Aggregating and evaluating ambiguity in decision models

by

Zachary Van Oosten

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

Waterloo, Ontario, Canada, 2026

© Zachary Van Oosten 2026

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: Manel Baucells
Professor, Darden School of Business
University of Virginia

Supervisors: Yi Shen
Associate Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Ruodu Wang
Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Internal Members: Max Nendel
Associate Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Alexander Schied
Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: Hossein Abouee Mehrizi
Professor, Dept. of Management Science and Engineering
University of Waterloo

Author's declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of contributions

I am the sole author of Chapters 1, 5, and 6.

Chapter 2 is joint work with Yi Shen and Ruodu Wang, and is forthcoming in a peer-reviewed journal:

> Shen, Y., **Van Oosten, Z.** and Wang, R. (2026). Partial Law Invariance and Risk Measures. *Management Science,* forthcoming.

Chapter 3 is joint work with Ruodu Wang, and has been submitted to a peer-reviewed journal. The arXiv reference is below:

> **Van Oosten, Z.** and Wang, R. (2025). Choquet rank-dependent utility with an exogenous unambiguous source. arXiv

Chapter 4 is joint work with Ruodu Wang. The arXiv reference is below:

> **Van Oosten, Z.** and Wang, R. (2025). Aggregate then evaluate. arXiv

Abstract

This thesis develops new decision models under ambiguity that accommodate arbitrary reductions in the absence of ambiguity, that is, models not restricted to expected utility (EU) under pure risk (no ambiguity). A central challenge is the identifiability problem: from preferences alone, one cannot distinguish which features of the representation functional reflect the decision maker's evaluation of pure risk from those that reflect their reaction to ambiguity. To resolve this, we exogenously specify a family of unambiguous events and employ the property of partial law invariance: the DM is indifferent between unambiguous acts with the same distribution. The resulting models developed across the four chapters outlined below are relevant to decision theory, quantitative risk management, mathematical finance, and operations research.

Chapter 2 studies partial law invariance in the context of risk measures. We fully characterize partially law-invariant coherent risk measures via a novel representation formula and extend the classical Kusuoka representation. Additionally, we propose new risk measures, including partially law-invariant versions of Expected Shortfall and entropic risk measures, along with tractable formulas for their calculation.

Chapter 3 axiomatizes the Choquet rank-dependent utility (CRDU) model, which cleanly separates pure risk from ambiguity and reduces to rank-dependent utility in the absence of ambiguity. We show that the coupling of ambiguity perception and ambiguity attitude can be fully characterized by the matching probability, and that the supermodularity of this matching probability gives CRDU a distributionally robust interpretation.

Chapter 4 distinguishes two conceptual frameworks for processing ambiguity: evaluate-then-aggregate (ETA), which first evaluates an act under each plausible model and then aggregates, and aggregate-then-evaluate (ATE), which first reduces ambiguity to a single representative distribution before evaluation. As most existing ambiguity models fall within the ETA framework, we develop and axiomatize the Choquet ATE model that generalizes both Choquet expected utility and CRDU while accommodating arbitrary pure-risk evaluations. Afterward, we provide a rich analysis of the interplay between ambiguity attitudes and risk attitudes.

Chapter 5 develops a distributionally robust optimization model whose ambiguity set is derived from a Bayesian second-order belief, providing a clear separation between pure risk, ambiguity perception, and ambiguity attitude. This chapter is motivated by the observation that the smooth ambiguity model does not satisfactorily generalize to arbitrary pure-risk evaluations. A canonical construction based on distorted second-order beliefs is introduced as a tractable instance of this model, with accompanying algorithms and numerical illustrations in the context of portfolio optimization.

Acknowledgements

I would like to thank my supervisors, Prof. Yi Shen and Prof. Ruodu Wang, both of whom have been incredibly supportive throughout my PhD. I would also like to express my sincerest gratitude to Prof. Ruodu Wang for being an exceptional academic mentor. He has cultivated in me a passion for this science, and I will always remember with great fondness the many thought-provoking conversations we shared over the past four years.

Thank you to my committee members, Prof. Manel Baucells, Prof. Max Nendel, Prof. Alexander Schied, and Prof. Hossein Abouee Mehrizi, for their willingness to serve on my committee and for the time they devoted to reading this thesis.

Over the past four years, I have had the pleasure of meeting many people in academic life. I owe a special thanks to the following three individuals, each of whom went above and beyond in supporting my development as a researcher. I would like to thank Prof. Peter Wakker for his generous and detailed feedback on much of the work presented in this thesis, and for his hospitality during my visit to the Erasmus School of Economics; Prof. Jean-Gabriel Lauzier for his support from the very start of my PhD; and Prof. Felix-Benedikt Liebrich for his helpful advice and for always making time to talk. Additionally, I would like to thank Prof. Greg Rice and Prof. Paul Marriott for the excellent courses.

I am thankful to the members of my research group. Those familiar with its size will appreciate that a complete list of names would require another page, but I owe special thanks to Benjamin, Chris, Qinghua, James, Qiuqi, Liyuan, Zhanyi, and Samuel.

I am grateful to my friends, who have made my life so much better and have been a constant source of support, joy, and encouragement throughout my time here. I would like to thank Nick and Tyler, who have been by my side long before I decided to come to Waterloo; Wes, Liane, and little Arlo, who have endured my ramblings at far too early an hour; Boris, Carlee, and Taran, who made Waterloo feel like home when for years it did not; Andrew, Sam, and Rish, who made my master's degree the experience it was; Ahmed, Evan, José, Matheus, Kelly, and Rick, the research group I was proud to be an unofficial member of; Maureen, David, and Daniel, whose kitchen table most of this document was written on; Mary Lou, the graduate coordinator who became a dear friend; Han, who has been an inspiration for me in ways he could never know; and Austin, the best part of coming to campus, who made my PhD what it was.

Most importantly, I thank my family: Mary Ann, Timothy, Samantha, and Donna, who have always been my greatest source of motivation. This document also serves as proof that I have, in fact, been doing something for the last four years, a point my sister has long questioned. Above all, thank you, Bryn, for your love and for always believing in me. No sentence I could write here would do you justice.

Dedication

This thesis is dedicated to my parents, Mary Ann and Timothy

Table of contents

Examining committee membership ii

Author's declaration iii

Statement of contributions iv

Abstract v

Acknowledgements vi

Dedication vii

List of figures xiii

List of tables xiv

1 Introduction 1

1.1 Models with arbitrary pure-risk projections 3

1.2 Structure of the thesis 5

2 Partial law invariance and risk measures 7

2.1 Introduction 7

2.2 Partial law invariance: Definition and motivating examples 11

2.2.1 Risk measures, preference relations, and partial law invariance . . . 11

2.2.2 Partial law invariance and model uncertainty . . . 13
2.2.3 Partial law invariance in decision models . . . 15
2.3 Coherent risk measures and the associated preferences . . . 17
2.3.1 Preferences and risk measures . . . 17
2.3.2 Representation of partially law-invariant coherent risk measures . . . 19
2.4 Multiple sources . . . 21
2.5 Extensions . . . 24
2.5.1 Coherent adjustments and an alternative representation . . . 24
2.5.2 Strong partial law invariance and Kusuoka-type representation . . . 26
2.5.3 Convex risk measures . . . 28
2.6 Explicit classes and applications . . . 30
2.6.1 Partially law-invariant ES . . . 30
2.6.2 Two constructions of partially law-invariant coherent risk measures 31
2.6.3 Numerical illustrations . . . 34
2.7 Conclusion . . . 36
2.8 Review on risk measures . . . 37
2.9 Additional results and proofs accompanying Section 2.2 . . . 39
2.10 Additional results and proofs accompanying Section 2.3 . . . 40
2.10.1 Additional results and proofs accompanying Section 2.3.1 . . . 40
2.10.2 Discussion and proof of Theorem 2.1 . . . 42
2.11 An additional example accompanying Section 2.4 . . . 43
2.12 Additional results and proofs accompanying Section 2.5 . . . 44
2.12.1 Additional results and proofs accompanying Section 2.5.1 . . . 44
2.12.2 Additivity on the subspace . . . 46
2.12.3 Uniqueness of the representation in Theorem 2.3 . . . 47
2.12.4 Additional results and proofs accompanying Section 2.5.2 . . . 51
2.12.5 Examples . . . 55
2.12.6 Additional results and proofs accompanying Section 2.5.3 . . . 57

2.13 Additional results and proofs accompanying Section 2.6 61
2.13.1 Additional results and proofs accompanying Section 2.6.1 61
2.13.2 Additional results and proofs accompanying Section 2.6.2 63
2.13.3 Additional results and proofs accompanying Section 2.6.3 66
2.14 Analysis in a finite probability space 67
2.14.1 Law-invariant coherent risk measures 67
2.14.2 Partially law-invariant coherent risk measures 68
2.14.3 Representation results 69

3 Choquet rank-dependent utility with an exogenous unambiguous source 72
3.1 Introduction 72
3.2 Preliminaries 75
3.2.1 Definitions and notation 75
3.2.2 The Savage setting with pure risk and the Choquet rank-dependent utility model 76
3.3 Axioms and representation 79
3.4 Ambiguity attitudes 83
3.4.1 Comparative and absolute ambiguity attitudes 83
3.4.2 Propensity for diversification 86
3.4.3 The robust interpretation 89
3.5 An exogenous global probability measure 90
3.6 Risk preferences given by the dual utility model 93
3.7 Conclusion 94
3.8 Results and proofs accompanying Section 3.3 95
3.9 Proofs accompanying Section 3.4 102
3.10 Results and proofs accompanying Section 3.5 108
3.11 Proofs accompanying Section 3.6 113
3.12 Modularity results for lattices 114

4 Aggregate then evaluate **116**

4.1 Introduction . . . 116

4.2 Preliminaries . . . 120

4.2.1 Standard definitions and notation . . . 120

4.2.2 The Savage setting . . . 121

4.2.3 The pure-risk setting, distributions, and statistics . . . 122

4.3 Two frameworks for separating pure risk and ambiguity . . . 125

4.3.1 The evaluate-then-aggregate framework . . . 125

4.3.2 The aggregate-then-evaluate framework . . . 127

4.4 Axioms and representation for the Choquet ATE model . . . 129

4.5 Risk and ambiguity attitudes . . . 135

4.5.1 Risk attitudes . . . 135

4.5.2 Comparative and absolute ambiguity attitudes . . . 139

4.5.3 Diversification seeking and aggregation concavity of act-to-distribution mappings . . . 141

4.5.4 The distributionally robust interpretation . . . 144

4.6 Conclusion . . . 144

4.7 Distorted capacities . . . 145

4.8 Proof accompanying Section 4.2 . . . 146

4.9 Proofs, discussions, and additional results accompanying Section 4.3 . . . 147

4.10 Proofs accompanying Section 4.4 . . . 152

4.11 Proofs and additional results accompanying Section 4.5 . . . 156

5 Distributionally robust optimization under Bayesian ambiguity sets **169**

5.1 Introduction . . . 169

5.2 Preliminaries . . . 172

5.2.1 General notation . . . 172

5.2.2 Our setting . . . 172

5.3 Decision functionals incorporating second-order models 175
5.3.1 The smooth model of ambiguity 175
5.3.2 Ambiguity-reduction mappings 177
5.3.3 The canonical example 179
5.4 Numerical examples 180
5.4.1 Parametrization 180
5.4.2 Convergence result 181
5.4.3 Alternative formulas and numerical results 182
5.5 Connections to existing literature 187
5.5.1 Unambiguous events 187
5.5.2 Bayesian consistency 188
5.5.3 Model misspecification 189
5.6 Conclusion 191
5.7 Additional results needed for the proofs 192
5.8 Proofs accompanying Section 5.2 194
5.9 Proofs accompanying Section 5.3 196
5.10 Proofs and additional results accompanying Section 5.4 198
5.11 Proofs and additional results accompanying Section 5.5 205
5.12 Omitted verifications 211

6 Conclusion and future work **213**

References **214**

List of Figures

1.1 The problem of identifiability 3

2.1 Varying β, while fixing $m_1 = 0$, $\sigma_1 = 0.1$, $\sigma_2 = 0.1$, $c = 0.5$, $\alpha = 0.95$; the red dotted curve connects the optimizers for different β 36

2.2 Varying σ_2, while fixing $m_1 = 0$, $\sigma_1 = 0.1$, $c = 0.5$, $\alpha = 0.95$, $\beta = 0.95$; the red dotted curve connects the optimizers for different σ_2 37

4.1 The ETA approach vs the ATE approach for the utility of act B 119

5.1 Worst-case Expected Shortfall 185

5.2 Worst-case entropic risk measure 186

5.3 Worst-case Value-at-Risk 187

List of Tables

5.1 Second-order models used in the numerical implementations 183

Chapter 1

Introduction

Decision theory is concerned with preference relations, i.e., total preorders over objects of choice, or *acts*, which express a decision maker's (DM's) personal ranking of those acts. These relations are traditionally characterized by *representation functionals*: real-valued functionals on the space of acts that translate the preference relation into the natural ordering of the reals. A *decision model* is understood as a class of preference relations whose representation functionals share a common structural form. A central aim of the field is to *axiomatize* decision models, that is, to identify conditions on preference relations that are normatively compelling and fully characterize representation functionals of the desired form. We now describe some common settings used in decision theory, each distinguished by its definition of the space of acts.

In the *pure-risk setting*, acts are probability distributions over an outcome space, and a representation functional is called a *statistic*. The most prominent decision model here is the expected utility (EU) model, axiomatized by von Neumann and Morgenstern [1947], which has long served as a cornerstone of the field. Despite its central role, the pure-risk setting is often too restrictive, as the space of probability distributions lacks the structure needed to define certain relevant operations, e.g., the outcome-mixture operation used to define diversified acts.

To move beyond the pure-risk setting, Savage [1954] reformulated acts as functions from states of the world to consequences, a setting we refer to as the *Savage setting*. By imposing normatively motivated axioms on a DM's preferences over such acts, Savage [1954] showed that (a) a subjective probability measure could be elicited for the DM, transforming the preference relation into the pure-risk setting and (b) within this elicited pure-risk setting, the DM's preferences are consistent with the EU model. This decision model is known as the subjective expected utility (SEU) model. The SEU model also motivates the

broader class of probabilistically sophisticated decision models, as described in Machina and Schmeidler [1992]. In such models, a representation functional is given by the composition of a probability measure and a statistic: the measure assigns a distribution to each act, and the statistic evaluates that distribution. A representation functional of this form is said to be *law-invariant*. Probabilistically sophisticated decision models arise across a range of fields, including finance, economics, actuarial science, and operations research. Notable examples include the mean-variance model of Markowitz [1952] in portfolio optimization, the rank-dependent utility (RDU) model of Quiggin [1982] in behavioral economics, and the coherent risk measures of Artzner et al. [1999] in financial risk management.

Though well-suited to many decision problems, the SEU model, and more broadly, probabilistically sophisticated decision models, have important limitations. Most notably, Ellsberg [1961] showed that when a DM possesses asymmetric information about event likelihoods, Savage's sure-thing principle, a key axiom of the SEU model, may fail to capture plausible behavior. Ellsberg's critique was pivotal in launching the formal study of decision-making under *ambiguity*, also called Knightian uncertainty or model uncertainty. In practice, ambiguity arises whenever the information available to a DM is insufficient to pin down a single probability measure. Moreover, neglecting ambiguity can lead to poor decisions with detrimental outcomes. The 2008 financial crisis serves as a compelling case in point: widespread failure to recognize the ambiguity inherent in prevailing risk models contributed to the catastrophic failure of the financial system (e.g., McNeil et al., 2015).

Several decision models have been developed to accommodate ambiguity. A non-exhaustive list of these models is: the Choquet expected utility (CEU) model of Schmeidler [1989], the maxmin expected utility (MEU) model of Gilboa and Schmeidler [1989], the α-maxmin model of Ghirardato et al. [2004], the smooth ambiguity model of Klibanoff et al. [2005], and the variational preferences model of Maccheroni et al. [2006]. While the majority of these decision models were originally axiomatized in the Anscombe Aumann (AA) setting (see Anscombe and Aumann, 1963), where acts are functions from states of the world to probability distributions over consequences, they can be formulated in the Savage setting, which is adopted throughout this thesis.

Each of the above models represents ambiguity through a specific component of the representation functional. Given a preference relation in one of these models, this component can be replaced with a new input representing the perception of no ambiguity. Crucially, the resulting preference relation will be an instance of the SEU model and can, therefore, be identified with a statistic consistent with the EU model. We refer to this statistic as the *pure-risk component*, and say that the above decision models have EU *pure-risk projections*. The main goal of this thesis is to study and axiomatize decision models whose pure-risk projections need not be consistent with EU.

1.1 Models with arbitrary pure-risk projections

Decision models under ambiguity with arbitrary pure-risk projections have appeared in the literature. The most prominent among these, motivated by the MEU model, are the *distributionally robust optimization (DRO)* models. To define a DRO model, one must first fix a class of statistics. A preference relation in the model is then specified by a statistic from that class and an *ambiguity set*, that is, a collection of probability measures. The corresponding representation functional evaluates an act by applying the statistic to its distribution under each measure in the ambiguity set and selecting the worst-case value. The pure-risk projection of this DRO model is determined by the chosen class of statistics. For instance, if this class is consistent with RDU, the resulting DRO model has an RDU pure-risk projection. When this class is only consistent with EU, the DRO model coincides with the MEU model. Axiomatizations of DRO models have been provided by Delage et al. [2019] and Fadina et al. [2024], though both take the ambiguity set as given, albeit in different ways. To my knowledge, the only existing axiomatizations that also elicit the ambiguity set from preferences are for the MEU model (which is only a single example of a DRO model).

While DRO models appear frequently in the chapters that follow, particularly in Chapters 2 and 5, the broader focus of this thesis is on decision models with arbitrary pure-risk projections that generalize, or provide alternatives to, the CEU and smooth models. A central challenge in the axiomatic work is uniquely identifying the pure-risk component from preferences. To see why, consider the representation functional ρ shown below in Figure 1.1, which utilizes the Donsker–Varadhan variational formula. Here, Δ denotes the set of probability measures, $\mu \in \Delta$, $\mathrm{H}(\nu|\mu)$ is the Kullback–Leibler divergence, and $\beta > 0$.

$$\rho(X) = \underbrace{-\beta \log\left(\int_\Omega \exp(-X/\beta)\mathrm{d}\mu\right)}_{\text{Risk-averse/ambiguity-neutral}} = \underbrace{\inf_{\nu\in\Delta}\left\{\int_\Omega X\mathrm{d}\nu + \beta\mathrm{H}(\nu|\mu)\right\}}_{\text{Risk-neutral/ambiguity-averse}}$$

Figure 1.1: The problem of identifiability

Assume we are only able to view the preference relation defined by the representation functional ρ in Figure 1.1. This preference relation is consistent with two decision models: the SEU model (in red) and the multiplier preferences model of Hansen and Sargent [2001] (in blue). In the SEU model, the DM perceives no ambiguity (or is ambiguity neutral)

and has the utility function $u(x) = -\exp(-x/\beta)$. In the multiplier preferences model, the DM perceives ambiguity, is ambiguity averse, and has a linear utility function. Since the utility function differs between the two interpretations, the pure-risk component cannot be identified without first specifying which decision model the preference relation is from. This identifiability issue is further complicated when the pure-risk component is not restricted to EU. Consider, for instance, a probabilistically sophisticated preference relation constructed from a statistic in the RDU model: from preferences alone, one cannot determine whether the probability distortion is a feature of the pure-risk component, or a consequence of the DM's reaction to a 'uniform' perception of ambiguity.

To resolve the identifiability problem, I employ the concept of *unambiguous events*. These are events equipped with a *likelihood mapping*: a function that behaves like a probability measure and assigns a unique likelihood to each unambiguous event. Acts that are measurable with respect to the unambiguous events are called *unambiguous acts*, and under appropriate assumptions, each unambiguous act can be associated with a unique distribution over consequences (an act in the pure-risk setting). The key observation is that, provided the unambiguous events are sufficiently rich, the unambiguous acts can be used to identify the pure-risk component of the DM's preferences.

To make use of unambiguous events, one must first determine how they are identified. There are three natural approaches.

(1) Determine the unambiguous events directly from preferences.

(2) Assume the unambiguous events are specified exogenously.

(3) Assume the perception of ambiguity is specified exogenously, and then derive the unambiguous events from it.

Approach (1) has a rich history in the literature. For instance, Ghirardato et al. [2004] and Cerreia-Vioglio et al. [2011] provide a definition of unambiguous events for very general classes of decision models. When restricted to the MEU model, their definition yields a clean result: an event is unambiguous when all measures in the ambiguity set assign it the same probability. However, this work was carried out in the AA setting. When translated to the Savage setting using the ideas in Ghirardato et al. [2003], the resulting decision models necessarily have EU pure-risk projections. Moreover, the mathematical machinery employed does not readily extend to arbitrary pure-risk components, rendering a similar approach infeasible.

Another direction taken for Approach (1) was Epstein and Zhang [2001], where the authors worked in the Savage setting. A notable advantage of this work is that it allows for

decision models with arbitrary pure-risk projections. However, the main obstacle to adopting their definition is that any probabilistically sophisticated preference relation deems all events unambiguous. To see why this is problematic, consider the multiplier preferences model of Hansen and Sargent [2001] shown in blue in Figure 1.1. This definition would classify every event as unambiguous, even though the DM faces a 'uniform' perception of ambiguity and strict aversion to it. For these reasons, I do not use Approach (1) in this thesis. For more on Approach (1), and the connections between Ghirardato et al. [2004] and Epstein and Zhang [2001], see Nehring [1999], Amarante and Filiz [2007] and Klibanoff et al. [2011].

In this thesis, I adopt Approach (2) for the axiomatic work. With the unambiguous events specified exogenously, a key axiom linking preferences to the unambiguous acts is *partial law invariance*, also called *risk conformity*.[1] Partial law invariance requires that the DM be indifferent between unambiguous acts that share the same distribution under the likelihood mapping, a natural condition reflecting that, in the absence of ambiguity, only the distribution of an act should matter. The notion of partial law invariance is inspired by, and closely related to, the work of Epstein [2004] and Baillon et al. [2025].

1.2 Structure of the thesis

The following four chapters are based on papers written during my PhD, where I served as lead author on each. Each chapter is self-contained and may be read independently; all notation introduced in a chapter is local to that chapter.

Chapter 2 presents a dedicated study of partial law invariance in the context of risk measures. Here, partial law invariance is examined with greater emphasis on its operational value. In the remaining chapters, by contrast, it serves as an idealized axiom used to motivate decision models that remain applicable even when, in practice, a rich collection of unambiguous events cannot be specified. As partial law invariance is a central axiom used in the axiomatic work of the subsequent chapters, this chapter provides a natural starting point for the thesis. Furthermore, this chapter provides interesting ambiguity sets for DRO decision models that do not assume a 'uniform' perception of ambiguity, unlike many ambiguity sets found in the literature. Beyond its role within the thesis, this chapter presents mathematical results relevant to the literature on risk measures.

Chapter 3 axiomatizes the Choquet rank-dependent utility (CRDU) model, a conceptual generalization of the CEU model whose pure-risk projection is consistent with the

[1] Both terminologies are used in the following chapters.

RDU model. We show that, in this model, the coupling of ambiguity perception and ambiguity attitude can be fully characterized by a matching probability, and that the supermodularity of this matching probability connects the CRDU model to a DRO model. The clear separation between pure risk and ambiguity further enables a rich analysis of the interplay between ambiguity attitudes and risk attitudes.

Chapter 4 describes two conceptual frameworks for decision models under ambiguity with arbitrary pure-risk projections: evaluate-then-aggregate (ETA), which first evaluates an act under each plausible model and then aggregates the results, and aggregate-then-evaluate (ATE), which first reduces ambiguity to a single representative distribution before evaluation. Since most existing decision models under ambiguity fall within the ETA framework, we develop and axiomatize the Choquet ATE model, which generalizes both the CEU and CRDU models and allows for arbitrary pure-risk projections. Similar to Chapter 3, we provide a detailed analysis of the interplay between ambiguity attitudes and risk attitudes within the Choquet ATE model.

Chapter 5 develops a DRO model with an arbitrary pure-risk projection, whose ambiguity set is derived from a Bayesian second-order belief. This DRO model is motivated by the observation that the smooth model does not generalize well to one that allows an arbitrary pure-risk projection. After motivating the model, we introduce a canonical construction based on distorted second-order beliefs, develop tractable algorithms, and illustrate them with portfolio optimization examples. Final remarks and avenues for future research are discussed in Chapter 6.

Chapter 2

Partial law invariance and risk measures

2.1 Introduction

Risk measures are popular tools in financial regulation and decision making. In finance, the usual interpretation of a risk measure is that it assigns a monetary amount to each financial position, representing the capital that should be held alongside the position to make it acceptable. In decision-theoretic terms, a risk measure is often interpreted as the unique certainty equivalent representing a preference relation over random variables, typically with some financially relevant properties. Risk measures such as the Value-at-Risk (VaR) and the Expected Shortfall (ES, also known as CVaR) are implemented in the financial regulatory framework of BCBS [2019], and they are also useful tools in optimization; see Föllmer and Schied [2016] and McNeil et al. [2015] for general treatments of risk measures in risk management and finance, and Shapiro et al. [2021] in the context of optimization.

Law invariance is the fundamental property that makes risk measures, as well as many decision models, applicable in practice. Once a reference probability model is specified, either through statistical estimation or expert opinion, law-invariant risk measures evaluate financial positions that share the same distribution under this probability model equally. This connects the theory of risk measures to statistics, thus allowing for a description of the probabilistic structure of financial positions, arguably a basic step in any quantitative financial analysis. In decision theory, law invariance is often known as probabilistic sophistication (Machina and Schmeidler [1992]).

