrm{VaR}_t(\lambda X_1 + (1-\lambda)X_2) = \lambda \mathrm{VaR}_t(X_1) + (1-\lambda)\mathrm{VaR}_t(X_2) \leqslant \mathrm{VaR}_t(Z + \lambda m_1 + (1-\lambda)m_2),$$

for all $t \in (0, 1)$. We thus have $Z + m \succeq_1 \lambda X_1 + (1-\lambda)X_2$. It follows that

$$\phi_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \rho(Z + \lambda m_1 + (1-\lambda)m_2) \leqslant \rho(Z + \max\{m_1, m_2\}) \leqslant \alpha + \varepsilon.$$

Since $\varepsilon$ is arbitrary, $\phi_Z(\lambda X_1 + (1-\lambda)X_2) \leqslant \alpha$. Therefore, $\phi_Z$ is comonotonic quasi-convex. $\square$

### 6.7.3 Writing cash-subadditive risk measures via monetary ones

Theorem 6.2 focuses on the representation of a general cash-subadditive risk measure in terms of some family of quasi-convex cash-subadditive risk measures. In general, we cannot write a cash-subadditive risk measure as the lower envelope of monetary risk measures, because cash additivity is preserved under infimum operations. Nevertheless, we can obtain a connection by allowing the monetary risk measures to be indexed by some functions $\beta$ which depend on $X$.

**Proposition 6.9.** *For a risk measure $\rho : \mathcal{X} \to \mathbb{R}$, the following statements are equivalent.*

*(i) $\rho$ is cash subadditive.*

*(ii) $\rho$ satisfies*

$$\rho(X) = \min_{\beta \in \mathcal{B}_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta(Q)\}, \quad \text{for all } X \in \mathcal{X}, \tag{6.17}$$

*where $\mathcal{B}_X$ is a set of mappings from $\mathcal{M}_f$ to $(-\infty, \infty]$ for all $X \in \mathcal{X}$, with $\mathcal{B}_{X_1} \subseteq \mathcal{B}_{X_2}$ for all $X_1, X_2 \in \mathcal{X}$ such that $X_1 \leqslant X_2$. Moreover, the set $\mathcal{B}_X$ can be chosen as*

$$\mathcal{B}_X = \{\beta : \mathcal{M}_f \to (-\infty, \infty],\ Q \mapsto \mathbb{E}_Q[Z] - \rho(Z) \mid Z \in \mathcal{X},\ Z \leqslant X\}, \quad \text{for all } X \in \mathcal{X}.$$

*(iii) $\rho$ satisfies*

$$\rho(X) = \min_{\psi \in \mathcal{C}_X} \psi(X), \quad \text{for all } X \in \mathcal{X},$$

*where $\mathcal{C}_X$ is a set of convex monetary risk measures on $\mathcal{X}$ for all $X \in \mathcal{X}$, with $\mathcal{C}_{X_1} \subseteq \mathcal{C}_{X_2}$ for all $X_1, X_2 \in \mathcal{X}$ such that $X_1 \leqslant X_2$.*

*(iv) $\rho$ satisfies*

$$\rho(X) = \min \left\{ \psi(X) \;\middle|\; \begin{array}{l} \psi \text{ is a convex monetary risk measure,} \\ \psi(X+c) \geqslant \rho(X+c) \text{ for all } c \geqslant 0 \end{array} \right\}, \quad \text{for all } X \in \mathcal{X}.$$

*Proof.* "(ii) ⇒ (iii)" is straightforward. We prove "(i) ⇒ (ii)", "(iii) ⇒ (iv)" and "(iv) ⇒ (i)".

(i) ⇒ (ii): Suppose that $\rho$ is cash subadditive. For all $X \in \mathcal{X}$, by Lemma 6.2, we have

$$\rho(X) = \min_{Z \in \mathcal{X}} \rho(Z + \text{ess-sup}(X - Z)).$$

Since the minimum above can be obtained by taking $Z = X$, we have

$$\rho(X) = \min_{Z \in \mathcal{X},\ Z \leqslant X} \rho(Z + \text{ess-sup}(X - Z)) \leqslant \min_{Z \in \mathcal{X},\ Z \leqslant X} \{\text{ess-sup}(X - Z) + \rho(Z)\} \leqslant \rho(X),$$

where the second last inequality is due to cash subadditivity of $\rho$ and the last inequality is implied by taking $Z = X$. Hence, we have

$$\begin{aligned} \rho(X) = \min_{Z \in \mathcal{X},\ Z \leqslant X} \{\text{ess-sup}(X - Z) + \rho(Z)\} &= \min_{Z \in \mathcal{X},\ Z \leqslant X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X - Z] + \rho(Z)\} \\ &= \min_{Z \in \mathcal{X},\ Z \leqslant X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta_Z(Q)\}, \end{aligned}$$

where $\beta_Z(Q) = \mathbb{E}_Q[Z] - \rho(Z)$ for all $Q \in \mathcal{M}_f$. Therefore, we show that $\{\beta_Z \mid Z \in \mathcal{X},\ Z \leqslant X\}$ is a desired family of convex functionals on $\mathcal{M}_f$.

(iii) $\Rightarrow$ (iv): Suppose that for $X \in \mathcal{X}$, there exists a family $\mathcal{C}_X$ of convex monetary risk measures on $\mathcal{X}$ such that $\rho(X) = \min_{\psi \in \mathcal{C}_X} \psi(X)$, where the sets $\mathcal{C}_{X_1} \subseteq \mathcal{C}_{X_2}$ for all $X_1, X_2 \in \mathcal{X}$ such that $X_1 \leqslant X_2$. It suffices to show that $\psi(X+c) \geqslant \rho(X+c)$ for all $\psi \in \mathcal{C}_X$ and all $c \geqslant 0$.

For all $m \geqslant 0$, since $\mathcal{C}_X \subseteq \mathcal{C}_{X+m}$ and $\psi$ is cash additive for all $\psi \in \mathcal{C}_{X+m}$, we have

$$\rho(X+m) = \min_{\psi \in \mathcal{C}_{X+m}} \psi(X+m) \leqslant \min_{\psi \in \mathcal{C}_X} \psi(X) + m = \rho(X) + m.$$

Thus $\rho$ is cash subadditive. It follows that for all $\psi \in \mathcal{C}_X$ and all $c \geqslant 0$, we have

$$\psi(X+c) = \psi(X) + c \geqslant \rho(X) + c \geqslant \rho(X+c).$$

(iv) $\Rightarrow$ (i): For all $X \in \mathcal{X}$ and $m \geqslant 0$, we have

$$\begin{aligned}
\rho(X+m) &= \min\left\{ \psi(X+m) \;\middle|\; \begin{array}{l} \psi \text{ is a convex monetary risk measure,} \\ \psi(X+m+c) \geqslant \rho(X+m+c) \text{ for all } c \geqslant 0 \end{array} \right\} \\
&\leqslant \min\left\{ \psi(X) + m \;\middle|\; \begin{array}{l} \psi \text{ is a convex monetary risk measure,} \\ \psi(X+c) \geqslant \rho(X+c) \text{ for all } c \geqslant 0 \end{array} \right\} \\
&= \rho(X) + m.
\end{aligned}$$

Thus $\rho$ is cash subadditive. □

A particularly interesting property is (iv) of Proposition 6.9, which says that a risk measure $\rho$ is cash subadditive if and only if, for each $X$, $\rho(X)$ is the minimum of $\psi(X)$ for all convex monetary risk measures $\psi$ dominating $\rho$ for risks of the type $X + c$ for $c \geqslant 0$.

*Remark* 6.7. It is straightforward to see that the equivalence result in Proposition 6.9 holds true if the maximum in (6.17) is replaced by a supremum. In that case, the set $\mathcal{M}_f$ can be conveniently replaced by $\mathcal{M}$.

### 6.7.4 Law-invariant cash-subadditive risk measures and VaR

We first connect law-invariant cash-subadditive risk measures to VaR defined in Section 6.2. It is well known that the class of VaR characterizes FSD via

$$X \succeq_1 Y \iff \mathrm{VaR}_t(X) \geqslant \mathrm{VaR}_t(Y) \text{ for all } t \in (0,1).$$

**Proposition 6.10.** *If $\rho : \mathcal{X} \to \mathbb{R}$ is a risk measure, then $\rho$ is law invariant and cash subadditive if and only if it satisfies*

$$\rho(X) = \min_{g \in \mathcal{G}_X} \sup_{t \in (0,1)} \{\mathrm{VaR}_t(X) - g(t)\}, \quad \textit{for all } X \in \mathcal{X},$$

*where $\mathcal{G}_X$ is a set of measurable functions from $(0,1)$ to $(-\infty, \infty]$ for all $X \in \mathcal{X}$, with $\mathcal{G}_{X_1} \subseteq \mathcal{G}_{X_2}$ for all $X_1, X_2 \in \mathcal{X}$ such that $X_2 \succeq_1 X_1$. Moreover, the set $\mathcal{G}_X$ can be chosen as*

$$\mathcal{G}_X = \{g : (0,1) \to (-\infty, \infty],\ t \mapsto \mathrm{VaR}_t(Z) - \rho(Z) \mid Z \in \mathcal{X},\ X \succeq_1 Z\}, \quad \textit{for all } X \in \mathcal{X}.$$

*Proof.* "$\Rightarrow$": Suppose that $\rho$ is a law-invariant cash-subadditive risk measure. For all $X \in \mathcal{X}$ and $Z \in \mathcal{X}$, define the risk measure

$$\phi_Z(X) = \inf\{\rho(Z + m) \mid m \in \mathbb{R},\ Z + m \succeq_1 X\}.$$

For all $m \in \mathbb{R}$ such that $Z + m \succeq_1 X$, since any law-invariant risk measure is FSD-consistent (e.g., Föllmer and Schied, 2016, Remark 4.58), we have $\rho(Z+m) \geqslant \rho(X)$. It follows that $\phi_Z(X) \geqslant \rho(X)$ for all $Z \in \mathcal{X}$. Noting that $\phi_X(X) = \rho(X)$, we have $\rho(X) = \min_{Z \in \mathcal{X}} \phi_Z(X)$. By definition of $\phi_Z$, we have

$$\begin{aligned}\phi_Z(X) &= \inf\{\rho(Z+m) \mid m \in \mathbb{R},\ \mathrm{VaR}_t(Z) + m \geqslant \mathrm{VaR}_t(X) \text{ for all } t \in (0,1)\} \\ &= \rho\left(Z + \sup_{t \in (0,1)} (\mathrm{VaR}_t(X) - \mathrm{VaR}_t(Z))\right).\end{aligned}$$

Similarly to the proof of Proposition 6.9, we have

$$\begin{aligned}\rho(X) &= \min_{Z \in \mathcal{X},\ X \succeq_1 Z} \sup_{t \in (0,1)} \rho(Z + \mathrm{VaR}_t(X) - \mathrm{VaR}_t(Z)) \\ &\leqslant \min_{Z \in \mathcal{X},\ X \succeq_1 Z} \sup_{t \in (0,1)} \{\mathrm{VaR}_t(X) - \mathrm{VaR}_t(Z) + \rho(Z)\} \leqslant \rho(X).\end{aligned}$$

It follows that

$$\rho(X) = \min_{Z \in \mathcal{X},\ X \succeq_1 Z} \sup_{t \in (0,1)} \{\mathrm{VaR}_t(X) - g_Z(t)\},$$

where $g_Z(t) = \mathrm{VaR}_t(Z) - \rho(Z)$ for all $t \in (0,1)$. Therefore, $\{g_Z \mid Z \in \mathcal{X},\ X \succeq_1 Z\}$ is a desired family of measurable functions on $(0,1)$.

"$\Leftarrow$": We first show that $\rho$ is cash subadditive. Indeed, for all $X \in \mathcal{X}$ and $m \geqslant 0$, we have $X + m \succeq_1 X$. Hence, $\mathcal{G}_X \subseteq \mathcal{G}_{X+m}$ and

$$\begin{aligned}\rho(X+m) &= \min_{g \in \mathcal{G}_{X+m}} \sup_{t \in (0,1)} \{\mathrm{VaR}_t(X) - g(t)\} + m \\ &\leqslant \min_{g \in \mathcal{G}_X} \sup_{t \in (0,1)} \{\mathrm{VaR}_t(X) - g(t)\} + m = \rho(X) + m.\end{aligned}$$

To show law invariance of $\rho$, for all $X, Y \in \mathcal{X}$ such that $X \overset{\mathrm{d}}{=} Y$, we have $X \succeq_1 Y$ and $Y \succeq_1 X$. It follows that $\mathcal{G}_X = \mathcal{G}_Y$ and thus $\rho$ is law invariant. $\square$

*Remark* 6.8. Although the functional

$$\phi_Z(X) = \inf\{\rho(Z+m) \mid m \in \mathbb{R},\ Z+m \succeq_1 X\}, \quad X \in \mathcal{X},$$

defined in the proof of Proposition 6.10 is monotone, cash subadditive and law invariant, $\phi_Z$ is not quasi-convex. This is because VaR does not satisfy quasi-convexity.

### 6.7.5 Certainty equivalents of RDEU with discount factor ambiguity

The rank-dependent expected utility (RDEU) of Quiggin (1982) is a popular behavioral decision model specified by the preference functional

$$\int_\Omega \ell(X)\,\mathrm{d}T \circ P, \quad X \in \mathcal{X},$$

where $\ell : \mathbb{R} \to \mathbb{R}$ is a strictly increasing and convex loss function (positive random variables represent losses), and $T : [0,1] \to [0,1]$ is a probability distortion function. We consider the choice of $T$ given by $T = \alpha T_1 + (1-\alpha)T_2$ where $T_1$ (resp. $T_2$) are increasing, differentiable and convex (resp. concave) probability distortion functions with $T_1(0) = T_2(0) = 0$ and $T_1(1) = T_2(1) = 1$. Following Carlier and Dana (2003), for an increasing, differentiable and convex distortion function $h : [0,1] \to [0,1]$ with $h(0) = 0$ and $h(1) = 1$, define

$$\mathrm{core}(h \circ P) = \{Q \in \mathcal{M}_f \mid Q(A) \geqslant h(P(A)) \text{ for all } A \in \mathcal{F}\}.$$

We have

$$\int_\Omega \ell(X)\,\mathrm{d}T \circ P = \alpha \min_{Q_1 \in \mathrm{core}(T_1 \circ P)} \mathbb{E}_{Q_1}[\ell(X)] + (1-\alpha) \max_{Q_2 \in \mathrm{core}(\widehat{T}_2 \circ P)} \mathbb{E}_{Q_2}[\ell(X)], \tag{6.18}$$

where $\widehat{T}_2 : x \mapsto 1 - T_2(1-x)$. Functional of the form (6.18) above belongs to the family of $\alpha$-maxmin expected utility as in Example 6.3. The certainty equivalent of the RDEU with ambiguity of a discount factor is given by

$$\begin{aligned}\rho(X) &= \sup_{\lambda \in I}\ \ell^{-1}\left(\int_\Omega \ell(\lambda X)\,\mathrm{d}T \circ P\right) \\ &= \min_{Q_1 \in \mathrm{core}(T_1 \circ P)} \max_{Q_2 \in \mathrm{core}(\widehat{T}_2 \circ P)} \sup_{\lambda \in I}\ \ell^{-1}\left(\alpha \mathbb{E}_{Q_1}[\ell(\lambda X)] + (1-\alpha)\mathbb{E}_{Q_2}[\ell(\lambda X)]\right),\end{aligned} \tag{6.19}$$

where the ambiguity set $I \subseteq [0,1]$. It is clear that if we take the loss function to be $\ell : x \mapsto e^{\gamma X}$ for $\gamma > 0$, then $\rho$ is a cash-subadditive risk measure, while $\rho$ pines down to a monetary risk measure without ambiguity of the discount factor $\lambda$.

Note that for all $\lambda \in I$, $Q_1 \in \mathrm{core}(T_1 \circ P)$ and $Q_2 \in \mathrm{core}(\widehat{T}_2 \circ P)$, the mapping

$$X \mapsto \ell^{-1}\left(\alpha \mathbb{E}_{Q_1}[\ell(\lambda X)] + (1-\alpha)\mathbb{E}_{Q_2}[\ell(\lambda X)]\right)$$

is quasi-convex and upper semi-continuous. Theorem 3.1 and Proposition 5.3 of Cerreia-Vioglio et al. (2011) showed an explicit representation of the certainty equivalent of the expected loss given by $\ell^{-1}(\mathbb{E}_P[\ell(\cdot)])$. In the proposition below, we show the representation result of a more general $\rho$ in a similar sense. Define $\bar{\ell} : [-\infty, \infty] \to [-\infty, \infty]$ as the extended-valued function with inverse function given by

$$\bar{\ell}^{-1}(x) = \begin{cases} \ell^{-1}(x), & x \in (\inf_{t\in\mathbb{R}} \ell(t), \infty), \\ -\infty, & x \in [-\infty, \inf_{t\in\mathbb{R}} \ell(t)], \\ \infty, & x = \infty. \end{cases}$$

Let $\ell^* : [-\infty, \infty] \to [-\infty, \infty]$ be the conjugate function of $\bar{\ell}$ given by

$$\ell^*(x) = \sup_{y \in [-\infty,\infty]} \{xy - \bar{\ell}(y)\}, \quad x \in [-\infty, \infty].$$

**Proposition 6.11.** *Let $\widetilde{Q} = \alpha Q_1 + (1-\alpha)Q_2$ for $Q_1 \in \mathrm{core}(T_1 \circ P)$ and $Q_2 \in \mathrm{core}(\widehat{T}_2 \circ P)$. For $X \in \mathcal{X}$, the risk measure $\rho$ in (6.19) adopts the following representation:*

$$\rho(X) = \min_{Q_1 \in \mathrm{core}(T_1 \circ P)} \max_{Q_2 \in \mathrm{core}(\widehat{T}_2 \circ P)} \sup_{\lambda \in I} \max_{Q \in \mathcal{M}_f} R(\mathbb{E}_Q[X], Q),$$

*where we have*

$$R(t, Q) = \ell^{-1}\left(\max_{x \geqslant 0}\left[\lambda x t - \mathbb{E}_{\widetilde{Q}}\left(\ell^*\left(x \frac{\mathrm{d}Q}{\mathrm{d}\widetilde{Q}}\right)\right)\right]\right) \quad \textit{for all } (t, Q) \in \mathbb{R} \times \mathcal{M}_f.$$

*Proof.* For all $X \in \mathcal{X}$, $\lambda \in I$, $Q_1 \in \mathrm{core}(T_1 \circ P)$ and $Q_2 \in \mathrm{core}(\widehat{T}_2 \circ P)$, by Theorem 3.1 of Cerreia-Vioglio et al. (2011), we have

$$\ell^{-1}\left(\alpha \mathbb{E}_{Q_1}[\ell(\lambda X)] + (1-\alpha)\mathbb{E}_{Q_2}[\ell(\lambda X)]\right) = \ell^{-1}\left(\mathbb{E}_{\widetilde{Q}}[\ell(\lambda X)]\right) = \max_{Q \in \mathcal{M}_f} R(\mathbb{E}_Q[X], Q),$$

where for all $(t, Q) \in \mathbb{R} \times \mathcal{M}_f$,

$$R(t, Q) = \inf\left\{\ell^{-1}\left(\mathbb{E}_{\widetilde{Q}}[\ell(\lambda X)]\right) \;\middle|\; \mathbb{E}_Q[X] = t\right\}.$$

It is clear by definition that $P$ is absolutely continuous with respect to $Q_1$ and $Q_2$ and thus $P$ is absolutely continuous with respect to $\widetilde{Q}$. It follows that all $Q \in \mathcal{M}_f$ are also absolutely continuous with respect to $\widetilde{Q}$. By Proposition 5.3 of Cerreia-Vioglio et al. (2011), we have

$$R(t, Q) = \ell^{-1}\left(\max_{x \geqslant 0}\left[\lambda x t - \mathbb{E}_{\widetilde{Q}}\left(\ell^*\left(x \frac{\mathrm{d}Q}{\mathrm{d}\widetilde{Q}}\right)\right)\right]\right). \qquad \square$$

### 6.7.6 An additional result on the inf-convolution in Section 6.5.3

In the proposition below, we consider general cash-subadditive risk measures and give a mathematical representation result of the inf-convolution $\square_{i=1}^n \rho_i$, whose penalty is given by the sum of the penalty functions of individual risk measures $\rho_1, \dots, \rho_n$.

**Proposition 6.12.** *Let $\rho_1, \dots, \rho_n : \mathcal{X} \to \mathbb{R}$ be cash-subadditive risk measures. Suppose that an optimal allocation of (6.16) exists. Define*

$$\mathcal{B}_X^i = \{\beta : \mathcal{M}_f \to (-\infty, \infty],\ Q \mapsto \mathbb{E}_Q[Z] - \rho_i(Z) \mid Z \in \mathcal{X},\ Z \leqslant X\}, \quad \textit{for all } X \in \mathcal{X}.$$

*We have*

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \min_{\beta \in \Sigma_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta(Q)\}, \quad \textit{for all } X \in \mathcal{X},$$

*where*

$$\Sigma_X = \bigcup_{(X_1, \dots, X_n) \in \mathbb{A}_n(X)} \mathcal{B}_{X_1}^1 + \dots + \mathcal{B}_{X_n}^n.$$[7]

*Proof.* We only prove the case $n = 2$ and the cases $n \geqslant 3$ are similar. Suppose that $\rho_1$ and $\rho_2$ are cash-subadditive risk measures. It is clear that $\square_{i=1}^2 \rho_i$ as shown in (6.16) is a cash-subadditive risk measure. By Proposition 6.9, we have

$$\mathop{\square}_{i=1}^{2} \rho_i(X) = \min_{\beta \in \mathcal{B}_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta(Q)\}, \quad \text{for all } X \in \mathcal{X},$$

where

$$\mathcal{B}_X = \{\beta : \mathcal{M}_f \to (-\infty, \infty],\ Q \mapsto \mathbb{E}_Q[Z] - \mathop{\square}_{i=1}^{2} \rho_i(Z) \mid Z \in \mathcal{X},\ Z \leqslant X\}, \quad \text{for all } X \in \mathcal{X}.$$

For all $Z, X \in \mathcal{X}$ with $Z \leqslant X$, $(Z_1, Z_2) \in \mathbb{A}_2(Z)$, $(X_1, X_2) \in \mathbb{A}_2(X)$ with $Z_1 \leqslant X_1$, $Z_2 \leqslant X_2$, and $Q \in \mathcal{M}_f$, we have $\square_{i=1}^2 \rho_i(Z) \leqslant \rho_1(Z_1) + \rho_2(Z_2)$ by definition. It follows that

$$\begin{aligned}
\mathop{\square}_{i=1}^{2} \rho_i(X) &= \min_{Z \in \mathcal{X},\ Z \leqslant X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \mathbb{E}_Q[Z] + \mathop{\square}_{i=1}^{2} \rho_i(Z)\} \\
&\leqslant \min_{(Z_1, Z_2) \in \mathcal{Z}_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \mathbb{E}_Q[Z] + \rho_1(Z_1) + \rho_2(Z_2)\} = \min_{\beta \in \Sigma_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta(Q)\},
\end{aligned}$$

[7]The sum of two sets $\mathcal{B}_1$ and $\mathcal{B}_2$ is defined as $\mathcal{B}_1 + \mathcal{B}_2 = \{\beta_1 + \beta_2 \mid \beta_1 \in \mathcal{B}_1, \beta_2 \in \mathcal{B}_2\}$.

where

$$\mathcal{Z}_X = \{(Z_1, Z_2) \in \mathbb{A}_2(Z) \mid Z_1 \leqslant X_1,\ Z_2 \leqslant X_2,\ (X_1, X_2) \in \mathbb{A}_2(X)\}, \quad X \in \mathcal{X}.$$

On the other hand, for all $X \in \mathcal{X}$ and $\beta \in \mathcal{B}_X$, there exists $Z \in \mathcal{X}$ with $Z \leqslant X$, such that $\beta(Q) = \mathbb{E}_Q[Z] - \square_{i=1}^2 \rho_i(Z)$ for all $Q \in \mathcal{M}_f$. Since an optimal allocation with respect to $\square_{i=1}^2 \rho_i(Z)$ exists, written as $(Z_1^*, Z_2^*) \in \mathbb{A}_2(Z)$, we have $\square_{i=1}^2 \rho_i(Z) = \rho_1(Z_1^*) + \rho_2(Z_2^*)$. Hence,

$$\beta(Q) = \mathbb{E}_Q[Z] - \rho_1(Z_1^*) + \rho_2(Z_2^*).$$

It is also clear that there exists $(X_1, X_2) \in \mathbb{A}_2(X)$ such that $Z_1^* \leqslant X_1$ and $Z_2^* \leqslant X_2$. It follows that $\beta \in \Sigma_X$ and thus $\mathcal{B}_X \subseteq \Sigma_X$. This implies that

$$\mathop{\square}_{i=1}^{2} \rho_i(X) \geqslant \min_{\beta \in \Sigma_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta(Q)\}, \quad \text{for all } X \in \mathcal{X}.$$

Therefore, we have

$$\mathop{\square}_{i=1}^{2} \rho_i(X) = \min_{\beta \in \Sigma_X} \max_{Q \in \mathcal{M}_f} \{\mathbb{E}_Q[X] - \beta(Q)\}, \quad \text{for all } X \in \mathcal{X}. \qquad \square$$

# Chapter 7

# E-backtesting

## 7.1 Introduction

Forecasting risk measures is important for financial institutions to calculate capital reserves for risk management purposes. Regulators are responsible to monitor whether risk forecasts are correctly reported by conducting hypothesis tests known as backtests (see e.g., Christoffersen, 2011; McNeil et al., 2015, for general treatments). Regulatory backtests have several features distinct from traditional testing problems; see Acerbi and Szekely (2014) and Nolde and Ziegel (2017). First, risk forecasts and realized losses arrive sequentially over time. Second, due to frequently changing portfolio positions and the complicated temporal nature of financial data, the losses and risk predictions are neither independent nor identically distributed, and they do not follow any standard time-series models. Third, the tester (e.g., a regulator) is concerned about risk measure underestimation, which means high insolvency risk, whereas overestimation (i.e., being conservative) is secondary or acceptable. Fourth, the tester does not necessarily accurately know the underlying model used by a financial institution to produce risk predictions.

In financial practice, a well-adopted simple approach exists for backtesting the Value-at-Risk (VaR), which is the so-called three-zone approach based on binomial tests described in BCBS (2013); this approach is *model-free* in the sense that one directly tests the risk forecast without testing any specific family of models. The Basel Committee on Banking Supervision BCBS (2016) has replaced VaR by the Expected Shortfall (ES) as the standard regulatory measure for market risk, mostly due to the convenient properties of ES, in particular, being able to capture tail risk.[1] However,

[1]Quoting BCBS (2016, p.1): *Use of ES will help to ensure a more prudent capture of "tail risk" and capital*

| Literature | Parametric or dependence assumptions | Forecast structural assumptions | Fixed sample size | Asymptotic test | Reliance on VaR or distributional forecasts |
|---|---|---|---|---|---|
| MF00 | yes | yes | yes | yes | yes |
| AS14 | yes | yes | yes | yes | yes |
| DE17 | yes | yes | yes | yes | yes |
| NZ17 | yes | yes | yes | yes | yes |
| BD22 | yes | yes | yes | yes | no |
| HD22 | yes | yes | no | no | yes |
| This chapter | no | no | no | no | yes |

Table 7.1: Comparison of existing backtesting methods for ES

*Notes:* We use shortcuts MF00 for McNeil and Frey (2000), AS14 for Acerbi and Szekely (2014), DE17 for Du and Escanciano (2017), NZ17 for Nolde and Ziegel (2017), BD22 for Bayer and Dimitriadis (2022), and HD22 for Hoga and Demetrescu (2022). Parametric or dependence assumptions refer to those on loss distributions, time series models, stationarity, or strong mixing. Forecast structural assumptions refer to requirements on the forms and properties of risk forecasts. Acerbi and Szekely (2014) proposed three methods of backtesting ES; The first two methods do not require specific forms of ES forecasts, but the third method requires ES to be estimated as realized ranks.

as discussed by Gneiting (2011), ES is not elicitable, and backtesting ES is substantially more challenging than VaR. Table 7.1 summarizes the main features of existing methods backtesting ES. To the best of our knowledge, there is no model-free non-asymptotic backtesting method for ES. Moreover, except for Hoga and Demetrescu (2022), most of the backtesting methods in the existing literature only work for a fixed data size, and are thus not valid under option stopping, or equivalently, not anytime valid (see e.g., Chu et al., 1996, for discussions). This creates limitations to financial regulation practice where early rejections are highly desirable.

In this chapter, we develop a model-free backtesting method for risk measures, including ES, using the concepts of e-values and e-tests (Shafer, 2021; Vovk and Wang, 2021; Grünwald et al., 2020). E-tests have important advantages over classical statistical tests (p-tests) based on p-values.

*adequacy during periods of significant financial market stress.* See also Wang and Zitikis (2021) for an axiomatic justification of ES in financial regulation.

Wang and Ramdas (2022, Section 2) collects many reasons for using e-values and e-tests, regarding high-dimensional asymptotics, composite models, sequential (any-time valid) inference, information accumulation, and robustness to model misspecification and dependence; other advantages of e-values are illustrated by Grünwald et al. (2020), Vovk and Wang (2021) and Vovk et al. (2022). As a particularly relevant feature to our context, our proposed e-tests allow regulators to get alerted early as the e-process accumulates to a reasonably large value. This is different from scientific discoveries (such as genome studies) where a scientist may not be entitled to reject a hypothesis based on merely "substantial" evidence. Noticing this, a multi-zone approach similar to the three-zone approach can be developed by setting different e-value thresholds in financial regulation.