Although law invariance is a fundamental concept, it may not be universally desirable in all contexts. For instance, model misspecification, or model uncertainty, occurs when the assumed probabilistic model for various financial positions does not accurately represent the true, unknown, distribution of these positions. This can be problematic as the value of a law-invariant risk measure may differ significantly between the assumed model and the objective truth. For instance, during the 2008 financial crisis, the probability model used for default rates of mortgages was questionable, and the loss distribution for mortgage-backed securities was underestimated, leading to insufficient capital requirements for many companies to remain solvent (e.g., McNeil et al. [2015]). Although we take a risk measure perspective in this chapter, the problem of model uncertainty extends beyond this and can affect any (usually statistical) process that relies on estimating a probabilistic model, followed by computing a quantity based on that model. We will generally refer to uncertainty in the probabilistic model as "model uncertainty", which is also known as Knightian uncertainty or ambiguity in the literature. There are many avenues of research that account for model uncertainty in risk assessment in different ways. One area is distributionally robust optimization, in which optimization with risk measures is performed against a worst-case scenario; see e.g., El Ghaoui et al. [2003], Natarajan et al. [2008], and Zhu and Fukushima [2009]. Another area is to study risk measures formulated from multiple scenarios; see e.g., Delage and Li [2018], Delage et al. [2019] and Wang and Ziegel [2021]. Model uncertainty is deeply rooted in decision theory, often called ambiguity in that literature. The goal there is to model preferences of decision makers who do not know the precise likelihood of events that govern their outcomes. Ambiguity is the key ingredient in the influential work of Choquet expected utility and maxmin expected utility pioneered by Schmeidler [1989] and Gilboa and Schmeidler [1989], followed by, e.g., the smooth ambiguity model of Klibanoff et al. [2005] and the variational preferences of Maccheroni et al. [2006].

Most of the literature, including the works cited above, addresses model uncertainty through three interrelated approaches. The first approach models financial positions as real-valued random variables (measurable functions), starting with a fixed reference probability model and a law-invariant risk measure, such as VaR, ES, or a function that represents expected utility preferences, that captures the risk preferences under full confidence in the reference model (that is, pure risk). Model uncertainty is then incorporated by aggregating the risk measure across a family of alternative probability models, often close to the reference probability in some way. This aggregation is typically the worst-case aggregation, as is common in distributionally robust optimization. The second approach, as in Artzner et al. [1999] and Föllmer and Schied [2016], also models financial positions as measurable functions, but it defines a risk measure directly on the space of these functions without fixing a reference probability model. This method handles model uncertainty by never supposing a reference probability model from the start, and it does not need a con-

cept of pure risk (or preferences over pure risks). The third approach, common in decision theory, adopts the two-stage framework of Anscombe and Aumann [1963], by building preferences over state-dependent monetary lotteries. It begins with a law-invariant risk measure, typically one representing some expected utility preferences, but it is applied to the monetary lotteries, instead of the mappings from states to lotteries. Values of the risk measure over different states are aggregated in various ways such as in Schmeidler [1989], Gilboa and Schmeidler [1989], and Maccheroni et al. [2006]. In this setting, pure risk refers to constant mappings that map each state to the same lottery.

In this chapter, we take a novel approach. We model financial positions as measurable functions and begin with a reference probability model but not a law-invariant risk measure. To account for model uncertainty, risk measures are assumed to satisfy law invariance only partially, that is, only with respect to random outcomes depending on events with no model uncertainty. This approach to modeling risk measures can address situations where some events are subject to more uncertainty than others. Traditional approaches, such as distributionally robust optimization, cannot handle this setting, as these models of uncertainty cannot directly incorporate sub-σ-algebras.

The following example can motivate this idea. After estimating a probability model for various loss factors, it is natural that different factors will have different levels of model uncertainty. For example, this difference might result from a statistical inference procedure, where the confidence regions for parameters tied to certain loss factors could be larger or smaller than those for other loss factors, for instance, due to data scarcity. From this point, we can typically divide the loss factors into two categories: the loss factors for which we have no, or negligible, model uncertainty and those that have non-negligible model uncertainty. We will call all the events related to the loss factors in the first category the "model-certain source." A financial position is said to depend on this source if its outcome is determined by events within this source. The next step is to select a risk measure for decision making. Requiring law invariance would not be able to distinguish factors in or outside the model-certain source, which is not desirable. However, for financial positions depending on the model-certain source, it is reasonable to measure their risk level solely based on distributions given by the probability model, as these positions involve no model uncertainty; thus, a partial form of law invariance is desirable. With this guiding principle in mind, we introduce partial law invariance, based on a suitably chosen source, which is mathematically modeled by a sub-σ-algebra $\mathcal{G}$ of the whole model space. A risk measure is partially law invariant (with respect to $\mathcal{G}$) if it equates the risks of financial positions that (a) both belong to this source (i.e., $\mathcal{G}$-measurable) and (b) share identical distributions. This property reflects the discussion above with $\mathcal{G}$ being the model-certain source, but it is applicable beyond this interpretation.

The property of partial law invariance was inspired by, and closely related to, the recent paper of Baillon et al. [2025], where the authors formalize the idea of sources. In that paper, sources are defined as algebras of events, and the main idea is that uncertainty attitudes may differ between sources. We take a fixed source, corresponding to the source of risk in Baillon et al. [2025], allowing us to obtain mathematically tractable representation results. Our setting is later generalized to incorporate multiple sources and Section 2.4 discusses the connection between their setting and ours. A similar concept in decision theory involves having a σ-algebra with no ambiguity regarding the probability, as seen in the expected uncertainty utility model of Gul and Pesendorfer [2014].

We summarize the main contributions of the chapter below. As the first contribution, we introduce the concept of partial law invariance as a new and relevant property for risk measures and preference relations in Section 2.2.1, with two substantive motivating examples in Sections 2.2.2 and 2.2.3, and a connection to stochastic dominance in Section 2.3.1. As the second contribution, a main theoretical result is presented, Theorem 2.1 in Section 2.3.2, which is a representation for coherent risk measures (Artzner et al. [1999]) satisfying partial law invariance. Third, generalizing the idea of a single source in Sections 2.2–2.3, the case of multiple sources is studied with a representation result Theorem 2.2 in Section 2.4, where we also explain the differences between the source theory of Baillon et al. [2025] and our approach. Fourth, we offer several extensions of the above results. A technical tool, called coherent adjustments, in Section 2.5.1 leads to an alternative representation of coherent risk measures in Theorem 2.3. This enables us to derive Theorem 2.4, which extends the well-known Kusuoka's representation (Kusuoka [2001] and Frittelli and Rosazza Gianin [2005]) under a stronger notion of partial law invariance in Section 2.5.2. With no surprise to specialists in risk measures, most results can be smoothly generalized to the case of convex risk measures, and we present some of them in Section 2.5.3. Finally, we discuss some explicit classes of partially law-invariant risk measures. Formulas to compute the partially law-invariant ES are studied in Section 2.6.1. To provide some concrete example, two constructions of partially law-invariant coherent risk measures are discussed in Section 2.6.2. An application with numerical examples is presented in Section 2.6.3. Section 2.7 concludes the chapter. We will focus on an atomless probability space in the main part of the chapter. The analysis in the case of a finite probability space is presented in Section 2.14

2.2 Partial law invariance: Definition and motivating examples

2.2.1 Risk measures, preference relations, and partial law invariance

We, throughout, fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is our reference probability model. We denote by L^0 the set of real-valued measurable functions, where functions that agree $\mathbb{P}$-almost surely (a.s.) are identified, and by L^∞ and L^1 the subsets of random variables that are essentially bounded and integrable, respectively. A generic set $\mathcal{X} \subseteq L^0$ is interpreted as the collection of financial positions of interest, and $X \in \mathcal{X}$ represents the loss of the financial position for various states of the world $\omega \in \Omega$.[1] It will always be assumed that $\mathcal{X}$ contains the constant random variables, identified with elements of $\mathbb{R}$. A risk measure is a mapping $\rho : \mathcal{X} \to \mathbb{R}$, with the main interpretation of the amount of capital needed to make a financial loss X acceptable.

For a decision-theoretic formulation, let $\succsim$ be a total preorder over the set $\mathcal{X}$, representing the preferences of a decision maker over the financial positions. We will also make two regularity assumptions on $\succsim$.

(C1) For all $X \in \mathcal{X}$, $\{c \in \mathbb{R} : X \succsim c\}$ and $\{c \in \mathbb{R} : c \succsim X\}$ are closed.

(C2) If $a, b \in \mathbb{R}$ and $a > b$, then $b \succ a$. (Recall that elements of $\mathcal{X}$ represent losses.)

These assumptions allow us to guarantee the existence of a unique certainty equivalent, as described in the following standard result.

Proposition 2.1. *Under* (C1) *and* (C2)*, there exists a unique certainty equivalent for every $X \in \mathcal{X}$, that is, for every $X \in \mathcal{X}$, there exists a unique $c_X \in \mathbb{R}$ such that $X \sim c_X$.*

Proposition 2.1 implies that we can find a risk measure $\rho : \mathcal{X} \to \mathbb{R}$ to represent $\succsim$ via $X \succsim Y \iff \rho(X) \leq \rho(Y)$ (e.g., the certainty equivalent, but not necessarily). The reason for the reversal is to stay consistent with our convention that risk measures represent riskiness, thus the smaller the better. In what follows, we formulate properties for $\rho : \mathcal{X} \to \mathbb{R}$, unifying risk measures and preference relations.

[1] A positive value of X represents a loss, and negative value of X represents a profit; this is the standard loss/profit convention in the risk measure literature.

Now, we define our main concept of partial law invariance. Given a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, we define $\mathcal{X}(\mathcal{G}) = \{X \in \mathcal{X} : X \text{ is } \mathcal{G}\text{-measurable}\}$. To establish technical results, we will assume that $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, that is, $\{\mathbb{P}(A) : A \subseteq B;\ A \in \mathcal{G}\} = [0, \mathbb{P}(B)]$ for all $B \in \mathcal{G}$, although this assumption is not needed for defining our main concept of partial law invariance. For $X, Y \in \mathcal{X}$, we write $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$ if X and Y have the same distribution under $\mathbb{P}$, that is $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$ for all $x \in \mathbb{R}$. For $\rho : \mathcal{X} \to \mathbb{R}$ (resp. for $\succsim$ on $\mathcal{X}$), we consider the following property.

> $\mathcal{G}$-law invariance: $\rho(X) = \rho(Y)$ (resp. $X \sim Y$) for all $X, Y \in \mathcal{X}(\mathcal{G})$ that satisfy $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$.

When $\mathcal{G}$ is the largest σ-algebra $\mathcal{F}$, the obtained $\mathcal{F}$-law invariance is the widely studied *law invariance* of the risk measure literature (or, *probabilistic sophistication* in decision theory). As far as we know, $\mathcal{G}$-law invariance with $\mathcal{G} \neq \mathcal{F}$ has not been studied in the literature. It is easy to see that law invariance is stronger than $\mathcal{G}$-law invariance. When making a contrast, we will refer to $\mathcal{G}$-law invariance as *partial law invariance* and to $\mathcal{F}$-law invariance as *full law invariance*, where the σ-algebras should be clear from the context. We also say that a set $\mathcal{X} \subseteq L^0$ is $\mathcal{G}$-law invariant if $X \in \mathcal{X} \cap L^0(\mathcal{G})$ and $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y \in L^0(\mathcal{G})$ imply $Y \in \mathcal{X}$.

To get a better intuition for partial law invariance, let us consider monotonicity of preferences and risk measures. For $X, Y \in \mathcal{X}$, we write $X \geq_1 Y$ if $\mathbb{P}(X \leq x) \leq \mathbb{P}(Y \leq x)$ for all $x \in \mathbb{R}$, which is the usual first-order stochastic dominance (FSD). We can formulate two notions of monotonicity.

> Monotonicity: $X \succsim Y$ for all $X, Y \in \mathcal{X}$ with $Y \geq X$ (in the sense of $\mathbb{P}$-a.s. throughout).
>
> $\mathcal{G}$-FSD monotonicity: $X \succsim Y$ for all $X, Y \in \mathcal{X}(\mathcal{G})$ with $Y \geq_1 X$.

The following simple result clarifies the connection between $\mathcal{G}$-FSD monotonicity and $\mathcal{G}$-law invariance.

Proposition 2.2. *Suppose that the preference relation $\succsim$ is monotonic and $\mathcal{X}$ is $\mathcal{G}$-law invariant. Then $\succsim$ is $\mathcal{G}$-FSD monotonic if and only if $\succsim$ is $\mathcal{G}$-law invariant.*

Therefore, partial law invariance can be equivalently interpreted as consistency with respect to FSD over $\mathcal{G}$-measurable random variables.

The main motivations for partial law invariance, in the absence of full law invariance, are model uncertainty in risk evaluation, illustrated in Section 2.2.2, as well as ambiguity in decision making, illustrated in Section 2.2.3. In the following examples, and for the rest of the chapter, we write $\sigma(\cdot)$ for the σ-algebra generated by either a random variable, a collection of random variables, or a collection of sets in $\mathcal{F}$.

2.2.2 Partial law invariance and model uncertainty

In this section we elaborate on the role of partial law invariance in the context of model uncertainty.

Suppose a firm's portfolio manager is tasked with making investment decisions involving the financial positions in $\mathcal{X}$. Among the random variables in $\mathcal{X}$, the manager knows that the distribution of a financial position Z is accurate under the probability model $\mathbb{P}$; that is, there is no (or negligible) model uncertainty. For instance, Z may be the losses from a regular property insurance portfolio, for which there are enough data to confidently estimate its distribution. The distribution of the other random variables X under the probability model may be subject to a considerable amount of model uncertainty, such as the losses of a stock without a sufficiently long history of data. The manager needs to decide on a risk measure. In this context, law invariance naturally applies when dealing with positions $g(Z)$ that are functions of Z because the manager is confident about their distributions. However, for random variables X that are not $\sigma(Z)$-measurable, if $X \overset{\mathrm{d}}{=}_{\mathbb{P}} g(Z)$, it is unreasonable to insist that X and $g(Z)$ are equally risky due to the model uncertainty associated with X; the same can be said for two general random variables X and Y. In other words, it is natural to require $\sigma(Z)$-law invariance but not (full) law invariance.

Denote by $\mathcal{M}_1$ the set of probability measures on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to $\mathbb{P}$. The manager may choose a collection of probability measures $\mathcal{Q} \subseteq \mathcal{M}_1$ to reflect model uncertainty. Because Z is the loss whose distribution the manager is certain about, the set $\mathcal{Q}$ should satisfy that for all $\mu \in \mathcal{Q}$, $F_{\mu,Z} = F_{\mathbb{P},Z}$, where $F_{\mu,Z}$ is the cumulative distribution function (cdf) of Z under μ. For example, the manager could take

$$\mathcal{Q} = \{\mu \in \mathcal{M}_1 \mid F_{\mu,Z} = F_{\mathbb{P},Z} \text{ and } d(\mu, \mathbb{P}) < \epsilon\}, \tag{2.1}$$

where $\epsilon > 0$ and $d : \mathcal{M}_1 \times \mathcal{M}_1 \to [0, \infty]$ is some measure of divergence, such as a Wasserstein distance or the Kullback–Leibler divergence. For $X \in \mathcal{X}$, the manager needs to evaluate the risk of X under each $\mu \in \mathcal{Q}$, and then aggregate these evaluations in some way. For instance, if the manager uses the mean to quantify risk and the worst-case approach of

Wald [1945, 1949] (axiomatized by Gilboa and Schmeidler [1989]) to aggregate the risk evaluations, then the resulting risk measure is $\rho : X \mapsto \sup_{\mu\in\mathcal{Q}} \mathbb{E}^{\mu}[X]$. We can see that ρ is $\sigma(Z)$-law invariant because $\rho(X) = \mathbb{E}^{\mathbb{P}}[X]$ for X that is $\sigma(Z)$-measurable.

Using the mean as a risk measure is overly simplistic. A risk measure common in financial practice is the Expected Shortfall (ES also known as CVaR). ES serves as the standard risk measure for market risk in the banking regulatory framework of BCBS [2019]; see Rockafellar and Uryasev [2002] for its optimization and Wang and Zitikis [2021] for an axiomatization. For a given probability measure $\mu \in \mathcal{M}_1$, the ES at level $\alpha \in [0, 1)$ is given by

$$\mathrm{ES}^{\mu}_{\alpha}(X) = \min_{x\in\mathbb{R}} \left(x + \frac{1}{1-\alpha}\mathbb{E}^{\mu}[(X-x)^{+}] \right), \qquad X \in L^{\infty}.$$

The function $\mathrm{ES}^{\mu}_{\alpha}(X)$ can be defined as the average of the left-quantile function of X under μ from α to 1, and also on larger spaces in L^0; see e.g., Rockafellar and Uryasev [2002] and McNeil et al. [2015, Chapter 8]. For simplicity, we write $\mathrm{ES}_{\alpha} = \mathrm{ES}^{\mathbb{P}}_{\alpha}$. Following the guidelines of BCBS [2019], for each $\mu \in \mathcal{Q}$, suppose that the decision maker uses $\mathrm{ES}^{\mu}_{\alpha}$ at level $\alpha = 0.975$ to evaluate the risk of $X \in L^{\infty}$ under μ. The resulting risk measure is the worst-case ES (see Zhu and Fukushima [2009]), given by

$$\overline{\mathrm{ES}}^{\mathcal{Q}}_{\alpha}(X) = \sup_{\mu\in\mathcal{Q}} \mathrm{ES}^{\mu}_{\alpha}(X), \qquad X \in L^{\infty}. \tag{2.2}$$

The risk measure $\overline{\mathrm{ES}}^{\mathcal{Q}}_{\alpha}$ with $\mathcal{Q}$ in (2.1) is $\sigma(Z)$-law invariant. Therefore, considering robust risk optimization will naturally lead to partial law invariance.

Another common way to assess model uncertainty is through the use of multiple prior probability measures. Law-invariant risk measures under multiple probability measures are studied by Wang and Ziegel [2021] under the name of scenario-based risk measures, which are formulated with the following property, for a set $\mathcal{Q} \subseteq \mathcal{M}_1$.

$$\mathcal{Q}\text{-basedness: } \rho(X) = \rho(Y) \text{ for all } X, Y \in L^{\infty} \text{ satisfying } F_{\mu,X} = F_{\mu,Y} \text{ for all } \mu \in \mathcal{Q}.$$

This property means that if two risks are identically distributed under all probability measures of interest described by $\mathcal{Q}$, then they are assessed as equally risky.

It is evident that law invariance is a particular case of both $\mathcal{G}$-law invariance, by choosing $\mathcal{G} = \mathcal{F}$, and $\mathcal{Q}$-basedness, by choosing $\mathcal{Q} = \{\mathbb{P}\}$. Moreover, law invariance is stronger in general than both $\mathcal{G}$-law invariance and $\mathcal{Q}$-basedness (if $\mathbb{P} \in \mathcal{Q}$). What is not so obvious is how $\mathcal{G}$-law invariance and $\mathcal{Q}$-basedness are connected, which we address next. For $\mu \in \mathcal{M}_1$, let $D_{\mu} = \mathrm{d}\mu/\mathrm{d}\mathbb{P} \in L^1$, that is, its Radon–Nikodym derivative with respect to $\mathbb{P}$.

Proposition 2.3. *If $\rho : L^\infty \to \mathbb{R}$ is $\mathcal{Q}$-based, then it is $\mathcal{G}$-law invariant for any $\mathcal{G}$ independent of $\sigma(D_\mu : \mu \in \mathcal{Q})$.*

This connection between $\mathcal{Q}$-basedness and $\mathcal{G}$-law invariance reinforces the idea that partial law invariance of a risk measure can serve as an indicator of sources without model uncertainty. In this context, observe that $\mathcal{Q}$ can be viewed as a set of competing probability models. The independence between D_μ and $\mathcal{G}$ for all $\mu \in \mathcal{Q}$ implies $\mu|_{\mathcal{G}} = \mathbb{P}|_{\mathcal{G}}$ for each $\mu \in \mathcal{Q}$. This means that all the competing probabilities in $\mathcal{Q}$ agree on $\mathcal{G}$, which can be directly interpreted as an absence of model uncertainty on $\mathcal{G}$, as in our motivating example.

2.2.3 Partial law invariance in decision models

We will explore a connection between matching probabilities (defined below) and partial law invariance. The idea of matching probability is used to analyze decisions under ambiguity; see e.g., Dimmock et al. [2016] and Baillon et al. [2018]. Suppose an agent has a preference relation $\succsim$ over a set of random variables $\mathcal{X}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ that includes the indicator functions. The agent does not have access to $\mathbb{P}$, so events in $\mathcal{F}$ can be associated with ambiguity to this agent. Some events, such as a coin toss or the simulation of a random number, are not ambiguous to this agent, and we denote by $\mathcal{G}$ the σ-algebra generated by these unambiguous events. For a decision scientist to understand the ambiguity attitude of the agent, the common approach is to associate each event $A \in \mathcal{F}$ with a number $p_A \in [0,1]$ that corresponds to the specified probability of an equally preferred event from $\mathcal{G}$, that is, $p_A = \mathbb{P}(B)$ where $B \in \mathcal{G}$ and $\mathbb{1}_A \sim \mathbb{1}_B$. A questionnaire for the agent would look like "would you prefer getting \$1 if event A happens, or getting \$1 with probability p". Assuming some regularity of the preference, a unique $p_A \in [0,1]$ exists. This p_A is called the matching probability of A. For this procedure to be sensible, an implicit assumption is that all events $A, B \in \mathcal{G}$ with $\mathbb{P}(A) = \mathbb{P}(B)$ satisfy $\mathbb{1}_A \sim \mathbb{1}_B$. Note that p_A is a numerical representation of the preference of the agent on $\{\mathbb{1}_A : A \in \mathcal{F}\}$, and this numerical representation satisfies $\mathcal{G}$-law invariance, but it should not, in general, satisfy full law invariance, as discussed below.

We will now turn to some classical models in decision theory under ambiguity, such as those of Gilboa and Schmeidler [1989], Klibanoff et al. [2005] and Maccheroni et al. [2006]. These models also naturally give rise to partially law-invariant functionals and preferences. We first describe a general framework that allows these models to be translated into our setting. Let $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ denote the probability space representing the potential future states of the world. Suppose a decision-maker does not possess knowledge of $\mathbb{P}_0$ but has access to a lottery-generating device, modeled by $((0,1], \mathcal{B}((0,1]), \lambda)$, where λ denotes the

Lebesgue measure. Putting these two sources of uncertainty together, one arrives at the probability space

$$(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_0 \times (0,1], \mathcal{F}_0 \otimes \mathcal{B}((0,1]), \mathbb{P}_0 \times \lambda),$$

where $\mathcal{F}_0 \otimes \mathcal{B}((0,1])$ is the product σ-algebra. Elements of L^∞ are called the Savage acts, and the decision maker has a preference relation $\succsim$ represented by $\rho : L^\infty \to \mathbb{R}$. Let $\mathcal{G} = \sigma(\pi)$, where $\pi : \Omega \to (0,1]$ is given by the projection onto the second coordinate. We refer to the elements of $L^\infty(\mathcal{G})$ as lotteries. This way of modeling lotteries with the product space structure is used by Sarin and Wakker [1992, 1997] and Klibanoff et al. [2005]. Denote by

$$\Delta = \{\mu \in \mathcal{M}_1 : \mu(A \times B) = \mu(A \times (0,1])\, \lambda(B) \text{ for all } A \in \mathcal{F}_0 \text{ and } B \in \mathcal{B}((0,1])\},$$

which represent the potential probability measures the decision maker needs to consider; note that the second marginal is always λ since there is no ambiguity about the lotteries. Many examples in decision theory can be formulated via the above setting, and we list three below. In each model, $\ell : \mathbb{R} \to \mathbb{R}$ is a loss function, with the corresponding function $u(x) = -\ell(-x)$ being a von Neumann-Morgenstern utility function. (a) The maxmin expected utility of Gilboa and Schmeidler [1989], given by

$$\rho(X) = \max_{\mu \in \mathcal{S}} \int_\Omega \ell(X)\, \mathrm{d}\mu = -\min_{\mu \in \mathcal{S}} \int_\Omega u(-X)\, \mathrm{d}\mu,$$

where $\mathcal{S} \subseteq \Delta$; (b) The smooth ambiguity model of Klibanoff et al. [2005], given by

$$\rho(X) = \int_\Delta \phi \left(\int_\Omega \ell(X)\, \mathrm{d}\mu \right) Q(\mathrm{d}\mu) = -\int_\Delta \widetilde{\phi} \left(\int_\Omega u(-X)\, \mathrm{d}\mu \right) Q(\mathrm{d}\mu),$$

where ϕ is an increasing transformation, $\widetilde{\phi}(x) = -\phi(-x)$, and Q is a probability measure over Δ; (c) The variational preferences model of Maccheroni et al. [2006], given by

$$\rho(X) = \max_{\mu \in \Delta} \left(\int_\Omega \ell(X)\, \mathrm{d}\mu - c(\mu) \right) = -\min_{\mu \in \Delta} \left(\int_\Omega u(-X)\, \mathrm{d}\mu + c(\mu) \right),$$

where $c : \Delta \to [0, \infty]$ satisfies $c(\mu) = 0$ for some $\mu \in \Delta$. Because measures in Δ have the same second marginal λ, all of the above models satisfy $\rho(Z) = \mathbb{E}^{\mathbb{P}}[\ell(Z)]$ for each lottery $Z \in L^\infty(\mathcal{G})$, and hence they are $\mathcal{G}$-law invariant, but clearly not law invariant in general.

To connect back to matching probabilities, assume that the agent's preference relation $\succsim$ is given by the maxmin expected utility model and $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ is atomless. Standard ar-

guments can justify that $\succsim$ is law invariant on $\{\mathbb{1}_A : A \in \mathcal{F}\}$ (i.e., the matching probability $A \mapsto P_A$ is law invariant) if and only if $\mathcal{S} = \{\mathbb{P}_0 \times \lambda\}$, i.e., the agent has no ambiguity.

2.3 Coherent risk measures and the associated preferences

We will study partial law invariance in the framework of coherent risk measures and their relationship to preferences. We focus on coherent risk measures for three reasons: First, they are prominent in financial regulation and portfolio optimization (e.g., McNeil et al. [2015] and Föllmer and Schied [2016]). Second, they give rise to tractable representation results that provide insights into the structure and implications of partial law invariance. Third, they can be translated into the max-min preferences of Gilboa and Schmeidler [1989] with suitable interpretation. For the rest of the chapter, we will assume that $\mathcal{X} = L^\infty$, as is common in the risk measure literature.

2.3.1 Preferences and risk measures

A *coherent risk measure* (Artzner et al. [1999]) is a mapping $\rho : L^\infty \to \mathbb{R}$ satisfying the following properties.

Cash invariance: $\rho(X + c) = \rho(X) + c$ for all $X \in L^\infty$ and $c \in \mathbb{R}$;

Monotonicity: $\rho(X) \leq \rho(Y)$ for all $X, Y \in L^\infty$ with $X \leq Y$;

Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for all $X, Y \in L^\infty$ and $\lambda \in [0, 1]$;

Positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$ for all $X \in L^\infty$ and $\lambda \geq 0$.

Moreover, a *convex risk measure* is a mapping that satisfies the first three properties in the above list, and a *monetary risk measure* is a mapping that satisfies the first two properties in the above list; see Frittelli and Rosazza Gianin [2002] and Föllmer and Schied [2002]. For background on risk measures, see Section 2.8.

A common continuity assumption in the risk measure literature is Fatou continuity. For $\rho : L^\infty \to \mathbb{R}$, we say that ρ is *Fatou (continuous)* if for all uniformly bounded sequences $(X_n)_{n\in\mathbb{N}}$ in L^∞ that converge $\mathbb{P}$-a.s. to some $X \in L^\infty$, it holds that $\rho(X) \leq$

$\liminf_{n\to\infty} \rho(X_n)$. Fatou continuity is essential because a risk measure $\rho : L^\infty \to \mathbb{R}$ is Fatou and coherent if and only if there exists $\mathcal{S} \subseteq \mathcal{M}_1$ such that

$$\rho(X) = \sup_{\mu\in\mathcal{S}} \mathbb{E}^\mu[X], \qquad X \in L^\infty; \tag{2.3}$$

see Proposition 2.10 in Section 2.8 for a rigorous statement of this result. This representation is significant for two reasons. First, it links coherent risk measures to the maxmin expected utility model of Gilboa and Schmeidler [1989]. With a given loss function ℓ, the random variables $\ell(X)$ can be interpreted as a subjective loss; symmetrically, for a utility function u and random wealth X, $u(X)$ is known as the util of X. Applying coherent risk measures to these subjective losses (or "dis-utils") agrees with the maxmin expected utility model of Gilboa and Schmeidler [1989]. Second, the worst-case expectation is evaluated with respect to a set of probability measures instead of finitely additive probabilities (which is the case without Fatou continuity). This distinction highlights the practical relevance of Fatou coherent risk measures, as statistical procedures yield probability measures rather than additive probabilities.

Preference relations that are represented by coherent risk measures will be called *CRM preference relations*, and they are described by a few properties; see Drapeau and Kupper [2013] and Proposition 2.12 in Section 2.10. For CRM preference relations, $\mathcal{G}$-law invariance is connected to second-order stochastic dominance (SSD), which we show below. For $X, Y \in L^\infty$, we write $X \geq_2 Y$ if $\mathbb{E}[\ell(X)] \geq \mathbb{E}[\ell(Y)]$ for all increasing convex functions $\ell : \mathbb{R} \to \mathbb{R}$. In the literature, the common definition is that for two random variables Z and W representing gains, Z dominates W in SSD if $\mathbb{E}[u(Z)] \geq \mathbb{E}[u(W)]$ for all increasing concave functions u. This is equivalent to $-W \geq_2 -Z$, via the connection $\ell(x) = -u(-x)$, where $-W$ and $-Z$ represent losses. Next, we define SSD consistency.

$\mathcal{G}$-SSD monotonicity: $X \succsim Y$ for all $X, Y \in L^\infty(\mathcal{G})$ with $Y \geq_2 X$.

In general, $\mathcal{G}$-SSD monotonicity implies $\mathcal{G}$-FSD monotonicity because $\geq_1$ is stronger than $\geq_2$. These two notions of monotonicity are equivalent for CRM preferences, proved via standard arguments in the literature.

Proposition 2.4. *For a CRM preference relation $\succsim$, the following are equivalent: (i) $\mathcal{G}$-law invariance; (ii) $\mathcal{G}$-FSD monotonicity; (iii) $\mathcal{G}$-SSD monotonicity.*

In the rest of the chapter, we will focus on risk measures instead of preference relations. Given a monetary risk measure $\rho : L^\infty \to \mathbb{R}$, we call $\mathcal{A}_\rho = \{X \in L^\infty \mid \rho(X) \leq 0\}$ the *acceptance set* of ρ. Acceptance sets are a standard way to describe monetary risk measures; for some classical results regarding acceptance sets, see Section 2.8. The following

proposition relates partial law invariance and the acceptance set in the case of a monetary risk measure.

Proposition 2.5. *A monetary risk measure ρ is $\mathcal{G}$-law invariant if and only if $\mathcal{A}_\rho$ is $\mathcal{G}$-law invariant.*

2.3.2 Representation of partially law-invariant coherent risk measures

Our next task is to understand the structure of all $\mathcal{G}$-law invariant coherent risk measures. First, we need some notation. We denote by $\mathcal{M}(\mathcal{G})$ (resp. $\mathcal{M}_1(\mathcal{G})$) the set of finite signed measures (resp. probability measures) on $(\Omega, \mathcal{G})$ that are absolutely continuous with respect to $\mathbb{P}|_{\mathcal{G}}$. We write $\mathcal{M} = \mathcal{M}(\mathcal{F})$, and $\mathcal{M}_1 = \mathcal{M}_1(\mathcal{F})$ for simplicity and to stay consistent with the notation from earlier. If $\mu \in \mathcal{M}_1(\mathcal{G})$, we will continue to use the notation D_μ to denote the Radon–Nikodym derivative with respect to $\mathbb{P}|_{\mathcal{G}}$, as $\mathrm{d}\mu/\mathrm{d}\mathbb{P}|_{\mathcal{G}} \in L^1(\mathcal{G})$. For $\mathcal{S} \subseteq \mathcal{M}(\mathcal{G})$ or $\mathcal{S} \subseteq \mathcal{M}$, we write $\widehat{\mathcal{S}} = \{D_\mu : \mu \in \mathcal{S}\}$, which is contained in $L^1(\mathcal{G})$ or L^1. Furthermore, in both cases, $\mathcal{S}$ is said to be $\mathcal{G}$-law invariant if the set $\widehat{\mathcal{S}}$ is $\mathcal{G}$-law invariant (defined in Section 2.2.1). For notational simplicity, when $\mu \in \mathcal{M}_1(\mathcal{G})$ we define $\mathbb{E}^\mu$ on L^∞ by $\mathbb{E}^\mu[X] = \mathbb{E}[D_\mu X]$. For $\mathcal{S} \subseteq \mathcal{M}$, we write $\mathcal{S}^{\mathcal{G}} = \{\mu|_{\mathcal{G}} : \mu \in \mathcal{S}\} \subseteq \mathcal{M}(\mathcal{G})$. Similarly, for $\mathcal{B} \subseteq L^1$, we write $\mathcal{B}^{\mathcal{G}} = \{\mathbb{E}[X|\mathcal{G}] : X \in \mathcal{B}\} \subseteq L^1(\mathcal{G})$. There is no ambiguity when we use the notation $\widehat{\mathcal{R}}^{\mathcal{G}}$ because $(\widehat{\mathcal{R}})^{\mathcal{G}} = \widehat{(\mathcal{R}^{\mathcal{G}})}$. Given $\mu \in \mathcal{M}_1(\mathcal{G})$, let $\mathcal{E}(\mu) = \{\nu \in \mathcal{M}_1 : \nu|_{\mathcal{G}} = \mu\}$. Clearly, $\mathcal{E}(\mu)$ depends on $\mathcal{G}$, which will be implicitly fixed in all places. If $\mu \in \mathcal{M}_1$, we will denote $\mathcal{E}(\mu|_{\mathcal{G}})$ by $\mathcal{E}(\mu)$ for simplicity.

Let $\mathfrak{X}$ and $\mathfrak{Y}$ be two vector spaces. Given a dual pairing $(\mathfrak{X}, \mathfrak{Y}, \langle\ ,\ \rangle)$ let $w(\mathfrak{X}, \mathfrak{Y})$ denote the weak topology on $\mathfrak{X}$. This makes $(\mathfrak{X}, w(\mathfrak{X}, \mathfrak{Y}))$ a locally convex (Hausdorff) topological vector space. Given some $\mathcal{R} \subseteq \mathfrak{X}$, let $\mathrm{conv}(\mathcal{R})$ denote the convex hull of $\mathcal{R}$. It is well known that $(L^\infty, \mathcal{M}, \langle\ ,\ \rangle)$ is a dual pairing under $\langle X, \mu\rangle = \mathbb{E}^\mu[X] = \mathbb{E}\left[D_\mu X\right]$. It is also well known that $(L^\infty, L^1, \langle\ ,\ \rangle)$ is a dual pairing under $\langle X, Y\rangle = \mathbb{E}[XY]$. Indeed, $w\left(L^\infty, \mathcal{M}\right)$ and $w\left(L^\infty, L^1\right)$ coincide since $\left(\mathcal{M}, w\left(\mathcal{M}, L^\infty\right)\right) \cong \left(L^1, w\left(L^1, L^\infty\right)\right)$ through the mapping $\mu \mapsto D_\mu$.

For a Fatou coherent risk measure $\rho : L^\infty \to \mathbb{R}$, its *supporting set* is defined as:

$$\mathcal{S}_\rho = \left\{\mu \in \mathcal{M}_1 : \mathbb{E}^\mu[X] \leq \rho(X) \text{ for all } X \in L^\infty\right\}.$$

Using the subjective loss interpretation for the random variables, we see that the supporting set is the set of priors in the the maxmin expected utility model of Gilboa and Schmeidler [1989].

We are now ready to present a general representation for Fatou $\mathcal{G}$-law-invariant coherent risk measures reminiscent of representation (2.3).

Theorem 2.1. *For a mapping $\rho : L^\infty \to \mathbb{R}$, the following are equivalent.*

(i) The risk measure ρ is Fatou continuous, $\mathcal{G}$-law invariant, and coherent;

(ii) there exists a convex $\mathcal{S} \subseteq \mathcal{M}_1$ such that the $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-closure of $\mathcal{S}^\mathcal{G}$ is $\mathcal{G}$-law invariant and

$$\rho(X) = \sup_{\mu \in \mathcal{S}} \mathbb{E}^\mu[X], \qquad X \in L^\infty; \tag{2.4}$$

(iii) the $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-closure of $(\mathcal{S}_\rho)^\mathcal{G}$ is $\mathcal{G}$-law invariant.

We say that $\mathcal{S} \subseteq \mathcal{M}_1(\mathcal{G})$ is almost $\mathcal{G}$-law invariant if the $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-closure of $\mathcal{S}$ is $\mathcal{G}$-law invariant as in parts (ii)–(iii) of Theorem 2.1. It is well-known that a Fatou coherent risk measure ρ is law invariant if and only if its supporting set $\mathcal{S}_\rho \subseteq \mathcal{M}_1$ is law invariant; see Föllmer and Schied [2016, Theorem 4.59]. Theorem 2.1 is reminiscent of this result, but with some subtle differences; in particular, one needs to take the restriction of elements of $\mathcal{S}$, as well as a closure of the resulting set. The first operation is clearly needed, and the second one seems essential. For a discussion of this condition and a proof of Theorem 2.1, see Section 2.10.2.

Example 2.1. Let ρ be the mapping $X \mapsto \mathbb{E}^\nu[X]$ where $\nu \in \mathcal{E}(\mathbb{P})$. Clearly, ρ is $\mathcal{G}$-law invariant, and the only possible set $\mathcal{S}$ in (2.4) is given by $\{\nu\}$. We have $\mathcal{S}^\mathcal{G} = \{\mathbb{P}|_\mathcal{G}\}$, which is almost $\mathcal{G}$-law invariant.

Example 2.2. Using the subjective loss interpretation for the random variables, we see the preference relations related to the maxmin expected utility model of Gilboa and Schmeidler [1989] are $\mathcal{G}$-law invariant if and only if $\mathcal{S}^\mathcal{G}$ is almost $\mathcal{G}$-law invariant, where $\mathcal{S}$ is the set of priors. In the specific example given in Section 2.2.3, it is straightforward to show that $\mathcal{S}^\mathcal{G} = \{\tilde{\lambda}\}$, where $\tilde{\lambda}(\Omega_0 \times B) = \lambda(B)$ for all $B \in \mathcal{B}((0,1])$. Thus $\mathcal{S}^\mathcal{G}$ is almost $\mathcal{G}$-law invariant.

Theorem 2.1 also gives a feasible way to construct $\mathcal{G}$-law-invariant risk measures. We can take any law-invariant coherent risk measure $\tilde{\rho}$ on $L^\infty(\mathcal{G})$ (e.g., ES_α), and for each $\mu \in \mathcal{S}_{\tilde{\rho}} \subseteq \mathcal{M}_1(\mathcal{G})$ we take any set $\mathcal{S}(\mu) \subseteq \mathcal{E}(\mu)$. Setting

$$\rho(X) = \sup_{\nu \in \mathcal{S}} \mathbb{E}^\nu[X], \quad X \in L^\infty, \quad \text{where } \mathcal{S} = \bigcup_{\mu \in \mathcal{S}_{\tilde{\rho}}} \mathcal{S}(\mu), \tag{2.5}$$

yields a $\mathcal{G}$-law-invariant risk measure. Moreover, ρ coincides with $\tilde{\rho}$ on $L^\infty(\mathcal{G})$. For instance, the risk measure $\overline{\mathrm{ES}}_\alpha^\mathcal{Q}$ in Section 2.2.2 coincides with ES_α on $L^\infty(\mathcal{G})$.

2.4 Multiple sources

We first clarify the connection between partial law invariance and the source theory of Baillon et al. [2025]. In the source theory, the decision maker is assumed to have a specific form of preference relation, and on the source $\mathcal{G}$ of risk,[2] the decision maker is law invariant with respect to an objective probability. Our concept of partial law invariance assumes a probability $\mathbb{P}$ on the entire σ-algebra $\mathcal{F}$, which is interpreted as objective on $\mathcal{G}$. The requirement on $\mathcal{G}$ for the preference relation is the same in both frameworks.

The two frameworks differ on how random variables outside $\mathcal{G}$ are treated. In our framework, losses outside $L^\infty(\mathcal{G})$ are connected to those in $L^\infty(\mathcal{G})$ via financial properties like monotonicity and convexity, but we do not directly specify how their distributions are used in evaluating the risk. In the source theory, multiple sources are specified, and the decision maker is law invariant with respect to a (possibly different) subjective probability on each of them.[3]

Thus far, we have only considered one source (sub-σ-algebra) on which the risk measure is law invariant. In practical applications, there may be multiple sources of randomness on which the risk measure is law invariant. Extending our framework to multiple sources, we consider a similar idea as in Baillon et al. [2025], to let the decision maker be law invariant on each of the sources, but with respect to the common probability $\mathbb{P}$. This leads to the following formal property. Let $\Sigma(\mathcal{F})$ denote the set of all atomless sub-σ-algebras of $\mathcal{F}$ and $\mathfrak{G} \subseteq \Sigma(\mathcal{F})$. For a mapping $\rho : L^\infty \to \mathbb{R}$, we say that ρ is $\mathfrak{G}$-*law invariant* if it is $\mathcal{G}$-law invariant for each $\mathcal{G} \in \mathfrak{G}$. When $\mathfrak{G} = \{\mathcal{G}\}$, $\mathfrak{G}$-law invariance is the same as $\mathcal{G}$-law invariance.

A natural corollary of Theorem 2.1 characterizes all Fatou $\mathfrak{G}$-law invariant coherent risk measures.

Theorem 2.2. *Let $\mathfrak{G} \subseteq \Sigma(\mathcal{F})$. A mapping $\rho : L^\infty \to \mathbb{R}$ is a Fatou $\mathfrak{G}$-law-invariant coherent risk measure if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{S}} \mathbb{E}^\mu[X], \qquad X \in L^\infty$$

holds for some convex $\mathcal{S} \subseteq \mathcal{M}_1$ such that the $\mathcal{S}^{\mathcal{G}}$ is almost $\mathcal{G}$-law invariant for each $\mathcal{G} \in \mathfrak{G}$.

[2] Baillon et al. [2025] used algebras instead of σ-algebras, which is a technical point and does not affect intuition.

[3] In source theory, the natural property of monotonicity (i.e., $X \gtrsim Y$ if $X \leq Y$ pointwise, representing losses) across different sources (i.e., X and Y may not be measurable to the same source) seems to be difficult to embed.

Section 2.11 contains an example of a $\mathfrak{G}$-law-invariant coherent risk measure given by $\rho(X) = \sup_{\mathcal{G}\in\mathfrak{G}} \rho_{\mathcal{G}}(\mathbb{E}[X|\mathcal{G}])$, where each $\rho_{\mathcal{G}}$ is law invariant and coherent, along with its representation in the form of Theorem 2.2.

There are at least two stronger yet natural ways of specifying partial law invariance across multiple sources, which we discuss below.

(a) We may further require that the same risk evaluation applies across different $\mathcal{G} \in \mathfrak{G}$, leading to the condition that $\rho(X) = \rho(Y)$ for all $X, Y \in \bigcup_{\mathcal{G}\in\mathfrak{G}} L^\infty(\mathcal{G})$ that satisfy $X \overset{\mathrm{d}}{=} Y$. Note that $\bigcup_{\mathcal{G}\in\mathfrak{G}} L^\infty(\mathcal{G})$ is not necessarily a space or even a convex set, making the property potentially cumbersome to work with.

(b) To obtain a proper linear space, we can consider the sub-σ-algebra $\sigma_{\mathfrak{G}} = \sigma\left(\bigcup_{\mathcal{G}\in\mathfrak{G}} \mathcal{G}\right)$, and work with $\sigma_{\mathfrak{G}}$-law invariance.

It is clear that both (a) and (b) are stronger than $\mathfrak{G}$-law invariance, and moreover, (b) is stronger than (a). In Example 2.3 below, we illustrate a situation in which (b) is not desirable, whereas (a) may be reasonable.

Example 2.3. In the context of robust risk aggregation (see e.g., Embrechts et al. [2013, 2015], Bernard et al. [2014] and the references therein), the marginal distributions of d random losses $X_1, \dots, X_d$ are known, but their joint distribution is subject to substantial model uncertainty. In this case, it is natural to assume partial law invariance for $\mathfrak{G} = \{\sigma(X_i) : i = 1, \dots, d\}$, but not on the larger σ-algebra $\sigma(X_1, \dots, X_d)$, because the distributions of X_1 and X_2 are certain, but the distribution of $X_1 + X_2$ is not. A concrete model is described below. Let $(\Omega, \mathcal{F}) = \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ and $\mathbb{P}$ be the multivariate Gaussian measure with mean $m \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. Let $\tilde{\Sigma} \in \mathbb{R}^{d\times d}$ be given by $\tilde{\Sigma}_{ij} = \mathbb{1}_{\{i=j\}}\sigma_i\sigma_j$. For $\lambda \in [0,1]$, let $\mu_\lambda \in \mathcal{M}_1$ be the multivariate Gaussian measure with mean m and covariance matrix $\lambda\Sigma + (1-\lambda)\tilde{\Sigma}$. Let $X_1, \dots, X_d$ be the random variables given by the projections and $\mathfrak{G} = \{\sigma(X_i) : i = 1, \dots, d\}$. For $\alpha \in [0,1)$, define the risk measure

$$\rho(X) = \sup_{\lambda\in[0,1]} \mathrm{ES}_\alpha^{\mu_\lambda}(X), \quad X \in L^\infty.$$

For $i \in \{1, ..., d\}$, let ν_i be the Gaussian measure on $\mathbb{R}$ with mean m_i and variance σ_i^2. For all $\lambda \in [0,1]$ and $B \in \mathcal{B}(\mathbb{R})$, $\mu_\lambda\left(X_i^{-1}(B)\right) = \nu_i(B)$. Hence, ρ is a $\mathfrak{G}$-law invariant coherent risk measure, and further it satisfies property (a) but not property (b).

Our current formulation of $\mathfrak{G}$-law invariance has an advantage over property (a), that is, the flexibility for modeling the situation where different sources are associated with different levels of model uncertainty (ambiguity), illustrated by Example 2.4 below.

Example 2.4. Consider $\mathfrak{G} = \{\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_n\}$ with their pairwise intersection being generated by constants (an example would be that they are independent). Our interpretation is that $\mathcal{G}_0$ has no model uncertainty and each $\mathcal{G}_i$ is associated with different levels of model uncertainty. Suppose that the decision maker uses the multiplier preferences of Hansen and Sargent [2001] on each $L^\infty(\mathcal{G}_i)$, $i \in \{0, 1, \dots, n\}$, represented by

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1} \left(\mathbb{E}^\mu[\ell(X)] - \frac{1}{\theta_i} R(\mu, \mathbb{P}) \right) = \frac{1}{\theta_i} \log \mathbb{E}\left[e^{\theta_i \ell(X)} \right], \quad X \in L^\infty(\mathcal{G}_i), \tag{2.6}$$

where $\theta_i \geq 0$, ℓ is an increasing loss function and $R(\mu, \mathbb{P})$ is the relative entropy of μ with respect to $\mathbb{P}$. The second equality in (2.6) is known as the Donsker–Varadhan variational formula; see e.g., Strzalecki [2011] for this formulation of the multiplier preferences via $\ell(x) = -u(-x)$ as in Section 2.2.3. When $\theta_i = 0$, $\rho(X) = \mathbb{E}[\ell(X)]$, which corresponds to the limit as $\theta_i \downarrow 0$. The formulation (2.6) is well defined because the intersections of $L^\infty(\mathcal{G}_i)$ and $L^\infty(\mathcal{G}_j)$ for $i \neq j$ only contain constants $c \in \mathbb{R}$, for which $\rho(c) = \ell(c)$. In this model, θ_i represents the magnitude of model uncertainty in $\mathcal{G}_i$, with a larger value representing more severe model uncertainty; note that ρ in (2.6) is increasing in θ_i. We may assume $\theta_0 = 0$ because $\mathcal{G}_0$ has no model uncertainty (but it is not necessary for the discussions here). For X outside $\bigcup_{i=0}^n L^\infty(\mathcal{G}_i)$, we can take ρ arbitrarily. By (2.6), we can see that ρ satisfies $\mathfrak{G}$-law invariance, and it further satisfies property (a) if $\theta_0 = \theta_1 = \cdots = \theta_n$. An explicit convex risk measure on L^∞ satisfying (2.6) is presented in the next example, where the sources are independent.

In Example 2.4, even if model uncertainty exists in each source $\mathcal{G}_1, \dots, \mathcal{G}_n$, the risk measure is law invariant on each subspace as a result of robustification, discussed by Hansen and Sargent [2001].

Example 2.5. The risk measure ρ in (2.6) on $L^\infty(\mathcal{G}_i)$ can be written as an entropic risk measure (Föllmer and Schied [2002]) applied to $\ell(X)$. The entropic risk measure $\mathrm{ER}_\beta : L^\infty \to \mathbb{R}$ with parameter $\beta \geq 0$ is defined by

$$\mathrm{ER}_\beta(X) = \frac{1}{\beta} \log \mathbb{E}\left[e^{\beta X} \right], \text{ for } \beta > 0, \quad \text{and} \quad \mathrm{ER}_0 = \mathbb{E}.$$

The convex risk measures ER have unique properties among monetary risk measures, characterized by Embrechts et al. [2021], and they play a fundamental role in the recent characterization of monotone additive statistics by Mu et al. [2024]. Note that (2.6) reads as $\rho(X) = \mathrm{ER}_{\theta_i}(\ell(X))$ on $L^\infty(\mathcal{G}_i)$ for each $i \in \{0, 1, \dots, n\}$. Suppose that $\mathfrak{G} = \{\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_n\} \subseteq \Sigma(\mathcal{F})$ is a collection of pair-wise independent σ-algebras. Define

$\widehat{\rho}: L^\infty \to \mathbb{R}$ by

$$\widehat{\rho}(X) = \max_{i\in\{0,1,\dots,n\}} \mathrm{ER}_{\theta_i}(\mathbb{E}[\ell(X)|\mathcal{G}_i]).$$

By Jensen's inequality, $\mathrm{ER}_{\theta_i}(\ell(X)) \geq \mathrm{ER}_{\theta_i}(\mathbb{E}[\ell(X)]) = \mathbb{E}[\ell(X)]$ for all $X \in L^\infty$ and $i \in \{0, 1, \dots, n\}$. Hence, $\widehat{\rho}(X) = \mathrm{ER}_{\theta_i}(\ell(X))$ for all $X \in L^\infty(\mathcal{G}_i)$ and thus it satisfies (2.6). If $\ell(x) = x$, then $\widehat{\rho}$ is a convex risk measure because it is the maximum of several convex risk measures of the form $\mathrm{ER}_{\theta_i}(\mathbb{E}[X|\mathcal{G}_i])$.

2.5 Extensions

In this section, we present three extensions of partially law-invariant coherent risk measures. First, in Section 2.5.1, we introduce a technical tool called coherent adjustments, which admit interesting mathematical properties. Coherent adjustments allow us to reformulate Theorem 2.1 and serve as the foundation for the subsequent two extensions. Second, in Section 2.5.2, we strengthen partial law invariance to a property, which still generalizes law invariance, that allows us to extend the representation of Kusuoka [2001]. Finally, in Section 2.5.3, we extend some of the results regarding partially law-invariant coherent risk measures to partially law-invariant convex risk measures.

2.5.1 Coherent adjustments and an alternative representation

Next, we take a different route to introduce the concept of coherent adjustments, a useful technical tool that helps us further understand $\mathcal{G}$-law-invariant coherent risk measures. In particular, coherent adjustments make precise how the risk measures with supporting sets $\mathcal{S}(\mu)$ are built in (2.5).

Definition 2.1. Let $\mu \in \mathcal{M}_1(\mathcal{G})$, a *$\mu$-coherent adjustment* is a mapping $\tau : L^\infty \to \mathbb{R}$ satisfying

$\mathcal{G}$-invariance: $\tau(X + Y) = \tau(X)$ for all $X \in L^\infty$ and $Y \in L^\infty(\mathcal{G})$;

μ-monotonicity: $\mathbb{E}^\mu[X] + \tau(X) \leq \mathbb{E}^\mu[Y] + \tau(Y)$ for all $X, Y \in L^\infty$ with $X \leq Y$;

Convexity: $\tau(\lambda X + (1-\lambda)Y) \leq \lambda\tau(X) + (1-\lambda)\tau(Y)$ for all $X, Y \in L^\infty$ and $\lambda \in [0, 1]$;

Positive homogeneity: $\tau(\lambda X) = \lambda\tau(X)$ for all $X \in L^\infty$ and $\lambda \geq 0$.