The main contribution of this chapter is four-fold: First, we introduce the new notion of model-free e-statistics and propose e-backtesting methods. In particular, we obtain model-free e-statistics for ES in Section 7.2 (Theorem 7.1), allowing us to construct e-processes to backtest ES, as well as other risk measures in Section 7.3 (Theorem 7.2). The backtesting method for VaR and ES is discussed in Section 7.4. Second, with model-free e-statistics chosen, the next important step to construct an e-process by choosing a suitable betting process, which we address in Section 7.5. We propose three new methods to calculate the betting processes based on data. It turns out that these methods are asymptotically optimal (equivalent to an oracle betting process) under different situations (Theorem 7.3). Third, we characterize model-free e-statistics for the mean, the variance, VaR (Theorem 7.4), and ES (Theorem 7.5) in Section 7.6 by establishing a link between model-free e-statistics and identification functions. All model-free e-statistics for these functionals take similar forms as mixtures between 1 and a simple model-free e-statistic. Finally, through the simulation study and data analysis in Sections 7.7 and 7.8, we demonstrate detailed procedures of backtesting VaR and ES using e-values for practical operations of financial regulations. In addition to our main content, Section 7.10 shows the betting processes calculated via Taylor approximation for VaR and ES; Section 7.11 discusses the link between model-free e-statistics and identification functions in preparation for the characterization results in Section 7.6; except for Theorems 7.1 and Theorem 7.3, proofs of all results are relegated to Section 7.12; Section 7.13 contains some necessary details of our simulation and data analysis. To support our new methodology, extended simulation studies and data analyses are presented in the separate paper Wang et al. (2022) for the interested reader.

### 7.1.1 Related literature

Besides financial regulation, evaluating forecasting models and methods for major economic variables is also essential in the decision-making processes of government institutions and regulatory authorities. Earlier work on predictive ability tests and forecast selection includes Diebold and Mariano (1995), whose method was extended by West (1996), Clark and McCracken (2001), and Giacomini and White (2006). Unconditional backtests of VaR were considered by Kupiec (1995) on testing Bernoulli distributions, which were extended by Christoffersen (1998) to include testing independence of the VaR-violations. Engle and Manganelli (2004) tested conditional autoregressive VaR; Berkowitz et al. (2020) unified existing evaluation methods of VaR; and Ziggel et al. (2014) proposed a Monte Carlo simulation-based backtesting method for VaR.

Due to its increasing importance and challenging nature, there are ample studies in the more recent literature on backtesting ES with different approaches and limitations. McNeil and Frey (2000) proposed bootstrap tests with iid innovations; Acerbi and Szekely (2014, 2017) studied three backtesting methods under independent losses; Du and Escanciano (2017) designed parametric test using cumulative violations; Nolde and Ziegel (2017) studied comparative backtests among forecasting methods; Bayer and Dimitriadis (2022) built backtesting through a linear regression model; and Hoga and Demetrescu (2022) proposed sequential monitoring based on parametric distributions. Their main features are summarized in Table 7.1.

The literature on e-values has also been growing fast recently. E-values were used in the early literature in different disguises, although the term "e-value" was proposed by Vovk and Wang (2021). For instance, e-values and e-tests were essentially used in the work of Wald (1945) and Darling and Robbins (1967), and they are central to the ideas of testing by betting and martingales (Shafer et al., 2011; Shafer and Vovk, 2019) and universal inference (Wasserman et al., 2020). E-values are shown to be useful in multiple hypothesis testing with dependence (Vovk and Wang, 2021), parametric tests with optional sampling (Grünwald et al., 2020), false discovery rate control (Wang and Ramdas, 2022), and many other statistical applications. We refer to the recent survey paper of Ramdas et al. (2022) for recent progresses on e-values.

## 7.2 E-values and model-free e-statistics

### 7.2.1 Definition and examples

Let $\mathfrak{P}$ be the set of probability measures on $(\Omega, \mathcal{F})$. A (composite) hypothesis $H$ is a subset of $\mathfrak{P}$. A hypothesis $H$ is simple if it is a singleton. Following the terminology of Vovk and Wang (2021), an *e-variable* for $H \subseteq \mathfrak{P}$ is an extended random variable $E : \Omega \to [0, \infty]$ such that $\mathbb{E}^P[E] \leqslant 1$ for each $P \in H$. We denote by $\mathcal{E}_H$ the set of e-variables for a hypothesis $H$ and by $\mathcal{E}_P$ the set of e-variables for the simple hypothesis $\{P\}$. An *e-test* rejects the hypothesis $H$ if a realized e-variable, called an *e-value*, is larger than a threshold. A common rule of thumb is that an e-value of 10 represents strong evidence.[2] A non-negative stochastic process $(E_t)_{t \in K}$, $K \subseteq \mathbb{N}$, adapted to a given filtration, is an *e-process* for $H$ if $\mathbb{E}^P[E_\tau] \leqslant 1$ for all stopping times $\tau$ taking values in $K$ and each $P \in H$.

Let $d$ be a positive integer. The model space $\mathcal{M}$ is a set of distributions on $\mathbb{R}$. The value of the functional $\psi = (\rho, \phi_1, \ldots, \phi_{d-1}) : \mathcal{M} \to \mathbb{R}^d$ represents the collection of available statistical information, where $\rho$ is the risk prediction to be tested, and $\phi$ contains auxiliary information. If $d = 1$, then the only available information is the predicted value of $\rho$. Let $\mathcal{M}_q$, $\mathcal{M}_\infty$, and $\mathcal{M}_0$ represent the set of distributions on $\mathbb{R}$ with finite $q$-th moment for $q \in (0, \infty)$, that of compactly supported distributions on $\mathbb{R}$, and that of all distributions on $\mathbb{R}$, respectively. Define $\psi$ for a random variable $X \in L^0$ via $\psi(X) = \psi(F)$, where $F$ is the distribution of $X$. For level $p \in (0,1)$, denote the the lower $p$-quantile by $Q_p(F)$ for $F \in \mathcal{M}_0$. VaR and ES belong to the class of dual utilities in Yaari (1987) and Schmeidler (1989). For $\psi = (\rho, \phi_1, \ldots, \phi_{d-1})$, the natural domains of $\rho, \phi_1, \ldots, \phi_{d-1}$ are not necessarily identical, and the domain of $\psi$ is set to be their intersection. In the following Definitions 7.1 and 7.2, we introduce the key tool we use for our e-tests.

**Definition 7.1** (Model-free e-statistics)**.** Let $\mathcal{P} \subseteq \mathcal{M}$. An *$\mathcal{P}$-model-free e-statistic for $\psi : \mathcal{M} \to \mathbb{R}^d$* is a measurable function $e : \mathbb{R}^{d+1} \to [0, \infty]$ satisfying $\int_{\mathbb{R}} e(x, \psi(F))\, \mathrm{d}F(x) \leqslant 1$ for each $F \in \mathcal{P}$. If $\mathcal{P} = \mathcal{M}$, then we simply call $e$ a model-free e-statistic for $\psi$.

To consider $\mathcal{P}$-model-free e-statistics for $\psi$ in Definition 7.1, it suffices to consider the restriction of $\psi$ on $\mathcal{P}$. Using the language of e-variables, a model-free e-statistic for $\psi$ is a function $e$ such

[2] Thresholds may be chosen according to the rule of thumb of Jeffreys (1961) in the disguise of likelihood ratios: 3.2 (substantial), which roughly corresponds to a p-value of 0.05; 10 (strong), which roughly corresponds to a p-value of 0.01; 100 (decisive), which roughly corresponds to a p-value of 0.0001. See Vovk and Wang (2021) for details and comparisons of these recommendations.

that $e(X, \psi(F)) \in \mathcal{E}_{\mathbb{P}}$ for each $F \in \mathcal{P}$, where $X$ has distribution $F$ under $\mathbb{P}$. The term "model-free" reflects that the e-statistic is valid for all $F \in \mathcal{P}$.

**Definition 7.2** (Model-free e-statistics testing $\rho$)**.** Let $\mathcal{P} \subseteq \mathcal{M}$. For $\psi = (\rho, \phi) : \mathcal{M} \to \mathbb{R} \times \mathbb{R}^{d-1}$, a $\mathcal{P}$-model-free e-statistic $e : \mathbb{R}^{d+1} \to [0, \infty]$ for $\psi = (\rho, \phi)$ is *testing* $\rho$ if $\int_{\mathbb{R}} e(x, r, z)\, \mathrm{d}F(x) > 1$ for all $(r, z) \in \psi(\mathcal{P})$ and $F \in \mathcal{P}$ with $\rho(F) > r$. The e-statistic is *strictly* testing $\rho$ if $(x, r, z) \mapsto e(x, r, z)$ is decreasing in $r$. If $d = 1$, $e$ is simply called a $\mathcal{P}$-model-free e-statistic testing $\rho$.

Intuitively, Definition 7.1 addresses the validity of the e-test: if a true value of the risk prediction for $\psi = (\rho, \phi)$ is provided, then the e-statistic will have a mean that is no larger than 1. On the other hand, Definition 7.2 addresses the consistency of the e-test: if the risk $\rho$ is underestimated, then the e-statistic will have a mean that is larger than 1, regardless of whether the prediction of the auxiliary functional $\phi$ is truthful. The special case where $d = 1$, i.e., no auxiliary functional $\phi$ is involved, will be discussed in detail in Section 7.6.1.

We are only interested in forecast values $(r, z)$ in the set $\psi(\mathcal{P})$. Any forecast values outside $\psi(\mathcal{P})$ can be rejected directly.

Our idea of model-free e-statistics specifically addresses the underestimation of $\rho$, which is consistent with the motivation in regulatory backtesting. If an e-statistic is strictly testing a risk measure $\rho$, then an overestimation of the risk is rewarded: An institution being scrutinized by the regulator can deliberately report a higher risk value (which typically means higher capital reserve) to pass to the regulatory test, thus rewarding prudence.

First, we give a few examples of model-free e-statistics for some common risk measures. Throughout, we use the convention that $0/0 = 1$ and $1/0 = \infty$, and let $\mathbb{R}_+ = [0, \infty)$.

**Example 7.1** (Model-free e-statistic testing the mean)**.** Let $\mathcal{P}$ be the set of distributions on $\mathbb{R}_+$ in $\mathcal{M}_1$. Define the function $e(x, r) = x/r$ for $x, r \geqslant 0$. In this case, we have $\mathbb{E}[e(X, \mathbb{E}[X])] = 1$ for all random variables $X$ with distribution in $\mathcal{P}$. Moreover, for any such $X$, $\mathbb{E}[X] > r \geqslant 0$ implies $\mathbb{E}[e(X, r)] > 1$. Therefore, $e$ is a $\mathcal{P}$-model-free e-statistic strictly testing the mean.

**Example 7.2** (Model-free e-statistic for $(\mathrm{var}, \mathbb{E})$ testing the variance)**.** Consider $(\mathrm{var}, \mathbb{E}) : \mathcal{M}_2 \to \mathbb{R}^2$. For all random variables $X$ with distribution in $\mathcal{M}_2$, we have

$$\mathbb{E}[e(X, \mathrm{var}(X), \mathbb{E}[X])] = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\mathrm{var}(X)} = 1.$$

Moreover, since $z = \mathbb{E}[X]$ minimizes $\mathbb{E}[(X-z)^2]$ over $z \in \mathbb{R}$ and $\mathrm{var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2]$, $\mathrm{var}(X) > r \geqslant 0$ implies

$$\mathbb{E}[e(X,r,z)] = \frac{\mathbb{E}[(X-z)^2]}{r} \geqslant \frac{\mathrm{var}(X)}{r} > 1.$$

Hence, $e(x,r,z) = (x-z)^2/r$ for $x, z \in \mathbb{R}$ and $r \geqslant 0$ is a model-free e-statistic for $(\mathrm{var}, \mathbb{E})$ strictly testing the variance.

**Example 7.3** (Model-free e-statistic testing a quantile)**.** Take $p \in (0,1)$. Define the function

$$e_p^Q(x,r) = \frac{1}{1-p}\mathbb{1}_{\{x>r\}}, \quad x, r \in \mathbb{R}. \tag{7.1}$$

We have $\mathbb{E}[e(X, Q_p(F))] \leqslant 1$ for any random variables $X$ with distribution $F$. Moreover, $Q_p(F) > r$ implies $\mathbb{P}(X > r) > 1-p$, and hence $\mathbb{E}[e(X,r)] > 1$. Therefore, $e$ is a model-free e-statistic strictly testing the $p$-quantile.

**Example 7.4** (Model-free e-statistic testing an expected loss)**.** For some $a \in \mathbb{R}$, let $\ell : \mathbb{R} \to [a, \infty)$ be a function that is interpreted as a loss. Define the function $e(x,r) = (\ell(x)-a)/(r-a)$ for $x \in \mathbb{R}$ and $r \geqslant a$. Analogously to Example 7.1, $e$ is a model-free e-statistic strictly testing the expected loss $F \mapsto \int \ell \,\mathrm{d}F$ on its natural domain.

The choice of a model-free e-statistic $e$ for $\psi$ is not necessarily unique. For instance, a linear combination of $e$ with 1 with the weight between 0 and 1 is also a model-free e-statistic for $\psi$. Depending on the specific situation, either e-statistic may be useful in practice.

*Remark* 7.1. The functional $\psi = (\rho, \phi) = (\mathrm{var}, \mathbb{E})$ in Example 7.2 is an example of a Bayes pair, that is, a there exists a measurable function $L : \mathbb{R}^{d+1} \to \mathbb{R}$, called the *loss function*, such that

$$\phi(F) \in \arg\min_{z\in\mathbb{R}^d} \int L(z,x)\,\mathrm{d}F(x) \text{ and } \rho(F) = \min_{z\in\mathbb{R}^d} \int L(z,x)\,\mathrm{d}F(x), \quad F \in \mathcal{M}, \tag{7.2}$$

where $\int L(z,x)\,\mathrm{d}F(x)$ is assumed to be well-defined for each $z \in \mathbb{R}^d$, $F \in \mathcal{M}$ (see e.g., Fissler and Ziegel, 2016; Frongillo and Kash, 2021, and Chapter 4). Bayes pairs often admit model-free e-statistics. A typical example commonly used in risk management practice is $(\mathrm{ES}, \mathrm{VaR})$,[3] which we mainly focus on in this chapter.

Next, we see that, for $p \in (0,1)$ the function

$$e_p^{\mathrm{ES}}(x,r,z) = \frac{(x-z)_+}{(1-p)(r-z)}, \quad x \in \mathbb{R},\ z \leqslant r. \tag{7.3}$$

[3]We omit the probability level $p$ in $\mathrm{VaR}_p$, $\mathrm{ES}_p$ and $(\mathrm{VaR}_p, \mathrm{ES}_p)$ in the text (but never in equations) where we do not emphasis the probability level.

defines a model-free e-statistic for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ testing $\mathrm{ES}_p$. Recall the convention that $0/0 = 1$ and $1/0 = \infty$, and set $e_p^{\mathrm{ES}}(x, r, z) = \infty$ if $r < z$, which is a case of no relevance since $\mathrm{ES}_p(F) \geqslant \mathrm{VaR}_p(F)$ for any $F \in \mathcal{M}_1$.

**Theorem 7.1.** *The function $e_p^{\mathrm{ES}}$ is a model-free e-statistic for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ strictly testing $\mathrm{ES}_p$.*

*Proof.* By the VaR-ES relation of Rockafellar and Uryasev (2002), for any random variable $X$ with finite mean,

$$\mathrm{VaR}_p(X) \in \operatorname*{arg\,min}_{z \in \mathbb{R}} \left\{ z + \frac{1}{1-p} \mathbb{E}[(X-z)_+] \right\}, \tag{7.4}$$

and

$$\mathrm{ES}_p(X) = \min_{z \in \mathbb{R}} \left\{ z + \frac{1}{1-p} \mathbb{E}[(X-z)_+] \right\}. \tag{7.5}$$

This indicates that $(\mathrm{ES}_p, \mathrm{VaR}_p)$ is a Bayes pair by (4.2) with loss function $L : (z, x) \mapsto z + (x - z)_+/(1-p)$. Since $L(z, x) \geqslant z$, for $r \geqslant z$, we have that $e_p^{\mathrm{ES}}(x, r, z) = (L(z, x) - z)/(r - z) \geqslant 0$, and it is decreasing in $r$. Furthermore, for $z < r \leqslant \mathrm{ES}_p(X)$,

$$\mathbb{E}\left[\frac{L(z, X) - z}{r - z}\right] \geqslant \frac{\mathrm{ES}_p(X) - z}{r - z} \geqslant 1$$

with equality if and only if $r = \mathrm{ES}_p(X)$. □

While Examples 7.1-7.4 and Theorem 7.1 show that interesting model-free e-statistics exist, much more can be said about their general structure; see Section 7.6.

## 7.3 E-backtesting risk measures with model-free e-statistics

We next present a general methodology for backtesting risk measures via model-free e-statistics in a sequential setting.

Let $T$ be any time horizon, which may be fixed, infinite, or adaptive, i.e., depending on the data observed. The flexibility of infinite or data-dependent time horizons is a feature of e-tests, which allows us to address situations more general than the ones considered in the literature, e.g., Hoga and Demetrescu (2022), where $T$ is a pre-specified fixed time horizon and the tester has to start over when time $T$ is reached. For any positive integer $n$, denote by $[n] = \{1, \ldots, n\}$, and for $n = \infty$ it is $[n] = \mathbb{N}$, the set of positive integers.

Let the $\sigma$-algebra $\mathcal{F}_t$ represent all the available information up to time $t \in [T]$, such that $\mathcal{F}_m \subseteq \mathcal{F}_n$ for all $m \leqslant n$. Let $(L_t)_{t \in [T]}$ be a sequence of realized losses that are adapted to the filtration $(\mathcal{F}_t)_{t\in[T]}$. Denote by $\rho(L_t|\mathcal{F}_{t-1})$ and $\phi(L_t|\mathcal{F}_{t-1})$, respectively, the values of $\rho$ and $\phi$ applied to the conditional distribution of $L_t$ given $\mathcal{F}_{t-1}$. Let $r_t$ and $z_t$ be forecasts for $\rho(L_t|\mathcal{F}_{t-1})$ and $\phi(L_t|\mathcal{F}_{t-1})$ made at time $t-1$, respectively. Note that $\rho(L_t|\mathcal{F}_{t-1})$ and $\phi(L_t|\mathcal{F}_{t-1})$ are themselves random variables and $\mathcal{F}_{t-1}$-measurable for all relevant functionals of interest (see e.g., Fissler and Holzmann, 2022).

We assume that the risk forecasts $r_t$ and $z_t$ are obtained based on past market information and all other possible factors that may affect the decisions of risk predictors in financial institutions. For instance, the information may even include throwing a die or random events such as coffee spilling; all these events up to time $t-1$ are included in $\mathcal{F}_{t-1}$.

We test the following null hypothesis:

$$H_0: \quad r_t \geqslant \rho(L_t|\mathcal{F}_{t-1}) \text{ and } z_t = \phi(L_t|\mathcal{F}_{t-1}) \quad \text{for } t \in [T]. \tag{7.6}$$

Rejecting (7.6) implies, in particular, rejecting $r_t = \rho(L_t|\mathcal{F}_{t-1})$ and $z_t = \phi(L_t|\mathcal{F}_{t-1})$ for all $t \in [T]$. In the special case that $d = 1$ (i.e. we do not need auxiliary information from $\phi$), (7.6) becomes

$$H_0: \quad r_t \geqslant \rho(L_t|\mathcal{F}_{t-1}) \quad \text{for } t \in [T].$$

*Remark* 7.2. Since $\rho$ is the regulatory risk measure of interest, over-predicting $\rho$ is conservative. On the other hand, $\phi$ represents some additional statistical information and it may not relate to measuring financial risk. Hence, over-predicting $\phi$ is not necessarily conservative. See Example 7.5 below for a sanity check. Therefore, it is more natural to test an equality of the auxiliary information $z_t$ in (7.6) instead of an inequality; note also that this hypothesis is still more lenient than testing a specified loss distribution. For the case where a financial institution is conservative for both the risk measures $\rho$ and $\phi$, see Section 7.4.

For a nonnegative function $e: \mathbb{R}^{d+1} \to [0, \infty]$, let $X_t = e(L_t, r_t, z_t)$ for each t. We construct the following stochastic process: $M_0 = 1$ and

$$M_t(\boldsymbol{\lambda}) = (1 - \lambda_t + \lambda_t X_t) M_{t-1}(\boldsymbol{\lambda}) = \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s X_s), \quad t \in [T], \tag{7.7}$$

where the process $\boldsymbol{\lambda} = (\lambda_t)_{t\in[T]}$ is chosen such that $\lambda_t$ is a function of $(L_{s-1}, r_s, z_s)_{s\in[t]}$ and takes values in $[0,1]$ for $t \in [T]$.[4] Suppose that $e$ is a model-free e-statistic for $(\rho, \phi): \mathcal{M} \to \mathbb{R} \times \mathbb{R}^{d-1}$.

[4]More generally, we may allow $\lambda_t$ to be $\mathcal{F}_{t-1}$-measurable instead of $\sigma((L_{s-1}, r_s, z_s)_{s\in[t]})$-measurable, but this adds no further methodological value.

We have by definition that $X_t$ is an e-variable conditional on $\mathcal{F}_{t-1}$ under $H_0$ for all $t \in [T]$. As suggested by Vovk and Wang (2020), the only admissible (or unwasteful) way to combine these e-variables is through the martingale function (7.7). The e-process in (7.7) may be interpreted as the payoffs of a betting strategy against the null hypothesis $H_0$ as in Shafer and Vovk (2019). In this betting game, the initial capital is $M_0 = 1$ and all the money is invested at each step. The payoff per capital at each step is $1 - \lambda_t + \lambda_t X_t$ for $t \in [T]$. As a result, the player earns money at step $t$ if $X_t > 1$. In this sense, we call the process $\boldsymbol{\lambda}$ in (7.7) a *betting process*. The following theorem follows from Ville's well-known inequality (Ville, 1939), indicating that $(M_t(\boldsymbol{\lambda}))_{t\in\{0,\dots,T\}}$ in (7.7) is a non-negative supermartingale, so in particular an e-process under the null hypothesis in (7.6).

**Theorem 7.2.** *Suppose that $e$ is a model-free e-statistic for $(\rho, \phi) : \mathcal{M} \to \mathbb{R} \times \mathbb{R}^{d-1}$. Under $H_0$ in (7.6), $(M_t(\boldsymbol{\lambda}))_{t\in\{0,\dots,T\}}$ in (7.7) is a non-negative supermartingale with $M_0 = 1$, and for each $\alpha \in (0, 1)$,*

$$\mathbb{P}\left(\sup_{t\in\{0,\dots,T\}} M_s(\boldsymbol{\lambda}) \geqslant \frac{1}{\alpha}\right) \leqslant \alpha.$$

Based on Theorem 7.2, we will use the e-test that arises by using the e-variable $M_\tau(\boldsymbol{\lambda})$ where $M(\boldsymbol{\lambda})$ is given by (7.7) and $\tau$ is the stopping time $\min\{T, \inf\{t \geqslant 0 : M_t(\boldsymbol{\lambda}) \geqslant 1/\alpha\}\}$. This is common practice in testing with e-values.

*Remark* 7.3. Our framework of backtesting risk measures can be applied to a simpler hypothesis testing problem in a static setting. Suppose that $r$ and $z$ are fixed forecasts of risk measures $\rho : \mathcal{M} \to \mathbb{R}$ and $\phi : \mathcal{M} \to \mathbb{R}^d$, respectively, for some random variable $L$. Consider the following testing problem:

$$H_0 : r \geqslant \rho(L) \;\text{ and }\; z = \phi(L) \;\text{ or }\; \widetilde{H}_0 : (r, z) = (\rho(L), \phi(L)). \tag{7.8}$$

We observe iid samples $L_1, \dots, L_n$ from $L$ and assume that the observations arrive sequentially. Suppose that there exists a model-free e-statistic $e : \mathbb{R}^{d+1} \to [0, \infty]$ for $(\rho, \phi)$ testing $\rho$. We obtain e-values under the null hypotheses $H_0$ and $\widetilde{H}_0$ in (7.8) given by $e(L_i, r, z)$, $i \in [n]$. A simulation study on this setting is provided in the separate paper Wang et al. (2022).

## 7.4 E-backtesting Value-at-Risk and Expected Shortfall

To put our general ideas in the context of financial regulation, we focus on backtesting VaR and ES in this section. Let $L_t$ be the random loss at time $t$. For the case of backtesting VaR,

$(r_t)_{t\in[T]}$ are the forecasts for $\mathrm{VaR}_p(L_t|\mathcal{F}_{t-1})$, $p \in (0,1)$. As we see in Example 7.3, the function $e_p^Q(L_t, r_t)$ defined in (7.1) is an e-variable under the following null hypothesis that we are testing,

$$H_0: \quad r_t \geqslant \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}), \qquad t \in [T].$$

For the case of backtesting ES, $(r_t)_{t\in[T]}$ are the forecasts for $\mathrm{ES}_p(L_t|\mathcal{F}_{t-1})$ and $(z_t)_{t\in[T]}$ are the forecasts for $\mathrm{VaR}_p(L_t|\mathcal{F}_{t-1})$, $p \in (0,1)$. By Theorem 7.1, $e_p^{\mathrm{ES}}(L_t, r_t, z_t)$ is an e-variable under the following null hypothesis:

$$H_0: \quad r_t \geqslant \mathrm{ES}_p(L_t|\mathcal{F}_{t-1}) \text{ and } z_t = \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}), \qquad t \in [T]. \tag{7.9}$$

In practice, a financial institution may use a conservative model for risk management purposes, which leads to underestimation of both VaR and ES. In the following proposition, we illustrate that $e_p^{\mathrm{ES}}(L_t, r_t, z_t)$ is a valid e-variable in case both $\mathrm{VaR}_p(L_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_p(L_t|\mathcal{F}_{t-1})$ are over-predicted, together with their difference.

**Proposition 7.1.** *For $p \in (0,1)$, $(e_p^{\mathrm{ES}}(L_t, r_t, z_t))_{t\in[T]}$ are e-variables for*

$$H_0: \; z_t \geqslant \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}) \quad \textit{and} \quad r_t - z_t \geqslant \mathrm{ES}_p(L_t|\mathcal{F}_{t-1}) - \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}), \; t \in [T]. \tag{7.10}$$

The hypothesis $H_0$ in (7.10) is stronger than

$$H_0: z_t \geqslant \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}) \;\text{ and }\; r_t \geqslant \mathrm{ES}_p(L_t|\mathcal{F}_{t-1}). \tag{7.11}$$

In contrast to (7.10), we note that $e_p^{\mathrm{ES}}(L_t, r_t, z_t)$ is not necessarily an e-variable for (7.11). For instance, $\mathbb{E}[e_p^{\mathrm{ES}}(L_t, r_t, z_t)] = \infty$ if $r_t = z_t = \mathrm{ES}_p(L_t|\mathcal{F}_{t-1})$ and $\mathbb{P}(L_t > z_t) > 0$. It implies that over-predicting $\mathrm{VaR}_p(L_t|\mathcal{F}_{t-1})$ does not always leads to a smaller e-value realized by $e_p^{\mathrm{ES}}(L_t, r_t, z_t)$; in contrast, over-predicting $\mathrm{ES}_p(L_t|\mathcal{F}_{t-1})$ always reduces the resulting e-value. The following example illustrates that a poor VaR forecast could result in large e-values although it is obtained by over-prediction.

**Example 7.5.** For $p \in (0,1)$, a continuously distributed random variable $X$ with $a = \mathrm{VaR}_p(X) < \mathrm{ES}_p(X) = 1$ (this implies $\mathbb{P}(X \leqslant 1) < 1$) has a heavier tail than $Y$ with $\mathrm{VaR}_p(Y) = \mathrm{ES}_p(Y) = 1$ (this implies $\mathbb{P}(Y \leqslant 1) = 1$). Thus, intuitively, a forecaster producing the random loss $X$ is more conservative than that producing $Y$. This shows that over-predicting both $\mathrm{VaR}_p$ and $\mathrm{ES}_p$ does not always mean that the forecaster is more conservative about the risk. Our model-free e-statistic $e_p^{\mathrm{ES}}$ can detect this, because $\mathbb{E}[e_p^{\mathrm{ES}}(X, 1, a)] \leqslant 1$ while $\mathbb{E}[e_p^{\mathrm{ES}}(X, 1, 1)] = \infty$, thus correctly rejecting the forecast $(1,1)$ of $(\mathrm{ES}_p(X), \mathrm{VaR}_p(X))$ but not rejecting the truthful forecast $(1, a)$, although $a < 1$.

The following example collects some practical situations of conservative forecasts. In each case, $e_p^{\mathrm{ES}}(L_t, r_t, z_t)$ yields a valid e-variable.

**Example 7.6.** (i) $r_t = \mathrm{ES}_q(L_t|\mathcal{F}_{t-1})$ and $z_t = \mathrm{VaR}_q(L_t|\mathcal{F}_{t-1})$ for $q > p$:

$$\mathbb{E}[e_p^{\mathrm{ES}}(L_t, r_t, z_t)] = \frac{1-q}{1-p}\mathbb{E}[e_q^{\mathrm{ES}}(L_t, \mathrm{ES}_q(L_t|\mathcal{F}_{t-1}), \mathrm{VaR}_q(L_t|\mathcal{F}_{t-1}))] = \frac{1-q}{1-p} < 1.$$

In this situation, $\mathrm{VaR}_p$ and $\mathrm{ES}_p$ are over-predicted by lifting the confidence level $p$ to $q$. For instance, this may represent the output of a stress-testing scenario which amplifies the probability of extreme losses.

(ii) $r_t = c_1\mathrm{ES}_p(L_t|\mathcal{F}_{t-1})$ and $z_t = c_2\mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}) \geqslant 0$ for $c_1 \geqslant c_2 \geqslant 1$, as justified by Proposition 7.1. In particular, $\mathrm{VaR}_p$ and $\mathrm{ES}_p$ can be over-predicted by the same multiplicative factor.

(iii) $r_t = \mathrm{ES}_p(L_t|\mathcal{F}_{t-1}) + b_1$ and $z_t = \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}) + b_2$ for $b_1 \geqslant b_2 \geqslant 0$, as justified by Proposition 7.1. In particular, $\mathrm{VaR}_p$ and $\mathrm{ES}_p$ can be over-predicted by the same absolute amount.

*Remark* 7.4. Our e-backtesting method of ES and the method based on cumulative violations introduced in Du and Escanciano (2017) (we call it the cumulative violation method) have several major different features discussed as follows. First, the cumulative violation method requires distributional forecasts $\hat{u}_t(\hat{\theta})$ as input based on some parametric distribution; our e-backtesting method needs ES and VaR forecasts that can be arbitrarily reported. The arbitrary structure of the forecasts provides more flexibility in practice. Due to this nature, our e-backtesting method does not require special consideration and treatment of estimation effects as discussed in Du and Escanciano (2017) and Hoga and Demetrescu (2022). Second, the cumulative violation method focuses more on detecting model misspecification and is a two-sided test on both overestimation and underestimation of risk; our e-backtesting method is a one-sided test focusing only on the underestimation of ES. This means that we do not reject the null as long as ES is not underestimated even though the forecasts are obtained based on a wrong model or no specific model is assumed. Third, the cumulative violation method relies on a fixed sample size $T$ and relies on an asymptotic model, which means its statistical validity requires it to be only evaluated at the end of the sampling period $T$ that is large enough; our e-backtesting method is sequential and is valid at any stopping time, where detections can be achieved much earlier. This is desirable in risk management applications as early detection of insufficient risk predictions is valuable. Most other classical backtesting methodologies become invalid when evaluated before the end of the pre-specified time period set for testing; see Table 7.1.