Moreover, a *μ-convex adjustment* is a mapping that satisfies the first three properties in the above list, or $\tau = -\infty$ on L^∞. Given $\mu \in \mathcal{M}_1(\mathcal{G})$ we denote the set of all Fatou μ-coherent adjustments by $\mathrm{CA}(\mu)$. Let $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$, we say that $(\tau_\mu)_{\mu\in\mathcal{R}}$ is a *CA-assignment* if $\tau_\mu \in \mathrm{CA}(\mu)$ for all $\mu \in \mathcal{R}$.

Let $\mu \in \mathcal{M}_1(\mathcal{G})$. From the above properties, it is straightforward that the supremum of μ-coherent adjustments is again a μ-coherent adjustment. A simple example of a Fatou μ-coherent adjustment is $\tau = \mathbb{E}^\nu - \mathbb{E}^\mu$ for any $\nu \in \mathcal{E}(\mu)$, which is easy to check. In Section 2.12.1, Proposition 2.15, shows that, indeed, the supremum of the form $\tau = \mathbb{E}^\nu - \mathbb{E}^\mu$ contains all possible Fatou μ-coherent adjustments.

We also note that $\mathcal{G}$-invariance of τ is equivalent to $\tau(X) = 0$ for all $X \in L^\infty(\mathcal{G})$ in the presence of subadditivity.

Remark 2.1. The properties of μ-coherent adjustment τ appear similar to a generalized deviation measure D of Rockafellar et al. [2006], which is defined by four properties: location invariance (i.e., $D(X + c) = D(X)$ for $c \in \mathbb{R}$), subadditivity, positive homogeneity and $D(X) \geq 0$ for all X with strict $D(X) > 0$ when X is not constant. Deviation measures naturally appear in a decomposition of insurance premiums; see Nendel et al. [2021]. Certainly, location invariance is not satisfied by any monetary risk measure because it conflicts cash invariance. By choosing $\mathcal{G} = \{\varnothing, \Omega\}$ (in which case $\mathcal{M}_1(\mathcal{G}) = \{\mu\}$ and $D_\mu = 1 = \mathbb{E}[D_\nu|\mathcal{G}]$ for all $\nu \in \mathcal{M}_1$), $\mathcal{G}$-invariance is the same as location invariance; in this case, τ and D share three properties. However, there is an important difference between τ and D, as τ may take negative values. For instance, $\tau = \mathbb{E}^\nu - \mathbb{E}^\mu$ defines a μ-coherent adjustment, but not a generalized deviation measure.

The name "μ-coherent adjustment" is explained by Proposition 2.6.

Proposition 2.6. *Let $\rho : L^\infty \to \mathbb{R}$ and $\mu \in \mathcal{M}_1(\mathcal{G})$. The following are equivalent.*

(i) The mapping ρ is a Fatou coherent risk measure with $\rho|_{L^\infty(\mathcal{G})} = \mathbb{E}^\mu|_{L^\infty(\mathcal{G})}$;

(ii) there exists $\tau \in \mathrm{CA}(\mu)$ such that $\rho = \mathbb{E}^\mu + \tau$.

If we assume additivity on $L^\infty(\mathcal{G})$ and $\mathcal{G}$-law invariance for a Fatou coherent risk measure ρ, then ρ has a form $\mathbb{E}^\mathbb{P} + \tau$ for $\tau \in \mathrm{CA}(\mathbb{P}|_\mathcal{G})$ that is similar to Proposition 2.6 (ii); see Section 2.12.2.

For some $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ and CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$, by Proposition 2.6, the following risk measure

$$\rho(X) = \sup_{\mu\in\mathcal{R}} \left(\mathbb{E}^\mu[X] + \tau_\mu(X)\right), \quad X \in L^\infty, \tag{2.7}$$

is coherent, as it is the supremum of coherent risk measures. Theorem 2.3 below shows that this mapping is Fatou continuous and that all Fatou coherent risk measures have this representation.

To explain why μ-coherent adjustments help us extend a coherent risk measure on $L^\infty(\mathcal{G})$ to L^∞, let $\tilde{\rho} : L^\infty(\mathcal{G}) \to \mathbb{R}$ be a law-invariant coherent risk measure. For a convex $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ that is $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-dense in $\mathcal{S}_{\tilde{\rho}}$ and a CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$, let ρ be given by (2.7). It is easy to see that $\rho|_{L^\infty(\mathcal{G})} = \tilde{\rho}$, meaning that it is $\mathcal{G}$-law invariant.

The above ideas lead to the following alternative representation of partially law-invariant and Fatou coherent risk measures.

Theorem 2.3. *The mapping $\rho : L^\infty \to \mathbb{R}$ is a Fatou coherent risk measure if and only if there exists a convex set $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ and a CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ such that*

$$\rho(X) = \sup_{\mu\in\mathcal{R}} \left(\mathbb{E}^\mu[X] + \tau_\mu(X)\right), \quad X \in L^\infty. \tag{2.8}$$

Moreover, in (2.8), ρ is $\mathcal{G}$-law invariant if and only if $\mathcal{R}$ is almost $\mathcal{G}$-law invariant.

Remark 2.2. There is a straightforward way to switch between representations (2.4) and (2.8) for a Fatou and $\mathcal{G}$-law-invariant coherent risk measure. Given a convex set $\mathcal{S} \subseteq \mathcal{M}_1(\mathcal{G})$ that represents ρ in Theorem 2.1, define $\mathcal{R} = \mathcal{S}^\mathcal{G}$, which is almost $\mathcal{G}$-law invariant. For each $\mu \in \mathcal{R}$, define $\mathcal{S}(\mu) = \mathcal{E}(\mu) \cap \mathcal{S}$ and $\tau_\mu(X) = \sup_{\nu\in\mathcal{S}(\mu)}(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X])$. These can be used in representation (2.8). Conversely, let an almost $\mathcal{G}$-law invariant convex set $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ and a CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ representing ρ in Theorem 2.3 be given. For $\mu \in \mathcal{R}$, let $\mathcal{S}_{\tau_\mu}$ be the $w(\mathcal{M}, L^\infty)$-closed and convex set in Proposition 2.15 (see Section 2.12.1) that represents τ_μ. It is easy to show that $\mathcal{S} = \bigcup_{\mu\in\mathcal{R}} \mathcal{S}_{\tau_\mu}$ is convex and $S^\mathcal{G} = \mathcal{R}$. We can use $\mathcal{S}$ in representation (2.4). The above connection clarifies the construction in (2.5).

Uniqueness of representation (2.8) under a stronger notion of continuity is addressed in Section 2.12.3. Representation (2.8) has little resemblance to the classical representation result of Kusuoka [2001] for law-invariant coherent risk measures. The following section addresses this gap by further imposing a necessary and sufficient condition for a similar representation along that direction.

2.5.2 Strong partial law invariance and Kusuoka-type representation

Next, we introduce a new property, which is stronger than partial law invariance and is necessary and sufficient to reduce (2.8) to a Kusuoka-type representation; see Proposition 2.11 in Section 2.8.

Define $\ker(\mathcal{G}) = \{X \in L^\infty : \mathbb{E}[X|\mathcal{G}] = 0\}$. We say that $\rho : L^\infty \to \mathbb{R}$ is *strongly $\mathcal{G}$-law invariant* if for all $Z \in \ker(\mathcal{G})$ and $X, Y \in L^\infty(\mathcal{G})$ with $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$ we have

$$\rho(Z + X) = \rho(Z + Y).$$

Taking $Z = 0 \in \ker(\mathcal{G})$, we see that strong $\mathcal{G}$-law invariance implies $\mathcal{G}$-law invariance, hence the name. Since $\ker(\mathcal{F}) = \{0\}$, the risk measure ρ is law invariant if and only if ρ is strongly $\mathcal{F}$-law invariant. Therefore, strong $\mathcal{G}$-law invariance and $\mathcal{G}$-law invariance are both a generalization of law invariance. The following result gives an equivalent condition to strong $\mathcal{G}$-law invariance.

Lemma 2.1. *The mapping ρ is strongly $\mathcal{G}$-law invariant if and only if for all $X \in L^\infty$ and $Y \in L^\infty(\mathcal{G})$ with $Y \overset{\mathrm{d}}{=}_{\mathbb{P}} \mathbb{E}[X|\mathcal{G}]$, we have*

$$\rho(X) = \rho(X - \mathbb{E}[X|\mathcal{G}] + Y).$$

An example of strongly $\mathcal{G}$-law invariant risk measure ρ is $\rho(X) = \tilde{\rho}(\mathbb{E}[X|\mathcal{G}])$, where $\tilde{\rho} : L^\infty(\mathcal{G}) \to \mathbb{R}$ is law invariant. Examples of functionals satisfying $\mathcal{G}$-law invariance but not strong $\mathcal{G}$-law invariance will be provided in Section 2.12.5.

Before we give the main result of the section, we need some technical details. Given $X \in L^1$, $\mu \in \mathcal{M}_1$ and $\alpha \in [0, 1)$, define $F^{-1}_{\mu,X}(\alpha) = \inf\{x \in \mathbb{R} \mid F_{\mu,X}(x) \geq \alpha\}$ and $q_X(\alpha) = F^{-1}_{\mathbb{P},X}(\alpha)$. Given $\mu \in \mathcal{M}_1(\mathcal{G})$, we can define $Q \in \mathcal{M}_B$ (see Föllmer and Schied [2016, Theorem 4.62]) by

$$Q([0, t)) = \int_{[0,t)} (1 - s)\, \mathrm{d}q_{D_\mu}(s), \quad \text{for} \quad t \in [0, 1). \tag{2.9}$$

For $Q \in \mathcal{M}_B$, define $\mathcal{M}_Q = \{\mu \in \mathcal{M}_1(\mathcal{G}) : (2.9) \text{ holds}\}$, which is non-empty because $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, and

$$\sup_{\mu \in \mathcal{M}_Q} \mathbb{E}^\mu[X] = \int_{[0,1)} \mathrm{ES}_\alpha(\mathbb{E}[X|\mathcal{G}])\, Q(\mathrm{d}\alpha); \tag{2.10}$$

see Föllmer and Schied [2016, Theorem 4.62]. Let $\mathrm{CA}(Q) = \bigcap_{\mu \in \mathcal{M}_Q} \mathrm{CA}(\mu)$. This set is non-empty since it contains the zero map. We present the following result, generalizing the representation of Kusuoka [2001].

Theorem 2.4. *Let $\rho : L^\infty \to \mathbb{R}$, the following are equivalent.*

(i) The mapping ρ is a Fatou and strongly $\mathcal{G}$-law-invariant coherent risk measure;

(ii) there exist $\mathcal{D} \subseteq \mathcal{M}_B$ *and* $\tau_Q \in \mathrm{CA}(Q)$ *for each* $Q \in \mathcal{D}$ *such that*

$$\rho(X) = \sup_{Q \in \mathcal{D}} \left(\int_{[0,1)} \mathrm{ES}_\alpha \left(\mathbb{E}[X|\mathcal{G}] \right) \; Q(\mathrm{d}\alpha) + \tau_Q(X) \right). \tag{2.11}$$

In the setting $\mathcal{G} = \mathcal{F}$, the representation (2.11) reduces to the Kusuoka representation,

$$\rho(X) = \sup_{Q \in \mathcal{D}} \int_{[0,1)} \mathrm{ES}_\alpha(X) \; Q(\mathrm{d}\alpha), \; X \in L^\infty,$$

because in this case $\mathbb{E}[X|\mathcal{G}] = X$, $\tau_Q(X) = 0$, and strong $\mathcal{G}$-law invariance is law invariance. A simple example is to choose $\tau_Q = 0$ for all $Q \in \mathcal{D}$, which yields the risk measure $\rho : X \mapsto \tilde{\rho}(\mathbb{E}[X|\mathcal{G}])$ studied in Proposition 2.8 below.

2.5.3 Convex risk measures

This section discusses convex risk measures and their relation to the variational preferences model of Maccheroni et al. [2006]. We provide representations for partially law-invariant convex risk measures akin to Theorem 2.3. A risk measure $\rho : L^\infty \to \mathbb{R}$ is Fatou and convex if and only if there exists $\alpha : \mathcal{M}_1 \to (-\infty, \infty]$ such that

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1} \left(\mathbb{E}^\mu[X] - \alpha(\mu) \right), \quad X \in L^\infty;$$

see Föllmer and Schied [2016]. Using the subjective loss interpretation for the random variables, we see that applying convex risk measures to subjective losses agrees with the variational preferences model of Maccheroni et al. [2006] in the case that $\inf_{\mu \in \mathcal{M}_1} \alpha(\mu) = 0$.

Similar to CA-assignments, we say that $(\tau_\mu)_{\mu \in \mathcal{M}_1(\mathcal{G})}$ is a *CoA-assignment* if for all $\mu \in \mathcal{M}_1(\mathcal{G})$, τ_μ is a Fatou μ-convex adjustment and the mapping $\mu \mapsto \tau_\mu(0)$ is concave. The following result is the convex version of Theorem 2.3.

Theorem 2.5. *The mapping* $\rho : L^\infty \to \mathbb{R}$ *is a Fatou convex risk measure if and only if there exists a CoA-assignment* $(\tau_\mu)_{\mu \in \mathcal{M}_1(\mathcal{G})}$ *such that*

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1(\mathcal{G})} \left(\mathbb{E}^\mu[X] + \tau_\mu(X) \right) \quad X \in L^\infty.$$

Moreover, ρ *is* $\mathcal{G}$*-law invariant if and only if for all* $c > -\rho(0)$ *the set*

$$\{\mu \in \mathcal{M}_1(\mathcal{G}) \mid -\tau_\mu(0) \leq c\}$$

is almost $\mathcal{G}$-law invariant.

The risk measures ER in Example 2.5 are the most common examples of convex risk measures. We next discuss a few $\mathcal{G}$-law-invariant versions of ER, which shed some light on different ways to construct partially law-invariant convex risk measures. In the examples below, partial law invariance is not necessarily associated with the interpretation of model uncertainty.

We write the conditional version of ER (this is not a risk measure in our definition) as

$$\mathrm{ER}_\beta(X|\mathcal{G}) = \frac{1}{\beta} \log \mathbb{E}\left[e^{\beta X}|\mathcal{G}\right], \quad X \in L^\infty.$$

There are many ways to generalize the family of ER to satisfy $\mathcal{G}$-partial law-invariance, and we list three simple choices. The first is to use the mean of the conditional ER, that is,

$$\mathrm{ER}_\beta^{\mathcal{G}}(X) = \mathbb{E}[\mathrm{ER}_\beta(X|\mathcal{G})] = \mathbb{E}\left[\frac{1}{\beta} \log \mathbb{E}\left[e^{\beta X}|\mathcal{G}\right]\right], \quad X \in L^\infty.$$

The second is to compute the ER of the conditional mean, that is,

$$\widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}(X) = \mathrm{ER}_\beta(\mathbb{E}[X|\mathcal{G}]) = \frac{1}{\beta} \log \mathbb{E}\left[e^{\beta \mathbb{E}[X|\mathcal{G}]}\right], \quad X \in L^\infty.$$

Finally, we can also compute ER of the conditional ER, but this gives back the usual ER, that is

$$\mathrm{ER}_\beta(\mathrm{ER}_\beta(X|\mathcal{G})) = \frac{1}{\beta} \log \mathbb{E}\left[\exp\left(\beta \frac{1}{\beta} \log \mathbb{E}\left[e^{\beta X}|\mathcal{G}\right]\right)\right] = \mathrm{ER}_\beta(X), \quad X \in L^\infty.$$

(This is time-consistency of ER as a dynamic risk measure; see Kupper and Schachermayer [2009].) The last version is fully law invariant and not interesting to us. The first two versions are $\mathcal{G}$-law invariant and have intriguing and distinct properties. For instance, if X is $\mathcal{G}$-measurable, then

$$\mathrm{ER}_\beta^{\mathcal{G}}(X) = \mathbb{E}[X] \quad \text{and} \quad \widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}(X) = \mathrm{ER}_\beta(X);$$

if Y is independent of $\mathcal{G}$, then

$$\mathrm{ER}_\beta^{\mathcal{G}}(Y) = \mathrm{ER}_\beta(Y) \quad \text{and} \quad \widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}(Y) = \mathbb{E}[Y].$$

The above formulas show that $\mathrm{ER}_\beta^{\mathcal{G}}$ and $\widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}$ are not fully law invariant, and none of

them dominates the other. Moreover, if X and Y above are identically distributed and non-degenerate, then $\mathrm{ER}_\beta^{\mathcal{G}}(X) < \mathrm{ER}_\beta^{\mathcal{G}}(Y)$ and $\widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}(X) > \widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}(Y)$. In other words, $\mathrm{ER}_\beta^{\mathcal{G}}$ penalizes risks independent of $\mathcal{G}$, whereas $\widetilde{\mathrm{ER}}_\beta^{\mathcal{G}}$ penalizes risks measurable to $\mathcal{G}$. This shows different emphases of the two risk measures when assessing risks from different sources; note that here we do not interpret events outside $\mathcal{G}$ as events with ambiguity; indeed, all computations are carried with $\mathbb{P}$ that is known and fixed in this setting.

2.6 Explicit classes and applications

In this section, we propose a few explicit constructions of partially law-invariant risk measures and illustrate with numerical examples. Recall our notation $\mathcal{E}(\mathbb{P}) = \{\nu \in \mathcal{M}_1 : \nu|_{\mathcal{G}} = \mathbb{P}\}$, which will be used frequently below.

2.6.1 Partially law-invariant ES

This section gives tractable formulas for an important class of partially law-invariant risk measures, the $\mathcal{G}$-law invariant ES introduced in Section 2.2.2, given by

$$\overline{\mathrm{ES}}_\alpha^{\mathcal{Q}}(X) = \sup_{\mu \in \mathcal{Q}} \mathrm{ES}_\alpha^\mu(X), \qquad X \in L^\infty,$$

where $\mathcal{Q} \subseteq \mathcal{E}(\mathbb{P})$ and satisfies some regularity (to be made rigorous soon). Afterward, we numerically implement the results of this section and see how decisions in a simple investment vary as we change the parameters in the risk measure.

We focus on ES here for two reasons. First, ES is the standard regulatory risk measure in the banking sector (BCBS [2019]), and hence the most relevant for financial applications. Second, ES has special advantages in optimization (Rockafellar and Uryasev [2002]), and the partially law-invariant version of ES admits tractable formulas (Proposition 2.7 below).

To prepare for the formula of $\sup_{\mu \in \mathcal{Q}} \mathrm{ES}^\mu$, we first define two technical terms.

Definition 2.2. Let $\mathcal{X} \subseteq L^\infty$, we call $Y \in L^\infty$ the *essential supremum* of $\mathcal{X}$ if

(i) For all $X \in \mathcal{X}$, $X \le Y$;

(ii) if $Z \in L^\infty$ and $X \le Z$ for all $X \in \mathcal{X}$, then $Y \le Z$.

We will denote Y by ess-sup $\mathcal{X}$, as the essential supremum (if it exists) is $\mathbb{P}$-a.s. unique.

Definition 2.3. We say that $\mathcal{X} \subseteq L^1$ is $\mathcal{G}$-*convex* if for all $X, Y \in \mathcal{X}$ and $\lambda \in L^\infty(\mathcal{G})$ with $0 \le \lambda \le 1$, we have $\lambda X + (1-\lambda)Y \in \mathcal{X}$. We say that $\mathcal{R} \subseteq \mathcal{M}$ is $\mathcal{G}$-*convex* if $\widehat{\mathcal{R}}$ is $\mathcal{G}$-convex.

Given certain conditions on $\mathcal{Q} \subseteq \mathcal{E}(\mathbb{P})$, the $\mathcal{G}$-law invariant risk measure $\sup_{\mu\in\mathcal{Q}} \mathrm{ES}^\mu$ mentioned in Section 2.2.2 can be determined via a minimization formula, similar to the formulation of ES by Rockafellar and Uryasev [2002]. Below, a coherent conditional risk measure is a mapping from L^∞ to $L^\infty(\mathcal{G})$ satisfying the four axioms of coherence but with constants replaced by elements of $L^\infty(\mathcal{G})$; see Detlefsen and Scandolo [2005].

Proposition 2.7. *Let $\mathcal{Q} \subseteq \mathcal{E}(\mathbb{P})$ be $w(\mathcal{M}, L^\infty)$-compact and $\mathcal{G}$-convex. For $\alpha \in [0,1)$, we have*

$$\sup_{\mu\in\mathcal{Q}} \mathrm{ES}^\mu_\alpha(X) = \min_{x\in\mathbb{R}} \left(x + \frac{1}{1-\alpha}\mathbb{E}\left[\rho^{\mathcal{Q}}_{\mathcal{G}}\left((X-x)^+\right)\right]\right), \quad X \in L^\infty,$$

where $\rho^{\mathcal{Q}}_{\mathcal{G}} : L^\infty \to L^\infty(\mathcal{G})$ is the coherent conditional risk measure

$$\rho^{\mathcal{Q}}_{\mathcal{G}}(X) = \operatorname*{ess\text{-}sup}_{\mu\in\mathcal{Q}} \mathbb{E}^\mu[X|\mathcal{G}], \qquad X \in L^\infty.$$

Moreover, the set $\mathcal{Q}_\beta = \{\mu \in \mathcal{E}(\mathbb{P}) : \ D_\mu \le 1/(1-\beta)\}$ for $\beta \in [0,1)$ is $w(\mathcal{M}, L^\infty)$-compact and $\mathcal{G}$-convex.

Denote by $\mathrm{ES}_\beta(X|\mathcal{G}) = \rho^{\mathcal{Q}_\beta}_{\mathcal{G}}(X)$, $X \in L^\infty$, which is the conditional version of ES in Detlefsen and Scandolo [2005]. The benefit of using $\mathcal{Q}_\beta$ is that under tractable assumptions of the conditional distribution of X given $\mathcal{G}$, $\mathbb{E}[\mathrm{ES}_\beta((X-x)^+|\mathcal{G})]$ can be computed using conditional distributions. Section 2.6.3 offers a numerical illustration of this idea.

2.6.2 Two constructions of partially law-invariant coherent risk measures

We next present two approaches to constructing partially law-invariant coherent risk measures. The first class arises from mappings analogous to conditional expectations.

Let $\psi : L^\infty \to L^\infty(\mathcal{G})$ be a mapping such that $\psi(X) = X$ for all $X \in L^\infty(\mathcal{G})$; an example would be $\psi(X) = \mathbb{E}[X|\mathcal{G}]$ or, more generally, the coherent conditional risk measures. Consider risk measures ρ of the form

$$\rho(X) = \tilde{\rho}(\psi(X)), \ X \in L^\infty, \tag{2.12}$$

where $\tilde{\rho} : L^\infty(\mathcal{G}) \to \mathbb{R}$ is a law-invariant mapping. Clearly, ρ defined in (2.12) is $\mathcal{G}$-law invariant because $\rho(X) = \tilde{\rho}(\psi(X)) = \tilde{\rho}(X) = \tilde{\rho}(Y) = \rho(Y)$ for any $X, Y \in L^\infty(\mathcal{G})$ with $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$.

Remark 2.3. There is a large literature on conditional and dynamic risk measures, which also involve a fixed sub-σ-algebra $\mathcal{G}$; see Riedel [2004], Detlefsen and Scandolo [2005], and Filipović et al. [2012]. Conditional risk measures take values in random variables, different from our setting, and the motivation for partial law invariance is quite different from the above literature. Nevertheless, it is possible, as shown above, to use the theory of conditional coherent risk measures to construct $\mathcal{G}$-law-invariant coherent risk measures.

Functionals of the form (2.12) can be found in the two-stage market evaluation of Pelsser and Stadje [2014] and the market-consistent valuation of Dhaene et al. [2017]. Setting $\psi(X) = \mathbb{E}[X|\mathcal{G}]$, equation (2.12) offers a straightforward method for constructing partially law-invariant risk measures. Specifically,

$$\rho(X) = \tilde{\rho}(\mathbb{E}[X|\mathcal{G}]), \quad X \in L^\infty. \tag{2.13}$$

Coherent risk measures of the form (2.13) are characterized by the following result.

Proposition 2.8. *For a Fatou coherent risk measure $\rho : L^\infty \to \mathbb{R}$, equivalent are:*

(i) There exists a Fatou coherent risk measure $\tilde{\rho} : L^\infty(\mathcal{G}) \to \mathbb{R}$ such that (2.13) holds;

(ii) there exists $\tilde{\rho} : L^\infty(\mathcal{G}) \to \mathbb{R}$ such that (2.13) holds;

(iii) $\rho(X) = \rho(\mathbb{E}[X|\mathcal{G}])$ for all $X \in L^\infty$;

(iv) $\widehat{\mathcal{S}}_\rho \subseteq L^1(\mathcal{G})$.

Moreover, if any of the above conditions are satisfied, ρ is $\mathcal{G}$-law invariant if and only if $\tilde{\rho}$ is law invariant.

Example 2.6. The idea of using conditional expectations can be generalized further to incorporate different sub-σ-algebras (see Section 2.4). Let $\mathcal{H}$ be a sub-σ-algebra independent of $\mathcal{G}$. Let $\rho_1 : L^\infty(\mathcal{G}) \to \mathbb{R}$ and $\rho_2 : L^\infty(\mathcal{H}) \to \mathbb{R}$ both be coherent risk measures, with the former being law invariant. Define the following risk measure

$$\rho(X) = \rho_1(\mathbb{E}[X|\mathcal{G}]) \vee \rho_2(\mathbb{E}[X|\mathcal{H}]), \ X \in L^\infty,$$

where $a \vee b$ denotes the maximum between two real numbers a and b. To show that ρ above is $\mathcal{G}$-law invariant, we first notice that for X that is $\mathcal{G}$-measurable, we have $\rho(X) =$

$\rho_1(X) \vee \mathbb{E}[X]$. Coherence and law invariance of ρ_1 implies $\rho_1(X) \geq \mathbb{E}[X]$ for $X \in L^\infty(\mathcal{G})$; for this result, see Theorem 4.3 of Bäuerle and Müller [2006]. Therefore $\rho(X) = \rho_1(X)$ for all $X \in L^\infty(\mathcal{G})$, which implies that ρ is $\mathcal{G}$-law invariant.

The second class of partially law-invariant coherent risk measures arises from statistical functions and tail risk measures, similar to the example of robust ES.

Let $\mathcal{P}_c(\mathbb{R})$ denote the set of Borel probability measures on $\mathbb{R}$ with compact support. A *statistical function* is a mapping $\gamma : \mathcal{P}_c(\mathbb{R}) \to \mathbb{R}$. We say that a statistical function is *coherent* if the functional on L^∞ given by $X \mapsto \gamma(X_{\#}\mathbb{P})$ is coherent, where $X_{\#}\mathbb{P} \in \mathcal{P}_c(\mathbb{R})$ is the distribution of X under $\mathbb{P}$. We say that a statistical function is *Fatou* if the functional on L^∞ given by $X \mapsto \gamma(X_{\#}\mathbb{P})$ is Fatou.

Proposition 2.9. *Given a statistical function γ and $\mathcal{S} \subseteq \mathcal{E}(\mathbb{P})$, the risk measure $\rho : L^\infty \to \mathbb{R}$ given by*

$$\rho(X) = \sup_{\mu \in \mathcal{S}} \gamma(X_{\#}\mu), \quad X \in L^\infty \tag{2.14}$$

is Fatou, $\mathcal{G}$-law invariant, and coherent.

Using Proposition 2.9, we can establish a connection between a "bottom-up" approach to addressing model uncertainty and the current framework. Suppose that an agent begins with a probability space $(\Omega, \mathcal{G}, \mathbb{P}_0)$, where they are certain about their probabilistic model $\mathbb{P}_0$ on $\mathcal{G}$. The agent wishes to extend this model to a finer-σ-algebra $\mathcal{F} \supseteq \mathcal{G}$, on which they have model uncertainty. To start, she defines a reference probability $\mathbb{P}$ on $\mathcal{F}$, such that $\mathbb{P}|_{\mathcal{G}} = \mathbb{P}_0$, to identify null sets. From here, the agent chooses a set of competing probabilistic models $\mathcal{S} \subseteq \mathcal{E}(\mathbb{P}_0)$. For financial positions in $L^\infty(\mathcal{G})$, the agent intends to use the law-invariant risk measure $\rho : L^\infty(\mathcal{G}) \to \mathbb{R}$ given by $\tilde{\rho}(X) = \gamma(X_{\#}\mathbb{P}_0)$, $X \in L^\infty(\mathcal{G})$, where γ is a statistical function that is Fatou and coherent. The risk measure defined in equation (2.14) would be a natural risk measure to use on L^∞. According to Proposition 2.9, this risk measure is Fatou, $\mathcal{G}$-law invariant, and coherent. For a similar approach in the dynamical setting, see Epstein and Schneider [2003] and Hansen and Sargent [2022].

Example 2.7. For $\beta \in [0, 1)$, define $\mathcal{S} = \{\mu \in \mathcal{E}(\mathbb{P}) : \ D_\mu \leq 1/(1-\beta)\}$ and let $\tilde{\mathbb{E}} : \mathcal{P}_c(\mathbb{R}) \to \mathbb{R}$ be the expectation statistical function (mapping a distribution to its mean). Define $\rho : L^\infty \to \mathbb{R}$ by

$$\rho(X) = \sup_{\mu \in \mathcal{S}} \tilde{\mathbb{E}}(X_{\#}\mu) = \sup_{\mu \in \mathcal{S}} \mathbb{E}^\mu[X], \quad X \in L^\infty.$$

In this case, ρ is of the form in (2.4) and (2.14). As $\mathcal{S}^{\mathcal{G}} = \{\mathbb{P}|_{\mathcal{G}}\}$ is almost $\mathcal{G}$-law invariant, we can also verify that ρ is $\mathcal{G}$-law invariant from Theorem 2.1. Take any σ-algebra $\mathcal{H}$

independent of $\mathcal{G}$, such that $(\Omega, \mathcal{H}, \mathbb{P})$ is atomless (assumed to exist; this corresponds to a notion of conditional atomless property; see Remark 2.4 in Section 2.9). We claim

$$\rho(X) = \mathbb{E}[X], \quad X \in L^\infty(\mathcal{G}) \quad \text{and} \quad \rho(X) = \mathrm{ES}_\beta(X), \quad Y \in L^\infty(\mathcal{H}). \tag{2.15}$$

Furthermore, these are the extreme values for random variables with a particular distribution. That is, if $Q \in \mathcal{P}_c(\mathbb{R})$ and $X \in L^\infty$ with $X_\# \mathbb{P} = Q$, we claim

$$\tilde{\mathbb{E}}(Q) \le \rho(X) \le \tilde{\mathrm{ES}}_\beta(Q),$$

where $\tilde{\mathrm{ES}}_\beta : \mathcal{P}_c(\mathbb{R}) \to \mathbb{R}$ is the ES statistical function. The σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are extreme as $\rho(X) = \tilde{\mathbb{E}}(Q)$ if $X \in L^\infty(\mathcal{G})$ and $\rho(X) = \tilde{\mathrm{ES}}_\beta(Q)$ if $X \in L^\infty(\mathcal{H})$. For proofs of the above claims, see Section 2.13.2.

By allowing $\mathcal{S}$ to vary but maintaining the condition that $\mathcal{S}^{\mathcal{G}} = \{\mathbb{P}|_{\mathcal{G}}\}$, we allow the risk measure to behave "nicely" on $L^\infty(\mathcal{G})$, while leaving flexibility outside this σ-field. Certainly, it is not necessary to require $\mu \in \mathcal{E}(\mathbb{P})$ for all $\mu \in \mathcal{S}$ (like $\mathcal{S}$ in Example 2.7); according to Theorem 2.1, it is necessary and sufficient to take $\mathcal{S}^{\mathcal{G}}$ to be almost $\mathcal{G}$-law invariant. Indeed, generalizing Example 2.7, we can have ρ behave like a law-invariant risk measure $\widehat{\rho}$ on $L^\infty(\mathcal{G})$ and like a tail risk measure generated by $\widehat{\rho}$ on $L^\infty(\mathcal{H})$ in the sense of Liu and Wang [2021], this is illustrated by Example 2.8.

Example 2.8. Given a Fatou law-invariant coherent risk measure $\widehat{\rho} : L^\infty \to \mathbb{R}$, there exists a Fatou coherent statistical function $\gamma : \mathcal{P}_c(\mathbb{R}) \to \mathbb{R}$ such that $\widehat{\rho}(X) = \gamma(X_\# \mathbb{P})$ for all $X \in L^\infty$. For $\beta \in [0, 1)$, define $\mathcal{S} = \{\mu \in \mathcal{E}(\mathbb{P}) : \ D_\mu \le 1/(1-\beta)\}$. By Proposition 2.9, the risk measure ρ given by (2.14) is Fatou, $\mathcal{G}$-law invariant, and coherent. We claim that the risk measure ρ satisfies

$$\rho(X) = \widehat{\rho}(X), \quad X \in L^\infty(\mathcal{G}) \quad \text{and} \quad \rho(X) = \widehat{\rho}^\beta(X), \quad X \in L^\infty(\mathcal{H}),$$

where $\widehat{\rho}^\beta$ is the β-tail risk measure generated by $\widehat{\rho}$. See Section 2.13.2 for a proof of this claim.

2.6.3 Numerical illustrations

In this section, we numerically evaluate the formula from Proposition 2.7 to generate plots and explore how model uncertainty shapes portfolio allocations.

Let $(\Omega, \mathcal{F}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ and $\mathbb{P}$ be multivariate Gaussian with mean $m \in \mathbb{R}^2$ and covariance matrix $\Sigma \in \mathbb{R}^{2\times 2}$, where $\Sigma_{ij} = (c + (1-c)\mathbb{1}_{\{i=j\}})\sigma_i \sigma_j$ for $i, j \in \{1, 2\}$ (that is,

an equicorrelation matrix). Let X_1, X_2 be the random variables given by the projections, representing the single-period percentage losses of two financial positions. We interpret $\mathbb{P}$ as the initial estimated bivariate distribution of these losses. Assuming the distribution of X_1 is accurately specified (see Section 2.2.2), set $\mathcal{G} = \sigma(X_1)$. The objective is to evaluate the risk of portfolios of the form $\pi_1 X_1 + \pi_2 X_2$ using the risk measure $\rho_\beta = \sup_{\mu \in \mathcal{Q}_\beta} \mathrm{ES}_\alpha^\mu$, where (π_1, π_2) lies in the unit simplex of $\mathbb{R}^2$. The parameter $\beta \in [0, 1)$ represents the decision maker's confidence in $\mathbb{P}$ for the events outside $\mathcal{G}$. When $\beta = 0$, the risk measure ρ_0 coincides with ES_α, reflecting complete confidence in the reference probability model $\mathbb{P}$. As β increases, it signals a growing doubt in the model's accuracy outside of $\mathcal{G}$, leading to more conservative risk evaluations.

In Section 2.13.3, for $\beta \in [0, 1)$, we derive an integral formula for

$$f_\beta(\pi_1, \pi_2, x) = \mathbb{E}\left[\mathrm{ES}_\beta\left((\pi_1 X_1 + \pi_2 X_2 - x)^+ | X_1\right)\right], \tag{2.16}$$

which can be evaluated using numerical integration. By Proposition 2.7,

$$\rho_\beta\left(\pi_1 X_1 + \pi_2 X_2\right) = \min_{x \in \mathbb{R}} \left(x + \frac{1}{1-\alpha} f_\beta(\pi_1, \pi_2, x) \right),$$

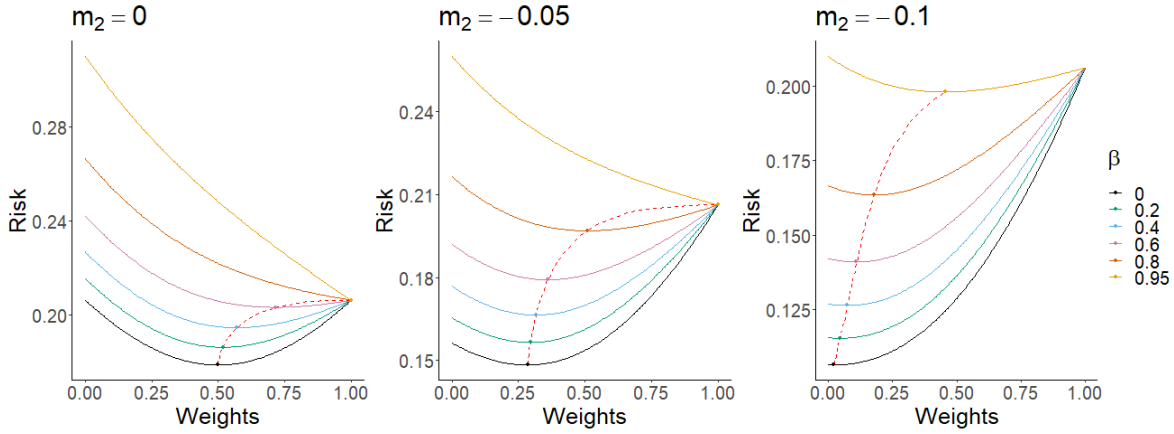
which can be computed numerically using a convex program.

Figures 2.1 and 2.2 display plots of the computed values of ρ_β for various portfolio weights, modeling assumptions, and values of β. The horizontal axes in all panels are given by π_1, the weight allocated to X_1. Since $\rho_\beta(X_1) = \mathrm{ES}_\alpha(X_1)$ for all $\beta \in [0, 1)$, the lines agree when $\pi_1 = 1$.

Figure 2.1 illustrates how the risk-minimizing portfolio evolves as β increases. When $m_2 = 0$ (left panel), an agent with $\beta = 0$ optimally selects an equal allocation between X_1 and X_2. This is consistent with the preference for pure diversification as both X_1 and X_2 share the same correctly-specified distribution. As β increases, the risk-minimizing portfolio shifts monotonically toward a full allocation to X_1. This shift reflects the agent's growing concern that the model may be misspecified outside the estimated distribution for X_1. The center and right panels ($m_2 = -0.05$ and -0.1) represent cases where X_2 is initially expected to yield a higher return. In these cases, the same shift towards a full investment into X_1 persists, albeit with reduced intensity as the perceived downside of model uncertainty is partially offset by the favorable mean of X_2.

Figure 2.2 considers a fixed $\beta = 0.95$ and varying σ_2. As expected, when σ_2 is very small, even the worst-case distributions within the "β-neighborhood" of the estimated distribution for X_2 are more favorable than the well-specified distribution of X_1. Consequently, the

Figure 2.1: Varying β, while fixing $m_1 = 0$, $\sigma_1 = 0.1$, $\sigma_2 = 0.1$, $c = 0.5$, $\alpha = 0.95$; the red dotted curve connects the optimizers for different β



agent continues to allocate fully to X_2, despite the high level of model uncertainty. When σ_2 is large, the risk-minimizing portfolio has an increasing weight allocated to X_1.

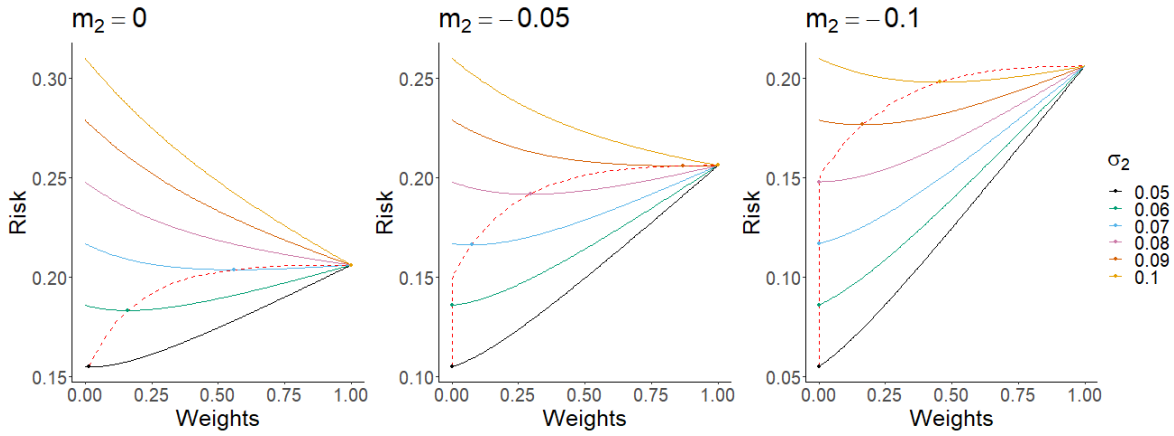
2.7 Conclusion

Our main purpose is to propose and study the concept of partial law invariance as a generalization of law invariance for risk measures and preference relations. This new property is appealing from a decision-theoretic perspective and interesting from a purely mathematical perspective. To summarize the main motivation, different economic sources have different levels of model uncertainty, and a law-invariant mapping or preference relation does not take these differences into account, thus calling for the concept of partial law invariance to capture this discrepancy.

The economic relevance of the new framework is illustrated throughout the chapter with examples. In addition to conceptual development, the new framework also leads to mathematical novelty in terms of new classes of risk measures and new representation results in forms not present in the literature. In particular, coherent and convex adjustments play an important role in understanding the new property and its representation in both the context of coherent and convex risk measures.

The possibility of characterizing partial law invariance to other classes of risk measures may be worth discussing. The star-shaped risk measures, studied by Castagnoli et al.

Figure 2.2: Varying σ_2, while fixing $m_1 = 0$, $\sigma_1 = 0.1$, $c = 0.5$, $\alpha = 0.95$, $\beta = 0.95$; the red dotted curve connects the optimizers for different σ_2



[2022], are a natural generalization of convex risk measures, and they satisfy the following property in addition to $\rho(0) = 0$.

Star-shapedness: $\rho(\lambda X) \leq \lambda \rho(X)$ for all $X \in L^\infty$ and $\lambda \in [0, 1]$.

Representations of law-invariant (quasi-)star-shaped risk measures have recently been studied by Han et al. [2025] and Laeven et al. [2024], which admit a quite different structure than the Kusuoka-type. Another research question is how to incorporate different probabilities, instead of projections of the same probability $\mathbb{P}$ as in our setting, in the context of multiple sources.

2.8 Review on risk measures

This section reviews risk measures in the framework of Artzner et al. [1999].

We say that a risk measure ρ is *Lebesgue (continuous)* if for all uniformly bounded sequences $(X_n)_{n\in\mathbb{N}}$ in L^∞ that converge $\mathbb{P}$-a.s. to some $X \in L^\infty$, it holds that

$$\lim_{n\to\infty} \rho(X_n) = \rho(X).$$

An example of a Lebesgue (thus Fatou) coherent risk measure is ES^μ_α for $\mu \in \mathcal{M}_1$ and $\alpha \in [0, 1)$.

Some well-known facts are collected below. A coherent risk measure ρ is Fatou if and only if ρ is $w(L^\infty, \mathcal{M})$-lower semicontinuous, a proof of which can be found in Föllmer and Schied [2016, Theorem 4.33]. A coherent risk measure has a dual representation in the following form; see, e.g., Föllmer and Schied [2016, Theorem 4.33]. Recall that for a Fatou coherent risk measure $\rho : L^\infty \to \mathbb{R}$, it's *supporting set* is given by

$$\mathcal{S}_\rho = \{\mu \in \mathcal{M}_1 : \mathbb{E}^\mu[X] \leq \rho(X) \text{ for all } X \in L^\infty\}.$$

Proposition 2.10 (Delbaen [2002])**.** *For $\rho : L^\infty \to \mathbb{R}$, the following are equivalent.*

(i) The mapping ρ is a Fatou coherent risk measure;

(ii) there exists a $w(\mathcal{M}, L^\infty)$-closed and convex set $\mathcal{S} \subseteq \mathcal{M}_1$ such that

$$\rho(X) = \sup_{\mu \in \mathcal{S}} \mathbb{E}^\mu[X], \quad X \in L^\infty. \tag{2.17}$$

In case (ii) holds, $\mathcal{S}$ is uniquely given by $\mathcal{S}_\rho$ (the supporting set of ρ). Moreover, $\mathcal{R} \subseteq \mathcal{M}_1$ gives representation (2.17) if and only if $\operatorname{conv}(\mathcal{R})$ is $w(\mathcal{M}, L^\infty)$-dense in $\mathcal{S}_\rho$.

Example 2.9. For $\mu \in \mathcal{M}_1$ and $\alpha \in [0, 1)$, ES^μ_α admits the dual representation

$$\mathrm{ES}^\mu_\alpha(X) = \sup_{\nu \in \mathcal{S}_{\alpha,\mu}} \mathbb{E}^\nu[X], \quad X \in L^\infty,$$

where $\mathcal{S}_{\alpha,\mu} = \{\nu \in \mathcal{M}_1 \mid (1-\alpha)D_\nu \leq D_\mu\}$; see e.g., Föllmer and Schied [2016, Theorem 4.52]. Since $\mathcal{S}_{\alpha,\mu}$ is $w(\mathcal{M}, L^\infty)$-closed and convex, $\mathcal{S}_{\mathrm{ES}^\mu_\alpha} = \mathcal{S}_{\alpha,\mu}$ holds.

Law-invariant coherent risk measures also admit a representation of Kusuoka [2001], and Fatou continuity is automatic for law-invariant convex risk measures, as shown by Jouini et al. [2006] and Svindland [2010]. Let $\mathcal{M}_B$ denote the probability measures on $[0, 1)$ equipped with the restricted Borel σ-algebra.

Proposition 2.11 (Kusuoka [2001], Jouini et al. [2006], Svindland [2010])**.** *For $\rho : L^\infty \to \mathbb{R}$, the following are equivalent.*

(i) The mapping ρ is a Fatou law-invariant coherent risk measure;

(ii) the mapping ρ is a law-invariant coherent risk measure;

(iii) there exists some $\mathcal{D} \subseteq \mathcal{M}_B$ *such that*

$$\rho(X) = \sup_{Q \in \mathcal{D}} \int_{[0,1)} \mathrm{ES}_\alpha(X)\, Q(\mathrm{d}\alpha),\ X \in L^\infty.$$

It is well known that the monetary risk measure $\rho : L^\infty \to \mathbb{R}$ is convex if and only if $\mathcal{A}_\rho$ is convex, and coherent if and only if $\mathcal{A}_\rho$ is a convex cone.

2.9 Additional results and proofs accompanying Section 2.2

Proof. Proof of Proposition 2.1: We claim the set $\{c \in \mathbb{R} : X \succsim c\}$ must be of the form $[a, \infty)$ for some $a \in \mathbb{R}$. To see this, let $x \in \{c \in \mathbb{R} : X \succsim c\}$ and $y > x$, then $X \succsim x \succ y$ and $y \in \{c \in \mathbb{R} : X \succsim c\}$ by (C2), which proves $\{c \in \mathbb{R} : X \succsim c\}$ is an interval with no right endpoint. Since $\{c \in \mathbb{R} : X \succsim c\}$ is closed by (C1), it must be of the form $[a, \infty)$ for some $a \in \mathbb{R}$. Using the same argument, we can show that the set $\{c \in \mathbb{R} : c \succsim X\}$ must be of the form $(\infty, b]$ for some $b \in \mathbb{R}$. As $\succsim$ is a total ordering, it must hold that $a \leq b$ since if this were not true, there would be $c \in \mathbb{R}$ that was not comparable to X. Assume that $a < b$, then $X \succsim a \succ b \succsim X$, which is clearly a contradiction. Thus, $a = b$ and we can define $c_X = a = b$. Then, by the definition of a and b, $c_X \sim X$. □

Proof. Proof of Proposition 2.2: Assume that $\succsim$ is $\mathcal{G}$-law invariant. Let $X, Y \in \mathcal{X}(\mathcal{G})$ with $X \geq_1 Y$. As $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, we can find a $\mathcal{G}$-measurable U such that U has a uniform distribution under $\mathbb{P}$. Define $\tilde{X} = F^{-1}_{\mathbb{P},X}(U)$ and $\tilde{Y} = F^{-1}_{\mathbb{P},Y}(U)$, thus $X \overset{\mathrm{d}}{=}_{\mathbb{P}} \tilde{X}$, $Y \overset{\mathrm{d}}{=}_{\mathbb{P}} \tilde{Y}$, and $\tilde{X} \geq \tilde{Y}$. By $\mathcal{G}$-law invariance and monotonicity, $Y \sim \tilde{Y} \succsim \tilde{X} \sim X$, thus $\succsim$ is $\mathcal{G}$-FSD mononotonic. The converse is straightforward. □

Proof. Proof of Proposition 2.3: Let $X, Y \in L^\infty(\mathcal{G})$ and $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$. For $\mu \in \mathcal{Q}$ and $x \in \mathbb{R}$ we have

$$\mu(X \leq x) = \mathbb{E}\left[D_\mu \mathbb{1}_{\{X \leq x\}}\right] = \mathbb{E}\left[D_\mu\right] \mathbb{E}\left[\mathbb{1}_{\{X \leq x\}}\right] = \mathbb{P}(X \leq x)$$

and similarly $\mu(Y \leq x) = \mathbb{P}(Y \leq x)$. Hence, X and Y are identically distributed under each element of $\mathcal{Q}$. Since ρ is $\mathcal{Q}$-based, we have $\rho(X) = \rho(Y)$. This implies that ρ is $\mathcal{G}$-law invariant. □

The next remark discusses how Proposition 2.3 is related to conditional non-atomicity proposed by Shen et al. [2019].

Remark 2.4. Let $\mathcal{Q}$ be a collection of probability measures on $(\Omega, \mathcal{F})$ and $\mathcal{M}_1(\mathcal{Q})$ be the set of probability measures dominating (in the sense of absolute continuity) each element of $\mathcal{Q}$, and we assume it is non-empty (it is always non-empty if $\mathcal{Q}$ is finite or $\mathcal{Q} \subseteq \mathcal{M}_1$). The set $\mathcal{Q}$ is said to be *conditionally atomless* if there exist $\nu \in \mathcal{M}_1(\mathcal{Q})$ and random variable X such that under ν, X is continuously distributed and independent of $\sigma(D_\mu : \mu \in \mathcal{Q})$. As discussed by Delbaen [2021], an equivalent statement for conditional non-atomicity of $\mathcal{Q}$ is that there exists an atomless σ-algebra $\mathcal{G}$ independent of $\sigma(D_\mu : \mu \in \mathcal{Q})$ under some $\nu \in \mathcal{M}_1(\mathcal{Q})$. Clearly, when $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, the condition that $\mathcal{G}$ is independent of $\sigma(D_\mu : \mu \in \mathcal{Q})$ in Proposition 2.3 is sufficient for conditional non-atomicity of $\mathcal{Q}$.

2.10 Additional results and proofs accompanying Section 2.3

2.10.1 Additional results and proofs accompanying Section 2.3.1

For a preference relation $\succsim$, we define the following properties:

Translation invariance: $X + c \succsim Y + c$. for all $X, Y \in L^\infty$ with $X \succsim Y$ and $c \in \mathbb{R}$;

Convexity: $\lambda X + (1-\lambda)Y \succsim Z$ for all $X, Y, Z \in L^\infty$ with $X \succsim Z$ and $Y \succsim Z$ and $\lambda \in [0, 1]$;

Positive homogeneity: $\lambda X \succsim \lambda Y$ for all $X, Y \in L^\infty$ with $X \succsim Y$ and $\lambda \geq 0$;

Continuity from below: for all increasing sequences $(X_n)_{n\in\mathbb{N}}$ in L^∞ that converges to X and $Y \in L^\infty$ with $X_n \succsim Y$ for $n \in \mathbb{N}$, we have $X \succsim Y$.

The next proposition clarifies the connection between CRM preferences and these properties, and we provide a simple self-contained proof.

Proposition 2.12. *For a preference relation $\succsim$, the following are equivalent.*

(i) The preference relation $\succsim$ satisfies monotonicity, translation invariance, convexity, and positive homogeneity;

(ii) The preference relation $\succsim$ is a CRM preference relation.

Furthermore, if any of the above is true, $\succsim$ is continuous from below if and only if ρ is Fatou continuous.