## 7.5 Choosing the betting process

One of the essential steps in the testing procedure is choosing a betting process $\boldsymbol{\lambda} = (\lambda_t)_{t \in [T]}$ in (7.7). Throughout this section, $e$ is a $\mathcal{P}$-model-free e-statistic for $\psi = (\rho, \phi) : \mathcal{M} \to \mathbb{R} \times \mathbb{R}^{d-1}$ and $\mathcal{P} \subseteq \mathcal{M}$. We omit $\mathcal{P}$ when $\mathcal{P}$ is the domain of $\psi$. Any predictable process $\boldsymbol{\lambda}$ with values in $[0, 1]$ yields a supermartingale in (7.7) under $H_0$, and thus the testing procedure is valid at all stopping times by Theorem 7.2. However, the statistical power of the tests, and the growth of the process $(M_t(\boldsymbol{\lambda}))_{t \in \{0,\ldots,T\}}$ if the null hypothesis $H_0$ is false, heavily depends on a good choice of the betting process.

### 7.5.1 GRO, GREE, GREL and GREM methods

Our methods are related to maximizing expected log-capital originally proposed by Kelly (1956), adopted by Grünwald et al. (2020) in their GRO (growth-rate optimal) criterion, and studied for testing by betting by Shafer (2021) and Waudby-Smith and Ramdas (2023). For an e-variable $E$ and a probability measure $Q$ representing an alternative hypothesis, the key quantity to consider is $\mathbb{E}^Q[\log E]$, which is called the e-power of $E$ under $Q$ by Vovk and Wang (2022).

Let $T$ be the time horizon of interest, which either can be a finite integer or $\infty$. Let $Q_t$, $t \in [T]$, be specified probability measures representing alternative scenarios for the distribution of $(L_t, r_t, z_t)$ given the information contained in $\mathcal{F}_{t-1}$. Since $(r_t, z_t)$ is $\mathcal{F}_{t-1}$-measurable, the only relevant information from $Q_t$ is the conditional distribution of $L_t$ given $\mathcal{F}_{t-1}$. When choosing the betting process $(\lambda_t)_{t \in [T]}$, we fix an upper bound $\gamma \in (0, 1)$ and restrict $\lambda_t \in [0, \gamma]$ for all $t \in [T]$. The upper bound $\gamma$ is not restrictive and it only prevents some ill-behaving cases. We can safely set $\gamma = 1/2$ (see Remark 7.5 below). Below we formally introduce a few methods to determine the betting process.

1. **GRO** (growth-rate optimal): The optimal betting process maximizing the log-capital growth rate is given by

$$\lambda_t^{\mathrm{GRO}} = \lambda_t^{\mathrm{GRO}}(r, z) = \underset{\lambda \in [0,\gamma]}{\arg\max}\, \mathbb{E}^{Q_t}[\log(1 - \lambda + \lambda e(L_t, r, z)) | \mathcal{F}_{t-1}], \quad (r, z) \in \mathbb{R} \times \mathbb{R}^{d-1}. \quad (7.12)$$

The optimal $\lambda_t^{\mathrm{GRO}}$ in (7.12) can be calculated through a convex program since the function $\lambda \mapsto \log(1 - \lambda + \lambda e(L_t, r, z))$ is concave. This requires the knowledge of the conditional distribution $Q_t$ of $L_t$ given $\mathcal{F}_{t-1}$. In practice, $Q_t$ is unknown to the tester, and one may need to choose a model to approximate $Q_t$. If the probability measure $Q_t$ is unknown but from a certain family, one can use the method of mixtures or mixture martingales (see e.g., de la Peña et al., 2004, 2009) of alternative scenarios to obtain an e-process close to that based on the unknown true model. The optimizer $\lambda_t^{\mathrm{GRO}}$ may not be unique in some special cases, e.g., $e(L_t, r, z)$ is the constant 1 or the expectation in (7.12) is $\infty$, but in most practical cases it is unique.

2. **GREE** (growth-rate for empirical e-statistics): In case the alternative $Q_t$ is unknown, one possibility is to define a betting process based on the empirical probability measures $\widehat{Q}_{t-1}$ of the sample $(L_s, r_s, z_s)_{s\leqslant t-1}$, solving the following optimization problem:

$$\begin{aligned}\lambda_t^{\mathrm{GREE}} &= \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \mathbb{E}^{\widehat{Q}_{t-1}}[\log(1-\lambda+\lambda e(L_t, r_t, z_t))|\mathcal{F}_{t-1}] \\ &= \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1}\log(1-\lambda+\lambda e(L_s, r_s, z_s)).\end{aligned} \tag{7.13}$$

Since (7.13) uses the empirical distribution of the e-statistics $e(L_s, r_s, z_s)$, we call the method based on (7.13) the GREE method. The problem (7.13) can be solved directly via convex programming. This approach is closely related to the GRAPA method introduced in Waudby-Smith and Ramdas (2023, Section B.2).

3. **GREL** (growth-rate for empirical losses): Another alternative method is to choose the betting process solving

$$\begin{aligned}\lambda_t^{\mathrm{GREL}} = \lambda_t^{\mathrm{GREL}}(r,z) &= \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \mathbb{E}^{\widehat{Q}_{t-1}}[\log(1-\lambda+\lambda e(L_t, r, z))|\mathcal{F}_{t-1}] \\ &= \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1}\log(1-\lambda+\lambda e(L_s, r, z)), \quad (r,z)\in\mathbb{R}\times\mathbb{R}^{d-1},\end{aligned} \tag{7.14}$$

where $\widehat{Q}_{t-1}$ is the empirical distribution of the sample $(L_s)_{s\leqslant t-1}$. Since $\lambda_t^{\mathrm{GREL}}$ in (7.14) is calculated based on the empirical distribution of the losses, we call this method the GREL method. The betting process at $t$ in (7.14) is a function of the risk predictions $r$ and $z$, where $r$ and $z$ are usually chosen as the latest risk predictions $r_t$ and $z_t$, respectively. The problem (7.14) can be solved by convex programming, similarly to (7.13). The idea of constructing a betting process depending on predictions has previously been explored by Henzi and Ziegel (2022). By definition, the GREE and GREL methods are equivalent when the risk forecasts $r_t$ and $z_t$ are constant for all $t \in [T]$.

4. **GREM** (growth-rate for empirical mixing): The GREE and GREL methods are asymptotically optimal in different practical situations; see Theorem 7.3 for a rigorous statement and Example 7.7 for an illustration. It is usually difficult to identify the most suitable method based on observations of the losses and forecasts arriving sequentially. Motivated by this, we propose the GREM method, for which we calculate the e-process by taking the mixture of the GREE and GREL methods:

$$M_t(\boldsymbol{\lambda}^{\mathrm{GREM}}) = \frac{M_t(\boldsymbol{\lambda}^{\mathrm{GREE}})}{2} + \frac{M_t(\boldsymbol{\lambda}^{\mathrm{GREL}})}{2},$$

where $M$ is defined in (7.7), $\boldsymbol{\lambda}^{\mathrm{GREM}} = (\lambda_t^{\mathrm{GREM}})_{t\in[T]}$, $\boldsymbol{\lambda}^{\mathrm{GREE}} = (\lambda_t^{\mathrm{GREE}})_{t\in[T]}$, and $\boldsymbol{\lambda}^{\mathrm{GREL}} = (\lambda_t^{\mathrm{GREL}})_{t\in[T]}$. By Lemma 1 of Vovk and Wang (2020), there exists a betting process for the GREM method. More precisely,

$$\lambda_t^{\mathrm{GREM}} = \frac{M_{t-1}(\boldsymbol{\lambda}^{\mathrm{GREE}})\lambda_t^{\mathrm{GREE}} + M_{t-1}(\boldsymbol{\lambda}^{\mathrm{GREL}})\lambda_t^{\mathrm{GREL}}}{M_{t-1}(\boldsymbol{\lambda}^{\mathrm{GREE}}) + M_{t-1}(\boldsymbol{\lambda}^{\mathrm{GREL}})}.$$

The GREM method is asymptotically optimal for all practical cases when either the GREE or the GREL method is optimal (see Theorem 7.3).

An alternative and simple way to get an approximate for (7.13) and (7.14) is to use a Taylor expansion $\log(1+y) \approx y - y^2/2$ at $y = 0$ and the first-order condition. This leads to

$$\lambda_t^{\mathrm{GREL}} \approx 0 \vee \frac{\sum_{s=1}^{t-1} e(L_s, r, z) - t + 1}{\sum_{s=1}^{t-1} (e(L_s, r, z) - 1)^2} \wedge \gamma \tag{7.15}$$

for the GREL method (7.14), and we replace $(r, z)$ in (7.15) by $(r_s, z_s)$ for the GREE method (7.13). The special cases of (7.15) for VaR and ES are given in Section 7.10.

*Remark* 7.5. We restrict the betting process below the upper bound $\gamma$ to avoid the e-process collapsing to 0. For an illustration, suppose that for each $t$, $X_t = e(L_t, r, z)$ given $\mathcal{F}_{t-1}$ takes value 0 with a small probability and value 2 with a large probability, so its expected value is larger than 1 and the null hypothesis is not true. As long as we do not observe 0 up to time $t$, the empirical distribution is concentrated at 2, leading to an optimal strategy $\lambda_t^{\mathrm{GREE}} = 1$ if there is no upper bound $\gamma < 1$. This betting process yields an e-process that becomes 0 as soon as we observe a 0 from $X_t$ and therefore should be avoided. In all our numerical and data experiments, the optimal $\lambda_t$ from each method is typically quite small ($< 0.1$) for tail risk measures like VaR and ES. Hence, it is harmless to set $\gamma = 1/2$ by default.

### 7.5.2 Optimality of betting processes

Next, we discuss the optimality of the betting process. We first define an intuitive notion of asymptotic optimality. For asymptotic results discussed in this section, we will assume an infinite time horizon; that is, we consider $t \in \mathbb{N}$. All statements on probability and convergence are with respect to the true probability generating the data.

**Definition 7.3.** For $(L_{t-1}, r_t, z_t)_{t\in\mathbb{N}}$ adapted to $(\mathcal{F}_{t-1})_{t\in\mathbb{N}}$ and a given function $e : \mathbb{R}^{d+1} \to [0, \infty]$,

(i) two betting processes $\boldsymbol{\lambda} = (\lambda_t)_{t\in\mathbb{N}}$ and $\boldsymbol{\lambda}' = (\lambda'_t)_{t\in\mathbb{N}}$ are *asymptotically equivalent*, denoted by $\boldsymbol{\lambda} \simeq \boldsymbol{\lambda}'$, if

$$\frac{1}{T}(\log M_T(\boldsymbol{\lambda}) - \log M_T(\boldsymbol{\lambda}')) \xrightarrow{L^1} 0 \quad \text{as } T \to \infty,$$

where $M$ is defined in (7.7);

(ii) a betting process $\boldsymbol{\lambda}$ is *asymptotically optimal* if $\boldsymbol{\lambda} \simeq (\lambda_t^{\mathrm{GRO}}(r_t, z_t))_{t\in\mathbb{N}}$.

Intuitively, the asymptotic equivalence between two betting processes means that the long-term growth rates of the two resulting e-processes are the same. Furthermore, the asymptotic optimality of a betting process is defined by asymptotic equivalence using the GRO method as a benchmark because the GRO method is the best-performing method if we know the full distributional information of the losses.

The following proposition characterizes the situations where the betting processes in the GRO method do not reach 0 and 1. In our formulation, $\boldsymbol{\lambda}$ is not allowed to reach 1 due to the upper bound $\gamma < 1$, but we nevertheless give a theoretical condition that the unconstrained optimizer is less than 1.

**Proposition 7.2.** *For $(r, z) \in \mathbb{R} \times \mathbb{R}^{d-1}$ and $t \in \mathbb{N}$ and any optimizer $\lambda_t^{\mathrm{GRO}}$ of (7.12), the following statements hold.*

*(i) $\lambda_t^{\mathrm{GRO}}(r, z) > 0$ if and only if $\mathbb{E}^{Q_t}[e(L_t, r, z)|\mathcal{F}_{t-1}] > 1$.*

*(ii) With $\gamma = 1$ in (7.12), $\lambda_t^{\mathrm{GRO}}(r, z) < 1$ if and only if $\mathbb{E}^{Q_t}[1/e(L_t, r, z)|\mathcal{F}_{t-1}] > 1$.*

Below we present an assumption for the asymptotic analysis. The condition is very weak because the interesting case in backtesting is when $\mathbb{E}^{Q_t}[\log(e(L_t, r, z))]$ is small. Denote by $\psi^*(\mathcal{P}) \subseteq \psi(\mathcal{P})$ as the set of all values $(r, z) \in \psi(\mathcal{P})$ such that $e(x, r, z) < \infty$ for all $x \in \mathbb{R}$.

**Assumption 1.** *For all* $(r, z) \in \psi^*(\mathcal{P})$, $\sup_{t\in\mathbb{N}} \mathbb{E}^{Q_t}[\log(e(L_t, r, z))] < \infty$.

The following theorem addresses the asymptotic optimality of the GREE, GREL and GREM methods in different situations with its proof put in Section 7.5.3.

**Theorem 7.3.** *For* $(L_{t-1}, r_t, z_t)_{t\in\mathbb{N}}$ *adapted to* $(\mathcal{F}_{t-1})_{t\in\mathbb{N}}$ *such that* $(r_t, z_t)$ *takes values in* $\psi^*(\mathcal{P})$ *and* $e : \mathbb{R}^{d+1} \to [0, \infty]$, *under Assumption 1, the following statements hold.*

*(i)* $(\lambda_t^{\mathrm{GREE}})_{t\in\mathbb{N}}$ *is asymptotically optimal if* $(e(L_t, r_t, z_t))_{t\in\mathbb{N}}$ *is iid and* $(r_t, z_t)_{t\in\mathbb{N}}$ *is deterministic.*

*(ii)* $(\lambda_t^{\mathrm{GREL}}(r_t, z_t))_{t\in\mathbb{N}}$ *is asymptotically optimal if* $(L_t)_{t\in\mathbb{N}}$ *is iid and either:*

*(a)* $(r_t, z_t)_{t\in\mathbb{N}}$ *takes finitely many possible values in* $\mathbb{R}^d$.

*(b)* $(r_t, z_t)$, $t \in \mathbb{N}$, *are in a common compact set,* $e(x, r, z)$ *is continuous in* $(r, z)$, *and* $(r_t, z_t) \xrightarrow{\mathrm{p}} (r_0, z_0)$ *as* $t \to \infty$ *for some* $(r_0, z_0) \in \mathbb{R}^d$.

*(iii)* $(\lambda_t^{\mathrm{GREM}})_{t\in\mathbb{N}}$ *is asymptotically optimal if either* $(\lambda_t^{\mathrm{GREE}})_{t\in\mathbb{N}}$ *or* $(\lambda_t^{\mathrm{GREL}}(r_t, z_t))_{t\in\mathbb{N}}$ *is asymptotically optimal.*

The asymptotic optimality results in Theorem 7.3 are based on strong, and perhaps unrealistic, assumptions; they are imposed for technical reasons. Nevertheless, we obtain some useful insight on the comparison between the GREE and GREL. Intuitively, the GREE method should outperform the GREL method when the model-free e-statistics $e(L_t, r_t, z_t)$, $t \in \mathbb{N}$, are iid and $(r_t, z_t)$ is not informative about how to choose $\lambda_t$ (i.e., they are noises), while the GREL method should outperform the GREE method when the losses $L_t$, $t \in \mathbb{N}$, are iid and $(r_t, z_t)$ is informative about how to choose $\lambda_t$; recall that GREL uses the information of $(r_t, z_t)$ whereas GREE does not. Moreover, we expect the asymptotic optimality results to hold (approximately) without strong assumptions on the risk predictions $(r_t, z_t)_{t\in\mathbb{N}}$ as imposed in Theorem 7.3. We illustrate the insights for the comparison between the GREE and GREL methods through Example 7.7 below. In most practical cases, we do not know clear patterns of the losses and forecasts as they arrive sequentially over time. In this sense, the GREM method is recommended because Theorem 7.3 suggests that it would perform well in all the cases where either the GREE or the GREL method is asymptotically optimal.

**Example 7.7.** Let the size of training data be $l = 10$, the sample size for testing be $n = 1,000$, and $Z_1, \ldots, Z_{n+l}$ be iid samples simulated from the standard normal distribution. We report the average performance of backtesting methods over $1,000$ simulations.

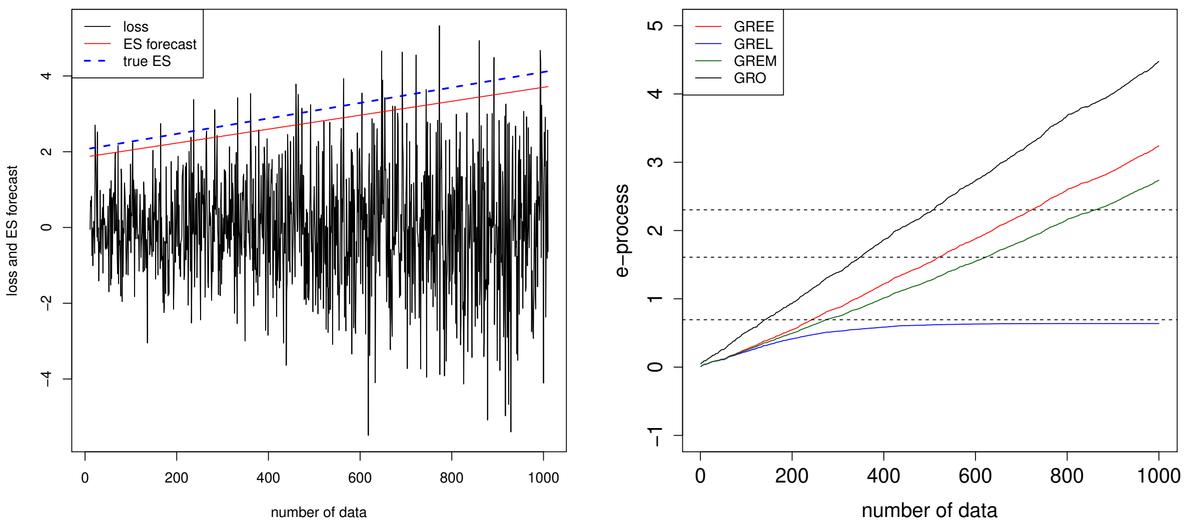


Figure 7.1: Realized losses and ES forecasts with a linear extending business (left panel); average log-transformed e-processes obtained by different methods over $1,000$ simulations (right panel)

(a) The iid condition of the whole model-free e-statistics implies that the GREE method works better than the GREL method when losses and risk forecasts exhibit co-movements over time. Such situations are common in the financial market; for instance, risk forecasts will increase over time when a company is extending its business. Assume that $L_t = (1 + t/(n + l))Z_t$ for $t \in [n+l]$. This model represents the case where the financial institution's investment generates iid cash flow but the institution increases the investment amount over time. Following the increasing trend of the investment, the risk forecaster announces the under-estimated forecasts of $\mathrm{VaR}_{0.95}(L_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_{0.95}(L_t|\mathcal{F}_{t-1})$ as $z_t = 1.48(1+t/(n+l))$ and $r_t = 1.86(1+t/(n+l))$, respectively, for $t \in [n + l]$. Figure 7.1 plots the realized losses $L_t$, ES forecasts $r_t$, and the e-processes obtained by the GRO, GREE, GREL and GREM methods for $t = l + 1, \ldots, n + l$. We observe from Figure 7.1 that the GREE e-process dominates the GREL e-process. This is consistent with the result of Theorem 7.3 by noting the co-movements of the losses and the VaR and ES forecasts which makes the model-free e-statistics $(e_p(L_t, r_t, z_t))_{t\in[n+l]}$ iid.

(b) Another example of co-movements between the losses and forecasts is where a company exhibits a non-linear business cycle. Take the random losses to be $L_t = Z_t(1 + \sin(\theta t))$ for $t \in [n + l]$, where $\theta = 0.01$. The risk forecasts of $\mathrm{VaR}_{0.95}(L_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_{0.95}(L_t|\mathcal{F}_{t-1})$ also have a similar trend but are under-estimated. Namely, we have $z_t = 1.48(1 + \sin(\theta t))$ for VaR and $r_t =$

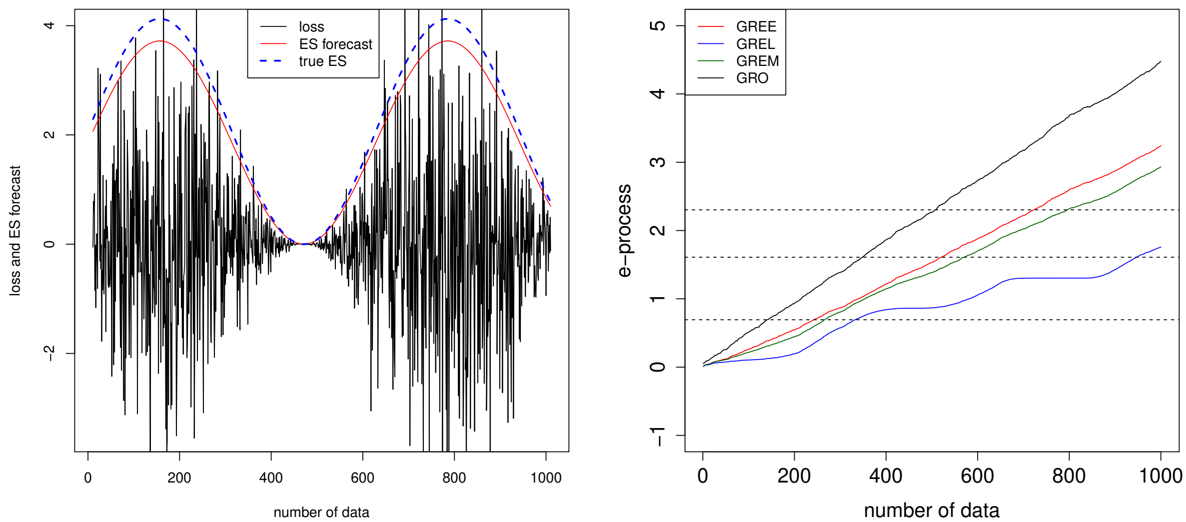


Figure 7.2: Realized losses and ES forecasts with a non-linear business cycle (left panel); average log-transformed e-processes obtained by different methods over $1,000$ simulations (right panel)

$1.86(1+\sin(\theta t))$ for ES. The losses and forecasts, and the average log e-processes for different methods are plotted in Figure 7.2. Similarly to (i), we also observe better performance of the GREE method than GREL because of the overall iid pattern of the whole e-statistics.

(c) It is expected from Theorem 7.3 that the GREL method will dominate the GREE method when the losses exhibit an iid pattern and there is no clear evidence of co-movements between losses and risk forecasts. Let the random losses be $Z_1,\dots,Z_{n+l}$; thus, they are iid. Suppose that the risk forecaster announces the forecasts of $\mathrm{VaR}_{0.95}(Z_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_{0.95}(Z_t|\mathcal{F}_{t-1})$ to be $z_t = 1.64+\epsilon_t$ and $r_t = 2.06+\epsilon_t$, respectively, for $t\in[n+l]$, where $\epsilon_1,\dots,\epsilon_{n+l}$ are iid samples uniformly distributed on the support $\{\pm i/10 : i = 0,\dots,5\}$. In this case, the forecaster is able to obtain risk forecasts close to the true values but is subject to a forecasting error $(\epsilon_t)_{t\in[n+l]}$. Figure 7.3 plots the realized losses $Z_t$, ES forecasts $r_t$, and the corresponding e-processes obtained by the GRO, GREE, GREL and GREM methods for $t = l+1,\dots,n+l$. We observe from Figure 7.3 that the GREL method outperforms the GREE method. This example shows that the GREL method is able to detect evidence against risk forecasts due to downward fluctuations of the forecasts, while GREE does not perform well in this case because it only uses historical forecasts whose average is close to the true value.

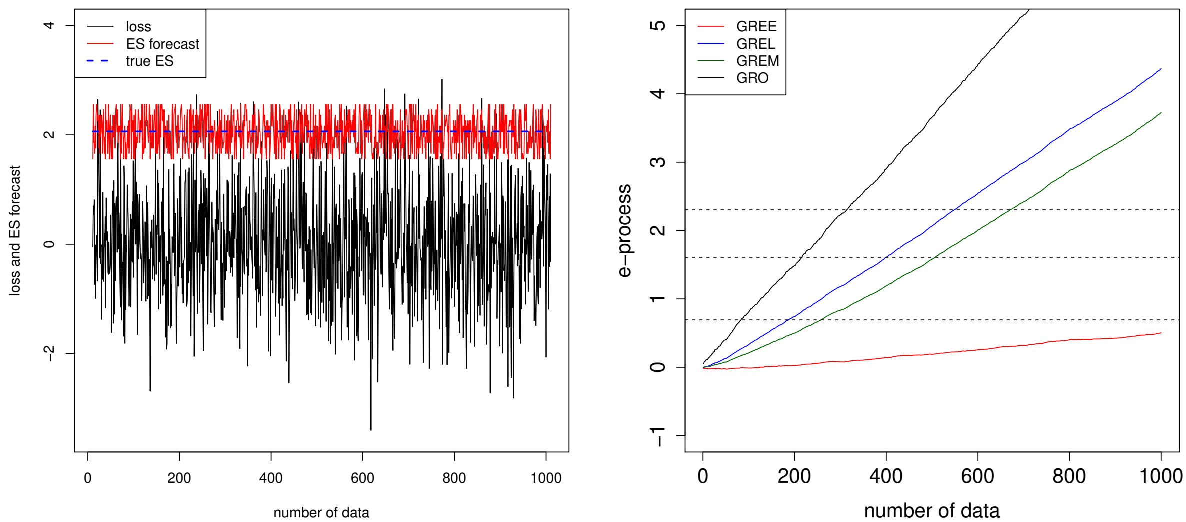


Figure 7.3: Realized losses and ES forecasts with iid losses (left panel); average log-transformed e-processes obtained by different methods over 1,000 simulations (right panel)

### 7.5.3 Proof of Theorem 7.3

We first present a proposition used in the proof of Theorem 7.3. Its proof is put in Section 7.12. This proposition says that under the iid assumption, the betting process computed from empirical distributions is asymptotically equivalent to that computed from the true distribution.

**Proposition 7.3.** *Let $X_1, X_2, \dots$ be nonnegative iid random variables with $\mathbb{E}[\log(X_1)] < \infty$. Let*

$$\lambda_t = \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \frac{1}{t-1}\sum_{s=1}^{t-1}\log(1-\lambda+\lambda X_s); \quad \lambda^* = \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \mathbb{E}[\log(1-\lambda+\lambda X_t)], \quad t\in\mathbb{N}.$$

*We have $T^{-1}\sum_{t=1}^{T}(\log(1-\lambda_t+\lambda_t X_t)) - \log(1-\lambda^*+\lambda^* X_t)) \xrightarrow{L^1} 0$ as $T\to\infty$.*

Proposition 7.3 gives a simplified illustration of the asymptotic optimality of the GREE method, which uses historical model-free e-statistics as iid input. A rigorous statement of this point is already presented in Theorem 7.3.

*Proof of Theorem 7.3.* For (i), because $(r_t, z_t)_{t\in\mathbb{N}}$ is deterministic and $(e(L_t, r_t, z_t))_{t\in\mathbb{N}}$ is iid, we have for all $t\in\mathbb{N}$,

$$\lambda_t^{\mathrm{GRO}}(r_t, z_t) = \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \mathbb{E}^{Q_t}[\log(1-\lambda+\lambda e(L_t, r_t, z_t))|\mathcal{F}_{t-1}] = \operatorname*{arg\,max}_{\lambda\in[0,\gamma]} \mathbb{E}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}] =: \lambda^*,$$

where $X_t = e(L_t, r_t, z_t)$. The statement thus follows directly from Proposition 7.3.

For (ii), we first show the result for any fixed $(r, z) \in \psi^*(\mathcal{P})$. This follows directly from Proposition 7.3 by taking $X_t = e(L_t, r, z)$, $\lambda_t = \lambda_t^{\mathrm{GREL}}(r, z)$, and $\lambda^* = \lambda_t^{\mathrm{GRO}}(r, z)$ for $t \in \mathbb{N}$.

(a) Suppose that $(r_t, z_t)_{t\in\mathbb{N}}$ takes finitely many possible values in $\mathbb{R}^{\mathbb{N}}$. Let $(M_t)_{t\in\mathbb{N}}$ be defined in (7.7), $\boldsymbol{\lambda}^{\mathrm{GREL}} = (\lambda_t^{\mathrm{GREL}}(r_t, z_t))_{t\in\mathbb{N}}$ and $\boldsymbol{\lambda}^{\mathrm{GRO}} = (\lambda^{\mathrm{GRO}}(r_t, z_t))_{t\in\mathbb{N}}$. We have

$$\frac{1}{T}(\log(M_T(\boldsymbol{\lambda}^{\mathrm{GREL}})) - \log(M_T(\boldsymbol{\lambda}^{\mathrm{GRO}}))) \xrightarrow{L^1} 0$$

by taking mixtures of all possible values of $(r_t, z_t)_{t\in\mathbb{N}}$ that are finitely many.

(b) It suffices to show the result for $d = 1$ and the general case holds similarly. Since $e(x, r)$ is continuous in $r$ and $r_t$, $t \in \mathbb{N}$, are in a common compact set, $e(x, r)$ is uniformly continuous with respect to $r$. Define $M(r, \lambda) = \mathbb{E}^{Q_t}[\log(1 - \lambda + \lambda e(L_t, r))|\mathcal{F}_{t-1}]$ with $r \in \mathbb{R}$, $\lambda \in [0, \gamma]$ and $L_t \sim Q_t$ for $t \in \mathbb{N}$.