Proof. Proof: (i) $\Rightarrow$ (ii): Define the mapping $\rho : L^\infty \to \mathbb{R}$ by $\rho(X) = c_X$. We know that ρ represents the preference relation $\succsim$. As $\succsim$ is monotonic, it is clear that ρ is monotonic. Let $X \in L^\infty$, as $X \sim \rho(X)$, by translation invariance, $\rho(X + c) \sim X + c \sim \rho(X) + c$ for all $c \in \mathbb{R}$. By assumption (C2), $\rho(X + c) = \rho(X) + c$ for all $c \in \mathbb{R}$. Define $\mathcal{A}_\rho = \{X \in L^\infty : \rho(X) \leq 0\} = \{X \in L^\infty : X \succsim 0\}$. It is easy to see that $\mathcal{A}_\rho$ is a convex cone as the preference relation $\succsim$ is convex and positively homogenous. Thus, by Föllmer and Schied [2016, Proposition 4.6], ρ is a coherent risk measure. (ii) $\Rightarrow$ (i): This is straightforward.

To prove the last statement, let $\rho : L^\infty \to \mathbb{R}$ be a coherent risk measure that represents the preference relation $\succsim$.

We claim that the preference relation $\succsim$ is continuous from below if and only if for all increasing sequences $(X_n)_{n\in\mathbb{N}}$ in L^∞ that converges to X, we have $\rho(X_n) \uparrow \rho(X)$. For the forward, assume that $(X_n)_{n\in\mathbb{N}}$ in L^∞ is an increasing sequence that converges to X. Let $l = \lim_{k\to\infty} \rho(X_k)$, which exists as $(\rho(X_n))_{n=1}^\infty$ is increasing. Since $\rho(X_n) \leq l$ for all $n \in \mathbb{N}$ and $\rho(l) = l$, $X_n \succsim l$ for all $n \in \mathbb{N}$. Therefore, $X \succsim l$ and

$$\rho(X) \leq l = \lim_{k\to\infty} \rho(X_k) \leq \rho(X),$$

which implies $\rho(X_n) \uparrow \rho(X)$. For the converse, assume that $(X_n)_{n\in\mathbb{N}}$ in L^∞ is an increasing sequence that converges to X and $Y \in L^\infty$ with $X_n \succsim Y$ for $n \in \mathbb{N}$. Therefore, it holds that $\rho(X_n) \leq \rho(Y)$ for all $n \in N$ and

$$\rho(X) = \lim_{n\to\infty} \rho(X_n) \leq \rho(Y),$$

meaning that $X \succsim Y$.

As ρ is monetary, it is a well-known fact that ρ is Fatou continuous if and only if for all increasing sequences $(X_n)_{n\in\mathbb{N}}$ in L^∞ that converges to X, we have $\rho(X_n) \uparrow \rho(X)$, which proves the claim. □

Proof. Proof of Proposition 2.4: (i) $\Leftrightarrow$ (ii): Follows from Proposition 2.2. (i) $\Leftrightarrow$ (iii): The functional $\rho : L^\infty \to \mathbb{R}$ defined by $\rho(X) = c_X$ is coherent as the preference relation $\succsim$ is a CRM preference relation. We say that ρ is $\mathcal{G}$-SSD monotonic if $\rho(X) \geq \rho(Y)$ for all $X, Y \in L^\infty(\mathcal{G})$ with $X \geq_2 Y$. It is well-known that a convex risk measure is law-invariant if and only if it is $\mathcal{F}$-SSD monotonic; see e.g., Mao and Wang [2020, Theorem 2.1 and Proposition 3.2]. Restricting this to $L^\infty(\mathcal{G})$, the preference relation $\succsim$ is $\mathcal{G}$-law invariant if

and only if ρ is $\mathcal{G}$-SSD monotonic. It is clear that ρ is $\mathcal{G}$-SSD monotonic if and only if the preference relation $\succsim$ is $\mathcal{G}$-SSD monotonic. □

Proof. Proof of Proposition 2.5: The forward is trivial. For the converse note that $\{X \in L^\infty \mid \rho(X) \leq c\} = \mathcal{A}_\rho + c$. Therefore, the level sets of ρ are $\mathcal{G}$-law invariant; the claim is straightforward to prove after this observation. □

2.10.2 Discussion and proof of Theorem 2.1

A proof of Theorem 2.1 follows from an analysis of the supporting set of ρ, which will be presented in a few separate results.

Proposition 2.13. *Let $\rho : L^\infty \to \mathbb{R}$ be a Fatou coherent risk measure and $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$. Then, $\tilde{\rho}$ is a Fatou coherent risk measure. Furthermore, $\mathcal{S}_\rho^{\mathcal{G}} \subseteq \mathcal{S}_{\tilde{\rho}}$ and $\mathcal{S}_\rho^{\mathcal{G}}$ is $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-dense in $\mathcal{S}_{\tilde{\rho}}$.*

Proof. Proof: The fact that $\tilde{\rho}$ is a Fatou coherent risk measure is clear. Since $\mathcal{S}_\rho^{\mathcal{G}}$ is convex and for $X \in L^\infty(\mathcal{G})$,

$$\tilde{\rho}(X) = \rho(X) = \sup_{\mu \in \mathcal{S}_\rho} \mathbb{E}^\mu[X] = \sup_{\mu \in \mathcal{S}_\rho} \mathbb{E}[\mathbb{E}[D_\mu|\mathcal{G}]X] = \sup_{\mu \in \mathcal{S}_\rho^{\mathcal{G}}} \mathbb{E}^\mu[X],$$

we have that $\mathcal{S}_\rho^{\mathcal{G}}$ is $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-dense in $\mathcal{S}_{\tilde{\rho}}$ by Proposition 2.10. □

It is natural to wonder when $\mathcal{S}_\rho^{\mathcal{G}} = \mathcal{S}_{\tilde{\rho}}$ in Proposition 2.13. Proposition 2.14 offers sufficient conditions for this equality.

Lemma 2.2. *The mapping $R : \mathcal{M} \to \mathcal{M}(\mathcal{G}) : \mu \mapsto \mu|_{\mathcal{G}}$ is $w(\mathcal{M}, L^\infty)/w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-continuous.*

Proof. Proof: Let $\mu_\lambda, \mu \in \mathcal{M}$ with $\lim_\lambda \mu_\lambda = \mu$ in $w(\mathcal{M}, L^\infty)$. Fix some $X \in L^\infty(\mathcal{G})$. We have

$$\mathbb{E}^{R(\mu_\lambda)}[X] = \mathbb{E}^{\mu_\lambda}[X] \to \mathbb{E}^\mu[X] = \mathbb{E}^{R(\mu)}[X].$$

Therefore $R(\mu_\lambda) \to R(\mu)$ in $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$. □

Proposition 2.14. *Let $\rho : L^\infty \to \mathbb{R}$ be a Fatou coherent risk measure and $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$. The following are sufficient conditions for $\mathcal{S}_\rho^{\mathcal{G}} = \mathcal{S}_{\tilde{\rho}}$.*

(i) *The risk measure ρ satisfies $\rho(\mathbb{E}[X|\mathcal{G}]) \leq \rho(X)$ for all $X \in L^\infty$ (equivalently, $\widehat{\mathcal{S}}_\rho^{\mathcal{G}} \subseteq \widehat{\mathcal{S}}_\rho$).*

(ii) *The risk measure ρ is Lebesgue continuous.*

Proof. Proof: (i) For $\mu \in \mathcal{S}_{\tilde{\rho}}$, we have

$$\mathbb{E}^\mu[X] = \mathbb{E}[D_\mu \mathbb{E}[X|\mathcal{G}]] \leq \rho(\mathbb{E}[X|\mathcal{G}]) \leq \rho(X),$$

for all $X \in L^\infty$. Therefore, the extension of μ to $\mathcal{F}$, i.e., the measure $\hat{\mu} : \mathcal{F} \to [0,1]$ given by $A \mapsto \mathbb{E}^\mu[\mathbb{1}_A]$, is in $\mathcal{S}_\rho$. Since $\hat{\mu}|_{\mathcal{G}} = \mu$, we have $\mathcal{S}_{\tilde{\rho}} \subseteq \mathcal{S}_\rho^{\mathcal{G}}$. (ii) The restriction $\mu \mapsto \mu|_{\mathcal{G}}$ is $w(\mathcal{M}, L^\infty)/w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-continuous by Lemma 2.2 and $\mathcal{S}_\rho$ is $w(\mathcal{M}, L^\infty)$-compact (see Föllmer and Schied [2016, Corollary 4.38]), and hence we have that $\mathcal{S}_\rho^{\mathcal{G}}$ is $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-compact. Therefore, $\mathcal{S}_\rho^{\mathcal{G}}$ is closed since $w(L^1, L^\infty)$ is Hausdorff, which implies $\mathcal{S}_\rho^{\mathcal{G}} = \mathcal{S}_{\tilde{\rho}}$. □

Remark 2.5. Although Proposition 2.14 only provides sufficient conditions for $\mathcal{S}_\rho^{\mathcal{G}} = \mathcal{S}_{\tilde{\rho}}$, we did not find any counter-example in which $\mathcal{S}_\rho^{\mathcal{G}} \neq \mathcal{S}_{\tilde{\rho}}$, which seems to be a challenging question. This relates to whether we can omit the operation of taking a closure in Theorem 2.1. If $\mathcal{S}_\rho^{\mathcal{G}} = \mathcal{S}_{\tilde{\rho}}$ always holds, then that operation can be omitted.

We are ready to prove Theorem 2.1.

Proof. Proof of Theorem 2.1: (i) ⇒ (iii): Since $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$ is law-invariant, we have that $\mathcal{S}_{\tilde{\rho}}$ is $\mathcal{G}$-law invariant. By Proposition 2.13, $\mathcal{S}_\rho^{\mathcal{G}}$ is almost $\mathcal{G}$-law invariant. (iii) ⇒ (ii): By Proposition 2.10, representation (2.4) holds with $\mathcal{S}_\rho$. (ii) ⇒ (i): It is clear that ρ is a Fatou coherent risk measure. It is also clear that $\mathcal{S}^{\mathcal{G}}$ is convex and supports $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$. Since $\mathcal{S}^{\mathcal{G}}$ is almost $\mathcal{G}$-law invariant, $\tilde{\rho}$ is law invariant. This implies that ρ is $\mathcal{G}$-law invariant. □

2.11 An additional example accompanying Section 2.4

Example 2.10. Let $\mathfrak{G} \subseteq \Sigma(\mathcal{F})$ be a collection of pair-wise independent σ-algebras. For each $\mathcal{G} \in \mathfrak{G}$, let $\rho_{\mathcal{G}} : L^\infty(\mathcal{G}) \to \mathbb{R}$ be a Fatou law-invariant coherent risk measure. Define the risk measure

$$\rho(X) = \sup_{\mathcal{G} \in \mathfrak{G}} \rho_{\mathcal{G}}(\mathbb{E}[X|\mathcal{G}]), \quad X \in L^\infty.$$

Using a similar argument as Example 2.6 in Section 2.3.2, one can show that ρ is a Fatou $\mathfrak{G}$-law-invariant coherent risk measure. To see how this example relates to Theorem 2.2,

for each $\mathcal{G} \in \mathfrak{G}$, let $\iota_{\mathcal{G}} : \mathcal{M}_1(\mathcal{G}) \to \mathcal{M}_1$ be the embedding given by $[\iota_{\mathcal{G}}\mu](A) = \mathbb{E}^{\mu}[\mathbb{1}_A]$, $A \in \mathcal{F}$. Also, denote the supporting set of $\rho_{\mathcal{G}}$ by $\mathcal{S}_{\mathcal{G}}$. Define $\mathcal{S} = \operatorname{conv}\left(\bigcup_{\mathcal{G}\in\mathfrak{G}} \iota_{\mathcal{G}}(\mathcal{S}_{\mathcal{G}})\right)$. By Proposition 2.8, $\sup_{\mu\in\iota_{\mathcal{G}}(\mathcal{S}_{\mathcal{G}})} \mathbb{E}^{\mu}[X] = \rho_{\mathcal{G}}(\mathbb{E}[X|\mathcal{G}])$ for all $X \in L^{\infty}$. Thus, $\rho(X) = \sup_{\mu\in\mathcal{S}} \mathbb{E}^{\mu}[X]$, $X \in L^{\infty}$. Since $\mathfrak{G}$ is independent, for all $\mathcal{G} \in \mathfrak{G}$, we have $\mathcal{S}^{\mathcal{G}} = \mathcal{S}_{\mathcal{G}}$, which is $w(\mathcal{M}(\mathcal{G}), L^{\infty}(\mathcal{G}))$-closed and $\mathcal{G}$-law invariant. Therefore, the set $\mathcal{S}$ is a valid candidate for the set found in Theorem 2.2.

2.12 Additional results and proofs accompanying Section 2.5

2.12.1 Additional results and proofs accompanying Section 2.5.1

Proof. Proof of Proposition 2.6: (i)⇒(ii): For $X \in L^{\infty}$ define $\tau(X) = \rho(X) - \mathbb{E}[D_{\mu}X]$. The fact that τ is a μ-coherent adjustment is easily verified. To check that τ is Fatou, let $(X_n)_{n=1}^{\infty}$ be a uniformly bounded sequence in L^{∞} which converges $\mathbb{P}$-a.s. to some $X \in L^{\infty}$. By the Dominated Convergence Theorem (DCT) we have $\lim_{n\to\infty} \mathbb{E}[D_{\mu}X_n] = \mathbb{E}[D_{\mu}X]$. Therefore,

$$\begin{aligned}\liminf_{n\to\infty} \tau(X_n) = \liminf_{n\to\infty} (\rho(X_n) - \mathbb{E}[D_{\mu}X_n]) &\geq \liminf_{n\to\infty} \rho(X_n) - \lim_{n\to\infty} \mathbb{E}[D_{\mu}X_n] \\ &\geq \rho(X) - \mathbb{E}[D_{\mu}X] = \tau(X).\end{aligned}$$

The claim follows since $\rho(X) = \mathbb{E}[D_{\mu}X] + \tau(X)$ for all $X \in L^{\infty}$.

(ii)⇒(i): The facts that ρ is a coherent risk measure and that its restriction is $\mathbb{E}^{\mu}$ are easily verified. Fatou continuity follows an argument similar to the one given above. □

We provide some facts regarding the representation of μ-coherent adjustments.

Proposition 2.15. *Let $\tau : L^{\infty} \to \mathbb{R}$ and $\mu \in \mathcal{M}_1(\mathcal{G})$. The following are equivalent.*

(i) The mapping $\tau \in \mathrm{CA}(\mu)$;

(ii) there exists a $w(\mathcal{M}, L^{\infty})$-closed and convex set $\mathcal{S}_{\tau} \subseteq \mathcal{E}(\mu)$ such that

$$\tau(X) = \sup_{\nu\in\mathcal{S}_{\tau}} (\mathbb{E}^{\nu}[X] - \mathbb{E}^{\mu}[X]), \quad X \in L^{\infty}; \tag{2.18}$$

(iii) there exists $\mathcal{R} \subseteq \mathcal{E}(\mu)$ such that $\tau(X) = \sup_{\nu\in\mathcal{R}} (\mathbb{E}^{\nu}[X] - \mathbb{E}^{\mu}[X])$, $X \in L^{\infty}$.

In case (ii) holds, the closed and convex set $\mathcal{S}_\tau$ is uniquely given by

$$\{\nu \in \mathcal{M}_1 : \mathbb{E}^\nu[X] - \mathbb{E}^\mu[X] \leq \tau(X) \text{ for all } X \in L^\infty\}.$$

Proof. Proof: (i) $\Rightarrow$ (ii): Assume $\tau \in \mathrm{CA}(\mu)$. By Proposition 2.6, the mapping $\rho = \mathbb{E}^\mu + \tau$ is a Fatou coherent risk measure. We have

$$\mathbb{E}^\mu[X] + \tau(X) = \sup_{\nu \in \mathcal{S}_\rho} \mathbb{E}^\nu[X], \ X \in L^\infty.$$

By Proposition 2.13, $\mathcal{S}_\rho^{\mathcal{G}}$ must be contained in the supporting set of $\rho|_{L^\infty(\mathcal{G})} = \mathbb{E}^\mu|_{L^\infty(\mathcal{G})}$, and therefore $\mathcal{S}_\rho^{\mathcal{G}} = \{\mu\}$ (meaning $\mathcal{S}_\rho \subseteq \mathcal{E}(\mu)$). Note that

$$\mathbb{E}^\mu[X] + \tau(X) = \mathbb{E}^\mu[X] + \sup_{\nu \in \mathcal{S}_\rho} \left(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X]\right), \quad X \in L^\infty,$$

implying $\tau(X) = \sup_{\nu \in \mathcal{S}_\rho} \left(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X]\right)$, $X \in L^\infty$. Therefore (2.18) holds with $\mathcal{S}_\tau = \mathcal{S}_\rho$. Furthermore,

$$\begin{aligned}\mathcal{S}_\tau = \mathcal{S}_\rho &= \{\nu \in \mathcal{M}_1 \mid \mathbb{E}^\nu[X] \leq \rho(X) \text{ for all } X \in L^\infty\} \\ &= \{\nu \in \mathcal{M}_1 \mid \mathbb{E}^\nu[X] - \mathbb{E}^\mu[X] \leq \tau(X) \text{ for all } X \in L^\infty\}.\end{aligned}$$

The uniqueness follows directly from Proposition 2.10.

(ii) $\Rightarrow$ (iii): Obvious. (iii) $\Rightarrow$ (i): Let $\mathcal{R} \subseteq \mathcal{M}_1$ have the stated properties. Proving τ is a μ-coherent adjustment is straightforward. Let $(X_n)_{n=1}^\infty$ be a bounded sequence that converges $\mathbb{P}$-a.s. to X. By the DCT, $\lim_{n\to\infty} \mathbb{E}\left[(D_\nu - D_\mu) X_n\right] = \mathbb{E}\left[(D_\nu - D_\mu) X\right]$. Therefore,

$$\begin{aligned}\tau(X) = \sup_{\nu \in \mathcal{R}} \mathbb{E}\left[(D_\nu - D_\mu) X\right] = \sup_{\nu \in \mathcal{R}} \liminf_{n\to\infty} \mathbb{E}\left[(D_\nu - D_\mu) X_n\right] &\leq \liminf_{n\to\infty} \sup_{\nu \in \mathcal{R}} \mathbb{E}\left[(D_\nu - D_\mu) X_n\right] \\ &= \liminf_{n\to\infty} \tau(X_n),\end{aligned}$$

proving the Fatou continuity of τ. □

Corollary 2.1. *Let $\mu \in \mathcal{M}_1(\mathcal{G})$ and $\tau : L^\infty \to \mathbb{R}$ be a μ-coherent adjustment. The following are equivalent.*

(i) The mapping $\tau \in \mathrm{CA}(\mu)$;

(ii) the mapping τ is $w(L^\infty, \mathcal{M})$-lower semicontinuous;

(iii) *the mapping $\mathbb{E}^\mu + \tau$ is Fatou;*

(iv) *the mapping $\mathbb{E}^\mu + \tau$ is $w(L^\infty, \mathcal{M})$-lower semicontinuous.*

Proof. Proof: (i) $\Rightarrow$ (ii): This follows from the representation given in Proposition 2.15 and by the fact that the supremum of continuous functions is lower semicontinuous. (ii) $\Rightarrow$ (iv): The sum of lower semicontinuous functions is lower semicontinuous. (iv) $\Rightarrow$ (iii): This follows since $\mathbb{E}^\mu + \tau$ is a coherent risk measure that is lower $w(L^\infty, \mathcal{M})$-lower semicontinuous. (iii) $\Rightarrow$ (i): Since $\mathbb{E}^\mu + \tau$ is Fatou, we can use a similar argument as Proposition 2.15 to show that τ has the representation (2.18). Thus by Proposition 2.15, the mapping τ is Fatou. □

Proof. Proof of Theorem 2.3: Let ρ be a Fatou coherent risk measure. For $\mu \in \mathcal{S}_\rho^{\mathcal{G}}$, we define $\mathcal{S}(\mu) = \mathcal{E}(\mu) \cap \mathcal{S}_\rho$, note that $\mathcal{S}_\rho = \bigcup_{\mu \in \mathcal{S}_\rho^{\mathcal{G}}} \mathcal{S}(\mu)$. We have

$$\begin{aligned}
\rho(X) = \sup_{\nu \in \mathcal{S}_\rho} \mathbb{E}\left[D_\mu X\right] &= \sup_{\nu \in \mathcal{S}_\rho} \left(\mathbb{E}\left[\mathbb{E}[D_\nu|\mathcal{G}]X\right] + \mathbb{E}[(D_\nu - \mathbb{E}[D_\nu|\mathcal{G}])X]\right) \\
&= \sup_{\mu \in \mathcal{S}_\rho^{\mathcal{G}}} \sup_{\nu \in \mathcal{S}(\mu)} \left(\mathbb{E}\left[D_\mu X\right] + \mathbb{E}\left[(D_\nu - D_\mu)X\right]\right) \\
&= \sup_{\mu \in \mathcal{S}_\rho^{\mathcal{G}}} \left(\mathbb{E}^\mu[X] + \tau_\mu(X)\right), \quad X \in L^\infty,
\end{aligned}$$

where $\tau_\mu(X) = \sup_{\nu \in \mathcal{S}(\mu)}(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X])$, $X \in L^\infty$. The fact that $\tau_\mu \in \mathrm{CA}(\mu)$ follows from Proposition 2.15. It is straightforward to see that $\mathcal{S}_\rho^{\mathcal{G}}$ is convex. For the converse, note that (2.8) is the supremum of coherent risk measures (by Proposition 2.6; see (2.7)) and thus a coherent risk measure. By Corollary 2.1, $\mathbb{E}^\mu + \tau_\mu$ is $w(L^\infty, \mathcal{M})$-lower semicontinuous for all $\mu \in \mathcal{R}$. As the supremum of $w(L^\infty, \mathcal{M})$-lower semicontinuous and coherent risk measures, ρ is also one, and hence it is Fatou. Let $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$. We have that $\mathcal{R}$ in (2.8) is dense in $\mathcal{S}_{\tilde{\rho}}$, and the last claim now follows from the same argument in the proof of Theorem 2.1. □

2.12.2 Additivity on the subspace

In this section, we discuss coherent risk measures that are not only subadditive but also additive on $L^\infty(\mathcal{G})$, which turn out to have a strong connection to coherent adjustments.

For $\rho : L^\infty \to \mathbb{R}$, we say that ρ is *$\mathcal{G}$-additive* if $\rho(X + Y) = \rho(X) + \rho(Y)$ for all $X, Y \in L^\infty(\mathcal{G})$. Additivity has the interpretation as risk neutrality under a particular probability, an idea that goes back to at least de Finetti [1931]; see Wakker [2010, p. 23]

for discussions on this. Hence, $\mathcal{G}$-additivity reflects that the risk measure is risk neutral on the subset $L^\infty(\mathcal{G})$ of L^∞. In other words, such a risk measure is neutral for risks without uncertainty; recall that in the context of Section 2.2.2, random losses in $L^\infty(\mathcal{G})$ do not have uncertainty in their distributions.

For $X \in L^\infty$, we say that ρ is *linear along* X if $\rho(\lambda X) = \lambda\rho(X)$ for all $\lambda \in \mathbb{R}$; see Bellini et al. [2021] and Liebrich and Munari [2022] for more discussions on this technical property. The following proposition gives necessary and sufficient conditions for dropping the supremum in representation (2.8) in the case of $\mathcal{G}$-law invariance.

Proposition 2.16. *Let $\rho : L^\infty \to \mathbb{R}$ be a Fatou coherent risk measure. The following are equivalent.*

(i) The mapping ρ is $\mathcal{G}$-additive and $\mathcal{G}$-law invariant;

(ii) the mapping ρ is linear along a nonconstant $X \in L^\infty(\mathcal{G})$ and $\mathcal{G}$-law invariant;

(iii) there exists $\tau \in \mathrm{CA}(\mathbb{P}|_{\mathcal{G}})$ such that

$$\rho(X) = \mathbb{E}[X] + \tau(X),\ X \in L^\infty.$$

Proof. Proof: (i)⇒(ii): Since ρ is $\mathcal{G}$-additive and coherent, it is linear on $L^\infty(\mathcal{G})$. Since $L^\infty(\mathcal{G})$ contains non-constant random variables, (ii) holds. (ii)⇒(iii): The fact that $\rho|_{L^\infty(\mathcal{G})} = \mathbb{E}|_{L^\infty(\mathcal{G})}$ follows from Theorem 4.5 and Theorem 4.7 of Bellini et al. [2021]. Using Proposition 2.6, we prove the claim. (iii)⇒(i): This is clear by definition. □

2.12.3 Uniqueness of the representation in Theorem 2.3

A natural question would be, under what conditions is representation (2.8) unique? When the coherent risk measure is Lebesgue continuous, there is a definite answer to this. We first need some results before we can state and prove the theorem regarding uniqueness.

Proposition 2.17. *If $X_\lambda \to X$ in $w\,(L^1, L^\infty)$ then $\mathbb{E}[X_\lambda|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ in $w\,(L^1, L^\infty)$.*

Proof. Proof: Let $X_\lambda \to X$ in $w\,(L^1, L^\infty)$ and $Y \in L^\infty$, we have

$$\mathbb{E}[\mathbb{E}[X_\lambda|\mathcal{G}]Y] = \mathbb{E}[X_\lambda \mathbb{E}[Y|\mathcal{G}]] \to \mathbb{E}[X\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]Y].$$

Implying that $\mathbb{E}[X_\lambda|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ in $w\,(L^1, L^\infty)$. □

Corollary 2.2. *The set $L^1(\mathcal{G})$ is $w\,(L^1, L^\infty)$-closed in L^1.*

Proof. Proof: Let $X_\lambda \in L^1(\mathcal{G})$ and $X_\lambda \to X$ in $w(L^1, L^\infty)$. This also implies that $X_\lambda \to \mathbb{E}[X|\mathcal{G}]$ in $w(L^1, L^\infty)$ by Proposition 2.17. Since $w(L^1, L^\infty)$ is Hausdorff, $X = \mathbb{E}[X|\mathcal{G}] \implies X \in L^1(\mathcal{G})$. Therefore $L^1(\mathcal{G})$ is $w(L^1, L^\infty)$-closed in L^1. □

If $(\mathfrak{X}, \mathfrak{Y}, \langle\ ,\ \rangle)$ is a dual pairing, we say that $\mathcal{R} \subseteq \mathfrak{X}$ is $w(\mathfrak{X}, \mathfrak{Y})$-precompact if the $w(\mathfrak{X}, \mathfrak{Y})$-closure of $\mathcal{R}$ is $w(\mathfrak{X}, \mathfrak{Y})$-compact.

The next proposition, parallel to Proposition 2.15, characterizes Lebesgue-continuous μ-coherent adjustments.

Proposition 2.18. *Let $\tau : L^\infty \to \mathbb{R}$ and $\mu \in \mathcal{M}_1(\mathcal{G})$. The following are equivalent.*

(i) The mapping τ is a Lebesgue-continuous μ-coherent adjustment;

(ii) there exists a $w(\mathcal{M}, L^\infty)$-compact and convex set $\mathcal{S}_\tau \subseteq \mathcal{E}(\mu)$ such that

$$\tau(X) = \sup_{\nu \in \mathcal{S}_\tau} \left(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X]\right), \quad X \in L^\infty; \tag{2.19}$$

(iii) there exists $w(\mathcal{M}, L^\infty)$-precompact $\mathcal{R} \subseteq \mathcal{E}(\mu)$ such that

$$\tau(X) = \sup_{\nu \in \mathcal{R}} \left(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X]\right) \quad X \in L^\infty.$$

In case (ii) holds, the set $\mathcal{S}_\tau$ is uniquely given by

$$\left\{\nu \in \mathcal{M}_1 : \mathbb{E}^\nu[X] - \mathbb{E}^\mu[X] \leq \tau(X) \text{ for all } X \in L^\infty\right\}.$$

Proof. Proof: Everything follows the same line of reasoning as Proposition 2.15, except for proving Lebesgue continuity in (iii) ⇒ (i). Let $(X_n)_{n=1}^\infty$ be a sequence in L^∞ of uniformly bounded random variables that converge $\mathbb{P}$-a.s. to $X \in L^\infty$. Since $X \mapsto \sup_{\nu \in \mathcal{R}} \mathbb{E}[D_\nu X]$ is a Lebesgue-continuous coherent risk measure, we have

$$\begin{aligned}\lim_{n\to\infty} \tau(X_n) = \lim_{n\to\infty} \sup_{\nu\in\mathcal{R}} \mathbb{E}\left[\left(D_\nu - D_\mu\right) X_n\right] &= \lim_{n\to\infty} \sup_{\nu\in\mathcal{R}} \mathbb{E}\left[D_\nu X_n\right] - \lim_{n\to\infty} \mathbb{E}[D_\mu X_n] \\ &= \sup_{\nu\in\mathcal{R}} \mathbb{E}[D_\nu X] - \mathbb{E}[D_\mu X] = \tau(X).\end{aligned}$$

This proves the Lebesgue continuity of τ. □

Let $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ be $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-closed and convex, and assume that $(\tau_\mu)_{\mu\in\mathcal{R}}$ is a CA-assignment. We say that $(\tau_\mu)_{\mu\in\mathcal{R}}$ is *Lebesgue continuous* if τ_μ is Lebesgue continuous for all $\mu \in \mathcal{R}$. We say that $(\tau_\mu)_{\mu\in\mathcal{R}}$ is *regular* if

(a) For any fixed $X \in L^\infty$, concavity of $\mu \mapsto \tau_\mu(X)$ holds; that is, for all $\mu_1, \mu_2 \in \mathcal{R}$ and $\lambda \in (0,1)$ we have

$$\lambda\tau_{\mu_1}(X) + (1-\lambda)\tau_{\mu_2}(X) \leq \tau_{\lambda\mu_1+(1-\lambda)\mu_2}(X).$$

See e.g., Wang et al. [2020] for concavity for distributional mixtures in the context of Choquet integrals.

(b) A weak form of semicontinuity holds: If $(\mu_\lambda)_{\lambda\in\Lambda}$ is a net in $\mathcal{R}$ that converges to $\mu \in \mathcal{R}$ in $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$, then for all $X \in L^\infty$,

$$\tau_\mu(X) \geq \liminf_\lambda \tau_{\mu_\lambda}(X).$$

See Bredon [1993] for background information on nets, subnets, and their convergence.

Proposition 2.19. *Let $\rho : L^\infty \to \mathbb{R}$ be a Lebesgue-continuous coherent risk measure. There exists a unique $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-compact and convex set $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ and a unique regular Lebesgue-continuous CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ such that ρ can be represented by equation (2.8).*

Proof. Proof: Let $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$, by Proposition 2.14, we have that $\mathcal{S}_\rho^\mathcal{G} = \mathcal{S}_{\tilde{\rho}}$ is compact. Let $(\tau_\mu)_{\mu\in\mathcal{S}_\rho^\mathcal{G}}$ be the same as in the construction of representation (2.8) in the proof of Theorem 2.3. We claim that for all $\mu \in \mathcal{S}_\rho^\mathcal{G}$, $\mathcal{S}(\mu)$ is $w(\mathcal{M}, L^\infty)$-compact and convex. Fix some $\mu \in \mathcal{S}_\rho^\mathcal{G}$, convexity of $\mathcal{S}(\mu)$ follows from the convexity of $\mathcal{S}_\rho$ and the convexity of $\mathcal{E}(\mu)$. For compactness note that $\mathcal{S}(\mu)$ is $w(\mathcal{M}, L^\infty)$-precompact as $\mathcal{S}(\mu) \subseteq \mathcal{S}_\rho$. Therefore, compactness of $\mathcal{S}(\mu)$ would follow if we prove that $\mathcal{S}(\mu)$ is $w(\mathcal{M}, L^\infty)$-closed. This can be seen from the fact that $\mathcal{S}_\rho$ and $\mathcal{E}(\mu)$ are both $w(\mathcal{M}, L^\infty)$-closed. From the construction in 2.3, $\tau_\mu(X) = \sup_{\nu\in\mathcal{S}(\mu)} (\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X])$, $X \in L^\infty$, thus, by Proposition 2.18, τ_μ is Lebesgue continuous.

We will now show that $(\tau_\mu)_{\mu\in\mathcal{S}_\rho^\mathcal{G}}$ is regular. First, to show concavity (a), let $\mu_1, \mu_2 \in \mathcal{S}_\rho^\mathcal{G}$, $\nu_1 \in \mathcal{S}(\mu_1), \nu_2 \in \mathcal{S}(\mu_2)$, $X \in L^\infty$, and $\lambda \in (0,1)$. Define $\upsilon = \lambda\mu_1 + (1-\lambda)\mu_2$. Since $\mathcal{S}_\rho$ is convex, we have $\lambda\nu_1 + (1-\lambda)\nu_2 \in \mathcal{S}(\upsilon)$, and thus

$$\mathbb{E}\left[(D_{\nu_1} - D_{\mu_1})\,\lambda X\right] + \mathbb{E}\left[(D_{\nu_2} - D_{\mu_2})\,(1-\lambda)X\right] \leq \tau_\upsilon(X). \tag{2.20}$$

Taking the supremum of the left side of (2.20) over $(\nu_1, \nu_2) \in \mathcal{S}(\mu_1) \times \mathcal{S}(\mu_2)$ obtains the result. To show semicontinuity (b), let $(\mu_\lambda)_{\lambda\in\Lambda}$ be a net in $\mathcal{S}_\rho^\mathcal{G}$ and $\mu \in \mathcal{S}_\rho^\mathcal{G}$ such that $\mu_\lambda \to \mu$ in $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$ and let $X \in L^\infty$. For each λ choose $\nu_\lambda^* \in \mathcal{S}(\mu_\lambda)$ such that $\tau_{\mu_\lambda}(X) = \mathbb{E}[(D_{\nu_\lambda^*} - D_{\mu_\lambda})X]$, which is possible since $\mathcal{S}(\mu_\lambda)$ is $w(\mathcal{M}, L^\infty)$-compact. Since

the net $(\nu^*_\lambda)_{\lambda\in\Lambda}$ takes values in the $w(\mathcal{M},L^\infty)$-compact set $\mathcal{S}_\rho$, there exists a convergent subnet $(\nu^*_\gamma)_{\gamma\in\Gamma}$. Denote the limit of this subnet by $\nu\in\mathcal{S}_\rho$. For $Y\in L^\infty(\mathcal{G})$, by using the tower property and the convergence, we get

$$\mathbb{E}\left[\mathbb{E}[D_\nu|\mathcal{G}]Y\right]=\mathbb{E}\left[D_\nu Y\right]=\lim_\gamma \mathbb{E}\left[D_{\nu^*_\gamma}Y\right]=\lim_\gamma \mathbb{E}\left[D_{\mu_\gamma}Y\right]=\mathbb{E}[D_\mu Y].$$

Thus $\nu\in\mathcal{S}(\mu)$ and

$$\begin{aligned}\tau_\mu(X)\geq \mathbb{E}\left[(D_\nu-D_\mu)X\right]=\lim_\gamma \mathbb{E}\left[\left(D_{\nu^*_\gamma}-D_{\mu_\gamma}\right)X\right]&=\lim_\gamma \tau_{\mu_\gamma}(X)\\ &=\liminf_\gamma \tau_{\mu_\gamma}(X)\geq \liminf_\lambda \tau_{\mu_\lambda}(X).\end{aligned}$$

This shows that $(\tau_\mu)_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}$ is regular.

Now we prove the uniqueness of $(\tau_\mu)_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}$ by assuming that there exists another regular Lebesgue-continuous CA-assignment $(\tilde\tau_\mu)_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}$ such that

$$\rho(X)=\sup_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}\left(\mathbb{E}^\mu[X]+\tilde\tau_\mu(X)\right),\quad X\in L^\infty.$$

We will show that $\tau_\mu=\tilde\tau_\mu$ for all $\mu\in\mathcal{S}^{\mathcal{G}}_\rho$. For each $\mu\in\mathcal{S}^{\mathcal{G}}_\rho$, let $\mathcal{R}(\mu)\subseteq\mathcal{M}_1$ be the $w(\mathcal{M},L^\infty)$-compact convex set that gives representation (2.19) in Proposition 2.18 for $\tilde\tau_\mu$. Let $\tilde{\mathcal{S}}=\bigcup_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}\mathcal{R}(\mu)\subseteq\mathcal{M}_1$. Note that

$$\begin{aligned}\sup_{\nu\in\tilde{\mathcal{S}}}\mathbb{E}^\nu[X]=\sup_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}\left(\sup_{\nu\in\mathcal{R}(\mu)}\mathbb{E}^\nu[X]\right)&=\sup_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}\left(\mathbb{E}^\mu[X]+\sup_{\nu\in\mathcal{R}(\mu)}\left(\mathbb{E}^\nu[X]-\mathbb{E}^\mu[X]\right)\right)\\ &=\sup_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}\left(\mathbb{E}^\mu[X]+\tilde\tau_\mu(X)\right)=\rho(X),\quad X\in L^\infty.\end{aligned}$$

Therefore $\mathrm{conv}(\tilde{\mathcal{S}})$ is $w(\mathcal{M},L^\infty)$-dense in $\mathcal{S}_\rho$ by Proposition 2.10. We will show that regularity of $(\tilde\tau_\mu)_{\mu\in\mathcal{S}^{\mathcal{G}}_\rho}$ implies that $\tilde{\mathcal{S}}$ is convex and $w(\mathcal{M},L^\infty)$-closed. Let $\nu_1\in\mathcal{R}(\mu_1)$, $\nu_2\in\mathcal{R}(\mu_2)$, and $\lambda\in(0,1)$. We have, for $X\in L^\infty$,

$$\begin{aligned}&\mathbb{E}\left[(\lambda D_{\nu_1}+(1-\lambda)D_{\nu_2}-\lambda D_{\mu_1}-(1-\lambda)D_{\mu_2})X\right]\\ &=\mathbb{E}\left[(D_{\nu_1}-D_{\mu_1})\lambda X\right]+\mathbb{E}\left[(D_{\nu_2}-D_{\mu_2})(1-\lambda)X\right]\\ &\leq\tilde\tau_{\mu_1}(\lambda X)+\tilde\tau_{\mu_2}((1-\lambda)X)\leq\tau_{\lambda D_{\mu_1}+(1-\lambda)D_{\mu_2}}(X).\end{aligned}$$

Therefore, $\lambda\nu_1+(1-\lambda)\nu_2\in\mathcal{R}(\lambda\mu_1+(1-\lambda)\mu_2)\subseteq\tilde{\mathcal{S}}$ by Proposition 2.18. This implies that

$\tilde{\mathcal{S}}$ is convex. Let $(\nu_\lambda)_{\lambda\in\Lambda}$ be a net taking values in $\tilde{\mathcal{S}}$ and $\nu \in \mathcal{M}_1$ such that $\lim_\lambda \nu_\lambda = \nu$ in $w(\mathcal{M}, L^\infty)$, and define $\mu_\lambda = \nu_\lambda|_{\mathcal{G}}$ and $\mu = \nu|_{\mathcal{G}}$. By Proposition 2.17, $\lim_\lambda \mu_\lambda = \mu$ in $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$. Since $\mathcal{S}_\rho^{\mathcal{G}}$ is $w(\mathcal{M}(\mathcal{G}), L^\infty(\mathcal{G}))$-closed, we have that $\mu \in \mathcal{S}_\rho^{\mathcal{G}}$. For $X \in L^\infty$,

$$\mathbb{E}\left[(D_\nu - D_\mu)\, X\right] = \lim_\lambda \mathbb{E}\left[(D_{\nu_\lambda} - D_{\mu_\lambda})\, X\right] \leq \liminf_\lambda \tilde{\tau}_{\mu_\lambda}(X) \leq \tilde{\tau}_\mu(X),$$

implying that $\nu \in \mathcal{R}(\mu) \subseteq \tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}$ is $w(\mathcal{M}, L^\infty)$-closed. Therefore $\tilde{\mathcal{S}} = \mathcal{S}_\rho$, which means $\mathcal{S}(\mu) = \mathcal{R}(\mu)$ for all $\mu \in \mathcal{S}_\rho^{\mathcal{G}}$, implying $\tilde{\tau}_\mu = \tau_\mu$ for all $\mu \in \mathcal{S}_\rho^{\mathcal{G}}$. □

2.12.4 Additional results and proofs accompanying Section 2.5.2

Proof. Proof of Lemma 2.1: The only if is clear. Let $Y_1, Y_2 \in L^\infty(\mathcal{G})$ with $Y_1 \overset{\mathrm{d}}{=}_{\mathbb{P}} Y_2$ and $Z \in \ker(\mathcal{G})$. Define $X = Z + Y_1$, we have $\mathbb{E}[X|\mathcal{G}] = Y_1$ therefore

$$\rho(Z + Y_1) = \rho(X) = \rho(X - \mathbb{E}[X|\mathcal{G}] + Y_2) = \rho(Z + Y_2).$$

Therefore ρ is strongly $\mathcal{G}$-law invariant. □

The proof of Theorem 2.4 relies on several results we present below.

Lemma 2.3. *Föllmer and Schied [2016, Lemma A.32] Let $\mathcal{J} \subseteq \mathcal{F}$ be a sub-σ-algebra such that $(\Omega, \mathcal{J}, \mathbb{P})$ is atomless. For all $X \in L^1(\mathcal{J})$ there exists a uniform $(0,1)$ random variable $U \in L^\infty(\mathcal{J})$ such that $X = F^{-1}_{\mathbb{P},X}(U)$ $\mathbb{P}$-a.s.*

Lemma 2.4. *Let $X \in L^\infty$ (resp. L^1), $Y \in L^1$ (resp. L^∞) and $Z \in L^1(\mathcal{G})$ (resp. $L^\infty(\mathcal{G})$) with $Y \overset{\mathrm{d}}{=}_{\mathbb{P}} Z$. Then there exists $W \in L^\infty(\mathcal{G})$ (resp. $L^1(\mathcal{G})$) such that $X \overset{\mathrm{d}}{=}_{\mathbb{P}} W$ and $\mathbb{E}[XY] = \mathbb{E}[WZ]$.*

Proof. Proof:

Let $X \in L^\infty$, $Y \in L^1$ and $Z \in L^1(\mathcal{G})$ with $Y \overset{\mathrm{d}}{=}_{\mathbb{P}} Z$. Since $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, by Lemma 2.3, there exists a uniform on $(0,1)$ random variable $U_Z \in L^\infty(\mathcal{G})$ such that $F^{-1}_{\mathbb{P},Y}(U_Z) = Z$ $\mathbb{P}$-a.s. For all $p \in [0,1]$, define $V_p = (p - U_Z)\mathbb{1}_{\{U_Z \leq p\}} + U_Z \mathbb{1}_{\{U_Z > p\}}$. It is easy to see that V_p is a $\mathcal{G}$-measurable uniform $(0,1)$ random variable. For $p \in [0,1]$, define $W_p = F^{-1}_{\mathbb{P},X}(V_p)$. We have $W_p \in L^\infty(\mathcal{G})$ and $W_p \overset{\mathrm{d}}{=}_{\mathbb{P}} X$. Given a sequence $(p_n)_{n=1}^\infty$ with $p_n \to p$, it is clear that $W_{p_n} Z \to W_p Z$ $\mathbb{P}$-a.s.

For all $n \in \mathbb{N}$, $W_{p_n} \overset{\mathrm{d}}{=}_{\mathbb{P}} X$, implying that there exists an $M > 0$ such that

$$\mathbb{P}(\sup_{n\in\mathbb{N}} |W_{p_n}| \leq M) = 1.$$

Thus, $\sup_{n\in\mathbb{N}} |W_{p_n} Z| \leq M|Z| \in L^1(\mathcal{G})$, and $(W_{p_n} Z)_{n=1}^{\infty}$ is uniformly integrable. All of this combines to show that $\lim_{n\to\infty} \mathbb{E}[W_{p_n} Z] = \mathbb{E}[W_p Z]$. Therefore, the function

$$\zeta : [0,1] \to \mathbb{R} : p \mapsto \mathbb{E}[W_p Z]$$

is continuous. Note that

$$\zeta(1) = \mathbb{E}\left[F^{-1}_{\mathbb{P},X}(1-U_Z)F^{-1}_{\mathbb{P},Y}(U_Z)\right] \leq \mathbb{E}[XY] \leq \mathbb{E}\left[F^{-1}_{\mathbb{P},X}(U_Z)F^{-1}_{\mathbb{P},Y}(U_Z)\right] = \zeta(0).$$

Thus, by the intermediate value theorem, there exists a $p_0 \in [0,1]$ such that $\mathbb{E}[W_{p_0} Z] = \mathbb{E}[XY]$. Let $X \in L^1$, $Y \in L^\infty$ and $Z \in L^\infty(\mathcal{G})$ with $Y \overset{\mathrm{d}}{=}_{\mathbb{P}} Z$. The only thing that has to be checked is the uniform integrability condition. Let $M = \|Z\|_\infty + 1$ and $K > 0$, we have

$$\mathbb{E}\left[|W_{p_n} Z|; |W_{p_n} Z| > K\right] \leq M\mathbb{E}\left[|W_{p_n}|; |W_{p_n}| > K/M\right] = M\mathbb{E}\left[|X|; |X| > K/M\right]$$

for all $n \in \mathbb{N}$. This shows that $(W_{p_n} Z)_{n=1}^{\infty}$ is uniformly integrable. □

For a $\mathcal{G}$-law invariant set $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$, we say that the CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ is *$\mathcal{G}$-law invariant* if for all $\mu, \nu \in \mathcal{R}$ with $D_\mu \overset{\mathrm{d}}{=}_{\mathbb{P}} D_\nu$, it holds that $\tau_\mu = \tau_\nu$.

Proposition 2.20. *Let ρ be a Fatou coherent risk measure. The following are equivalent.*

(i) The risk measure ρ is strongly $\mathcal{G}$-law invariant;

(ii) the set $\mathcal{S}_\rho^{\mathcal{G}}$ is $\mathcal{G}$-law invariant and for all $\mu \in \mathcal{S}_\rho$ and $Y \in L^1(\mathcal{G})$ with $\mathbb{E}[D_\mu|\mathcal{G}] \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$, it holds that

$$D_\mu - \mathbb{E}[D_\mu|\mathcal{G}] + Y \in \widehat{\mathcal{S}}_\rho.$$

(iii) There exists a convex $\mathcal{G}$-law invariant set $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ and a $\mathcal{G}$-law invariant CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ such that

$$\rho(X) = \sup_{\mu\in\mathcal{R}}(\mathbb{E}^\mu[X] + \tau_\mu(X)), \quad X \in L^\infty.$$

Proof. Proof: (i) ⇒ (ii) : We will prove that $\mathcal{S}_\rho^{\mathcal{G}}$ is $\mathcal{G}$-law invariant after proving the latter

claim. Let $\mu \in \mathcal{S}_\rho$ and let $Y \in L^1(\mathcal{G})$ with $\mathbb{E}[D_\mu|\mathcal{G}] \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$ such that

$$D_\mu - \mathbb{E}[D_\mu|\mathcal{G}] + Y \notin \widehat{\mathcal{S}}_\rho.$$

By the Hahn-Banach separation Theorem, there exists $X \in L^\infty$ such that

$$\rho(X) < \mathbb{E}[(D_\mu - \mathbb{E}[D_\mu|\mathcal{G}] + Y)X].$$

By Lemma 2.4, find $Z \in L^\infty(\mathcal{G})$ with $\mathbb{E}[X|\mathcal{G}] \overset{\mathrm{d}}{=}_{\mathbb{P}} Z$ such that $\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[D_\mu|\mathcal{G}]Z]$. We have

$$\begin{aligned}
\mathbb{E}[(D_\mu - \mathbb{E}[D_\mu|\mathcal{G}] + Y)X] &= \mathbb{E}[(D_\mu - \mathbb{E}[D_\mu|\mathcal{G}])X] + \mathbb{E}[YX] \\
&= \mathbb{E}[(D_\mu - \mathbb{E}[D_\mu|\mathcal{G}])X] + \mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]] \\
&= \mathbb{E}[D_\mu(X - \mathbb{E}[X|\mathcal{G}])] + \mathbb{E}[\mathbb{E}[D_\mu|\mathcal{G}]Z] \\
&= \mathbb{E}[D_\mu(X - \mathbb{E}[X|\mathcal{G}])] + \mathbb{E}[D_\mu Z] \\
&= \mathbb{E}[D_\mu(X - \mathbb{E}[X|\mathcal{G}] + Z)] \\
&= \mathbb{E}^\mu[X - \mathbb{E}[X|\mathcal{G}] + Z].
\end{aligned}$$

Therefore, $\rho(X) < \mathbb{E}^\mu[X - \mathbb{E}[X|\mathcal{G}] + Z] \leq \rho(X - \mathbb{E}[X|\mathcal{G}] + Z)$, implying that ρ is not strongly $\mathcal{G}$-law invariant. We will now prove that $\mathcal{S}_\rho^{\mathcal{G}}$ is $\mathcal{G}$-law invariant. Let $\mu \in \mathcal{S}_\rho$ and $Y \in L^1(\mathcal{G})$ such that $\mathbb{E}[D_\mu|\mathcal{G}] \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$. We have $D_\mu - \mathbb{E}[D_\mu] + Y \in \widehat{\mathcal{S}}_\rho$ and $Y = \mathbb{E}[D_\mu - \mathbb{E}[D_\mu|\mathcal{G}] + Y|\mathcal{G}] \in \widehat{\mathcal{S}}_\rho^{\mathcal{G}}$.

(ii) $\Rightarrow$ (iii) : Since ρ is a Fatou coherent risk measure, by Theorem 2.3, we have

$$\rho(X) = \sup_{\mu \in \mathcal{S}_\rho^{\mathcal{G}}} (\mathbb{E}^\mu[X] + \tau_\mu(X)), \quad X \in L^\infty,$$

where $\tau_\mu(X) = \sup_{\nu \in \mathcal{S}(\mu)}(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X])$, $X \in L^\infty$ and $\mathcal{S}(\mu) = \mathcal{E}(\mu) \cap \mathcal{S}_\rho$. The set $\mathcal{S}_\rho^{\mathcal{G}}$ is $\mathcal{G}$-law invariant by assumption and is convex as $\mathcal{S}_\rho$ is convex. Let $\mu_1, \mu_2 \in \mathcal{S}_\rho^{\mathcal{G}}$ with $D_{\mu_1} \overset{\mathrm{d}}{=}_{\mathbb{P}} D_{\mu_2}$ and $\nu_1 \in \mathcal{S}(\mu_1)$. By assumption, there exists $\upsilon \in \mathcal{S}_\rho$ with $D_\upsilon = D_{\nu_1} - D_{\mu_1} + D_{\mu_2}$. This implies that $\upsilon \in \mathcal{E}(\mu_2)$ and thus $\upsilon \in \mathcal{S}(\mu_2)$. Yet we have $D_\upsilon - D_{\mu_2} = D_{\nu_1} - D_{\mu_1}$, implying that for all $X \in L^\infty$,

$$\tau_{\mu_2}(X) = \sup_{\upsilon \in \mathcal{S}(\mu_2)} \mathbb{E}[(D_\upsilon - D_{\mu_2})X] \geq \sup_{\upsilon \in \mathcal{S}(\mu_1)} \mathbb{E}[(D_\upsilon - D_{\mu_1})X] = \tau_{\mu_1}(X).$$

By symmetry, we get the reverse inequality.