Let $Q_t$ be the empirical probability measure $\widehat{Q}_{t-1}$ for all $t \in \mathbb{N}$. We now prove that for all $t \in \mathbb{N}$ and $\epsilon > 0$, there exists $\delta_1 > 0$, such that for all $|r - r'| < \delta_1$, $|\lambda_t^{\mathrm{GREL}}(r) - \lambda_t^{\mathrm{GREL}}(r')| \leqslant \epsilon$. Suppose that the negated statement is true. Hence there exists $t \in \mathbb{N}$ and $\epsilon_0 > 0$, such that for all $\delta > 0$, there exist $|r_\delta - r'_\delta| < \delta$, $|\lambda_t^{\mathrm{GREL}}(r_\delta) - \lambda_t^{\mathrm{GREL}}(r'_\delta)| > \epsilon_0$. Because $\lambda_t^{\mathrm{GREL}}(r) = \arg\max_{\lambda\in[0,1]} M(r, \lambda)$ and $M(r, \lambda)$ is strictly concave in $\lambda$, we have

$$\min\left\{M(r_\delta, \lambda_t^{\mathrm{GREL}}(r_\delta)) - M(r_\delta, \lambda_t^{\mathrm{GREL}}(r'_\delta)),\ M(r'_\delta, \lambda_t^{\mathrm{GREL}}(r'_\delta)) - M(r'_\delta, \lambda_t^{\mathrm{GREL}}(r_\delta))\right\} > l$$

for some $l > 0$. By uniform continuity of $M(r, \lambda)$ with respect to $r$, there exists $\delta_0 > 0$, such that for all $|r - r'| < \delta_0$,

$$\max\left\{\left|M(r, \lambda_t^{\mathrm{GREL}}(r)) - M(r', \lambda_t^{\mathrm{GREL}}(r))\right|, \left|M(r, \lambda_t^{\mathrm{GREL}}(r')) - M(r', \lambda_t^{\mathrm{GREL}}(r'))\right|\right\} < l.$$

Therefore,

$$\begin{aligned}2l &< M(r_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r_{\delta_0})) - M(r_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r'_{\delta_0})) + M(r'_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r'_{\delta_0})) - M(r'_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r_{\delta_0}))\\ &\leqslant \left|M(r_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r_{\delta_0})) - M(r'_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r_{\delta_0}))\right| + \left|M(r_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r'_{\delta_0})) - M(r'_{\delta_0}, \lambda_t^{\mathrm{GREL}}(r'_{\delta_0}))\right| < 2l.\end{aligned}$$

This leads to a contradiction.

Similarly, we can show that there exists $\delta_2 > 0$, such that for all $|r - r'| < \delta_2$, $|\lambda^{\mathrm{GRO}}(r) - \lambda^{\mathrm{GRO}}(r')| \leqslant \epsilon$ by taking $Q_t$ to the probability measure $Q$ for the iid random variables $L_t$, $t \in \mathbb{N}$. Take $\hat{\delta} = \min\{\delta_1, \delta_2\}$. Because $r_t \xrightarrow{\mathrm{P}} r_0$, for all $\eta > 0$, there exists $N \in \mathbb{N}$, such that $Q(|r_t - r_0| < \hat{\delta}) > 1 - \eta$ for all $t > N$. It follows that

$$\max\left\{Q(|\lambda_t^{\mathrm{GREL}}(r_t) - \lambda_t^{\mathrm{GREL}}(r_0)| > \epsilon),\ Q(|\lambda^{\mathrm{GRO}}(r_t) - \lambda^{\mathrm{GRO}}(r_0)| > \epsilon)\right\} \leqslant Q(|r_t - r_0| \geqslant \hat{\delta}) < \eta.$$

Since we also have $\lambda_t^{\mathrm{GREL}}(r_0) \xrightarrow{\mathrm{P}} \lambda^{\mathrm{GRO}}(r_0)$ as $t \to \infty$, it is clear that $\lambda_t^{\mathrm{GREL}}(r_t) \xrightarrow{\mathrm{P}} \lambda^{\mathrm{GRO}}(r_0)$ and $\lambda^{\mathrm{GRO}}(r_t) \xrightarrow{\mathrm{P}} \lambda^{\mathrm{GRO}}(r_0)$ as $t \to \infty$. By boundedness of the betting processes, we have $\lambda_t^{\mathrm{GREL}}(r_t) \xrightarrow{L^1} \lambda^{\mathrm{GRO}}(r_0)$ and $\lambda^{\mathrm{GRO}}(r_t) \xrightarrow{L^1} \lambda^{\mathrm{GRO}}(r_0)$ as $t \to \infty$. The result thus holds by (7.16).

For (iii), write $\boldsymbol{\lambda}^{\mathrm{GREE}} = (\lambda_t^{\mathrm{GREE}})_{t\in\mathbb{N}}$ and $\boldsymbol{\lambda}^{\mathrm{GREM}} = (\lambda_t^{\mathrm{GREM}})_{t\in\mathbb{N}}$. It suffices to notice that $M_T(\boldsymbol{\lambda}^{\mathrm{GREM}}) \geqslant \max\{M_T(\boldsymbol{\lambda}^{\mathrm{GREL}}), M_T(\boldsymbol{\lambda}^{\mathrm{GREE}})\}/2$, and by taking a limit as $T \to \infty$ we obtain the asymptotic optimality of $\boldsymbol{\lambda}^{\mathrm{GREM}}$ from that of $\boldsymbol{\lambda}^{\mathrm{GREE}}$ or $\boldsymbol{\lambda}^{\mathrm{GREL}}$. $\square$

## 7.6 Characterizing model-free e-statistics

In this section, we present several results on the characterization of model-free e-statistics. The main practical message is that the two e-statistics which we introduced, $e_p^Q$ and $e_p^{\mathrm{ES}}$ in Section 7.2, are essentially the only useful choices for VaR and (ES,VaR), respectively, in building up e-processes in Section 7.3. The reader more interested in applications may skip this section in the first reading, while keeping in mind the above practical message.

### 7.6.1 Necessary conditions for the existence of model-free e-statistics

Not all functionals $\rho$ on $\mathcal{M}$ admit model-free e-statistics that are solely based on the information of $\rho$. Below we give a necessary condition for a model-free e-statistic testing $\rho$ to exist. A functional $\rho : \mathcal{M} \to \mathbb{R}$ is *monotone* if $\rho(F) \leqslant \rho(G)$ for all $F \leqslant_1 G$, where $\leqslant_1$ is the usual stochastic order; namely, $F \leqslant_1 G$ if and only if $F \geqslant G$ pointwise on $\mathbb{R}$. We also say that $\rho$ is *uncapped* if for each $F \in \mathcal{M}$ and $r > \rho(F)$, there exists $\bar{F} \in \mathcal{M}$ such that $\bar{F} \geqslant_1 F$ and $\rho(\bar{F}) = r$. All monetary risk measures (Föllmer and Schied, 2016) are monotone and uncapped. A functional $\rho : \mathcal{M} \to \mathbb{R}$ is *quasi-convex* if $\rho(\lambda F + (1-\lambda)G) \leqslant \max\{\rho(F), \rho(G)\}$ for all $\lambda \in [0,1]$ and $F, G \in \mathcal{M}$. Similarly, $\rho$ is *quasi-concave* if $-\rho$ is quasi-convex, and $\rho$ is *quasi-linear* if it is both quasi-convex and quasi-concave.

**Proposition 7.4.** *Suppose that $\rho : \mathcal{M} \to \mathbb{R}$ is monotone and uncapped. If there exists a model-free e-statistic testing $\rho$, then $\rho$ is quasi-convex.*

When $\mathcal{M}$ is convex, quasi-convexity of $\rho$ is equivalent to the condition that the set $\{F \in \mathcal{M} : \rho(F) \leqslant r\}$ is convex for each $r \in \mathbb{R}$. The requirement in Proposition 7.4 rules out a large class of coherent risk measures including ES.[5] As is shown in the following proposition, if the e-statistic testing $\rho$ is strict, then $\rho$ is necessarily quasi-linear, which is stronger than the quasi-convexity in Proposition 7.4, and this result does not require that $\rho$ is monotone and uncapped.

**Proposition 7.5.** *If there exists a model-free e-statistic $e : \mathbb{R}^2 \to [0, \infty]$ strictly testing $\rho : \mathcal{M} \to \mathbb{R}$, then $\rho$ is quasi-linear.*

We say a functional $\rho : \mathcal{M} \to \mathbb{R}$ has *convex level sets (CxLS)* if the set $\{F \in \mathcal{M} : \rho(F) = r\}$ is convex for each $r \in \mathbb{R}$. Quasi-linearity of $\rho$ is stronger than the condition that $\rho$ has CxLS, and they are equivalent when $\mathcal{M}$ is convex and $\rho$ is monotone. Functionals with CxLS have been studied extensively in the recent literature due to their connection to elicitability and backtesting (Gneiting, 2011; Ziegel, 2016). For a recent summary of related results, see Wang and Wei (2020).

### 7.6.2 Characterizing model-free e-statistics for common risk measures

In this section we offer several characterization results of model-free e-statistics for some common risk measures using identification functions. The link between model-free e-statistics and identification functions is presented in Section 7.11. The following two propositions characterize all continuous model-free e-statistics testing the mean and for $(\mathrm{var}, \mathbb{E})$ testing the variance; see also Examples 7.1 and 7.2.

**Proposition 7.6** (Model-free e-statistics for the mean of bounded random variables)**.** *Let $a \in \mathbb{R}$ and $\mathcal{P}$ be the set of distributions in $\mathcal{M}_1$ with support in $[a, \infty)$. All continuous $\mathcal{P}$-model-free e-statistics $e'$ testing the mean are of the form*

$$e'(x, r) = 1 + h(r)\frac{x - r}{r - a}, \quad x \geqslant a,\ r \geqslant a,$$

*where $h$ is a continuous function on $[a, \infty)$ with $0 < h \leqslant 1$. Moreover, the functions $h$ and $r \mapsto (r - a)/h(r)$ are increasing if and only if $e'$ is strictly testing the mean.*

**Proposition 7.7** (Model-free e-statistics for the variance)**.** *All continuous model-free e-statistics $e'$ for $\psi = (\mathrm{var}, \mathbb{E})$ testing* var *are of the form*

$$e'(x, r, z) = 1 + h(r, z)\frac{(z - x)^2 - r}{r}, \quad x, z \in \mathbb{R},\ r \geqslant 0,$$

*where $h$ is a continuous function on $[0, \infty) \times \mathbb{R}$ with $0 < h \leqslant 1$. Moreover, the functions $r \mapsto h(r, z)$ and $r \mapsto r/h(r, z)$ are increasing for all $z \in \mathbb{R}$ if and only if $e'$ is strictly testing* var.

[5] In particular, all comonotonic-additive coherent risk measures except for the mean are monotone and uncapped but not quasi-convex (see e.g., Wang et al. 2020, Theorem 3).

The simplest model-free e-statistic for VaR is given in (7.1) in Example 7.3. The following proposition exhausts all model-free e-statistics testing VaR with an additional continuity requirement. We say a model-free e-statistic $e$ is *non-conservative* if $\int_{\mathbb{R}} e(x, \psi(F))\, \mathrm{d}F(x) = 1$ for each $F \in \mathcal{P}$.

**Theorem 7.4** (Model-free e-statistics for VaR)**.** *Let $p \in (0,1)$ and $\mathcal{P}$ be the set of all distributions with a quantile continuous at $p$. All $\mathcal{P}$-model-free e-statistics $e'$ that are continuous except at $x = r$, non-conservative, and testing $\mathrm{VaR}_p$ are of the form*

$$e'(x, r) = 1 + h(r) \frac{p - \mathbb{1}_{\{x \leqslant r\}}}{1 - p},$$

*where $h$ is a continuous function on $\mathbb{R}$ with $0 < h \leqslant 1$. The function $h$ is constant if and only if $e'$ is strictly testing $\mathrm{VaR}_p$.*

Next, we consider the model-free e-statistics for $(\mathrm{VaR}, \mathrm{ES})$ testing ES. It is straightforward that $\mathrm{ES}_p$ is monotone, uncapped, and $\{F \in \mathcal{M} : \mathrm{ES}_p(F) \leqslant r\}$ is not convex.[6] Hence, Proposition 7.4 implies that there does not exist a model-free e-statistic testing $\mathrm{ES}_p$ using solely the information of $\mathrm{ES}_p$. A similar point was made in Acerbi and Szekely (2017) that $\mathrm{ES}_p$ is not backtestable in some specific sense. As is shown in Theorem 7.1, there exists a model-free e-statistic $e_p^{\mathrm{ES}}$ testing $\mathrm{ES}_p$ using the information of $(\mathrm{ES}_p, \mathrm{VaR}_p)$. The following theorem characterizes all model-free e-statistics for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ testing ES, which is slightly more than $e_p^{\mathrm{ES}}$.

**Theorem 7.5.** *Let $p \in (0,1)$ and $\mathcal{P}$ be the set of all distributions with finite mean and a quantile continuous at $p$. All continuous and non-conservative $\mathcal{P}$-model-free e-statistics $e'$ for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ testing $\mathrm{ES}_p$ are of the form*

$$e'(x, r, z) = 1 + h(r, z) \left( \frac{(x - z)_+}{(1 - p)(r - z)} - 1 \right), \quad x \in \mathbb{R},\ z < r,$$

*where $h$ is a continuous function such that $0 < h \leqslant 1$. Moreover, the functions $r \mapsto h(r, z)$ and $r \mapsto (r - z)/h(r, z)$ are increasing for all $z < r$ if and only if $e'$ is strictly testing $\mathrm{ES}_p$.*

Theorems 7.4 and 7.5 illustrate the essential roles of $e_p^Q$ and $e_p^{\mathrm{ES}}$ among all possible choices of e-statistics for VaR and the pair (ES,VaR). All choices of e-statistics for $\mathrm{VaR}_p$ have the form $1 - \lambda + \lambda e_p^Q(x, r)$, and all those for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ has the form $1 - \lambda + \lambda e_p^{\mathrm{ES}}(x, r, z)$, where $\lambda$ is a function taking values in $[0, 1]$. Therefore, in view of (7.7), the e-statistics $e$ can be without loss of

[6] It might be interesting to note that $\mathrm{ES}_p$ is concave on $\mathcal{M}$, implying that the set $\{F \in \mathcal{M} : \mathrm{ES}_p(F) \geqslant r\}$ is convex for each $r$; see Theorem 3 of Wang et al. (2020).

generality chosen as $e_p^Q$ for $\mathrm{VaR}_p$ and $e_p^{\mathrm{ES}}$ for $(\mathrm{ES}_p, \mathrm{VaR}_p)$, and $\lambda$ can be chosen separately depending on the risk forecasts.

## 7.7 Simulation studies

In this section, we provide simulation studies on backtesting the Value-at-Risk and the Expected Shortfall. This illustrates the details of our backtesting methodology numerically. Furthermore, we examine how different factors affect the quality of the backtesting procedure, especially the impact of the choice of the betting process in (7.7). We evaluate the backtesting performance when the risk measures are under-reported, over-reported, or reported exactly by the risk forecaster.

For all e-tests, we report evidence against the forecasts when the e-process exceeds thresholds 2, 5, or 10.[7] We call such evidence a detection. From the practical viewpoint, the three thresholds we choose form four zones for levels of alerts to financial institutions. This is in a similar sense to the standard three-zone approach for backtesting VaR in the financial industry.

*Remark* 7.6. In classical statistical terminology, what we call a *detection* is a *rejection* of the null hypothesis based on our e-test with thresholds 2, 5, and 10, respectively. Since the threshold of 2 has a guaranteed significance level of 50%, it would be unconventional to speak of a rejection of the null hypothesis. However, having detected evidence of size 2 with the e-test may be a useful early warning that risk predictions might not be prudent enough. Recall that Jeffrey's threshold of e-values for "substantial" evidence is 3.2 and for "decisive" evidence is 10; see Shafer (2021) and Vovk and Wang (2021) for more discussions on observing moderately large e-values.

The simulation and data analysis in Sections 7.7 and 7.8, together with those in Section 7.13, illustrate our main methodology. They are supplemented by extended results and discussions in a separate paper Wang et al. (2022).[8]

[7]We usually do not compare the e-value thresholds with p-value thresholds in traditional p-tests. However, if necessary, the reader can obtain significance levels in p-tests by taking inverse (significance levels 50%, 20%, and 10% from the e-value thresholds 2, 5, and 10, respectively), but this conversion generally loses statistical evidence.

[8]Wang et al. (2022) includes detailed descriptions and results for e-tests with iid observations, stationary time series data, detecting structural change of time series, analysis with NASDAQ index on an extended data period, optimized portfolios, results for $\mathrm{VaR}_{0.95}$ and $\mathrm{ES}_{0.875}$, and comparison between the GREE, GREL and GREM methods.

### 7.7.1 Backtests via stationary time series

We apply our e-backtesting procedure to a setting with time series. For comparison, we use the same setup as in Nolde and Ziegel (2017) and simulate data from an AR(1)–GARCH(1, 1) process for daily negated log-returns of a financial asset:

$$L_t = \mu_t + \sigma_t Z_t, \quad \mu_t = -0.05 + 0.3X_{t-1}, \quad \sigma_t^2 = 0.01 + 0.1\epsilon_{t-1}^2 + 0.85\sigma_{t-1}^2, \quad t \in \mathbb{N}$$

where $\{Z_t\}_{t\in\mathbb{N}}$ is a sequence of iid innovations following a skewed-t distribution with shape parameter $\nu = 5$ and skewness parameter $\gamma = 1.5$. In total, $1,000$ independent simulations are produced, each of which includes a sample of size 500 used for backtesting. A rolling window of size 500 is applied for risk estimation at each time spot $t$.

For forecasting, we assume that the data follow an AR(1)–GARCH(1, 1) process $\{L_t\}_{t\in\mathbb{N}}$ with $L_t = \mu_t + \sigma_t Z_t$, where $\{Z_t\}_{t\in\mathbb{N}}$ is assumed to be a sequence of iid innovations with mean 0 and variance 1, following a normal, t-, or skewed-t distribution. Thus, the forecaster has a correct time-series structure with possibly incorrect innovation. Here, $\{\mu_t\}_{t\in\mathbb{N}}$ and $\{\sigma_t\}_{t\in\mathbb{N}}$ are adapted to $(\mathcal{F}_{t-1})_{t\in\mathbb{N}}$. The details of the forecasting procedure are described in Section 7.13.1. The risk forecaster deliberately under-reports, over-reports, or reports the exact point forecasts of $\mathrm{VaR}_p$ or $(\mathrm{ES}_p, \mathrm{VaR}_p)$ she obtains.

For backtesting, the e-processes in (7.7) are calculated with the betting process $(\lambda_t)_{t\in[T]}$ chosen by the GREM method using Taylor approximation via (7.15). The results for the GREE and GREL methods and their comparison are demonstrated in Section 7.13.2. We detect evidence against the forecasts when the e-processes exceed thresholds 2, 5, or 10. We first present results for backtesting $\mathrm{VaR}_{0.99}$. The percentage of detections, the average number of days taken to detect evidence against the forecasts (conditional on detection occurring), and the average final log-transformed e-values are shown in Tables 7.2 and 7.3.

As expected, the results show that evidence against normal and t-innovations is more likely to be detected than against skewed-t innovations, which is the true model. The percentage of detections for exact skewed-t forecasts, or the Type I error, is 0.5% for threshold 10. Under-reporting VaR leads to earlier detections than reporting the exact VaR forecasts and the converse holds true for over-reporting.

Results on backtests of ES are reported in Tables 7.4 and 7.5. The results in Table 7.4 confirm our intuition that under-reporting or using a wrong innovation can be detected with a large

| | normal | | | t | | | skewed-t | | |
|---|---|---|---|---|---|---|---|---|---|
| threshold | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| $-10\%$ | 99.6 | 98.0 | 94.3 | 97.8 | 88.8 | 76.1 | 46.0 | 14.4 | 5.9 |
| exact | 97.0 | 87.9 | 75.0 | 86.8 | 60.9 | 40.2 | 17.8 | 2.5 | 0.5 |
| $+10\%$ | 86.1 | 62.4 | 41.4 | 62.5 | 26.3 | 11.9 | 6.3 | 0.4 | 0 |

Table 7.2: Percentage of detections (%) for $\mathrm{VaR}_{0.99}$ forecasts over $1,000$ simulations of time series and 500 trading days using the GREM method

| | normal | | | | t | | | | skewed-t | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| threshold | 2 | 5 | 10 | | 2 | 5 | 10 | | 2 | 5 | 10 | |
| $-10\%$ | 116 | 185 | 228 | (5.489) | 156 | 238 | 285 | (3.390) | 230 | 284 | 332 | (0.3707) |
| exact | 158 | 239 | 287 | (3.311) | 200 | 284 | 322 | (1.759) | 219 | 244 | 217 | ($-0.07467$) |
| $+10\%$ | 196 | 277 | 316 | (1.858) | 224 | 305 | 351 | (0.7341) | 183 | 227 | – | ($-0.2135$) |

Table 7.3: The average number of days taken to detect evidence against $\mathrm{VaR}_{0.99}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days using the GREM method; numbers in brackets are average final log-transformed e-values

probability, whereas forecasts from the true model and their more conservative versions appear the opposite. Moreover, under-reporting (resp. over-reporting) both of ES and VaR and under-reporting (resp. over-reporting) only ES have similar performance in terms of probability of detection and time of detection. The average time to detection (Table 7.5) is useful for risk management since early warnings (threshold 2) are often issued after about a fourth of the sampling time, and decisive warnings (threshold 10) after about half of the considered trading days.

### 7.7.2 Monitoring structural change of time series

We examine the power of our e-backtesting method to monitor the structural change of simulated time series data. We refer to Chu et al. (1996) and Berkes et al. (2004) for earlier work on monitoring the structural change of data sets. For a comparison with the results in Hoga and Demetrescu (2022), we use the same setup as described in their Section 6, and call their method the

| | normal | | | t | | | skewed-t | | |
|---|---|---|---|---|---|---|---|---|---|
| threshold | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| $-10\%$ ES | 99.8 | 99.5 | 98.5 | 98.4 | 88.8 | 77.1 | 47.6 | 16.1 | 6.2 |
| $-10\%$ both | 99.8 | 99.5 | 98.1 | 98.5 | 91.4 | 82.0 | 48.0 | 15.7 | 6.5 |
| exact | 99.3 | 95.7 | 88.3 | 88.1 | 63.9 | 43.1 | 18.8 | 4.0 | 0.8 |
| $+10\%$ both | 95.2 | 80.4 | 61.9 | 64.9 | 27.6 | 9.9 | 7.1 | 1.0 | 0 |
| $+10\%$ ES | 94.8 | 79.8 | 62.1 | 70.0 | 34.9 | 15.6 | 7.9 | 1.1 | 0.1 |

Table 7.4: Percentage of detections (%) for $\mathrm{ES}_{0.975}$ forecasts over $1,000$ simulations of time series and 500 trading days using the GREM method

sequential monitoring method. We simulate the losses $\{L_t\}_{t\in\mathbb{N}}$ following the GARCH$(1,1)$ process:

$$L_t = -\sigma_t Z_t, \quad \sigma_t^2 = 0.00001 + 0.04X_{t-1}^2 + \beta_t\sigma_{t-1}^2,$$

where $\{Z_t\}_{t\in\mathbb{N}}$ is a sequence of iid innovations following a skewed-t distribution with shape parameter $\nu = 5$ and skewness parameter $\gamma = 0.95$, $\beta_t = 0.7 + 0.25\mathbb{1}_{\{t>b^*\}}$ and $b^*$ represents the time after which the model is subject to a structural change. We simulate 250 presampled data for forecasting risk measures and another 250 data for backtesting.

We choose the probability level for $\mathrm{VaR}_p$ and $\mathrm{ES}_p$ to be $p = 0.95$. Via the presampled data, the forecaster obtains the forecasts of $\mathrm{VaR}_{0.95}(L_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_{0.95}(L_t|\mathcal{F}_{t-1})$ using empirical VaR and ES of the residuals and the estimated model parameters $\hat{\boldsymbol{\theta}} = (\hat\omega, \hat\alpha, \hat\beta)$. See Section 7.13.3 for details of the forecasting procedure. Due to the model-free nature, we only use the losses and forecasts $(L_t, r_t, z_t)$ for our e-backtesting method, while the sequential monitoring method also uses the estimated volatility $\sigma_t(\hat{\boldsymbol{\theta}})$ by assuming the GARCH model of the losses. As suggested by Hoga and Demetrescu (2022), the Monte Carlo simulations detector with a rolling window performs the best among others for both VaR and ES monitoring. Therefore, we take this method for comparison with ours. We choose the size $m = 50$ of the rolling window. The significance level of the sequential monitoring method is set to be 5%, while we choose the rejection threshold of our e-backtesting method to be $1/5\% = 20$.

Figure 7.4 plots the average results we get based on $10,000$ simulations, where the betting processes of e-backtesting are chosen by the GREE, GREL or GREM method. The top panels plot the percentage of detections over the total $10,000$ simulations, including those before and

| | normal | | | | t | | | | skewed-t | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| threshold | 2 | 5 | 10 | | 2 | 5 | 10 | | 2 | 5 | 10 | |
| −10% ES | 89 | 137 | 176 | (6.671) | 151 | 223 | 277 | (3.428) | 224 | 271 | 264 | (0.5072) |
| −10% both | 94 | 146 | 189 | (6.347) | 141 | 215 | 271 | (3.720) | 218 | 256 | 251 | (0.4679) |
| exact | 129 | 201 | 247 | (4.311) | 185 | 267 | 311 | (1.953) | 198 | 195 | 251 | (−0.04676) |
| +10% both | 171 | 250 | 292 | (2.737) | 217 | 283 | 289 | (0.8275) | 141 | 206 | – | (−0.2072) |
| +10% ES | 168 | 247 | 297 | (2.702) | 207 | 286 | 298 | (0.9865) | 147 | 158 | 165 | (−0.2323) |

Table 7.5: The average number of days taken to detect evidence against $\mathrm{ES}_{0.975}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days using the GREM method; "–" represents no detection; numbers in brackets are average final log-transformed e-values

after the structural changes at $t = b^* + 1$, while the bottom panels show the average number of trading days from the structural changes at $t = b^* + 1$ to detections through backtesting, given that detections occur after $t = b^* + 1$. We call this quantity the average run length (ARL) as in Hoga and Demetrescu (2022). As expected, all three methods of e-backtesting are dominated by the sequential monitoring method since they do not rely on the model information. However, the GREE, GREL and GREM methods exhibit reasonable performance for all values of $b^*$. From the ARL plots, the GREE, GREL and GREM methods detect evidence against the forecasts around 0 to 30 days later than the sequential monitoring method. Near $b^* = 0$ and $b^* = 250$, the GREE, GREL and GREM methods yield similar detection percentages as the sequential monitoring method.

## 7.8 Financial data analysis

### 7.8.1 The NASDAQ index

We calculate the negated percentage log-returns using data of the NASDAQ Composite index from Jan 16, 1996 to Dec 31, 2021. An AR(1)-GARCH(1, 1) model is fitted to the data with a moving estimation window of 500 data points. The e-processes in (7.7) are calculated with the betting process $(\lambda_t)_{t\in[T]}$ chosen by the GREE, GREL or GREM method. Different from the backtesting methods used in Section 7.7.1, for each $t \in [T]$, the empirical mean in (7.15) is calculated using a moving window of data in the past 500 days. This choice is made to reflect the practice of

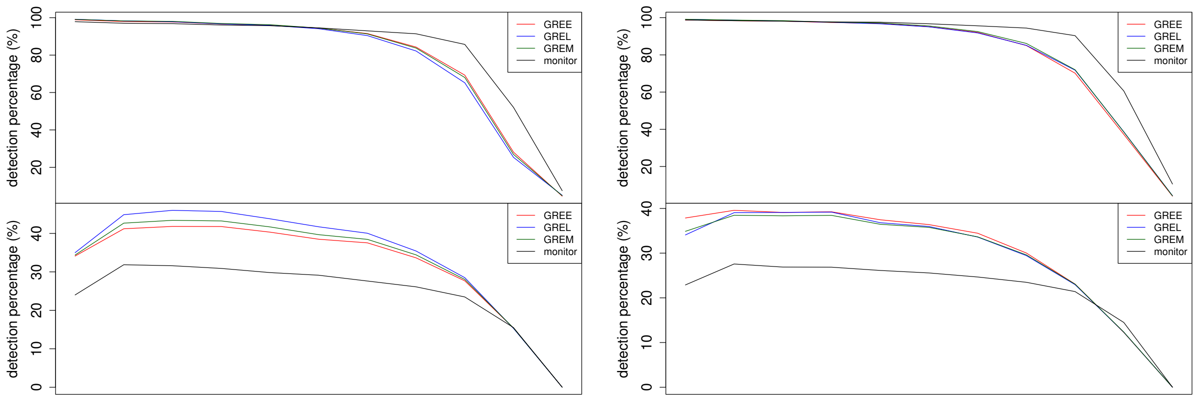


Figure 7.4: Percentage of detections (%) of $\mathrm{VaR}_{0.95}$ (left panels) and $\mathrm{ES}_{0.95}$ (right panels) forecasts over $10,000$ simulations of time series and 250 trading days with structural changes at $b^*$ (top panels); ARLs of backtesting procedures (bottom panels); black lines ("monitor") represent the results of the sequential monitoring method

risk modeling where more recent data represent better the current market and economic conditions. Therefore, the first 500 forecasts use 500 data points each, and we start the backtesting procedure after the first 500 forecasts are available, thus after the first $1,000$ data points. The sample size for backtesting is 5536, corresponding to forecasts of risk measures from Jan 3, 2000 to Dec 31, 2021. We plot the negated log-returns and the forecasts of $\mathrm{ES}_{0.975}$ fitted by normal, t-, and skewed-t distributions for the innovations over time in Figure 7.5. In addition to the parametric methods, we also plot the empirical risk forecasts with a non-parametric rolling window approach in Figure 7.5.

We present backtesting results using data from Jan 3, 2005 to Dec 31, 2021 to examine the impact of the $2007-2008$ financial crisis. Figure 7.6 shows the e-processes over time. Table 7.6 demonstrates the average $\mathrm{ES}_{0.975}$ forecasts and the number of days taken to detect evidence against the forecasts, where the second last rows contain the results for ES forecasts deliberately over-reported by 10% assuming skewed-t innovations as a forecasting model that is prudent.

We observe most of the detections in Table 7.6 happen around $500-700$ trading days after Jan 3, 2005, where significant losses occurred during the financial crisis. Correspondingly, there are sharp jumps of the e-processes in Figure 7.6 at around $500-700$ trading days. In general, we observe that detections for lower thresholds 2 and 5 are significantly earlier than those for the

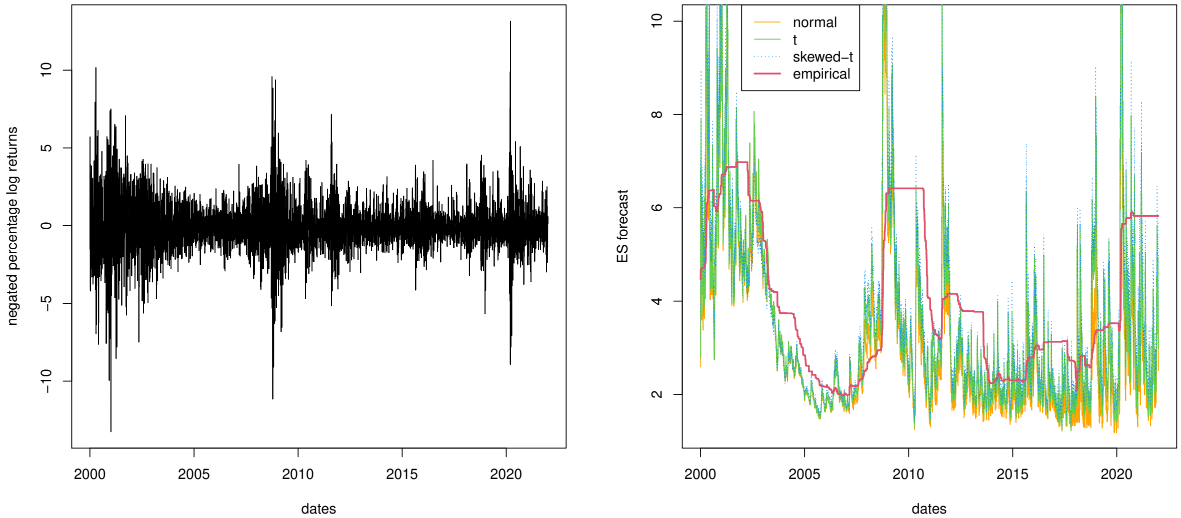


Figure 7.5: Negated percentage log-returns of the NASDAQ Composite index (left panel); $\mathrm{ES}_{0.875}$ and $\mathrm{ES}_{0.975}$ forecasts fitted by normal distribution (right panel) from Jan 3, 2000 to Dec 31, 2021

final threshold 10. This features one of the advantages of our e-backtesting procedure in practice: Our procedure is inherently sequential, and thus, no extra effort is required to allow for monitoring of predictive performance in comparison to testing only at the end of a sampling period. This allows regulators to get alerted much earlier than using the traditional p-tests when e-processes exceed the first threshold 2 or further exceed 5. The backtesting procedure may be stopped when an e-process exceeds 10, which indicates a "decisive" failure of the underlying model used by the financial institution.

The GREL method performs better than the GREE method in this case except for the empirical forecasts. This may be because the sharp increase of losses upon the occurrence of the financial crisis violates the growth trend and co-movements of the losses and the risk forecasts, making the GREE method not favorable compared with the GREL method as discussed in Example 7.7. It seems from the result that the GREL method is more likely to detect evidence against the risk forecasts for extreme events (e.g., financial crisis) causing an abnormally sharp increase in losses. The GREL method does not perform well in detecting evidence against the empirical forecasts for both VaR and ES. This is expected because the empirical forecasts and the betting process of the GREL method are both obtained only by the information of the empirical distribution of losses, making GREL lack additional information to reject the empirical forecasts. Compared with the

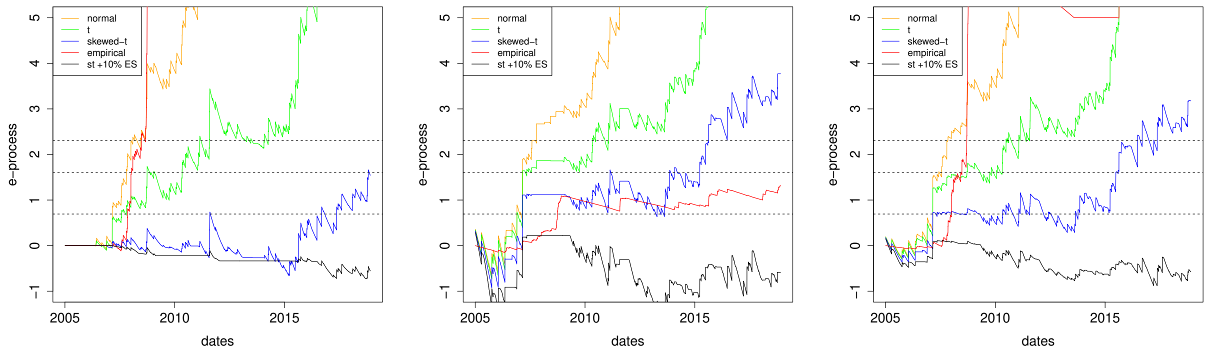


Figure 7.6: Log-transformed e-processes testing $\mathrm{ES}_{0.975}$ with respect to the number of days for the NASDAQ index from Jan 3, 2005 to Dec 31, 2021; left panel: GREE method, middle panel: GREL method, right panel: GREM method

GREE and GREL methods, the performance of the GREM method is more stable in different cases with all cases of underestimation detected. Therefore, the GREM method is recommended as a default choice for implementation.

### 7.8.2 Optimized portfolios

Apart from the NASDAQ index, we perform the e-backtesting procedure on data of a portfolio of $n = 22$ stocks from Jan 5, 2001 to Dec 31, 2021. Suppose that a bank invests in the above portfolio. After each trading day at time $t \in [T]$, the weights

$$\mathbf{w}_t = (w_t^1, \dots, w_t^n) \in \Delta_n = \left\{ (w_1, \dots, w_n) \in [0,1]^n : \sum_{i=1}^{n} w_i = 1 \right\}$$

are determined by a mean-variance criterion. Specifically, the bank solves the following optimization problem:[9]

$$\max_{\mathbf{w}_t \in \Delta_n} \mathbb{E}[-\mathbf{w}_t^\top \mathbf{L}_t] - \frac{\gamma}{2} \mathrm{var}(-\mathbf{w}_t^\top \mathbf{L}_t),$$

where $\mathbf{L}_t = (L_t^1, \dots, L_t^n)$ is the vector of negated percentage log-returns for all stocks in the portfolio modeled by an AR(1)-GARCH(1,1) process. The bank reports VaR and ES of the weighted portfolio by assuming $\mathbf{w}_t^\top \mathbf{X}_t$ to be normal, t-, or skewed-t distributed, respectively. Some of the

[9] We use the mean-variance strategy to illustrate our method for its simplicity, despite its performance may not be empirically satisfactory; see e.g., DeMiguel et al. (2009). Recall that our backtesting method does not require knowledge of the trading strategy or the statistical model, and can be applied to any trading strategy.

| threshold | | GREE 2 | GREE 5 | GREE 10 | GREL 2 | GREL 5 | GREL 10 | GREM 2 | GREM 5 | GREM 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| normal | (**2.676**) | 540 | 704 | 756 | 479 | 540 | 650 | 540 | 610 | 713 |
| t | (**2.997**) | 650 | 941 | 1545 | 479 | 540 | 1344 | 540 | 933 | 1381 |
| skewed-t | (**3.202**) | 1661 | 3477 | – | 540 | 1545 | 2676 | 540 | 2639 | 2889 |
| st +10% ES | (**3.522**) | – | – | – | – | – | – | – | – | – |
| empirical | (**3.656**) | 719 | 758 | 876 | 941 | 3823 | – | 756 | 862 | 931 |

Table 7.6: Average $\mathrm{ES}_{0.975}$ forecasts (boldface in brackets) and the number of days taken to detect evidence against the forecasts for the NASDAQ index from Jan 3, 2005 to Dec 31, 2021; "–" means no detection is detected till Dec 31, 2021

assumptions in the estimation procedure are simplistic, and hence we do not expect to obtain precise risk forecasts. Suppose that a financial institution reports its risk forecasts based on the naive approach described above. We are more likely to get detections if the simplistic assumptions lead to underestimation. The detailed setup and the list of stocks can be seen in Section 7.13.4.

Table 7.7 shows the average forecasts of $\mathrm{ES}_{0.975}$ and backtesting results with different innovation distributions. The e-processes are plotted in Figure 7.7. The portfolio data differ from the simulated time series in the sense that the random losses and risk predictions exhibit much more complicated temporal dependence. Detections are obtained in most of the cases for thresholds 2 and 5 before large losses come in during the financial crisis in 2008. This demonstrates one of the practical advantages of our method, that is, due to the model-free nature, our e-backtesting method is able to detect evidence against risk forecasts when losses and risk forecasts exhibit complicated temporal dependence. This enables regulations for most real portfolio investments in financial markets, where model assumptions (e.g., stationarity) made by previous literature on backtesting ES are less likely to hold.

Between the two methods, the GREL method works better than the GREE method for $\mathrm{ES}_{0.975}$ in most of the cases. However, there is no clear general guidance on which method dominates the other due to the complexity of the strategy, which may not be known. As such, we recommend the GREM method in general.

| | | GREE | | | GREL | | | GREM | | |
|---|---|---|---|---|---|---|---|---|---|---|
| threshold | | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| normal | (**1.224**) | 547 | 730 | 767 | 438 | 541 | 541 | 461 | 541 | 714 |
| t | (**1.385**) | 767 | 934 | 3036 | 1009 | 2207 | 2411 | 778 | 2207 | 2502 |
| skewed-t | (**1.435**) | 767 | – | – | 469 | 1009 | 2502 | 541 | 2411 | 2972 |
| st +10% ES | (**1.578**) | – | – | – | – | – | – | – | – | – |

Table 7.7: Average $\mathrm{ES}_{0.975}$ forecasts (boldface in brackets) and the number of days taken to detect evidence against the forecasts for portfolio data from Jan 3, 2005 to Dec 31, 2021; "–" means no detection is detected till Dec 31, 2021

## 7.9 Concluding remarks

The e-backtesting method proposed in this chapter is the first model-free and non-asymptotic backtest for ES, the most important risk measure in financial regulation implemented by BCBS (2016). Our methodology contributes to the backtesting issues of ES, which have been a central point of discussions in the risk management literature (e.g., Nolde and Ziegel, 2017; Du and Escanciano, 2017; Hoga and Demetrescu, 2022, and the references therein). Our methods are constructed using the recently developed notions of e-values and e-processes, which are shown to be promising in many application domains of statistics other than risk management. Some topics on which e-values become useful include sequential testing (Shafer, 2021; Grünwald et al., 2020), multiple testing and false discovery control (Vovk and Wang, 2021; Wang and Ramdas, 2022), probability forecast evaluation (Henzi and Ziegel, 2022), meta-analysis in biomedical sciences (ter Schure and Grünwald, 2021), and composite hypotheses (Waudby-Smith and Ramdas, 2023). Our work connects two active areas of research through theoretical results and methodologies, and we expect more techniques from either world to be applicable to solve problems from the other.

Our e-test procedures feature advantages of e-values, including validity for all stopping times and feasibility for no assumptions on the underlying models. Central to our proposed backtesting method, the notion of model-free e-statistics is introduced, which is useful also for traditional testing problems (Remark 7.3), although the main focus of the chapter is backtesting. The characterization results in Section 7.6 give guidelines on how to choose model-free e-statistics to build the e-processes.

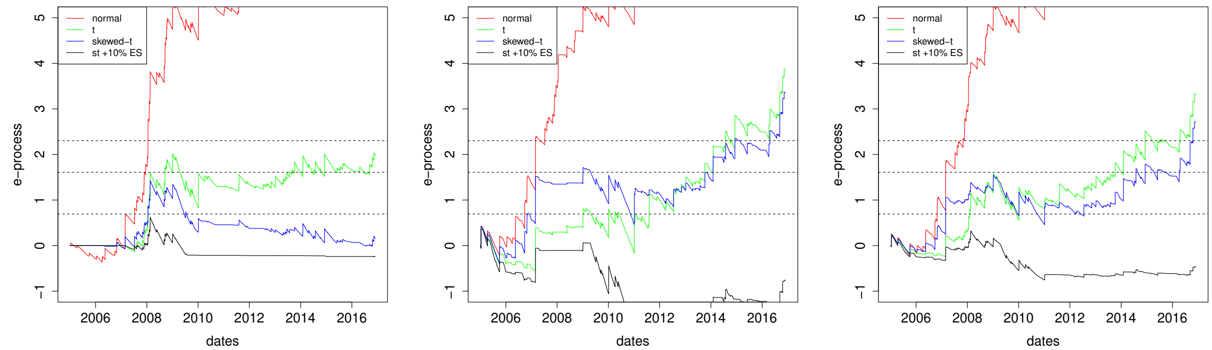


Figure 7.7: Log-transformed e-processes testing $\mathrm{ES}_{0.975}$ with respect to the number of days for portfolio data from Jan 3, 2005 to Dec 31, 2021; left panel: GREE method, middle panel: GREL method, right panel: GREM method

Remarkably, for VaR and ES, the unique forms of model-free e-statistics are identified, leaving little doubt on how to choose them for real-data applications.

If the sample size of a test is fixed, and accurate forecasts for the risk model are available and to be tested together with forecasts for the risk measure, then traditional model-based methods may be recommended to use in practice, as they often have better power than our e-backtests. In the more realistic situations where the sample size is not fixed, or no models are to be tested along with the risk measure forecasts, our e-backtests are useful, and their multifaceted attractive features are illustrated by our study.

As for any other new statistical methodology, e-backtests have their own limitations, challenges, and possible extensions. As the main limitation, since e-backtests require very little information on the underlying model, they could be less powerful than traditional model-based or p-value-based approaches. Therefore, there is a trade-off between flexibility and power that a risk practitioner has to keep in mind. For future directions, an important task is to obtain theoretically optimal betting processes using some data-driven procedures under practical assumptions. The methodology can be extended to more general risk measures and economic indices useful in different contexts, each demanding its own model-free e-statistics and backtesting procedure. Another future direction interesting to us is a game-theoretic framework in which the financial institution actively decides its optimal forecasting strategy by providing the least possible risk forecasts that are barely sufficient to pass a regulatory backtest. The intuition is that, since e-backtests are robust to model assumptions, they should be less vulnerable to this type of adverse strategies of the financial institutions compared to some model-based tests, but a full theoretical analysis is needed before any concrete conclusion can be drawn.

## 7.10 Taylor approximation formulas for GREE and GREL

We give formulas for the betting processes of the GREE and GREL methods for VaR and ES via Taylor approximation. For the GREL method, the special case of VaR, that is, taking $e = e_p^Q$ in (7.15), yields

$$\lambda_t^{\mathrm{GREL}} \approx 0 \vee \frac{(1-p)\left((t-1)p - \sum_{s=1}^{t-1} \mathbb{1}_{\{L_s \leqslant r\}}\right)}{(t-1)p^2 + (1-2p)\sum_{s=1}^{t-1} \mathbb{1}_{\{L_s \leqslant r\}}} \wedge \gamma.$$

For the special case of ES, taking $e = e_p^{\mathrm{ES}}$ in (7.15), the approximation is

$$\lambda_t^{\mathrm{GREL}} \approx 0 \vee \frac{(1-p)(r-z)\left(\sum_{s=1}^{t-1}(L_s - z)_+ - (t-1)(1-p)(r-z)\right)}{\sum_{s=1}^{t-1}((L_s - z)_+ - (1-p)(r-z))^2} \wedge \gamma.$$

The corresponding formulas for the GREE method are obtained by replacing $r$ and $(r, z)$ by $r_s$ and $(r_s, z_s)$ in the $s$-th summand in above formulas, respectively.

## 7.11 Link between model-free e-statistics and identification functions

The link between model-free e-statistics and identification functions is useful for deriving the characterization results of model-free e-statistics in Section 7.6.2. An integrable function $V : \mathbb{R}^{d+1} \to \mathbb{R}^d$ is said to be an $\mathcal{M}$-*identification function* for a functional $\psi : \mathcal{M} \to \mathbb{R}^d$ if $\int_{\mathbb{R}} V(x, \psi(F))\, \mathrm{d}F(x) = \mathbf{0}$ for all $F \in \mathcal{M}$. Furthermore, $V$ is said to be *strict* if

$$\int_{\mathbb{R}} V(x, y)\, \mathrm{d}F(x) = \mathbf{0} \iff y = \psi(F)$$

for all $F \in \mathcal{M}$ and $y \in \mathbb{R}^d$ (Fissler and Ziegel, 2016). We say that $\psi$ is *identifiable* if there exists a strict $\mathcal{M}$-identification function for $\psi$.

There is a connection between model-free e-statistics strictly testing $\rho$ and identification functions. Let $e : \mathbb{R}^2 \to \mathbb{R}$ be a model-free e-statistic strictly testing $\rho : \mathcal{M} \to \mathbb{R}$. For $F \in \mathcal{M}$ and $r \geqslant \rho(F) > r'$ it holds that

$$\int e(x, r)\, \mathrm{d}F(x) \leqslant 1 < \int e(x, r')\, \mathrm{d}F(x),$$

and hence, $1-e(x,r)$ is often a strict identification function for $\rho$. Since identifiability of a functional coincides with eliciability under some assumptions detailed in Steinwart et al. (2014), Proposition 7.5 is not surprising since elicitable functionals are known to have CxLS.

**Proposition 7.8.** *Let $e : \mathbb{R}^{d+1} \to [0,\infty]$ be a non-conservative model-free e-statistic for $\psi = (\rho,\phi) : \mathcal{M} \to \mathbb{R}^d$, and assume that $\phi$ has a $\mathcal{P}$-identification function $v$. We have $V(x,r,z) = (v(x,z), 1-e(x,r,z))^\top$ is a $\mathcal{P}$-identification function for $\psi$.*

*Proof.* Let $F \in \mathcal{P}$. By assumption,

$$\int V(x,\rho(F),\phi(F))\,\mathrm{d}F(x) = \left(\int v(x,\phi(F))\,\mathrm{d}F(x), 1 - \int e(x,\rho(F),\phi(F))\,\mathrm{d}F(x)\right)^\top = \mathbf{0}. \qquad \square$$

The connection of model-free e-statistics to identification functions is useful because under some regularity conditions there are characterization results for all possible identification functions for a functional (Fissler, 2017; Dimitriadis et al., 2023). Below, we use these results to derive characterizations of model-free e-statistics. Roughly speaking, given a model-free e-statistic $e$ strictly testing $\rho$, then all other possible model-free e-statistics $e'$ strictly testing $\rho$ must be of the form

$$e'(x,r) = 1 + h(r)(e(x,r) - 1)$$

for some non-negative function $h$. Clearly, $h$ must fulfill further criteria to ensure that $e'$ is a model-free e-statistic strictly testing $\rho$.

A further consequence of these considerations is that for a functional $\rho$ with model-free e-statistic strictly testing $\rho$, there must be an identification function $V(x,r)$ that is bounded below by $-1$. This rules out a number of functionals including the expectation without further conditions on $\mathcal{M}$.

## 7.12 Omitted proofs of all results

*Proof of Proposition 7.1.* Suppose that $H_0$ in (7.10) holds. By the VaR-ES relation in (7.4) and (7.5),

$$\mathbb{E}[e_p^{\mathrm{ES}}(L,r,z)] = \frac{\mathbb{E}\left[(L-z)_+\right]}{(1-p)(r-z)} \leqslant \frac{\mathbb{E}\left[(L-\mathrm{VaR}_p(L))_+\right]}{(1-p)(\mathrm{ES}_p(L)-\mathrm{VaR}_p(L))} = 1.$$

Hence, $e_p^{\mathrm{ES}}(L,r,z)$ is an e-variable for (7.10). $\square$

*Proof of Proposition 7.2.* For all $(r,z) \in \mathbb{R} \times \mathbb{R}^{d-1}$ and $t \in \mathbb{N}$, write $X_t = e(L_t, r, z)$.

(i) For the "$\Leftarrow$" direction, suppose that $\lambda_t^{\mathrm{GRO}}(r,z) = 0$. By Taylor expansion at $\lambda = 0$ and continuity of $\lambda \mapsto \mathbb{E}^{Q_t}[(X_t - 1)\lambda | \mathcal{F}_{t-1}]$,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}] = \mathbb{E}^{Q_t}[(X_t-1)|\mathcal{F}_{t-1}] + \mathrm{o}(\lambda).$$

Taking $\lambda \downarrow 0$ yields that $\mathbb{E}^{Q_t}[X_t|\mathcal{F}_{t-1}] \leqslant 1$. For the "$\Rightarrow$" direction, suppose that $\mathbb{E}^{Q_t}[X_t|\mathcal{F}_{t-1}] \leqslant 1$. It follows that

$$\mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}] \leqslant \mathbb{E}^{Q_t}[(X_t-1)\lambda|\mathcal{F}_{t-1}] \leqslant 0.$$

By strict concavity of $\lambda \mapsto \mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}]$, we have $\lambda_t^{\mathrm{GRO}}(r,z) = 0$, where the upper bound 0 is obtained.

(ii) For the "$\Leftarrow$" direction, suppose that $\lambda_t^{\mathrm{GRO}}(r,z) = 1$. It is clear that $Q_t(X_t = 0) = 0$. It follows by continuity of $\lambda \mapsto \mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}]$ and $\lambda \mapsto \mathbb{E}^{Q_t}[(X_t-1)/(1-\lambda+\lambda X_t)]$ that

$$0 \leqslant \frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}]\bigg|_{\lambda=1} = \mathbb{E}^{Q_t}\left[\frac{X_t-1}{X_t}\bigg|\mathcal{F}_{t-1}\right].$$

Hence, $\mathbb{E}^{Q_t}[1/X_t|\mathcal{F}_{t-1}] \leqslant 1$. For the "$\Rightarrow$" direction, suppose that $\mathbb{E}^{Q_t}[1/X_t|\mathcal{F}_{t-1}] \leqslant 1$. It follows that

$$\begin{aligned}\mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t) - \log(X_t)|\mathcal{F}_{t-1}] &= \mathbb{E}^{Q_t}\left[\log\left(\frac{1-\lambda}{X_t}+\lambda\right)\bigg|\mathcal{F}_{t-1}\right] \\ &\leqslant \mathbb{E}^{Q_t}\left[(1-\lambda)\left(\frac{1}{X_t}-1\right)\bigg|\mathcal{F}_{t-1}\right] \leqslant 0.\end{aligned}$$

By strict concavity of $\lambda \mapsto \mathbb{E}^{Q_t}[\log(1-\lambda+\lambda X_t)|\mathcal{F}_{t-1}]$, we have $\lambda_t^{\mathrm{GRO}}(r,z) = 1$, where the upper bound $\log(X_t)$ is obtained. $\square$

The following lemma will be used in the proof of Proposition 7.3.

**Lemma 7.1.** *If $M, M_t : [0,1] \to L^0$ are convex for all $t \in \mathbb{N}$, $M$ is continuous and $M_t(\lambda) \xrightarrow{\text{a.s.}} M(\lambda)$ as $t \to \infty$ for all $\lambda \in [0,1]$, then $\sup_{\lambda\in[0,1]} |M_t(\lambda) - M(\lambda)| \xrightarrow{\mathrm{P}} 0$ as $t \to \infty$.*

*Proof.* For all $t \in \mathbb{N}$, define an affine function $\psi_t : [0,1] \to L^0$ such that $\psi_t(0) = M_t(0)$ and $\psi_t(1) = M_t(1)$. This is clear that $\psi_t$ converges uniformly to the affine function $\psi : [0,1] \to L^0$ such that $\psi(0) = M(0)$ and $\psi(1) = M(1)$. Therefore, replacing $M_t$ by $M_t - \psi_t$ and $M$ by $M - \psi$, we assume without loss of generality that $M_t(0) = M(0) = M_t(1) = M(1) = 0$ for all $t \in \mathbb{N}$.

For all $\eta > 0$, take $\epsilon = \eta/4$. By continuity of $M$, there exists $\delta_0 > 0$, such that $|M(\lambda) - M(\lambda')| < \epsilon$ for all $|\lambda - \lambda'| < \delta$. By convexity of $M$, there exists $\tilde{\lambda} \in (0,1)$, such that $M$ is decreasing on $[0, \tilde{\lambda}]$

and increasing on $[\tilde{\lambda}, 1]$. Define $K = (M(1) - M(\tilde{\lambda}))/(1 - \tilde{\lambda})$ and $\delta = \min\{\epsilon/K, \delta_0\}$. There exist $\lambda_1 < \cdots < \lambda_{I-1}$ for $I \in \mathbb{N} \setminus \{1\}$, such that $\lambda_{i+1} - \lambda_i < \delta$ for all $i \in \{0, \ldots, I-1\}$, where we write $\lambda_0 = 0$ and $\lambda_I = 1$. It follows that $\delta \leqslant 1 - \lambda_i$ for all $i \in \{0, \ldots, I-1\}$.

Because $M_t(\lambda) \xrightarrow{\text{a.s.}} M(\lambda)$ as $t \to \infty$ for all $\lambda \in [0, 1]$, there exists an event $A$ with $\mathbb{P}(A) = 1$ as follows: There exists $T \in \mathbb{N}$, such that for all $t > T$, we have $|M_t(\lambda_i) - M(\lambda_i)| < \epsilon$ for all $i = 0, \ldots, I$. For all $\lambda \in [0, 1]$, there exists $i \in \{0, \ldots, I\}$, such that $\lambda \in [\lambda_i, \lambda_{i+1}]$. Without loss of generality, we assume $0 \leqslant \lambda_i \leqslant \lambda \leqslant \lambda_{i+1} \leqslant \tilde{\lambda}$. The case of $\tilde{\lambda} \leqslant \lambda_i \leqslant \lambda \leqslant \lambda_{i+1} \leqslant 1$ can be shown analogously by symmetry. If $A$ holds, then

$$\begin{aligned}
|M_t(\lambda) - M(\lambda)| &\leqslant |M_t(\lambda) - M(\lambda_i)| \\
&< \epsilon + \frac{M(1) - M(\lambda_{i+1}) + \epsilon}{1 - \lambda_{i+1}} \delta + |M(\lambda_i) - M(\lambda_{i+1})| \\
&< 3\epsilon + K \min\left\{\frac{\epsilon}{K}, \delta_0\right\} \leqslant 4\epsilon = \eta.
\end{aligned}$$

Therefore, we have $\mathbb{P}(\sup_{\lambda \in [0,1]} |M_t(\lambda) - M(\lambda)| \geqslant \eta) = 0$ for all $t > T$. It follows that $\sup_{\lambda \in [0,1]} |M_t(\lambda) - M(\lambda)| \xrightarrow{\text{P}} 0$ as $t \to \infty$. $\square$

*Proof of Proposition 7.3.* Write $M(\lambda) = \mathbb{E}[\log(1 - \lambda + \lambda X_t)]$ and

$$M_t(\lambda) = \frac{1}{t-1} \sum_{s=1}^{t-1} \log(1 - \lambda + \lambda X_s) \quad \text{for } \lambda \in [0, \gamma], \ t \in \mathbb{N}.$$

By the strong law of large numbers, we have $M_t(\lambda) \xrightarrow{\text{a.s.}} M(\lambda)$ as $t \to \infty$ for all $\lambda \in [0, \gamma]$. Since the functions $M$ and $M_t$ are concave for all $t \in \mathbb{N}$, by Lemma 7.1, we have $\sup_{\lambda \in [0,\gamma]} |M_t(\lambda) - M(\lambda)| \xrightarrow{\text{a.s.}} 0$. For all $\epsilon > 0$, we have

$$\sup_{\lambda : |\lambda - \lambda^*| \geqslant \epsilon} M(\lambda) \leqslant M(\lambda^*)$$

by the definition of $\lambda^*$ and the concavity of $M$. For all $t \in \mathbb{N}$, we have $M_t(\lambda_t) \geqslant M_t(\lambda^*)$ by the definition of $\lambda_t$. Therefore, we have by Theorem 5.7 of van der Vaart (1998) that $\lambda_t \xrightarrow{\text{P}} \lambda^*$ as $t \to \infty$. Because $\lambda_t$ is bounded for all $t \in \mathbb{N}$, $\{\lambda_t\}_{t \in \mathbb{N}}$ is uniformly integrable. It follows that $\lambda_t \to \lambda^*$ with respect to the $L^1$-norm as $t \to \infty$, denoted by $\lambda_t \xrightarrow{L^1} \lambda^*$; see e.g., Resnick (2019, Theorem 6.6.1).

Next, we show that

$$\frac{1}{T} \sum_{t=1}^{T} (\log(1 - \lambda_t + \lambda_t X_t) - \log(1 - \lambda^* + \lambda^* X_t)) \xrightarrow{L^1} 0 \tag{7.16}$$

as $T \to \infty$. For all $\epsilon_1 > 0$ and $t \in \mathbb{N}$, by continuity and the monotone convergence theorem, there exists $\delta > 0$, such that $\log(1 - \delta) > -\epsilon_1$ and $\mathbb{E}[\log(1 + (X_t - 1)\delta) | X_t > 1] < \epsilon_1$. If $\lambda^* = 0$, we have

$\lambda_t \xrightarrow{\mathrm{p}} 0$ as $t \to \infty$. Hence, there exists $N > 0$, such that for all $t > N$, $\mathbb{P}(\lambda_t > \delta) < \epsilon_1$. We write $x_- = \min\{x, 0\}$ and $x_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. It follows that

$$\begin{aligned}\mathbb{E}[\log(1-\lambda_t+\lambda_t X_t)_-] \geqslant \mathbb{E}[\log(1-\lambda_t)] &\geqslant \log(1-\gamma)\mathbb{P}(\delta < \lambda_t \leqslant \gamma) + \log(1-\delta)\mathbb{P}(0 \leqslant \lambda_t \leqslant \delta)\\ &\geqslant (\log(1-\gamma)-1)\epsilon_1,\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[\log(1-\lambda_t+\lambda_t X_t)_+]\\ &= \mathbb{E}[\log(1-\lambda_t+\lambda_t X_t)(\mathbb{1}_{\{X_t>1, 0\leqslant\lambda_t\leqslant\delta\}} + \mathbb{1}_{\{X_t>1,\delta<\lambda_t\leqslant\gamma\}})]\\ &\leqslant \mathbb{E}[\log(1+(X_t-1)\delta)|X_t>1]\mathbb{P}(0\leqslant\lambda_t\leqslant\delta) + \mathbb{E}[\log(1+(X_t-1)\gamma)|X_t>1]\mathbb{P}(\delta<\lambda_t\leqslant\gamma)\\ &\leqslant (1+\mathbb{E}[\log(1+(X_t-1)\gamma)|X_t>1])\epsilon_1.\end{aligned}$$

It is clear that $\sup_{t\in\mathbb{N}}\mathbb{E}[\log(1+(X_t-1)\gamma)|X_t>1]$ is bounded because $\sup_{t\in\mathbb{N}}\mathbb{E}[\log(X_t)] < \infty$. Hence, $\mathbb{E}[|\log(1+\lambda_t+\lambda_t X_t)|] < M_1\epsilon_1$ for some $M_1 > 0$ and for all $t > N$. Therefore, there exists $N_1 > 0$, such that for all $T > N_1$,

$$\begin{aligned}\mathbb{E}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\log(1-\lambda_t+\lambda_t X_t)\right|\right] &\leqslant \frac{1}{T}\sum_{t=1}^{N}\mathbb{E}[|\log(1-\lambda_t+\lambda_t X_t)|] + \frac{1}{T}\sum_{t=N+1}^{T}\mathbb{E}[|\log(1-\lambda_t+\lambda_t X_t)|]\\ &\leqslant \epsilon_1 + \frac{(T-N)M_1}{T}\epsilon_1.\end{aligned}$$

Hence (7.16) holds for $\lambda^* = 0$.

If $\lambda^* \in (0, \gamma]$, we write $K = \max\{1/\lambda^*, 1/(1-\lambda^*)\} < \infty$. Thus

$$\frac{X_t - 1}{1+(X_t-1)\lambda^*} \in [-K, K] \quad \text{for all } t \in \mathbb{N}.$$

For all $\epsilon_2 > 0$, it is clear by continuity that there exists $\delta > 0$, such that $\log(1-\delta/K) > -\epsilon_2$. Because $\lambda_t \xrightarrow{L^1} \lambda^*$ as $t \to \infty$, there exists $N > 0$, such that $\mathbb{P}(|\lambda_t - \lambda^*| > \delta) < \epsilon_2$ and $\mathbb{E}[|\lambda_t - \lambda^*|] < \epsilon_2$ for all $t > N$. It follows that for all $t > N$,

$$\begin{aligned}&\mathbb{E}\left[(\log(1-\lambda_t+\lambda_t X_t) - \log(1-\lambda^*+\lambda^* X_t))_-\right]\\ &= \mathbb{E}\left[\log\left(1+\frac{X_t-1}{1+(X_t-1)\lambda^*}(\lambda_t-\lambda^*)\right)_-\right]\\ &= \mathbb{E}\left[\log\left(1+\frac{X_t-1}{1+(X_t-1)\lambda^*}(\lambda_t-\lambda^*)\right)\left(\mathbb{1}_{\{X_t\geqslant 1, 0\leqslant\lambda_t\leqslant\lambda^*\}} + \mathbb{1}_{\{X_t<1,\lambda^*<\lambda_t\geqslant\gamma\}}\right)\right]\\ &\geqslant \mathbb{P}(|\lambda_t-\lambda^*|\leqslant\delta)\log(1-\delta/K) + \mathbb{P}(|\lambda_t-\lambda^*|>\delta)(\mathbb{E}[\log(1+(X_t-1)\lambda_t)|X_t\geqslant 1, 0\leqslant\lambda_t<\lambda^*-\delta]\\ &+ \mathbb{E}[\log(1+(X_t-1)\lambda_t)|X_t<1, \lambda^*+\delta\leqslant\lambda_t<\gamma] - \mathbb{E}[\log(1-\lambda^*+\lambda^* X_t)])\\ &\geqslant (\log(1-\gamma) - \mathbb{E}[\log(1-\lambda^*+\lambda^* X_t)] - 1)\epsilon_2,\end{aligned}$$

and

$$\mathbb{E}\left[(\log(1-\lambda_t+\lambda_t X_t)-\log(1-\lambda^*+\lambda^* X_t))_+\right]=\mathbb{E}\left[\log\left(1+\frac{X_t-1}{1+(X_t-1)\lambda^*}(\lambda_t-\lambda^*)\right)_+\right]$$
$$\leqslant K\mathbb{E}[|\lambda_t-\lambda^*|]<K\epsilon_2.$$

Because $\mathbb{E}[\log(1-\lambda^*+\lambda^*X_t)]$ is bounded, we have $E\left[|\log(1-\lambda_t+\lambda_tX_t)-\log(1-\lambda^*+\lambda^*X_t)|\right]<M_2\epsilon_2$ for some $M_2>0$ and for all $t>N$. Similar argument as the case of $\lambda^*=0$ leads to (7.16). □

*Proof of Proposition 7.4.* Let $e$ be a model-free e-statistic testing $\rho$. Write $\mathcal{M}_r(\rho)=\{F\in\mathcal{M}:\rho(F)\leqslant r\}$. Take $F,G\in\mathcal{M}_r(\rho)$ satisfying $\rho(F)=\rho(G)=r$. We have $\int e(x,r)\,\mathrm{d}F(x)\leqslant 1$ and $\int e(x,r)\,\mathrm{d}G(x)\leqslant 1$, and hence $\int e(x,r)\,\mathrm{d}(\lambda F+(1-\lambda)G)(x)\leqslant 1$. Thus, $\rho(\lambda F+(1-\lambda)G)\leqslant r$. Next, for any $F,G\in\mathcal{M}_r(\rho)$, without loss of generality we assume $q:=\rho(F)\geqslant\rho(G)$. Take $\bar{G}\in\mathcal{M}$ such that $\bar{G}\geqslant_1 G$ and $\rho(\bar{G})=q$. From the above analysis, we know that $\lambda F+(1-\lambda)\bar{G}\in\mathcal{M}_q(\rho)\subseteq\mathcal{M}_r(\rho)$. Since $\lambda F+(1-\lambda)G\leqslant_1\lambda F+(1-\lambda)\bar{G}$, we have $\lambda F+(1-\lambda)G\in\mathcal{M}_r(\rho)$. □

*Proof of Proposition 7.5.* Take $F,G\in\mathcal{M}$, $r\in\mathbb{R}$, $\lambda\in[0,1]$, and write $H_\lambda=\lambda F+(1-\lambda)G$. First, suppose that $\rho(F),\rho(G)\leqslant r$. Since $(x,r)\mapsto e(x,r)$ is decreasing in $r$, we have $\int e(x,r)\,\mathrm{d}F(x)\leqslant 1$ and $\int e(x,r)\,\mathrm{d}G(x)\leqslant 1$, and hence $\int e(x,r)\,\mathrm{d}H_\lambda(x)\leqslant 1$ for all $\lambda\in[0,1]$. This implies $\rho(H_\lambda)\leqslant r$. Further, suppose that $\rho(F),\rho(G)\geqslant r$. Assume that $\rho(H_\lambda)<r$. Write $q=\rho(H_\lambda)$. There exists $\epsilon>0$, such that $q+\epsilon<r$. Since $(x,r)\mapsto e(x,r)$ is decreasing in $r$, $\int e(x,q+\epsilon)\,\mathrm{d}H_\lambda(x)\leqslant 1$, $\int e(x,q+\epsilon)\,\mathrm{d}F(x)>1$, and $\int e(x,q+\epsilon)\,\mathrm{d}G(x)>1$. This leads to a contradiction. Therefore, $\rho(H_\lambda)\geqslant r$. Summarizing the above arguments, $\rho(F),\rho(G)\leqslant r$ implies $\rho(H_\lambda)\leqslant r$, and $\rho(F),\rho(G)\geqslant r$ implies $\rho(H_\lambda)\geqslant r$. This gives the quasi-linearity of $\rho$. □

*Proof of Proposition 7.6.* Let $e'(x,r)$ be a continuous non-conservative model-free e-statistic testing the mean. By Proposition 7.8 in Section 7.11, $1-e'(x,r)$ is an $\mathcal{P}$-identification function for the mean. The function $V(x,r)=x-r$ is a strict $\mathcal{P}$-identification function for the mean which satisfies Dimitriadis et al. (2023, Assumption (S.5)). By Dimitriadis et al. (2023, Theorem S.1), there is a function $\tilde{h}:(a,\infty)\to\mathbb{R}$ such that

$$1-e'(x,r)=\tilde{h}(r)(x-r),\quad x\geqslant a, r>a.$$

It is clear that $h(r)=\tilde{h}(r)(r-a)$ has to be continuous. The condition $0\leqslant h\leqslant 1$ arises since $e'(x,r)\geqslant 0$. Since $e'$ is testing the mean, it has to hold that $h>0$. The condition that $e'$ is strictly testing implies the remaining conditions. □

*Proof of Proposition 7.7.* Let $e'(x, r, z)$ be a continuous non-conservative model-free e-statistic for $(\mathrm{var}, \mathbb{E})$ testing var. By Proposition 7.8 in Section 7.11, $(x-z, 1-e'(x, r, z))^\top$ is an $\mathcal{M}_2$-identification function for $(\mathrm{var}, \mathbb{E})$. The function $V(x, r, z) = (x-z, r-(x-z)^2)^\top$ is a strict $\mathcal{M}_2$-identification function for $(\mathrm{var}, \mathbb{E})$ which satisfies Dimitriadis et al. (2023, Assumption (S.5)). By Dimitriadis et al. (2023, Theorem S.1), there are functions $h_1$ and $h_2$ such that

$$e'(x, r, z) = 1 + h_1(r, z)(z-x) + h_2(r, z)\frac{(z-x)^2 - r}{r}, \quad x, z \in \mathbb{R},\ r > 0. \tag{7.17}$$

Since $e'$ is continuous, one can see that $h_2$ has to be continuous by choosing $x = z$, but then also $h_1$ has to be continuous. Putting $t = z - x$ and considering the resulting quadratic function in $t$, one sees that $rh_1^2(r, z) \leqslant 4h_2(r, z)(1 - h_2(r, z))$ is necessary and sufficient for $e'(x, r, z) \geqslant 0$.

For all $r > 0$, $z \in \mathbb{R}$, $p \in (0, 1)$, and $\epsilon_1, \epsilon_2 > 0$, consider a random variable

$$X = \left(z + \epsilon_1 + \left(\frac{r(1-p)}{p} + \frac{\epsilon_2}{p^2}\right)^{1/2}\right)\mathbb{1}_A + \left(z + \epsilon_1 - \left(\frac{rp}{1-p} + \frac{\epsilon_2}{(1-p)^2}\right)^{1/2}\right)\mathbb{1}_{A^c},$$

where $A$ is a set with $\mathbb{P}(A) = p$ and $A^c$ denotes the complement of $A$. It follows that $\mathbb{E}[X] = z + \epsilon_1$ and $\mathrm{var}(X) = r + \epsilon_2/(p(1-p)) > r$. By (7.17) and since $e'$ is testing var, we have

$$\begin{aligned} 1 < \mathbb{E}[e'(X, r, z)] &= 1 + h_1(r, z)(z - \mathbb{E}[X]) + h_2(r, z)\frac{\mathbb{E}[(z-X)^2] - r}{r} \\ &= 1 - h_1(r, z)\epsilon_1 + h_2(r, z)\left(\frac{\epsilon_1^2}{r} + \frac{\epsilon_2}{rp(1-p)}\right). \end{aligned} \tag{7.18}$$

Arbitrariness of $\epsilon_1$ and $\epsilon_2$ yields that $h_1 \leqslant 0$. Similarly, for all $r > 0$, $z \in \mathbb{R}$, $p \in (0, 1)$, and $\epsilon_1, \epsilon_2 > 0$, consider the random variable

$$X = \left(z - \epsilon_1 + \left(\frac{r(1-p)}{p} + \frac{\epsilon_2}{p^2}\right)^{1/2}\right)\mathbb{1}_A + \left(z - \epsilon_1 - \left(\frac{rp}{1-p} + \frac{\epsilon_2}{(1-p)^2}\right)^{1/2}\right)\mathbb{1}_{A^c}.$$

It yields the condition

$$1 + h_1(r, z)\epsilon_1 + h_2(r, z)\left(\frac{\epsilon_1^2}{r} + \frac{\epsilon_2}{rp(1-p)}\right) > 1,$$

which implies that $h_1 \geqslant 0$. It follows that $h_1(r, z) = 0$ for all $r > 0$ and $z \in \mathbb{R}$. Thus $0 = rh_1^2(r, z) \leqslant 4h_2(r, z)\,(1 - h_2(r, z))$, and thus $0 \leqslant h_2 \leqslant 1$. Since $(\mathbb{E}[(z-X)^2] - r)/r > 0$ for all $X$ with finite variance, all $z \in \mathbb{R}$ and $r < \mathrm{var}(X)$, the first line of (7.18) yields that $h_2 > 0$. It is also clear that $h_1 = 0$ and $0 < h_2 \leqslant 1$ is sufficient for $e'$ to be a model-free e-statistic for $(\mathrm{var}, \mathbb{E})$ testing var. Substituting $h_2$ by $h$ yields the assertion. The final statement is directly obtained from the same argument as the proof of Proposition 7.6. □

*Proof of Theorem 7.4.* Let $e'(x, r)$ be an e-statistic satisfying the stated conditions. By Proposition 7.8 in Section 7.11, $1 - e'(x, r)$ is a $\mathcal{P}$-identification function for $\mathrm{VaR}_p$. The function $\mathbb{1}_{\{x \leqslant r\}} - p$ is a strict $\mathcal{P}$-identification function for $\mathrm{VaR}_p$ which satisfies Dimitriadis et al. (2023, Assumption (S.5)). By Dimitriadis et al. (2023, Theorem S.1), for any $F \in \mathcal{P}$ it has to hold that

$$\int e'(x, r)\,\mathrm{d}F(x) = \int 1 + h(r)\frac{p - \mathbb{1}_{\{x \leqslant r\}}}{1 - p}\,\mathrm{d}F(x).$$

Since $e'$ is assumed to be continuous except for points in a set of Lebesgue measure zero, it satisfies Dimitriadis et al. (2023, Assumption (S.7)). All Dirac measures can be approximated by distributions in $\mathcal{P}$ with compact support. Therefore, $\mathcal{P}$ satisfies Dimitriadis et al. (2023, Assumption (S.6)). Together this implies that the stated form of $e'$ holds for almost all $(x, r)$. Due to the continuity assumption, we obtain it for all $(x, r)$.

The condition $0 \leqslant h \leqslant 1$ ensures that $e'(x, r) \geqslant 0$. In order to obtain that $e'$ is testing, it is necessary and sufficient that $h > 0$. Fix $x \in \mathbb{R}$. For $r \geqslant x$, the function $r \mapsto e'(x, r)$ is decreasing if and only if $h$ is increasing; for $r < x$, the same function is decreasing if and only if $h$ is decreasing. Since these considerations hold for any $x \in \mathbb{R}$, $h$ has to be constant. $\square$

We first show the following auxiliary lemma for the proof of Theorem 7.5.

**Lemma 7.2.** *Let $p \in (0, 1)$ and $\mathcal{P}$ be the set of all distributions with finite mean with a quantile continuous at $p$. All $\mathcal{P}$-model-free e-statistics $e'$ for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ that are non-conservative and continuous except at points with $x = z$ are of the form*

$$e'(x, r, z) = 1 + h_1(r, z)\frac{\mathbb{1}_{\{x \leqslant z\}} - p}{1 - p} + h_2(r, z)\left(\frac{(x - z)_+}{(1 - p)(r - z)} - 1\right), \quad x \in \mathbb{R},\ z < r, \tag{7.19}$$

*where $h_1, h_2$ are continuous functions such that $h_2 \geqslant 0$ and*

$$-1 + h_2(r, z) \leqslant h_1(r, z) \leqslant \frac{1 - p}{p}\left(1 - h_2(r, z)\right), \quad z < r.$$

*Proof of Lemma 7.2.* Let $e'(x, r, z)$ be an e-statistic satisfying the stated conditions. By Proposition 7.8 in Section 7.11, $(\mathbb{1}_{\{x \leqslant z\}} - p, 1 - e'(x, r, z))^\top$ is a $\mathcal{P}$-identification function for $(\mathrm{ES}_p, \mathrm{VaR}_p)$, since the function $\mathbb{1}_{\{x \leqslant z\}} - p$ is a $\mathcal{P}$-identification function for $\mathrm{VaR}_p$. The strict $\mathcal{P}$-identification function $(\mathbb{1}_{\{x \leqslant z\}} - p, (x - z)_+ - (1 - p)(r - z))$ for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ satisfies Dimitriadis et al. (2023, Assumption (S.5)). With exactly the same arguments as in the proof of Theorem 7.4, we now obtain (7.19) first in an integrated version, and then pointwise for all $(x, z, r)$ with $z < r$. Considering separately the cases that $x \leqslant z$ and $x > z$, we find that necessary and sufficient conditions for $e'(x, r, z) \geqslant 0$ are $h_2 \geqslant 0$ and $-1 + h_2 \leqslant h_1 \leqslant ((1 - p)/p)\,(1 - h_2)$. $\square$

*Proof of Theorem 7.5.* Suppose that $(x, r, z) \mapsto e'(x, r, z)$ is a model-free e-statistic for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ testing $\mathrm{ES}_p$ that is non-conservative on $\mathcal{P}$ and continuous except at $x = z$. For all $r \in \mathbb{R}$, $z \leqslant r$, $\epsilon > 0$, and for some $q \in (p, 1]$, consider a random variable

$$X = ((r - z)(1 - p)/(1 - q) + \epsilon)\,\mathbb{1}_A + z,$$

where $\mathbb{P}(A) = 1 - q$. It follows that $\mathbb{E}[(X - z)_+] = (1 - p)(r - z) + \epsilon(1 - q)$ and $\mathrm{ES}_p(X) = r + \epsilon(1 - q)/(1 - p) > r$, since $\mathrm{VaR}_p(X) = z$. By Lemma 7.2, there exist continuous functions $h_1$ and $h_2$, such that

$$\begin{aligned} 1 < \mathbb{E}[e'(x, r, z)] &= 1 + h_1(r, z)\frac{\mathbb{P}(X \leqslant z) - p}{1 - p} + h_2(r, z)\left(\frac{\mathbb{E}[(X - z)_+]}{(1 - p)(r - z)} - 1\right) \\ &= 1 + h_1(r, z)\frac{q - p}{1 - p} + h_2(r, z)\left(\frac{(1 - q)\epsilon}{(1 - p)(r - z)}\right). \end{aligned} \tag{7.20}$$

Arbitrariness of $\epsilon$ implies that $h_1 \geqslant 0$ and $h_2 \geqslant 0$, with at least one of the two inequalities being strict. Similarly, for all $r \in \mathbb{R}$, $z \leqslant r$, $\epsilon > 0$, and for some $q \in [0, p)$, take

$$X = (r - z + \epsilon)\,\mathbb{1}_A + \epsilon\mathbb{1}_B + z,$$

where $\mathbb{P}(A) = 1 - p$ and $\mathbb{P}(B) = p - q$ and $A \cap B = \varnothing$. We have $\mathbb{E}[(X - z)_+] = (1 - p)(r - z) + \epsilon(1 - q)$ and $\mathrm{ES}_p(X) = r + \epsilon > r$. Thus, condition (7.20) also has to hold for $q \in [0, p)$, which implies that $h_1 \leqslant 0$ and $h_2 \geqslant 0$, with at least one of the two inequalities being strict. According to the previous arguments, we have $h_1 = 0$ and $h_2 > 0$. This implies that $(x, r, z) \mapsto e'(x, r, z)$ is also continuous at points with $x = z$. The inequality $-1 + h_2 \leqslant h_1 \leqslant ((1 - p)/p)\,(1 - h_2)$ in Lemma 7.2 yields $h_2 \leqslant 1$. Substituting $h_2$ by $h$ completes the proof. The final statement is obtained by the same argument as the proof of Proposition 7.6. □

## 7.13 Supplementary simulation and data analysis

### 7.13.1 Forecasting procedure for stationary time series data

This section describes the details of the forecasting procedure for VaR and ES in Section 7.7.1. We assume that the data generated above follow an AR(1)–GARCH(1, 1) process $\{L_t\}_{t\in\mathbb{N}}$ with $L_t = \mu_t + \sigma_t Z_t$, where $\{Z_t\}_{t\in\mathbb{N}}$ is assumed to be a sequence of iid innovations with mean 0 and variance 1, and $\{\mu_t\}_{t\in\mathbb{N}}$ and $\{\sigma_t\}_{t\in\mathbb{N}}$ are $\mathcal{F}_{t-1}$-measurable. Specifically, we have

$$\mu_t = c + \psi X_{t-1} \quad \text{and} \quad \sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \quad t \in \mathbb{N}. \tag{7.21}$$

We first assume the innovations $Z_t$ to follow a normal, t-, or skewed-t distribution, and estimate $\{\hat{\mu}_t\}_{t\in\mathbb{N}}$ and $\{\hat{\sigma}_t^2\}_{t\in\mathbb{N}}$ through obtaining the maximum likelihood estimators of $(c, \psi, \alpha_0, \alpha_1, \beta)$ under the assumption on the distribution of $Z_t$. For t- and skewed-t distributions, parameters are estimated by the maximum likelihood method via the standardized residuals $\{(L_t - \hat{\mu}_t)/\hat{\sigma}_t\}_{t\in\mathbb{N}}$. For a risk measure $\rho$ ($\rho = \mathrm{VaR}_p$ or $\rho = \mathrm{ES}_p$), the value of $\rho(Z_t)$ can be calculated explicitly for the assumed parametric models (see e.g., McNeil et al., 2015; Nolde and Ziegel, 2017; Patton et al., 2019). For estimation, the estimated parameters are plugged into these formulas resulting in estimates $\widehat{\rho(Z_t)}$. The final risk predictions are then $\hat{\mu}_t + \hat{\sigma}_t\widehat{\rho(Z_t)}$, where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are computed from (7.21) with the estimated parameters. Table 7.8 shows the average of the forecasts of VaR and ES at different levels over all $1,000$ trials and all trading days, where the last line shows the average forecasts of VaR and ES using the true information of the data generating process.

| | $\overline{\mathrm{VaR}}_{0.95}$ | $\overline{\mathrm{VaR}}_{0.99}$ | $\overline{\mathrm{VaR}}_{0.875}$ | $\overline{\mathrm{ES}}_{0.875}$ | $\overline{\mathrm{VaR}}_{0.975}$ | $\overline{\mathrm{ES}}_{0.975}$ |
|---|---|---|---|---|---|---|
| normal | 0.619 | **0.906** | 0.411 | **0.620** | **0.752** | **0.910** |
| t | **0.534** | **0.999** | **0.300** | **0.576** | **0.722** | **1.065** |
| skewed-t | 0.676 | 1.281 | 0.369 | 0.727 | 0.922 | 1.358 |
| true | 0.674 | 1.271 | 0.368 | 0.723 | 0.918 | 1.343 |

Table 7.8: Average point forecasts of VaR and ES at different levels over $1,000$ simulations of time series and 500 trading days; values in boldface are underestimated by at least 10% compared with values in the last line

### 7.13.2 Comparing GREE and GREL methods for stationary time series

This section serves as a supplement to Section 7.7.1 by demonstrating the results of backtesting VaR and ES using the GREE and GREL methods through Taylor approximation in (7.15). Meanwhile, we compare the performance of the GREE and GREL methods. The results of VaR are shown in Tables 7.9 and 7.10 and those for ES are shown in Tables 7.11 and 7.12. The GREL method is better than the GREE method in terms of percentage of detections in all cases of VaR and ES. This is consistent with the result in Theorem 7.3 because for the time series data, the losses used by the GREL method are relatively closer to an iid pattern compared to the whole e-statistics used by the GREE method. This is also not a contradiction to the slightly longer expected time to detection conditional on the detection of the GREL method, noting that the GREL method detects more often.

| GREE | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | normal | | | t | | | skewed-t | | |
| threshold | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| $-10\%$ | 99.0 | 96.5 | 92.1 | 95.5 | 84.2 | 72.4 | 38.3 | 13.7 | 5.6 |
| exact | 95.0 | 84.4 | 72.0 | 79.5 | 55.4 | 37.2 | 14.0 | 2.8 | 0.8 |
| $+10\%$ | 80.5 | 56.8 | 38.2 | 52.3 | 22.0 | 9.8 | 5.3 | 0.6 | 0 |
| GREL | | | | | | | | | |
| | normal | | | t | | | skewed-t | | |
| threshold | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| $-10\%$ | 99.7 | 98.3 | 94.5 | 98.2 | 88.7 | 76.7 | 51.0 | 15.1 | 6.6 |
| exact | 97.8 | 88.6 | 75.9 | 87.4 | 62.4 | 39.8 | 24.3 | 3.2 | 0.6 |
| $+10\%$ | 87.7 | 65.0 | 43.9 | 67.5 | 28.9 | 13.6 | 10.8 | 0.4 | 0.1 |

Table 7.9: Percentage of detections (%) for $\mathrm{VaR}_{0.99}$ forecasts over $1,000$ simulations of time series and 500 trading days

| | GREE | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | normal | | | | t | | | | skewed-t | | | |
| threshold | 2 | 5 | 10 | | 2 | 5 | 10 | | 2 | 5 | 10 | |
| −10% | 123 | 186 | 228 | (5.475) | 159 | 236 | 278 | (3.327) | 206 | 260 | 300 | (0.2856) |
| exact | 164 | 239 | 283 | (3.236) | 197 | 272 | 311 | (1.638) | 189 | 229 | 265 | (−0.1012) |
| +10% | 197 | 268 | 300 | (1.734) | 217 | 280 | 318 | (0.5933) | 158 | 224 | – | (−0.1706) |

| | GREL | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | normal | | | | t | | | | skewed-t | | | |
| threshold | 2 | 5 | 10 | | 2 | 5 | 10 | | 2 | 5 | 10 | |
| −10% | 116 | 185 | 233 | (5.338) | 158 | 241 | 293 | (3.290) | 239 | 281 | 330 | (0.3492) |
| exact | 160 | 241 | 295 | (3.240) | 196 | 286 | 332 | (1.736) | 233 | 238 | 289 | (−0.1463) |
| +10% | 189 | 284 | 330 | (1.849) | 226 | 304 | 358 | (0.7599) | 230 | 211 | 377 | (−0.3472) |

Table 7.10: The average number of days taken to detect evidence against $\mathrm{VaR}_{0.99}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days; numbers in brackets are average final log-transformed e-values

| | GREE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | normal | | | t | | | skewed-t | | |
| threshold | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| $-10\%$ ES | 99.7 | 98.9 | 97.9 | 96.7 | 85.6 | 73.8 | 39.9 | 14.9 | 6.1 |
| $-10\%$ both | 99.8 | 98.9 | 97.3 | 97.9 | 89.2 | 79.3 | 41.1 | 13.6 | 6.4 |
| exact | 98.6 | 93.2 | 86.1 | 83.0 | 59.8 | 41.1 | 13.5 | 3.2 | 1.1 |
| $+10\%$ both | 91.1 | 75.7 | 59.8 | 54.2 | 22.3 | 10.2 | 5.1 | 0.8 | 0 |
| $+10\%$ ES | 91.0 | 76.1 | 60.2 | 60.6 | 28.3 | 13.8 | 6.0 | 1.1 | 0.1 |

| | GREL | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | normal | | | t | | | skewed-t | | |
| threshold | 2 | 5 | 10 | 2 | 5 | 10 | 2 | 5 | 10 |
| $-10\%$ ES | 99.9 | 99.2 | 98.2 | 97.5 | 86.2 | 74.1 | 49.3 | 16.6 | 6.5 |
| $-10\%$ both | 99.9 | 99.0 | 97.6 | 98.1 | 89.2 | 78.8 | 49.9 | 17.4 | 6.6 |
| exact | 99.2 | 94.6 | 85.7 | 87.6 | 62.4 | 41.9 | 25.3 | 5.6 | 0.9 |
| $+10\%$ both | 94.1 | 77.3 | 56.2 | 66.0 | 31.6 | 12.5 | 11.5 | 1.5 | 0.1 |
| $+10\%$ ES | 94.4 | 76.7 | 57.4 | 72.7 | 36.1 | 16.9 | 13.1 | 1.7 | 0.2 |

Table 7.11: Percentage of detections (%) for $\mathrm{ES}_{0.975}$ forecasts over $1,000$ simulations of time series and 500 trading days

| | GREE | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | normal | | | | t | | | | skewed-t | | | |
| threshold | 2 | 5 | 10 | | 2 | 5 | 10 | | 2 | 5 | 10 | |
| $-10\%$ ES | 95 | 141 | 181 | (6.596) | 158 | 226 | 274 | (3.365) | 213 | 250 | 234 | (0.3991) |
| $-10\%$ both | 95 | 152 | 194 | (6.341) | 149 | 220 | 269 | (3.674) | 207 | 242 | 224 | (0.3594) |
| exact | 139 | 201 | 250 | (4.278) | 193 | 265 | 307 | (1.822) | 208 | 233 | 220 | $(-0.1011)$ |
| $+10\%$ both | 177 | 249 | 292 | (2.669) | 219 | 267 | 286 | (0.6485) | 106 | 109 | – | $(-0.1843)$ |
| $+10\%$ ES | 174 | 248 | 295 | (2.625) | 210 | 266 | 288 | (0.8098) | 90 | 96 | 156 | $(-0.2005)$ |

| | GREL | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | normal | | | | t | | | | skewed-t | | | |
| threshold | 2 | 5 | 10 | | 2 | 5 | 10 | | 2 | 5 | 10 | |
| $-10\%$ ES | 97 | 147 | 189 | (6.344) | 155 | 231 | 282 | (3.238) | 235 | 264 | 277 | (0.4990) |
| $-10\%$ both | 99 | 154 | 201 | (5.963) | 146 | 221 | 271 | (3.511) | 223 | 263 | 278 | (0.4565) |
| exact | 134 | 209 | 258 | (4.027) | 191 | 266 | 318 | (1.892) | 208 | 233 | 220 | $(-0.09266)$ |
| $+10\%$ both | 174 | 257 | 291 | (2.577) | 217 | 289 | 298 | (0.8661) | 186 | 207 | 70 | $(-0.3171)$ |
| $+10\%$ ES | 173 | 254 | 296 | (2.557) | 215 | 282 | 301 | (1.007) | 189 | 185 | 271 | $(-0.3653)$ |

Table 7.12: The average number of days taken to detect evidence against $\mathrm{ES}_{0.975}$ forecasts conditional on detection over $1,000$ simulations of time series and 500 trading days; "–" represents no detection; numbers in brackets are average final log-transformed e-values

### 7.13.3 Forecasting procedure for time series with structural change

This section provides details for the forecasting procedure for time series data with structural change in Section 7.7.2. After a burn-in period of length $1,000$, 500 data points are simulated, within which 250 presampled data $L_1, \dots, L_{250}$ are for forecasting risk measures and the rest 250 data $L_{251}, \dots, L_{500}$ are for backtesting. The forecaster obtains the estimates $\hat{\boldsymbol{\theta}}$ of the model parameters $\boldsymbol{\theta} = (\omega, \alpha, \beta_t)$ once using the standard Gaussian QML for the presampled 250 losses. For $t \in \{251, \dots, 500\}$, the forecasts of $\mathrm{VaR}_{0.95}(L_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_{0.95}(L_t|\mathcal{F}_{t-1})$ are obtained by $z_t = \sigma_t(\hat{\boldsymbol{\theta}})\widehat{\mathrm{VaR}}_{0.95}$ and $r_t = \sigma_t(\hat{\boldsymbol{\theta}})\widehat{\mathrm{ES}}_{0.95}$, respectively, where $\widehat{\mathrm{VaR}}_{0.95}$ and $\widehat{\mathrm{ES}}_{0.95}$ are empirical forecasts of VaR and ES using presampled residuals $L_1/\sigma_1(\hat{\boldsymbol{\theta}}), \dots, L_{250}/\sigma_{250}(\hat{\boldsymbol{\theta}})$. We choose the size

$m = 50$ of the rolling window and a leg $d = 5$ of autocorrelations.[10]

#### 7.13.4 Detailed setup of data analysis for optimized portfolios

For the data of optimized portfolios we use in Section 7.8.2, the first 500 data points are used for the initial forecast and another 500 data points are for computing the first value of the betting process of the backtesting procedure. The final sample for backtesting contains 4280 negated percentage log-returns from Jan 3, 2005 to Dec 31, 2021. The selected stocks are those with the largest market caps in the 11 S&P 500 sectors divided by the GICS level 1 index as of Jan 3, 2005. The list of selected stocks is shown in Table 7.13.

| **Communication Services** | **Customer Discretionary** | **Consumer Staples** |
|---|---|---|
| Verizon Communications Inc. | Time Warner Inc. | The Procter & Gamble Co. |
| AT&T Inc. | The Home Depot, Inc. | Walmart Inc. |
| **Energy** | **Financials** | **Health Care** |
| Exxon Mobil Corp. | Citigroup Inc. | Johnson & Johnson |
| Chevron Corp. | Bank of America Corp. | Pfizer Inc. |
| **Industrials** | **Information Technology** | **Materials** |
| United Parcel Service Inc. | International Business Machines Corp. | EI du Pont de Nemours and Co. |
| General Electric Co. | Microsoft Corp. | The Dow Chemical Co. |
| **Real Estate** | **Utilities** | |
| Weyerhaeuser Co. | Exelon Corp. | |
| Simon Property Group Inc. | The Southern Co. | |

Table 7.13: 22 selected stocks in S&P 500 sectors divided by GICS level 1 as of Jan 3, 2005 for the portfolio

For forecasting, we assume $L_t^i = \mu_t^i + \sigma_t^i Z_t^i$ with $\mu_t^i$ and $\sigma_t^i$ defined to be the same as (7.21) for all $i \in [n]$. The innovations $\{Z_t^i\}_{t\in\mathbb{N}}$ are iid with respect to time with mean 0 and variance 1 for $i \in [n]$, assumed to be normal, t-, or skewed-t distributed. If a stock delists from the S&P 500 during the period from Jan 3, 2005 to Dec 31, 2021, it is removed from the portfolio as soon as it delists with all of its weight redistributed to the other stocks in the portfolio. The bank reports the VaR and the ES of the weighted portfolio by assuming $\mathbf{w}_t^\top \mathbf{X}_t$ to be normal, t-, or skewed-t distributed, respectively,

[10]The leg $d = 5$ is not necessary for the Monte Carlo simulations detector. We choose it to be consistent with the simulation setting of Hoga and Demetrescu (2022) for comparison.

with the mean $\sum_{i=1}^{n} w_t^i \mu_t^i$ and the variance $(w_t^1\sigma_t^1, \ldots, w_t^n\sigma_t^n)^\top \Sigma_t (w_t^1\sigma_t^1, \ldots, w_t^n\sigma_t^n)$, where $\Sigma_t$ is the covariance matrix of $(Z_t^1, \ldots, Z_t^n)$. The assumption is true when the innovations follow the normal distribution or follow the t-distribution with the same degree of freedom. We use this assumption to approximate the true distribution of the portfolio for t-distributions of different degrees of freedom and skewed-t distributions. The parameters of the t- and skewed-t distributions of the weighted portfolio are estimated by the maximum likelihood method assuming the negated percentage log-return of the portfolio to be the AR(1)-GARCH(1, 1) process with innovations belonging to the same class of distribution. Figure 7.8 shows the negated log-returns of the portfolio and the forecasts of $\mathrm{ES}_{0.975}$ over time assuming different innovation distributions.

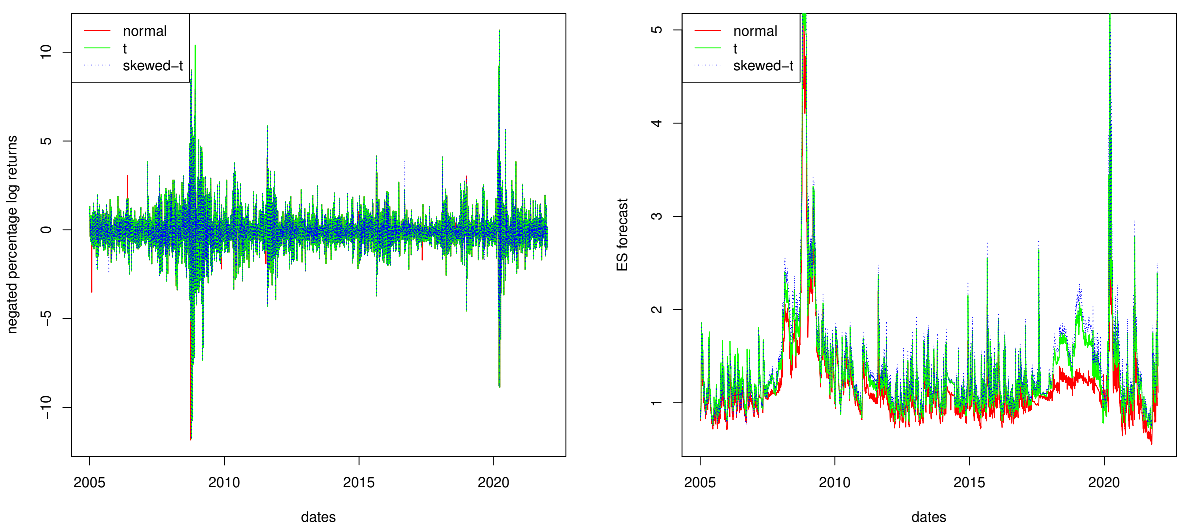


Figure 7.8: Portfolio data fitted by different distribution from Jan 3, 2005 to Dec 31, 2021; left panel: negated percentage log-returns, right panel: $\mathrm{ES}_{0.975}$ forecasts

# Chapter 8

# Concluding remarks and future work

## 8.1 Concluding remarks

Most of the important topics on riskmetrics are within the categories of characterization, optimization, and statistical approaches. My Ph.D. thesis research aims to explore various topics on riskmetrics to get myself well prepared for potential future research. The following are several summaries and remarks for all major chapters of this thesis.

Chapter 2 serves as a theoretical foundation of my thesis research. The class of riskmetrics we study, distortion riskmetrics, is general enough to include several common risk measures and deviation measures. Besides characterization, we study and summarize properties of distortion riskmetrics including finiteness, convexity, and continuity. Most of the results are not ground breaking, but are good as a starting point to study other interesting applications and deep theoretical extensions of distortion riskmetrics. Several followup questions on distortion riskmetrics have already been done, including risk sharing problems and robust optimization. In the future, it will be interesting to study distortion riskmetrics in other contexts, such as optimal insurance. Some specific classes of distortion riskmetrics are also worth exploring, such as the difference between two distortion risk measures, which measures the disagreement between two couterparties. Distortion riskmetrics are also not general enough to include all common risk functionals. We can also consider extending this class to a setting based on RDEU.

Chapter 3 is a natural extension of our study on distortion riskmetrics in Chapter 2 to distributionally robust optimization, a hot topic in operations research. The core message of this chapter is that we are able to find a unifying result leading to the equivalence between the original problem to its convex counterpart. Taking the advantage of distortion riskmetrics as a large class of objective functional, we can solve many practical optimization problems in finance and operations research. However, the theory is not complete yet in the sense that we only obtain a sufficient condition for the equivalence to hold. It is important to find a necessary one on the uncertainty set, so that we can obtain a complete characterization of the equivalence result. The sufficient property we proposed, closedness under concentration, does not cover all common practical cases where we can freely convert from the original problem to its convex counterpart, some optimal insurance problems as examples. Future research can also focus on obtaining more general sufficient conditions holding as least under common setups such as moment constraints and Wasserstein balls.

Chapters 4-6 are all characterization of various riskmetrics. Chapter 4 proposes the notions of Bayes pair and Bayes risk measure under the context of statistical elicitation. We characterize ES as the only coherent Bayes risk measure and entropic risk measure as the only elicitable Bayes risk measure. Chapter 5 characterizes several classes of riskmetrics in the context of optimal insurance design problems. In particular, we find the only objective the insured and the insurer use must be an ES/$\mathbb{E}$-mixture when the Pareto optimal set of contracts is with a deductible. Chapter 6 studies characterization of cash-subadditive risk measures. We find a general cash-subadditive risk measure can be represented by the lower envelope of a family of cash-subadditive and quasi-convex risk measures. We propose a new property called quasi-star-shapedness and find that our characterization result also holds true with quasi-star-shapedness and normalization. A characterization problem is a good starting point to study further extensions of riskmetrics. For instance, it is attracting to study characterization of riskmetrics under optimal insurance for a fixed distribution rather than a class of random losses, since practical problems of insurance design are usually under the situation where a loss distribution is given. This will make the problem and result much more practical. We can use the result to further deduce insured's risk attitude according to their optimal decision, which is very helpful for insurance companies. Moreover, it is surprising that the property of quasi-star-shapedness has sound economic interpretations of ambiguity aversion. Future research can focus on possible links between this property and deep economic problems.

Chapter 7 studies a new topic on riskmetrics about risk forecast and backtest, which has long been great interest and challenge to financial regulators and researchers. We propose a model-free and non-asymptotic backtesting method that works for various riskmetrics including VaR and ES. Through our study, we create an interesting link between backtesting riskmetrics and e-values, a recently developed powerful tool as an alternative to p-values. Our characterization of model-free e-statistics and optimality results of e-backtesting methods provide useful guidelines for practitioners to deal with real financial regulation situations. Given the importance of backtest and the promising potential extensions of e-values, the direction of this chapter is worth exploring and focusing on in the future. Our approach can be naturally extended to backtesting other riskmetrics beyond VaR and ES, provided that proper model-free e-statistics are obtained for specific riskmetrics. Considering the regulator and the financial institution as two players, we can study a game-theoretic framework of real backtesting problems. Moreover, as one of the limitations of e-values is its lack of power due to its model-free nature, we can improve the current methods by incorporating partial information of the underlying distribution. Some study can be done on this direction and comparisons can be made between e-value tests and p-value tests in terms of the power.

We discuss some specific future directions in the following section. All the notes are for possible future references.

## 8.2 Future work and open questions

### 8.2.1 On characterization

1. Chapter 2 has discussed about the characterization of distortion riskmetrics. The class of dual utilities or distortion risk measures have also been studied by earlier literature. We refer to Yaari (1987) and Schmeidler (1989) in economics and Denneberg (1994) and Wang et al. (1997) in actuarial science. However, it is still an open question about the characterization of rank-dependent expected utility (RDEU) in Quiggin (1982), defined under probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as
$$\int_\Omega \ell(X)\,\mathrm{d}T\circ\mathbb{P}, \quad X\in\mathcal{X},$$
where $\ell:\mathbb{R}\to\mathbb{R}$ is a strictly increasing and convex loss function (positive random variables represent losses), and $T:[0,1]\to[0,1]$ is a probability distortion function. It may be either RDEU is not natural, or it is still difficult to find a characterization. If the former is true, then why it is unnatural but so popular?

2. Motivated from Chapter 2, it might be interesting to consider the class of functionals which are differences of distortion risk measures, defined as
$$\rho_h(X)=\rho_{h_1}(X)-\rho_{h_2}(X), \quad X\in\mathcal{X}, \tag{8.1}$$

where $\rho_{h_1}, \rho_{h_2}$ are defined by (1.3), $h_1$ and $h_2$ are increasing and normalized, and $h = h_1 - h_2$. This class of functionals measure the disagreement between two risk attitudes. By the main result of Chapter 3, we are able to compute the upper and lower bounds of this measure of disagreement in (8.1). This class of distortion riskmetrics has also appeared in Lauzier et al. (2023), where an optimal allocation for risk sharing problems of distortion riskmetrics exisits when $h_i(1) = 0$ for all $i = 1, \dots, n$. But this condition does not guarantee the equivalence between sum optimality and Pareto optimality unless the objectives are inter-quantile-ranges (IQDs). Above work suggests that the functional (8.1) has some special roles in different contexts. We can explore different characterizations of (8.1) in different problems. It is also interesting to find other applications of this class of functionals to practice?

3. One limitation of our main result in Chapter 5 is that we have only characterized ES as an objective functional so that deductible insurance contracts are optimal for all distributions of random losses. This makes the result unrealistic although it provides interesting mathematical insights, because the insured and the insurer usually consider a fixed distribution of the insurance loss for a specific contract. In this sense, we are motivated to explore if similar results to that in Chapter 5 can be obtained for any fixed distribution of the underlying random loss. Namely, we have the following conjecture.

   **Conjecture 8.1.** *Suppose that $\rho$ and $\psi$ are law-invariant convex risk measures with $\rho(0) = \psi(0) = 0$. For any fixed $X \in \mathcal{X}_+$, the following statements hold true:*

   *(i)* $\mathcal{I}^X_{\rho,\psi} = \mathcal{I}_2$ *if and only if* $\rho(X) = \psi(X)$ *and* $\rho$ *and* $\psi$ *are convex distortion risk measures;*

   *(ii)* $\mathcal{I}^X_{\rho,\psi} = \mathcal{I}_0$ *if and only if* $\rho(X) = \psi(X) = \mathbb{E}[X]$.

   *(iii) For any fixed* $d \geqslant 0$ *and* $p \in [0, 1)$, *we have* $\mathcal{I}^X_{\rho,\psi} \supseteq \mathcal{I}^d_1$ *if and only if* $\rho(X) = \psi(X) = \mathrm{ES}^\lambda_p(X)$ *for some* $\lambda \geqslant 0$.

   The notations in Conjecture 8.1 follow from Chapter 5. The enhancement of the model seems slight, but it involves nontrivial theoretical analysis. More importantly, characterization results for any fixed distribution will provide much deeper economic insights in terms of evaluating the risk attitudes of decision makers given their optimal choices of insurance contracts. This further links our work to the inverse optimization problem in operations research; see e.g., Bertsimas et al. (2012) and Li (2021) for inverse optimization problems involving risk measures. Several other extensions are subject to further investigation. First, it is unknown whether Conjecture 8.1 is true when we change the inclusion $\mathcal{I}^X_{\rho,\psi} \supseteq \mathcal{I}^d_1$ to equation $\mathcal{I}^X_{\rho,\psi} = \mathcal{I}^d_1$. This will complete the theory with a concrete characterization. Second, in practical situation, the risk attitudes of the insured and the insurer are not necessarily the same. Hence, it will be interesting to consider the case where the two risk measures $\rho$ and $\psi$ are different. Third, our study does not consider coinsurance contracts, which are also very common in insurance practice. It is promising to find the potential link between coinsurance contracts and some specific forms of riskmetrics.

4. I am still not so clear about the following direction but would be interested to write it down for future deeper thoughts and explorations. As discussed in Chapter 6, comonotonic allocations turn out to be optimal in many cases with law-invariant convex or consistent risk measures. The result reflects that comonotonicity is closely related to risk aversion. However, in some (or most) other cases where the objective functionals are not convex (see e.g., Embrechts et al., 2018, for the case of quantile-based risk sharing), optimal allocations are negatively dependent. One natural question is whether negative dependence corresponds to risk seeking preferences. In fact, negative dependence is also very common in the real life (such as gamblings and lotteries). What risk attitude does negative dependence correspond to? Can we characterize some measures of preferences (such as RDEU) given that negative dependence is optimal?

### 8.2.2 On optimization

1. Beyond the framework of Chapter 3, it is also possible that a nonconvex optimization problem is equivalent to its convex counterpart if the uncertainty set is not closed under concentration. In this sense, we are motivated to explore DRO problems where closedness under concentration does not necessarily hold. For example, it has been shown in Chapter 3 that the Wasserstein uncertainty set is only closed under concentration under some special conditions. We can study more general cases where such conditions are loosed and derive the worst- and best-case distortion riskmetrics to see if they are still equal to the bounds for its convex counterpart. This will help weaken the sufficient condition of closedness under concentration for the equivalence result to hold, and therefore leads to more practical applications in more general scenarios. Moreover, we can extend the current framework to an optimal insurance problem with distributional uncertainty. In general, we aim to solve the following problem:

$$\min_{f\in\mathcal{I}} \sup_{F_X\in\mathcal{M}} \rho_h(X - f(X) + \pi(f(X))). \tag{8.2}$$

Liu and Mao (2022) discussed a special case of the problem (8.2) withdeductible contract, expected premium, moment constraint, and VaR/ES as the objective functionals. Although the uncertainty set for the random variable $X - f(X) + \pi(f(X))$ is not closed under concentration, it still holds true that the worst-case VaR and ES are the same under the setup of Liu and Mao (2022). It will be of great interest to study the more general version (8.2) and obtain a sufficient (or necessary) condition for the equivalence

$$\sup_{F_X \in \mathcal{M}} \rho_h(X - f(X) + \pi(f(X))) = \sup_{F_X \in \mathcal{M}} \rho_{h^*}(X - f(X) + \pi(f(X)))$$

to hold.

2. Following Chapter 3, it remains to be an open question that what is the necessary condition on the uncertainty set for the equivalence result between the nonconvex optimization problem and its convex counterpart to hold. Chapter 3 only demonstrates that if such an equivalence holds, the set of the optimizers should be closed under concentration. However, a necessary condition on the uncertainty set itself is still unknown. Such a necessary condition, if found, will demonstrate a complete characterization of the equivalence result between the original problem and its convex counterpart, and thus deepen our understanding of similar DRO problems in decision theory, finance, game theory, and operations research.

3. Optimal insurance under belief heterogeneity is a popular topic recently. Chi (2019) studied the following problem:

$$\max_{I \in \mathcal{C}} \mathbb{E}^{\mathbb{P}}[U(W - X + I(X) - (1+\rho)\mathbb{E}^{\mathrm{Q}}[I(X)])], \tag{8.3}$$

where $U$ is an increasing concave utility function and $\rho \geqslant 0$ is a safety loading coefficient. Chi (2019) found that the optimal solution of (8.3) is still of a deductible form when belief heterogeneity satisfies the so-called monotone hazard rate (MHR) condition, which is expressed as

$$\mathrm{Hr}(t) = \frac{\mathrm{Q}(X > t)}{\mathbb{P}(X > t)} \text{ is decreasing over } [0, \max\{M_{\mathbb{P}}(X), M_{\mathrm{Q}}(X)\}),$$

where $M_{\mathbb{P}}(X)$ and $M_{\mathrm{Q}}(X)$ represent the essential supremums of $X$ under probability measures $\mathbb{P}$ and Q respectively. However, it seems that the MHR condition is not necessary for the optimal contract to be of a deductible form. My numerical studies indicate that the optimal indemnity is still close to a deductible form under some special cases where MHR is not satisfied. For example, it turns out that we still get deductible insurance as the optimal contract when we take $\mathbb{P}$ and Q as shown in the following example:

**Example 8.1** (MHR condition is not satisfied)**.** One natural example of belief heterogeneity violating the MHR condition is where $X \overset{\mathbb{Q}}{\sim} \mathrm{U}[0,5]$ and $X \overset{\mathbb{P}}{\sim} F$ such that the probability density function $f$ of $F$ is given by

$$f(t) = \begin{cases} 0.3, & 0 < t < 1, \\ 0.25, & 1 < t < 2, \\ 0.2, & 2 < t < 3, \\ 0.15, & 3 < t < 4, \\ 0.1, & 4 < t < 5. \end{cases}$$

In this case, the insurer believes the random loss is uniformly distributed from 0 to 5, while the insured believes there would be lower probability to have large losses compared with small losses. The insured and insurer's beliefs on risks, as described above, are also commonly seen in real cases. However, the MHR condition is no longer satisfied since the ratio $\mathbb{Q}(X > t)/\mathbb{P}(X > t)$ is not deceasing with respect to $t$ over $[0, \max\{M_{\mathbb{P}}(X), M_{\mathbb{Q}}(X)\})$.

In light of the numerical findings, it is worth further exploration to find a more general condition than MHR at least for some special classes of loss distributions. Also, the optimization problem under belief heterogeneity beyond expected utility (EU) framework may also be interesting to investigate.

### 8.2.3 On elicitability and backtesting

1. It is natural to extend the backtesting method described in Chapter 7 to other riskmetrics beyond VaR and ES. Gini deviation (GD) and Gini coefficient (GC) are important measures of statistical dispersion widely applied in statistics and economics. The Gini deviation $\mathrm{GD} : L^1 \to \mathbb{R}$ is defined as

$$\mathrm{GD}(X) = \frac{1}{2}\mathbb{E}[|X - X'|],$$

where $X'$ is and iid copy of $X$. Alternatively, we can write GD as a function from $\mathcal{M}^1 \to \mathbb{R}$ as a signed Choquet integral

$$\mathrm{GD}(F) = \int_0^1 F^{-1}(t)(2t-1)\,\mathrm{d}t = \int_{\mathbb{R}} F(t)(1 - F(x))\,\mathrm{d}x;$$

see e.g., Wang et al. (2020, Example 1). The Gini coefficient $\mathrm{GC} : L^1_+ \to [0,1]$ is defined as

$$\mathrm{GC}(X) = \frac{\mathrm{GD}(X)}{\mathbb{E}[X]} = \frac{\int_0^\infty F(t)(1 - F(x))\,\mathrm{d}x}{\int_0^\infty (1 - F(x))\,\mathrm{d}x}.$$

It remains to be an open question about the elicitability and elicitation complexity of GD and GC. This creates a barrier in evaluating the estimation performance of GD and GC in practice. We are motivated to studying the elicitation complexity of GD and GC, and it is conjectured that the answer to this question may be infinity. Furthermore, we are interested in exploring more about hypothesis tests for the forecasts of GD and GC using e-values. E-testing GD and GC has promising applications in statistical and economic practice. Possible forms of the model-free e-statistics for GD have been derived. The main idea is to use a pair of two data points to form a model-free e-statistic for GD, which is of the following form

$$e(r, x_1, x_2) = \frac{|x_1 - x_2|}{2r}.$$

This e-statistic requires the input of iid copies of the data we have. This restricts the tractability of our method. An e-testing method for GD or GC without using iid copies is still not known.

2. The characterization results of model-free e-statistics in Chapter 7 are all based on the setup of one single time period. Since backtesting risk measures is usually done in a dynamic framework, it is natural to consider extensions of the current characterization results to a multi-dimensional setup in multiple time periods. Based on the results in Vovk and Wang (2022), we have proved that the product form of e-processes dominates other merging functions of e-values for general risk measures, expressed in the following conjecture:

**Conjecture 8.2.** *For all sequential e-variables $E_1, \dots, E_T$, let $S_t = F(E_1, \dots, E_t, 1, \dots, 1)$, $t \in [T]$, and $S_0 = 1$, where $F : [0, \infty]^T \to [0, \infty]$. If $(S_t)_{t\in[T]}$ is an e-process, then for $t \in [T]$, there exist $\lambda_s$ taking values in $[0, 1]$ that are functions of $(E_1, \dots, E_{s-1})$ for $s \in [t]$, such that*

$$S_t \leqslant \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s E_s).$$

*Proof.* The proof follows directly from Lemma 3 and Theorem 2 of Vovk and Wang (2022). $\square$

We have also obtained a characterization result for the form of multi-time-period e-processes backtesting ES; see the following conjecture:

**Conjecture 8.3.** *Let $F : [0, \infty]^{3T} \to [0, \infty]$ satisfy the following properties:*

*(i) $F$ is continuous.*

*(ii) $(r_1, \dots, r_T) \mapsto F(x_1, r_1, z_1, \dots, X_T, r_T, z_T)$ is strictly decreasing.*

*(iii) For all $(X_t)_{t\in[T]}$ adapted to $(\mathcal{F}_t)_{t\in[T]}$ such that $F_{X_1}, \dots F_{X_T} \in \mathcal{M}'$ and $t \in [T]$, if $r_t = \mathrm{ES}_p(X_t|\mathcal{F}_{t-1})$ and $z_t = \mathrm{VaR}_p(X_t|\mathcal{F}_{t-1})$, then $\mathbb{E}[S_t/S_{t-1}|\mathcal{F}_t] = 1$, where $S_t = F(X_1, r_1, z_1 \dots, X_t, r_t, z_t, 1, \dots, 1)$ for $t \in [T]$ and $S_0 = 1$.*

*For $t \in [T]$, $x_s \in \mathbb{R}$, $z_s < r_s$, and $s \in [t]$, there exist $\lambda_s : \mathbb{R}^2 \to \mathbb{R}$, such that*

$$F(x_1, r_1, z_1 \dots, x_t, r_t, z_t, 1, \dots, 1) = \prod_{s=1}^{t} (1 - \lambda_s(r_s, z_s) + \lambda_s(r_s, z_s) e_p^{\mathrm{ES}}(x_s, r_s, z_s)).$$

The following is a tentative idea for proofing the conjecture.

We prove this result by mathematical induction. For $t = 1$, it follows from Theorem 7.5 that $F(x_1, r_1, z_1, 1, \dots, 1) = 1 - \lambda_1 + \lambda_1 e_p^{\mathrm{ES}}(x_1, r_1, z_1)$ for some $\lambda_1$ that is a function of $r_1$ and $z_1$ taking values in $(0, 1]$. Next, suppose that for $t \geqslant 1$,

$$F(x_1, r_1, z_1 \dots, x_t, r_t, z_t, 1, \dots, 1) = \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s e_p^{\mathrm{ES}}(x_s, r_s, z_s)).$$

For all $t \in [T]$, we can write $\mathcal{F}_t = \sigma(A_1^t, A_2^t, \dots)$ for disjoint $A_1^t, A_2^t, \dots \in \mathcal{F}_t$. For all $(X_s)_{s\in[t]}$ adapted to $(\mathcal{F}_s)_{s\in[t]}$, define

$$F'(x_{t+1}, r_{t+1}, z_{t+1}) = \sum_{n:\mathbb{P}(A_n^t)>0} \frac{F(X_1, r_1, z_1 \dots, X_t, r_t, z_t, 1, x_{t+1}, r_{t+1}, z_{t+1}, 1, \dots, 1)}{F(X_1, r_1, z_1 \dots, X_t, r_t, z_t, 1, \dots, 1)} \frac{\mathbb{1}_{A_n^t}}{\mathbb{P}(A_n^t)}.$$

For all $F_{X_1}, \dots, F_{X_t} \in \mathcal{M}'$, if $r_{t+1} = \mathrm{ES}_p(X_{t+1}|\mathcal{F}_t)$ and $z_{t+1} = \mathrm{VaR}_p(X_{t+1}|\mathcal{F}_t)$, then by (iii), $\mathbb{E}[F'(X_{t+1}, r_{t+1}, z_{t+1})] = \mathbb{E}[S_{t+1}/S_t|\mathcal{F}_t] = 1$. By (i) and (ii), the function $(x_{t+1}, r_{t+1}, z_{t+1}) \mapsto F'(x_{t+1}, r_{t+1}, z_{t+1})$ is also continuous and strictly decreasing with respect to $r$. It follows that $F'$ is an $\mathcal{M}'$-model-free e-statistic for $(\mathrm{VaR}_p, \mathrm{ES}_p)$ strictly testing $\mathrm{ES}_p$ that is non-conservative on $\mathcal{M}'$. By Theorem 7.5, there exists $\lambda_{t+1}$ as a function of $(r_{t+1}, z_{t+1})$ taking values in $(0, 1]$, such that

$$F'(x_{t+1}, r_{t+1}, z_{t+1}) = 1 - \lambda_{t+1} + \lambda_{t+1} e_p^{\mathrm{ES}}(x_{t+1}, r_{t+1}, z_{t+1}),\ x_{t+1} \in \mathbb{R},\ z_{t+1} < r_{t+1},$$

and hence

$$F(x_1, r_1, z_1 \dots, x_{t+1}, r_{t+1}, z_{t+1}, 1, \dots, 1) = \prod_{s=1}^{t+1} (1 - \lambda_s + \lambda_s e_p^{\mathrm{ES}}(x_s, r_s, z_s)).$$

It is promising to explore further on this project about the forms of model-free e-statistics for other risk measures in the dynamic framework.

3. Given the close link between elicitability and backtestability, we are motivated to explore further on backtesting the general class of elicitable risk measures using e-values. As is discussed in Chapter 7, identification functions (closely related to elicitability) can usually be derived using model-free e-statistics. Therefore, it is promising to obtain a general form of model-free e-statistics for elicitable risk measures. After establishing such as result, we can construct a procedure for e-backtesting all elicitable risk measures. This will lead to great practical convenience in terms of backtesting various forms of risk measures in financial regulation. Besides elicitable risk measures, distortion risk measures may also be a potential general class of risk measures that we can backtest. Further investigations are needed on this direction.

4. Backtesting is important from the practical side in financial regulations. We are thinking about empirical projects on applying our e-backtesting model to more real financial datasets and see whether the method works well for various kinds of real financial data. Moreover, following the methodology of e-backtesting, we are also designing trading strategies of cryptocurrencies, a popular topic in financial technology. Real trading data of cryptocurrencies are applied to evaluate the performance of various trading strategies based on e-values. Moreover, as discussed in Chapter 7, procedures of backtesting risk measures involves both banks and regulators. The whole model can be formulated as a game between financial institutions and regulators, where a financial institution provides the forecasts $r_t$ of risk measures to pass the backtest and a regulator adopts a test martingale by choosing $\lambda_t$. There might be some equilibrium between the two players. It will be interesting to examine the optimal choices of the bank and the regulator in some equilibrium.

# Bibliography


Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**(7), 1505–1518.

Acerbi, C. and Szekely, B. (2014). Backtesting expected shortfall. *Risk Magazine*, **27**, 76–81.

Acerbi, C. and Szekely, B. (2017). General properties of backtestable statistics. *SSRN*: 2905109.

Anscombe, F. J. and Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, **34**(1), 199–205.

Armbruster, B. and Delage, E. (2015). Decision making under uncertainty when preference information is incomplete. *Management Science*, **61**(1), 111–128.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5), 941–973.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1997). Thinking coherently. *Risk*, 68–71.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Barrieu, P. and El Karoui, N. (2005). Inf-convolution of risk measures and optimal risk transfer. *Finance and Stochochatics*, **9**, 269–298.

Barrieu, P. and Scandolo, G. (2008). General Pareto Optimal Allocations and Applications to Multi-Period Risks1. *ASTIN Bulletin*, **38**(1), 105–136.

Basak, S. and Shapiro, A. (2001). Value-at-Risk-based risk management: Optimal policies and asset prices. *Review of Financial Studies*, **14**(2), 371–405.

Basak, S., Shapiro, A. and Teplá, L. (2006). Risk management with benchmarking. *Management Science*, **52**, 542–557.

BCBS (2013). Basel Committee on Banking Supervision (Jan. 2016). *Consultative Document: Fundamental Review of the Trading Book* Bank for International Settlements. https://www.bis.org/publ/bcbs219.pdf

BCBS (2016). Basel Committee on Banking Supervision (Jan. 2016). *Minimum Capital Requirements for Market Risk.* Bank for International Settlements. https://www.bis.org/bcbs/publ/d352.pdf

BCBS (2019). Basel Committee on Banking Supervision (Jan. 2019; revised Feb. 2019). *Minimum Capital Requirements for Market Risk.* Bank for International Settlements. https://www.bis.org/bcbs/publ/d457.pdf

Bayer, S. and Dimitriadis, T. (2022) Regression-based Expected Shortfall backtesting. *Journal of Financial Econometrics*, **20**(3), 437–471.

Belles-Sampera, J., Guillén, M. and Santolino, M. (2014). Beyond value-at-risk: GlueVaR distortion risk measures. *Risk Analysis*, **34**(1), 121–134.

Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, **15**(5), 725–733.

Bellini, F., Bignozzi, V. and Puccetti, G. (2018). Conditional expectiles, time consistency and mixture convexity properties. *Insurance: Mathematics and Economics*, **82**, 117–123.

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **54**(1), 41–48.

Bellini, F. and Peri, I. (2021). An axiomatization of $\Lambda$-quantiles. *SIAM Journal on Financial Mathematics*, **13**(1), SC26–SC38.

Ben-Tal, A. and Teboulle, M. (2007). An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, **17**(3), 449–476.

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004). Sequential change-point detection in GARCH$(p, q)$ models. *Econometric Theory*, **20**(6), 1140–1167.

Berkowitz, J., Christoffersen, P. and Pelletier D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, **57**(12), 2213–2227.

Bernard, C., Pesenti, S. M. and Vanduffel, S. (2020). Robust distortion risk measures. *SSRN*: 3677078.

Bernard, C. and Tian, W. (2009). Optimal reinsurance arrangements under tail risk measures. *Journal of Risk and Insurance*, **76**(3), 709–725.

Bertsimas D., Gupta V., Paschalidis I. C. (2012). Inverse optimization: A new perspective on the blacklitterman model. *Operations Research* **60**(6), 1389–1403.

Birghila, C. and Pflug, G. C. (2019). Optimal XL-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics and Economics*, **88**, 30–43.

Blanchet, J., Lam, H., Liu, Y. and Wang, R. (2020). Convolution bounds on quantile aggregation. *arXiv:* 2007.09320.

Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, **44**(2), 565–600.

Braun, M. and Muermann, A. (2004). The impact of regret on the demand for insurance. *Journal of Risk and Insurance*, **71**(4), 737–767.

Brehmer, J. and Gneiting, T. (2021). Scoring interval forecasts: Equal-tailed, shortest, and modal interval.

*Bernoulli*, forthcoming.

Brighi, B. and Chipot, M. (1994). Approximated convex envelope of a function. *SIAM Journal on Numerical Analysis*, **31**, 128–148.

Cai, J. and Chi, Y. (2020). Optimal reinsurance designs based on risk measures: A review. *Statistical Theory and Related Fields*, **4**(1), 1–13.

Cai, J., Li, J. and Mao, T. (2020). Distributionally robust optimization under distorted expectations. *SSRN*: 3566708.

Cai, J., Liu, H. and Wang, R. (2017). Pareto-optimal reinsurance arrangements under general model settings. *Insurance: Mathematics and Economics*, **77**, 24–37.

Cai, J., Liu, H. and Wang, R. (2018). Asymptotic equivalence of risk measures under dependence uncertainty. *Mathematical Finance*, **28**(1), 29–49.

Cai, J. and Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. *ASTIN Bulletin*, **37**(1), 93–112.

Cai, J., Tan, K. S., Weng, C. and Zhang, Y. (2008). Optimal reinsurance under VaR and CTE risk measures. *Insurance: mathematics and Economics*, **43**(1), 185–196.

Carlier, G. and Dana, R.-A. (2003). Core of convex distortions of a probability.. *Journal of Economic Theory*, **113**, 199–222.

Castagnoli, E., Cattelan, G., Maccheroni, F., Tebaldi, C. and Wang, R. (2022). Star-shaped risk measures. *Operations Research*. https://doi.org/10.1287/opre.2022.2303.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011). Risk measures: rationality and diversification. *Mathematical Finance*, **21**(4), 743–774.

Cheridito, P., Delbaen, F. and Kupper, M. (2004). Coherent and convex risk measures for bounded càdlàg processes. *Stochastic processes and their applications*, **112**(1), 1–22.

Chen, C., Iyengar, G. and Moallemi, C. C. (2013). An axiomatic approach to systemic risk. *Management Science*, **59**(6), 1373–1388.

Chen, L., He, S. and Zhang, S. (2011). Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, **59**(4), 847–865.

Chen, Y., Lin, L. and Wang, R. (2021). Ordered Risk Aggregation under Dependence Uncertainty. *arXiv*: 2104.07718.

Cheridito, P. and Li, T. (2009). Risk measures on Orlicz hearts. *Mathematical Finance*, **19**(2), 189–214.

Cherny, A. S. and Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, **22**(7), 2571–2606.

Cheung, K. C., Sung, K. C. J., Yam, S. C. P. and Yung, S. P. (2014). Optimal reinsurance under general law-invariant risk measures. *Scandinavian Actuarial Journal*, **2014**(1), 72–91.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox, *Econometrica*, **51**, 1065–1092.

Chi, Y. (2019). On the optimality of a straight deductible under belief heterogeneity. *ASTIN Bulletin*, **49**(1), 243–262.

Chong, W. F., Feng, R. and Jin, L. (2021). Holistic principle for risk aggregation and capital allocation. *Annals of Operations Research*, 1–34.

Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, **5**, 131–295.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, **39**(4), 841–862.

Christoffersen, P. (2011). *Elements of Financial Risk Management*. Academic Press.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, **105**(1), 85–110.

Chen, Y. (2021). Optimal reinsurance from the viewpoints of both an insurer and a reinsurer under the CVaR risk measure and Vajda condition. *ASTIN Bulletin*, **51**(2), 631–659.

Chi, Y. and Weng, C. (2013). Optimal reinsurance subject to Vajda condition. *Insurance: Mathematics and Economics*, **53**(1), 179–189.

Chu, J. C. S., Stinchcombe, M. and White, H. (1996). Monitoring structural change. *Econometrica*, **64**(5), 1045–1065.

Cifarelli, D. M. and Regazzini, E. (1996). De Finetti's contribution to probability and statistics. *Statistical Science*, 253–282.

Cont, R., Deguest, R. and He, X. D. (2013). Loss-based risk measures. *Statistics and Risk Modeling*, **30**, 133–167.

Cornilly, D., Rüschendorf, L. and Vanduffel, S. (2018). Upper bounds for strictly concave distortion risk measures on moment spaces. *Insurance: Mathematics and Economics*, **82**, 141–151.

Cui, W., Yang, J. and Wu, L. (2013). Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles. *Insurance: Mathematics and Economics*, **53**(1), 74–85.

Cummins, J. D. and Mahul, O. (2004). The demand for insurance with an upper limit on coverage. *Journal of Risk and Insurance*, **71**(2), 253–264.

Dana, R. A. and Scarsini, M. (2007). Optimal risk sharing with background risk. *Journal of Economic Theory*, **133**(1), 152–176.

Daníelsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Shin, H. S. (2001). An academic response to Basel II. *LSE Special Paper Series May 2001.*

Darling, D. A. and Robbins, H. (1967). Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences*, **58**(1), 66–68.

Davis, M. H. A. (2013). Consistency of risk measure estimates. *Risk Management & Analysis in Financial*

*Institutions eJournal. SSRN*: 2342279.

Davis, M. H. A. (2014). The Beaufort scale of predictability. In Podolskij, M., Stelzer, R., ThorbjÃžrnsen, S. and Almut E. D. Veraart, A.E.D. (Eds.) (2014). *The Fascination of Probability, Statistics and their Applications. In honour of Ole E. Barndorff-Nielsen*. Springer, 419–434.

Davis, M. H. A. (2016). Verification of internal risk measure estimates. *Statistics and Risk Modeling*, **33**(3-4), 67–93.

Davis, M. H. A. (2017). Discussion of "elicitability and backtesting: Perspectives for banking regulation" *Annals of Applied Statistics*, **11**(4), 1886–1887.

Dawid, A. P. (1984). Present position and potential developments: some personal views. Statistical theory: the prequential approach (with discussion). *Journal of the Royal Statistical Society: Series A*, **147**, 278–292.

Dawid, A. P. and Vovk, V. (1999). Prequential probability: principles and properties. *Bernoulli*, **5**, 125–162.

de la Peña, V. H., Klass, M. J. and Lai, T. L. (2004). Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Annals of Probability*, **32**(3), 1902–1933.

de la Peña, V. H., Lai, T. L. and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications. Probability and Its Applications*. First edition. Springer, Berlin.

Delage, E., Arroyo, S. and Ye, Y. (2014). The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. *Operations Research*, **62**(6), 1377–1393.

Delage, E., Guo, S. and Xu, H. (2022). Shortfall risk models when information on loss function is incomplete. *Operations Research*, **70**(6), 3511–3518..

Delage, E. and Li, Y. (2018). Minimizing risk exposure when the choice of a risk measure is ambiguous. *Management Science*, **64**(1), 327–344.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, **58**(3), 595–612.

Delbaen, F. (2002). Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics* (pp. 1–37). Springer, Berlin.

Delbaen, F. (2012). *Monetary Utility Functions*. Osaka University Press, Osaka.

Delbaen, F., Bellini, F., Bignozzi, V. and Ziegel, J. (2016). Risk measures with convex level sets. *Finance and Stochastics*, **20**(2), 433–453.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, **22**(5), 1915–1953.

Denneberg, D. (1990). Premium calculation: why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, **20**, 181–190.

Denneberg, D. (1994). *Non-additive Measure and Integral.* Springer Science & Business Media.

Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Wiley: Chichester, UK.

Deprez, O. and Gerber, H. U. (1985). On convex principles of premium calculation. *Insurance: Mathematics and Economics*, **4**(3), 179–189.

Dhaene, J., Kukush, A., Linders, D. and Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, **2**(2), 319–328.

Dhaene, J., Vanduffel, S., Goovaerts, M.J., Kaas, R., Tang, Q. and Vynche, D. (2006). Risk measures and comonotonicity: a review. *Stochastic Models*, **22**, 573–606.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, **13**, 253–263.

Dimitriadis, T., Fissler, T. and Ziegel, J. F. (2023). Osband's principle for identification functions. *Statistical Papers*. doi: https://doi.org/10.1007/s00362-023-01428-x.

Drapeau, S. and Kupper, M. (2013). Risk Preferences and their Robust Representation. *Mathematics of Operations Research*, **38** (1), 28–62.

Du, Z. and Escanciano, J. C. (2017). Backtesting expected shortfall: Accounting for tail risk. *Management Science*, **63**, 940–958.

El Karoui, N. and Ravanelli, C. (2009). Cash subadditive risk measures and interest rate ambiguity. *Mathematical Finance*, **19**(4), 562–590.

Embrechts, P., Liu, H., Mao, T. and Wang, R. (2020). Quantile-based risk sharing with heterogeneous beliefs. *Mathematical Programming Series B*, **181**(2), 319–347.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, **97**(2), 526–547.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Emmer, S., Kratz, M. and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, **18**(2), 31–60.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economic Statistics*, **22**(4), 367–381.

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstien metric: performance guarantees and tractable reformulations. *Mathematical Programming Series A*,

**171**(1), 115–166.

Farkas, W., Koch-Medina, P. and Munari, C. (2014). Beyond cash-additive risk measures: when changing the numeraire fails. *Finance Stochastic*, 18, 145–173.

Filipović, D. and Svindland, G. (2008). Optimal capital and risk allocations for law- and cash-invariant convex functions. *Finance and Stochastics*, **12**, 423–439.

Filipović, D. and Svindland, G. (2012). The canonical model space for law-invariant convex risk measures is $L^1$. *Mathematical Finance*, **22**(3), 585–589.

Fissler, T. (2017). *On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space*. PhD thesis, University of Bern.

Fissler, T., Hlavinová, J. and Rudloff, B. (2020). Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals. *arXiv*: `1910.07912v2`.

Fissler, T. and Holzmann, H. (2022). Measurability of functionals and of ideal point forecasts. *Electronic Journal of Statistics*, **16**(2), 5019–5034.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, **44**(4), 1680–1707.

Föllmer, H. and Knispel, T. (2011). Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics*, **11**, 333–351.

Föllmer, H. and Schied, A. (2002a). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**(4), 429–447.

Föllmer, H. and Schied, A.(2002b). Robust preferences and convex measures of risk (pp. 39–56). Springer, Berlin, Heidelberg.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance: An Introduction in Discrete Time*. Fourth edition. Walter de Gruyter.

Frittelli, M. and Maggis, M. (2011). Dual representation of quasiconvex conditional maps. *SIAM Journal of Financial Mathematics*, **2**, 357–382.

Frittelli, M., Maggis, M. and Peri, I. (2014). Risk measures on $\mathcal{P}(\mathbb{R})$ and value at risk with probability/loss function. *Mathematical Finance*, **24**(3), 442–463.

Frittelli, M. and Rosazza Gianin, E. (2002). Putting order in risk measures. *Journal of Banking and Finance*, **26**,1473–1486.

Frongillo, R. and Kash, I. A. (2021). Elicitation complexity of statistical properties. *Biometrika*, **108**(4), 857–879.

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocation and heavy-tailed risks. *Journal of Banking and Finance*, **83**, 70–84.

Gerber, H. U. (1974). On additive premium calculation principles. *ASTIN Bulletin*, **7**(3), 215–222.

Ghirardato, P., Maccheroni, F. and Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, **118**(2), 133–173.

Ghossoub, M. (2017). Arrow's Theorem of the deductible with heterogeneous beliefs. *North American Actuarial Journal*, **21**(1), 15–35.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.

Gilboa, I., Postlewaite, A. and Samuelson, L. (2019). What are axiomatizations good for? *Theory Decision*, 86(3–4), 339–359.

Gneiting T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494),746–762.

Gollier, C. (1996). Optimum insurance of approximate losses. *Journal of Risk and Insurance*, **63**(3), 369–380.

Gollier, C. (2013). The economics of optimal insurance design. In *Handbook of Insurance* (pp. 107–122). Springer, New York, NY.

Gollier, C. and Schlesinger, H. (1996). Arrow's theorem on the optimality of deductibles: a stochastic dominance approach. *Economic Theory*, **7**(2), 359–363.

Grant, S., Kajii, A. and Polak, B. (2000). Temporal resolution of uncertainty and recursive non-expected utility models. *Econometrica*, **68**(2), 425–434.

Grechuk, B., Molyboha, A. and Zabarankin, M. (2009). Maximum entropy principle with general deviation measures. *Mathematics of Operations Research*, **34**(2), 445–467.

Grünwald, P. D. and Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, **32**(4), 1367–1433.

Grünwald, P. D., de Heide, Rianne R. and Koolen, W. M. (2020). Safe testing. In *2020 Information Theory and Applications Workshop (ITA)*, pages 1–54.

Guo, S. and Xu, H. (2020). Statistical robustness in utility preference robust optimization models. *Mathematical Programming Series A*, **190**, 679–720.

Han, X., Wang, B., Wang, R. and Wu, Q. (2021). Risk concentration and the mean-Expected Shortfall criterion. *arXiv*: 2108.05066.

Henzi, A. and Ziegel, J. F. (2022). Valid sequential inference on probability forecast performance. *Biometrika*, **109**(3), 647–663.

Hofmann, A., Häfen, O. V. and Nell, M. (2019). Optimal insurance policy indemnity schedules with policyholders' limited liability and background risk. *Journal of Risk and Insurance*, **86**(4), 973–988.

Hoga, Y. and Demetrescu, M. (2022). Monitoring Value-at-Risk and Expected Shortfall Forecasts. *Management Science*, **69**(5), 2954–2971.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics.* Second edition. Wiley Series in Probability and Statistics. Wiley, New Jersey.

Huberman, G., Mayers, D. and Smith Jr, C. W. (1983). Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, 415–426.

Jarrow, R. (2002). Put option premiums and coherent risk measures. *Mathematical Finance*, **12**(2), 135–142.

Jeffreys, H. (1961). *Theory of Probability.* Third edition. Oxford University Press, Oxford.

Jia, G., Xia, J. and Zhao, R. (2021). Monetary risk measures. *arXiv*: `2012.06751`.

Jiang, R. and Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming Series A*, **158**(1), 291–327.

Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, **9**, 49–71.

Jouini, E., Schachermayer, W. and Touzi, N. (2008). Optimal risk sharing for law invariant monetary utility functions. *Mathematical Finance*, **18**(2), 269–292.

Kelly, J. L. (1976). A new interpretation of information rate. *Bell System Technical Journal*, **35**(4), 917–926.

Korpelevich, G. M. (1976). The extragradient method for finding saddle points and other problems. *Matecon*, **12**, 747–756.

Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5): 1056-1072.

Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and quantitiative robustness for law-invariant risk measures. *Finance and Stochastics*, **18**(2), 271–295.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives*, **3**(2), 73–84.

Kupper, M. and Schachermayer, W. (2009). Representation results for law invariant time consistent functions, *Mathematics and Financial Economics*, **2**, 189–210.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.

Lambert, N., Pennock, D. M. and Shoham, Y. (2008). Eliciting properties of probability distributions. *Proceedings of the 9th ACM Conference on Electronic Commerce*, 129–138.

Lauzier, J. G., Lin, L. and Wang, R. (2023). Risk sharing, measuring variability, and distortion riskmetrics. *arXiv*: `2302.04034`.

Li, H. and Wang, R. (2019). PELVE: Probability equivalent level of VaR and ES. *Journal of Econometrics*, **234**(1), 353–370.

Li, J. (2021). Inverse optimization of convex risk functions. *Management Science*, **67**(11), 7113–7141.

Li, L., Shao, H., Wang, R. and Yang, J. (2018). Worst-case Range Value-at-Risk with partial information. *SIAM Journal on Financial Mathematics*, **9**(1), 190–218.

Li, Y. (2018). Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research*, **66**(6), 1457–1759.

Liebrich, F. B. (2021). Risk sharing under heterogeneous beliefs without convexity. *arXiv*: `2108.05791`.

Liebrich, F-B. and Svindland, G. (2017). Model spaces for risk measures. *Insurance: Mathematics and Economics*, **77**, 150–165.

Liu, F., Cai, J., Lemieux, C. and Wang, R. (2020). Convex risk functionals: representation and applications. *Insurance: Mathematics and Economics*, **90**, 66–79.

Liu, F., Mao, T., Wang, R. and Wei, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519..

Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.

Liu, H. and Mao, T. (2022). Distributionally robust reinsurance with Value-at-Risk and Conditional Value-at-Risk. *Insurance: Mathematics and Economics*, **107**, 393-417.

Liu, P., Schied, A. and Wang, R. (2020). Distributional transforms, probability distortions, and their applications. *Mathematic of Operations Research*, **46**(4), 1490–1512.

Lo, A., Tang, Q. and Tang, Z. (2021). Universally marketable insurance under multivariate mixtures. *ASTIN Bulletin*, **51**(1), 221–243.

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, **74**(6), 1447–1498.

Mao, T., Hu, J. and Liu, H. (2018). The average risk sharing problem under risk measure and expected utility theory. *Insurance: Mathematics and Economics*, **83**, 170–179.

Mao, T., Wang, B. and Wang, R. (2019). Sums of uniform random variables. *Journal of Applied Probability*, **56**(3), 918–936.

Mao, T. and Wang, R. (2015). On aggregation sets and lower-convex sets. *Journal of Multivariate Analysis*, **138**, 170–181.

Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

Mao, T., Wang, R. and Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. *arXiv*: `2201.06370`.

Marinacci, M. (2002). Probabilistic sophistication and multiple priors. *Econometrica*, **70**(2), 755–764.

Marinacci, M. and Montrucchio, L. (2008). On concavity and supermodularity. *Journal of Mathematical Analysis and Applications*, **344**(2), 642–654.

Marshall, J. M. (1992). Optimum insurance with deviant beliefs. In *Contributions to Insurance Economics* (pp. 255–274). Springer, Dordrecht.

Mastrogiacomo, E. and Rosazza Gianin, E. (2015). Portfolio optimization with quasiconvex risk measures.

*Mathematics of Operations Research*, **40**(4), 1042–1059.

McNeil, A. J. and Frey, R. (2000). *Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance*, **7**(3–4), 271–300.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised edition. Princeton, NJ: Princeton University Press.

Moresco, M. R. and Righi, M. B. (2022). On the link between monetary and star-shaped risk measures. *Statistics and Probability Letters*, **184**, 109345.

Müller, A., Scarsini, M., Tsetlin, I. and Winkler. R. L. (2017). Between first and second-order stochastic dominance. *Management Science*, **63**(9), 2933–2947.

Natarajan, K., Pachamanova, D. and Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, **54**(3), 573–585.

Nemirovski, A. (2004). Prox-method with rate of convergence O(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, **15**(1), 229–251.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*, **11**(4), 1833–1874.

Osband, K. (1985). *Providing Incentives for Better Cost Forecasting*. Ph.D. Thesis, University of California, Berkeley.

Ouyang, Y. and Xu, Y. (2021). Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming Series A*, **185**(1), 1–35.

Patton, A., Ziegel, J. F. and Chen, R. (2019). Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics*, **211**(2), 388–413.

Pichler, A. (2013). The natural Banach space for version independent risk measures. *Insurance: Mathematics and Economics*, **53**(2), 405-415.

Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, **55**(1), 98–112.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122–136.

Puccetti, G. and Rüschendorf, L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, **236**(7), 1833–1840.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, **3**(4), 323–343.

Ramdas, A., Grünwald, P., Vovk, V. and Shafer, G. (2022). Game-theoretic statistics and safe anytime-valid inference. *arXiv*: `2210.01948`.

Raviv, A. (1979). The design of an optimal insurance policy. *American Economic Review*, **69**(1), 84–96.

Resnick, S. (2019). *A Probability Path*. First edition. Springer, Birkhäuser Boston, MA.

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and their Applications*, **112**(2), 185–200.

Righi, M. B. (2021). Star-shaped acceptability indexes. *arXiv*: 2110.08630.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

Rockafellar, R. T. and Royset, J. O. (2018). Superquantile/CVaR risk measures: second-order theory. *Annals of Operations Research*, **262**(1), 3–28.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.

Rockafellar, R. T. and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, **18**(1-2), 33–53.

Rockafellar, R. T., Uryasev, S. and Zabarankin, M. (2006). Generalized deviation in risk analysis. *Finance and Stochastics*, **10**, 51–74.

Rudin, W. (1987). *Real and Complex Analysis*. Tata McGraw-Hill Education.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.

Ruszczyński, A. and Shapiro, A. (2006). Optimization of convex risk functions. *Mathematics of Operations Research*, **31**(3), 433–452.

Simchi-Levi, D., Chen, X. and Bramel, J. (2005). *The Logic of Logistics. Theory, Algorithms, and Applications for Logistics and Supply Chain Management*. Third edition. New York, NY: Springer.

Schlesinger, H. (1981). The optimal level of deductibility in insurance contracts. *Journal of risk and insurance*, **48**(3), 465–481.

Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, **97**(2), 255–261.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Schlesinger, H. (1997). Insurance demand without the expected-utility paradigm. *Journal of Risk and Insurance*, **64**(1), 19–39.

Shafer, G. (2021). The language of betting as a strategy for statistical and scientific communication. *Journal of the Royal Statistical Society: Series A*, forthcoming.

Shafer, G., Shen, A., Vereshchagin, N. and Vovk, V. (2011). Test martingales, Bayes factors and p-values. *Statistical Science*, **26**(1), 84–101.

Shafer, G. and Vovk, V. (2019). *Game-Theoretic Foundations for Probability and Finance*, volume 455. John Wiley & Sons.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics.

Song, Y. and Yan, J. (2006). The representations of two types of functions on $L^\infty(\Omega, \mathcal{F})$ and $L^\infty(\Omega, \mathcal{F}, P)$. *Science in China Series A: Mathematics*, **49**, 1376–1382.

Song, Y. and Yan, J. (2009). Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. *Insurance: Mathematics and Economics*, **45**(3), 459–465.

Steinwart, I., Pasin, C., Williamson, R. and Zhang, S. (2014). Elicitation and identification of properties. *JMLR: Workshop and Conference Proceedings*, **35**, 1–45.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, **8**(4), 814–827.

ter Schure, J. and Grünwald, P. (2021). ALL-IN meta-analysis: breathing life into living systematic reviews. *arXiv*: `2109.12141`.

Tian, D. and Long, J. (2015). Quasiconvex risk statistics with scenario analysis. *Mathematics and Financial Economics*, **9**(2), 111–121.

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

Vajda, S. (1962). Minimum variance reinsurance. *ASTIN Bulletin*, **2**(2), 257–260.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Van Heerwaarden, A. E., Kaas, R. and Goovaerts, M. J. (1989). Optimal reinsurance in relation to ordering of risks. *Insurance: Mathematics and Economics*, **8**(1), 11–17.

Ville, J. (1939). Étude critique de la notion de collectif. *Thèses de l'entre-deux-guerres*.

Vovk, V., Wang, B. and Wang, R. (2021). Admissible ways of merging p-values under arbitrary dependence.*Annals of Statistics*, **50**(1), 351–375.

Vovk, V. and Wang, R. (2020). Merging sequential e-values via martingales. *arXiv*:, `2007.06382`.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination, and applications. *Annals of Statistics*, forthcoming.

Vovk, V. and Wang, R. (2021). Efficiency of nonparametric e-tests. *arXiv*:, `2208.08925`.

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, **16**(2), 117–186.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.

Wang, Q., Wang, R. and Ziegel, J. F. (2022). Simulation and data analysis for e-backtesting. *SSRN*: `4346325`.

Wang, R., Bignozzi, V. and Tsakanas, A. (2015). How superadditive can a risk measure be? *SIAM Jounral on Financial Mathematics*, **6**, 776–803.

Wang, R. and Ramdas, A. (2020). False discovery rate control with e-values. *Journal of the Royal Statistical Society, Series B*, **84**(3), 822–852.

Wang, R. and Wei, Y. (2020). Risk functionals with convex level sets. *Mathematical Finance*, **30**, 1337–1367.

Wang, R., Wei, Y. and Willmot, G. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematic of Operations Research*, **45**(3), 993–1015.

Wang, R., Xu, Z. Q. and Zhou, X. Y. (2019). Dual utilities on risk aggregation under dependence uncertainty. *Finance and Stochastics*, **23**(4), 1025–1048.

Wang, R. and Ziegel, J. (2015). Elicitable distortion risk measures: A concise proof. *Statistics and Probability Letters*, **100**, 172–175.

Wang, R. and Zitikis, R. (2020). Weak comonotonicity. *European Journal of Operational Research*, **282**(1), 386–397.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the expected shortfall. *Management Science*, **67**, 1329–1992.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, **26**, 71–92.

Wang, S., Young, V. R. and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**(2), 173–183.

Wang, W. and Xu, H. (2020). Robust spectral risk optimization when information on risk spectrum is incomplete. *SIAM Journal on Optimization*, **30**(4), 3198–3229.

Wasserman, L., Ramdas, A. and Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, **117**(29), 16880–16890.

Waudby-Smith, I. and Ramdas, A. (2023). Estimating means of bounded random variables by betting. *Journal of Royal Statistical Society, Series B*, forthcoming.

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, **16**, 419–441.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, **64**, 1067–1084.

Wiesemann, W., Kuhn, D. and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, **62**(6), 1203–1466.

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal*, **23**(2), 189-257.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, **42**(12), 1676–1690.

Wu, Q., Mao, T. and Hu, T. (2020). Optimized certainty equivalent with applications in the rank-dependent expected utility model. *SSRN*: 3694866.

Xia, J. (2013). Comonotonic convex preferences. *SSRN*: 2298884.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

Zheng, Y. and Cui, W. (2014). Optimal reinsurance with premium constraint under distortion risk measures. *Insurance: Mathematics and Economics*, **59**, 109–120.

Zhu, W. and Shao, H. (2018). Closed-form solutions for extreme-case distortion risk measures and applications to robust portfolio management. *SSRN*: 3103458.

Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155–1168.

Zhuang, S. C., Weng, C., Tan, K. S. and Assa, H. (2016). Marginal indemnification function formulation for optimal reinsurance. *Insurance: Mathematics and Economics*, **67**, 65–76.

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

Ziggel, D., Berens, T., Weiß, G. N. F. and Wied, D. (2014). A new set of improved Value-at-Risk backtests. *Journal of Banking & Finance*, **48**(48), 29–41.

Zymler, S., Kuhn, D. and Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, **137**(1–2), 167–198.