(iii) $\Rightarrow$ (i) : Let $\mu \in \mathcal{R}$, $X \in L^\infty$ and $Y \in L^\infty(\mathcal{G})$ such that $\mathbb{E}[X|\mathcal{G}] \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$. Define $W = X - \mathbb{E}[X|\mathcal{G}] + Y$. By Lemma 2.4, find $Z \in L^1(\mathcal{G})$ such $D_\mu \overset{\mathrm{d}}{=}_{\mathbb{P}} Z$ and $\mathbb{E}[D_\mu Y] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]]$. Let $\nu \in \mathcal{M}_1(\mathcal{G})$ be such that $D_\nu = Z$. Since $\mathcal{R}$ is $\mathcal{G}$-law invariant, $\nu \in \mathcal{R}$. We have

$$\begin{aligned}\mathbb{E}^\mu[W] + \tau_\mu(W) = \mathbb{E}[D_\mu Y] + \tau_\mu(X) = \mathbb{E}[D_\nu \mathbb{E}[X|\mathcal{G}]] + \tau_\nu(X) = \mathbb{E}[D_\nu X] + \tau_\nu(X) \\ = \mathbb{E}^\nu[X] + \tau_\nu(X).\end{aligned}$$

By Lemma 2.4, find $Z \in L^1(\mathcal{G})$ such $D_\mu \overset{\mathrm{d}}{=}_{\mathbb{P}} Z$ and $\mathbb{E}[D_\mu \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[ZY]$. Let $\nu \in \mathcal{M}_1(\mathcal{G})$ be such that $D_\nu = Z$. Since $\mathcal{R}$ is $\mathcal{G}$-law invariant, $\nu \in \mathcal{R}$. We have

$$\begin{aligned}\mathbb{E}^\mu[X] + \tau_\mu(X) = \mathbb{E}[D_\mu \mathbb{E}[X|\mathcal{G}]] + \tau_\mu(W) = \mathbb{E}[D_\nu Y] + \tau_\nu(W) = \mathbb{E}[D_\nu W] + \tau_\nu(W) \\ = \mathbb{E}^\nu[W] + \tau_\nu(W).\end{aligned}$$

Therefore, $\{\mathbb{E}^\mu[X] + \tau_\mu(X) : \mu \in \mathcal{R}\} = \{\mathbb{E}^\mu[W] + \tau_\mu(W) : \mu \in \mathcal{R}\}$, and $\rho(X) = \rho(W)$. $\square$

Proof. Proof of Theorem 2.4: (i) $\Rightarrow$ (ii): By Proposition 2.20, we have that there exist a $\mathcal{G}$-law-invariant convex $\mathcal{R} \subseteq \mathcal{M}_1(\mathcal{G})$ and a $\mathcal{G}$-law-invariant CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ such that

$$\rho(X) = \sup_{\mu\in\mathcal{R}} \left(\mathbb{E}^\mu[X] + \tau_\mu(X)\right), \quad X \in L^\infty.$$

By Föllmer and Schied [2016, Theorem 4.62], for $Q \in \mathcal{M}_B$ and $\mu \in \mathcal{M}_Q$, we have

$$\mathcal{M}_Q = \left\{\nu \in \mathcal{M}_1(\mathcal{G}) \mid D_\nu \overset{\mathrm{d}}{=} D_\mu\right\}.$$

Define $\mathcal{D} = \{Q \in \mathcal{M}_\mathcal{B} \mid \mathcal{M}_Q \cap \mathcal{R} \neq \varnothing\}$, and note that $\mathcal{R} = \bigcup_{Q\in\mathcal{D}} \mathcal{M}_Q$ since $\mathcal{R}$ is $\mathcal{G}$-law invariant. Define $\tau_Q = \tau_\mu$ for some $\mu \in \mathcal{M}_Q$ (thus all $\mu \in \mathcal{M}_Q$ since $(\tau_\mu)_{\mu\in\mathcal{R}}$ is law invariant), therefore

$$\begin{aligned}\rho(X) = \sup_{\mu\in\mathcal{R}} \left(\mathbb{E}^\mu[X] + \tau_\mu(X)\right) = \sup_{Q\in\mathcal{D}} \sup_{\mu\in\mathcal{M}_Q} \left(\mathbb{E}^\mu[X] + \tau_\mu(X)\right) \\ \overset{(2.10)}{=} \sup_{Q\in\mathcal{D}} \left(\int_{[0,1)} \mathrm{ES}_\alpha(\mathbb{E}[X|\mathcal{G}])Q(\mathrm{d}\alpha) + \tau_Q(X)\right).\end{aligned}$$

(ii) $\Rightarrow$ (i): The fact that ρ is cash invariant, convex, and positively homogenous is clear.

Let $X, Y \in L^\infty$ with $X \leq Y$ and $Q \in \mathcal{D}$. Then we have

$$\begin{aligned}\int_{[0,1)} \mathrm{ES}_\alpha\left(\mathbb{E}[X|\mathcal{G}]\right) Q(\mathrm{d}\alpha) + \tau_Q(X) &= \sup_{\mu \in \mathcal{M}_Q} \left(\mathbb{E}^\mu\left[X\right] + \tau_Q(X)\right) \\ &\leq \sup_{\mu \in \mathcal{M}_Q} \left(\mathbb{E}^\mu\left[Y\right] + \tau_Q(Y)\right) \\ &= \int_{[0,1)} \mathrm{ES}_\alpha\left(\mathbb{E}[Y|\mathcal{G}]\right) Q(\mathrm{d}\alpha) + \tau_Q(Y).\end{aligned}$$

Therefore ρ is monotonic as it is the supremum of monotonic functionals, so it is coherent.

Fix $Q \in \mathcal{D}$. The mapping from L^∞ to $\mathbb{R}$ given by

$$X \mapsto \int_{[0,1)} \mathrm{ES}_\alpha(\mathbb{E}[X|\mathcal{G}])Q(\mathrm{d}\alpha)$$

is Fatou continuous and coherent by Propositions 2.11 and 2.24 (see Section 2.13.2), and therefore $w(L^\infty, \mathcal{M})$-lower semicontinuous. Thus, the functional on L^∞ given by

$$X \mapsto \int_{[0,1)} \mathrm{ES}_\alpha(\mathbb{E}[X|\mathcal{G}])Q(\mathrm{d}\alpha) + \tau_Q(X)$$

is $w(L^\infty, \mathcal{M})$-lower semicontinuous by Corollary 2.1, therefore ρ is $w(L^\infty, \mathcal{M})$-lower semicontinuous as it is the supremum of $w(L^\infty, \mathcal{M})$-lower semicontinuous functionals. Since ρ is coherent, it is Fatou. Let $X \in L^\infty$ and $Y \in L^\infty(\mathcal{G})$ such that $\mathbb{E}[X|\mathcal{G}] \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$, let $Z = X - \mathbb{E}[X|\mathcal{G}] + Y$. For all $Q \in D$,

$$\begin{aligned}\int_{[0,1)} \mathrm{ES}_\alpha(\mathbb{E}[Z|\mathcal{G}])Q(\mathrm{d}\alpha) + \tau_Q(Z) &= \int_{[0,1)} \mathrm{ES}_\alpha(Y)Q(\mathrm{d}\alpha) + \tau_Q(X) \\ &= \int_{[0,1)} \mathrm{ES}_\alpha(\mathbb{E}[X|\mathcal{G}])Q(\mathrm{d}\alpha) + \tau_Q(X),\end{aligned}$$

where the last equality follows from the fact that ES_α is law invariant. Therefore, $\rho(Z) = \rho(X)$. □

2.12.5 Examples

This section will give examples of partially law-invariant and strongly partially law-invariant risk measures. We use the following lemma from Delbaen [2000, Section 4.3.i]

frequently in this section.

Lemma 2.5. *Let $\rho_1 : L^\infty \to \mathbb{R}$ and $\rho_2 : L^\infty \to \mathbb{R}$ be Fatou coherent risk measures. Then $\rho_1 \vee \rho_2$ is a Fatou coherent risk measure with supporting set equal to* $\operatorname{conv}(\mathcal{S}_{\rho_1} \cup \mathcal{S}_{\rho_2})$.

Let $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k) = ([0,1], \mathcal{B}([0,1]), \lambda)$ for $k = 1, 2$. For the rest of the section let $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$ and $\mathcal{G} = \sigma(\pi_1)$ where $\pi_1(\omega, \tilde{\omega}) = \omega$. Define the set $\mathcal{S}_1 = \{\mu \in \mathcal{M}_1 \mid D_\mu \in L^1(\mathcal{G}), D_\mu \leq 2\}$. This set is $w(\mathcal{M}, L^\infty)$-closed and convex and corresponds to the Fatou coherent risk measure $\rho_1 = \mathrm{ES}_{1/2}(\mathbb{E}[X|\mathcal{G}])$.

The following example gives a Fatou $\mathcal{G}$-law-invariant coherent risk measure that is not strongly $\mathcal{G}$-law invariant.

Example 2.11. Define $\nu \in \mathcal{M}_1$ by $D_\nu : \Omega \to \mathbb{R} : (\omega, \tilde{\omega}) \mapsto 4\omega\tilde{\omega}$. We claim that the Fatou coherent risk measure $\rho : L^\infty \to \mathbb{R} : X \mapsto \mathbb{E}[D_\nu X] \vee \rho_1(X)$, which has supporting set $\mathcal{S}_\rho = \operatorname{conv}(\mathcal{S}_1 \cup \{\nu\})$, is $\mathcal{G}$-law invariant but not strongly $\mathcal{G}$-law invariant. To see this, note $\mathbb{E}[D_\nu|\mathcal{G}](\omega, \tilde{\omega}) = 2\omega$. So the measure associated to $\mathbb{E}[D_\nu|\mathcal{G}]$ is in the set $\mathcal{S}_1$. Therefore, $\operatorname{conv}(\mathcal{S}_1 \cup \{\nu\})^{\mathcal{G}} = \operatorname{conv}(\mathcal{S}_1^{\mathcal{G}} \cup \{\nu\}^{\mathcal{G}}) = \mathcal{S}_1$, which is $\mathcal{G}$-law invariant. Thus, ρ is $\mathcal{G}$-law invariant by Theorem 2.1. Define the following random variables

$$X(\omega, \tilde{\omega}) = \omega, \quad Y(\omega, \tilde{\omega}) = 1 - \omega \quad \text{and} \quad Z(\omega, \tilde{\omega}) = \tilde{\omega} - 1/2.$$

We have that $Z \in \ker(\mathcal{G})$, $\mathbb{E}[D_\nu Z] = 1/6$, $X, Y \in L^\infty(\mathcal{G})$ and $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$, $\mathrm{ES}_\alpha(X) = \mathrm{ES}_\alpha(Y) = 3/4$, $\mathbb{E}[D_\nu X] = 2/3$ and $\mathbb{E}[D_\nu Y] = 1/3$. Thus, $\rho(Z+X) = 5/6$ while $\rho(Z+Y) = 3/4$, implying that ρ is not strongly $\mathcal{G}$-law invariant by definition.

The following example gives a Fatou strongly $\mathcal{G}$-law invariant coherent risk measure that is not in the form of (2.13).

Example 2.12. Define $\nu \in \mathcal{M}_1$ by $D_\nu : \Omega \to \mathbb{R} : (\omega, \tilde{\omega}) \mapsto 2\tilde{\omega}$. We claim that the Fatou coherent risk measure $\rho : L^\infty \to \mathbb{R} : X \mapsto \mathbb{E}[D_\nu X] \vee \rho_1(X)$ is strongly $\mathcal{G}$-law invariant. To see this let $Z \in \ker(\mathcal{G})$ and $X, Y \in L^\infty(\mathcal{G})$ with $X \overset{\mathrm{d}}{=}_{\mathbb{P}} Y$. We have

$$\begin{aligned}\rho(Z+X) = \rho_1(Z+X) \vee \mathbb{E}[D_\nu(Z+X)] &= \rho_1(X) \vee (\mathbb{E}[D_\nu Z] + \mathbb{E}[X]) \\ &= \rho_1(Y) \vee (\mathbb{E}[D_\nu Z] + \mathbb{E}[Y]) = \rho(Z+Y).\end{aligned}$$

Note that $\widehat{\mathcal{S}}_\rho$ is not contained in $L^1(\mathcal{G})$, thus ρ is not in the form of (2.13).

2.12.6 Additional results and proofs accompanying Section 2.5.3

To do this, we will need some results from convex analysis. Let $(\mathfrak{X}, \mathfrak{Y}, \langle \cdot, \cdot \rangle)$ be a dual pairing and $f : \mathfrak{X} \to \mathbb{R} \cup \{\infty\}$ such that $\operatorname{dom}(f) = \{x \in \mathfrak{X} \mid f(x) < \infty\} \neq \varnothing$, define

$$f^* : \mathfrak{Y} \to \mathbb{R} \cup \{\infty\} : y \mapsto \sup_{x \in \mathfrak{X}} \left(\langle x, y \rangle - f(x) \right),$$
$$f^{**} : \mathfrak{X} \to \mathbb{R} \cup \{\infty, -\infty\} : x \mapsto \sup_{y \in \mathfrak{Y}} \left(\langle x, y \rangle - f^*(y) \right).$$

The upper epigraph of f is given by $G_f = \{(x, c) \in \mathfrak{X} \times \mathbb{R} \mid f(x) \leq c\} \subseteq \mathfrak{X} \times \mathbb{R}$. Equip $\mathfrak{X} \times \mathbb{R}$ with the product topology $w(\mathfrak{X}, \mathfrak{Y}) \times E_{\mathbb{R}}$, where $E_{\mathbb{R}}$ is the standard Euclidean topology.

Proposition 2.21. *Let $f : \mathfrak{X} \to \mathbb{R} \cup \{\infty\}$ with $\operatorname{dom}(f) \neq \varnothing$.*

(i) The mapping f is convex if and only if G_f is convex;

(ii) the mapping f is $w(\mathfrak{X}, \mathfrak{Y})$-lower semicontinuous if and only if G_f is $(w(\mathfrak{X}, \mathfrak{Y}) \times E_{\mathbb{R}})$-closed;

(iii) for all $x \in \mathfrak{X}$,

$$f^{**}(x) = \sup \left(g(x) \mid g \leq f,\ g \text{ is } w(\mathfrak{X}, \mathfrak{Y})\text{-lower semicontinuous and convex} \right)$$

and $G_{f^{**}}$ is equal to the $(w(\mathfrak{X}, \mathfrak{Y}) \times E_{\mathbb{R}})$-closure of $\operatorname{conv}(G_f)$;

(iv) the mapping f is $w(\mathfrak{X}, \mathfrak{Y})$-lower semicontinuous and convex if and only if $f^{**} = f$.

See Zălinescu [2002] for a proof of Proposition 2.21.

First, we obtain a representation of μ-convex adjustments, similar to Proposition 2.15.

Proposition 2.22. *Let $\tau : L^\infty \to \mathbb{R}$, $\mu \in \mathcal{M}_1(\mathcal{G})$, and $\rho = \mathbb{E}^\mu + \tau$. The following are equivalent.*

(i) The mapping τ is a Fatou μ-convex adjustment;

(ii) the mapping ρ is a $w(L^\infty, L^1)$-lower semicontinuous convex risk measure with

$$\operatorname{dom}(\rho^*) \subseteq \mathcal{E}(\mu);$$

(iii) τ *can be represented by*

$$\tau(X) = \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^{\nu}[X] - \mathbb{E}^{\mu}[X] - \rho^*(\nu) \right), \quad X \in L^{\infty}; \tag{2.21}$$

(iv) *there exists* $\alpha : \mathcal{E}(\mu) \to \mathbb{R} \cup \{\infty\}$ *such that* τ *can be represented by*

$$\tau(X) = \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^{\nu}[X] - \mathbb{E}^{\mu}[X] - \alpha(\nu) \right), \quad X \in L^{\infty}.$$

Proof. Proof: (i)⇒(ii): It is clear that ρ is a convex risk measure and is Fatou (simple application of DCT). This implies that ρ is $w(L^{\infty}, L^1)$-lower semicontinuous. Note by Föllmer and Schied [2016, Theorem 4.33], $\mathrm{dom}(\rho^*) \subseteq \mathcal{M}_1$. Let $\nu \in \mathcal{M}_1 \backslash \mathcal{E}(\mu)$. There exists $A \in \mathcal{G}$ such that $\mathbb{E}[D_\nu \mathbb{1}_A] > \mathbb{E}[D_\mu \mathbb{1}_A]$. Thus, we have

$$\begin{aligned} \rho^*(\nu) = \sup_{X \in L^{\infty}} \left(\mathbb{E}^{\nu}[X] - \mathbb{E}^{\mu}[X] - \tau(X) \right) &\geq \sup_{\lambda \in \mathbb{R}} \left(\lambda \mathbb{E}^{\nu}[\mathbb{1}_A] - \lambda \mathbb{E}^{\mu}[\mathbb{1}_A] - \tau(\lambda \mathbb{1}_A) \right) \\ &= \sup_{\lambda \in \mathbb{R}} \left(\lambda (\mathbb{E}^{\nu}[\mathbb{1}_A] - \mathbb{E}^{\mu}[\mathbb{1}_A]) \right) - \tau(0) = \infty. \end{aligned}$$

Hence, $\nu \notin \mathrm{dom}(\rho)$. (ii)⇒(iii): Since ρ is convex and $w(L^{\infty}, L^1)$-lower semicontinuous. By Proposition 2.5, we have for $X \in L^{\infty}$,

$$\begin{aligned} \tau(X) = \rho(X) - \mathbb{E}^{\mu}[X] &= \sup_{\nu \in \mathrm{dom}(\rho^*)} \left(\mathbb{E}^{\nu}[X] - \rho^*(\nu) \right) - \mathbb{E}^{\mu}[X] \\ &= \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^{\nu}[X] - \rho^*(\nu) \right) - \mathbb{E}^{\mu}[X] \\ &= \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^{\nu}[X] - \mathbb{E}^{\mu}[X] - \rho^*(\nu) \right). \end{aligned}$$

(iii)⇒(iv): This is trivial. (iv)⇒(i): The fact that τ is a μ-convex adjustment is clear. Let $(X_n)_{n=1}^{\infty}$ be a uniformly bounded sequence in L^{∞} which converges $\mathbb{P}$-a.s. to some $X \in L^{\infty}$. We have

$$\begin{aligned} \tau(X) = \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}[(D_\nu - D_\mu)X] - \alpha(\nu) \right) &= \sup_{\nu \in \mathcal{E}(\mu)} \left(\liminf_{n \to \infty} \left(\mathbb{E}[(D_\nu - D_\mu)X_n] - \alpha(\nu) \right) \right) \\ &\leq \liminf_{n \to \infty} \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}[(D_\nu - D_\mu)X] - \alpha(\nu) \right) \\ &= \liminf_{n \to \infty} \tau(X_n). \end{aligned}$$

Thus proving the Fatou continuity of τ. □

For $f : \mathfrak{X} \to \mathbb{R} \cup \{\infty\}$ let $L_f^{\leq c} = \{x \in \mathfrak{X} \mid f(x) \leq c\}$, $L_f^{<c} = \{x \in \mathfrak{X} \mid f(x) < c\}$, and $L_f^{=c} = \{x \in \mathfrak{X} \mid f(x) = c\}$. We will also need the following two lemmas.

Lemma 2.6. *Let $(\mathfrak{X}, \mathfrak{Y}, \langle \cdot, \cdot \rangle)$ be a dual pairing and $f : \mathfrak{X} \to \mathbb{R} \cup \{\infty\}$ be convex with $\operatorname{dom}(\rho) \neq \varnothing$. Then $\overline{L_f^{\leq c}} = L_{f^{**}}^{\leq c}$ for all $c > -f^*(0) = \inf_{x \in \mathfrak{X}} f(x) = \inf_{x \in \mathfrak{X}} f^{**}(x)$.*

Proof. Proof: Let $c > -f^*(0)$. Since $f^{**}(x) \leq f(x)$ for all $x \in \mathfrak{X}$, $L_f^{\leq c} \subseteq L_{f^{**}}^{\leq c}$. We have $\overline{L_f^{\leq c}} \subseteq L_{f^{**}}^{\leq c}$ since f^{**} is $w(\mathfrak{X}, \mathfrak{Y})$-lower semicontinous as it is the supremum of $w(\mathfrak{X}, \mathfrak{Y})$-continuous functions (meaning $L_{f^{**}}^{\leq c}$ is $w(\mathfrak{X}, \mathfrak{Y})$-closed).

Let $\epsilon > 0$. We claim that $L_{f^{**}}^{\leq c-\epsilon} \subseteq \overline{L_f^{\leq c}}$. Clearly this is true if $L_{f^{**}}^{\leq c-\epsilon} = \varnothing$. Therefore assume that $L_{f^{**}}^{\leq c-\epsilon} \neq \varnothing$ and let $x \in L_{f^{**}}^{\leq c-\epsilon}$, which means $(x, c-\epsilon) \in G_{f^{**}}$. There exists $(x_\lambda, c_\lambda) \in G_f$ such that $\lim_\lambda (x_\lambda, c_\lambda) = (x, c-\epsilon)$ since $\overline{G_f} = G_{f^{**}}$. Find λ_0 such that for all $\lambda \geq \lambda_0$, $c_\lambda \leq c$. Therefore for all $\lambda \geq \lambda_0$ we have $x_\lambda \in L_f^{\leq c}$. Thus the tail net $(x_\lambda)_{\lambda \geq \lambda_0}$ is contained in $L_f^{\leq c}$ and converges to x, therefore $x \in \overline{L_f^{\leq c}}$. This shows the claim that $L_{f^{**}}^{\leq c-\epsilon} \subseteq \overline{L_f^{\leq c}}$. Since $L_{f^{**}}^{\leq c-\epsilon} \subseteq \overline{L_f^{\leq c}}$ for arbitrary $\epsilon > 0$, we have $L_{f^{**}}^{<c} \subseteq \overline{L_f^{\leq c}}$.

Clearly $L_{f^{**}}^{<c} \subseteq L_{f^{**}}^{\leq c}$, implying $\overline{L_{f^{**}}^{<c}} \subseteq L_{f^{**}}^{\leq c}$. Take $x \in L_{f^{**}}^{\leq c}$ and $y \in L_{f^{**}}^{<c}$, which is possible because $c > -f^*(0)$. Define

$$z_n = \left(1 - \frac{1}{n}\right) x + \frac{1}{n} y,$$

which is clearly in $L_{f^{**}}^{<c}$ by the convexity of f^{**}. Since $\lim_n z_n = x$, we have $x \in \overline{L_{f^{**}}^{<c}}$, showing that $\overline{L_{f^{**}}^{<c}} = L_{f^{**}}^{\leq c}$. Therefore

$$L_{f^{**}}^{\leq c} = \overline{L_{f^{**}}^{<c}} \subseteq \overline{L_f^{\leq c}} \subseteq L_{f^{**}}^{\leq c},$$

which completes the proof. □

Lemma 2.7. *Let $\rho : L^\infty \to \mathbb{R}$ be a Fatou convex risk measure. Then ρ is law invariant if and only if $L_{\rho^*}^{\leq c}$ is law invariant for all $c > -\rho(0)$.*

Proof. Proof: By Theorem 4.59 in Föllmer and Schied [2016], ρ is law invariant if and only if ρ^* is law invariant. Clearly, if ρ^* is law invariant, $L_{\rho^*}^{\leq c}$ is law invariant for all $c > -\rho(0)$. Now assume that $L_{\rho^*}^{\leq c}$ is law invariant for all $c > -\rho(0)$. Fix some $c > -\rho(0) = \inf_{\mu \in \mathcal{M}_1(\mathcal{G})} \rho^*(\mu)$,

find $N \in \mathbb{N}$ such that $c - \frac{1}{N} > -\rho(0)$. We have

$$L_{\rho^*}^{=c} = \bigcap_{n \geq N} \left(L_{\rho^*}^{\leq c} / L_{\rho^*}^{\leq c - \frac{1}{n}} \right)$$

is law invariant. Now let $c = -\rho(0)$, we have

$$L_{\rho^*}^{=c} = \bigcap_{n \geq 1} \left(L_{\rho^*}^{\leq c + \frac{1}{n}} \right)$$

is law invariant (since for $c < -\rho(0)$, $L_{\rho^*}^{\leq c} = \varnothing$). This shows that ρ^* is law invariant. □

Proof. Proof of Theorem 2.5: Let ρ be a Fatou convex risk measure. By Föllmer and Schied [2016, Theorem 4.18, Theorem 4.33], we have

$$\begin{aligned}
\rho(X) = \sup_{\nu \in \mathcal{M}_1} \left(\mathbb{E}^\nu[X] - \rho^*(\nu) \right) &= \sup_{\mu \in \mathcal{M}_1(\mathcal{G})} \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^\nu[X] - \rho^*(\nu) \right) \\
&= \sup_{\mu \in \mathcal{M}_1(\mathcal{G})} \left(\mathbb{E}^\mu[X] + \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X] - \rho^*(\nu) \right) \right) \\
&= \sup_{\mu \in \mathcal{M}_1(\mathcal{G})} \left(\mathbb{E}^\mu[X] + \tau_\mu(X) \right), \quad X \in L^\infty,
\end{aligned}$$

where $\tau_\mu(X) = \sup_{\nu \in \mathcal{E}(\mu)} \left(\mathbb{E}^\nu[X] - \mathbb{E}^\mu[X] - \rho^*(\nu) \right)$, $X \in L^\infty$. Note that for all $\mu \in \mathcal{M}_1(\mathcal{G})$, $\tau_\mu = -\infty$ on L^∞ or $\mathrm{dom}(\tau_\mu) = L^\infty$. By Proposition 2.22, τ_μ is a Fatou μ-convex adjustment. Define $N : \mathcal{M}_1(\mathcal{G}) \to \mathbb{R} \cup \{\infty\} : \mu \mapsto -\tau_\mu(0)$. We claim that N is convex. Let $\mu_1, \mu_2 \in \mathcal{M}_1(\mathcal{G})$ and $\lambda \in [0,1]$. Define $\upsilon = \lambda\mu_1 + (1-\lambda)\mu_2$, we have

$$\begin{aligned}
N(\upsilon) = \inf_{\nu \in \mathcal{E}(\upsilon)} \rho^*(\nu) &\leq \inf_{(\nu_1, \nu_2) \in \mathcal{E}(\mu_1) \times \mathcal{E}(\mu_2)} \rho^*\left(\lambda\nu_1 + (1-\lambda)\nu_2 \right) \\
&\leq \lambda \inf_{\nu_1 \in \mathcal{E}(\mu_1)} \rho^*(\nu_1) + (1-\lambda) \inf_{\nu_2 \in \mathcal{E}(\mu_2)} \rho^*(\nu_2) \\
&= \lambda N(\mu_1) + (1-\lambda) N(\mu_2),
\end{aligned}$$

where we are using the convexity of ρ^*. This completes the forward direction. The reverse direction is a simple application of Proposition 2.22, where one expands the convex adjustments by (2.21). To see the last claim, define $\tilde{\rho} = \rho|_{L^\infty(\mathcal{G})}$, note

$$\tilde{\rho}(X) = \sup_{\mu \in \mathcal{M}_1(\mathcal{G})} \left(\mathbb{E}^\mu[X] - N(\mu) \right) = N^*(X), \quad X \in L^\infty(\mathcal{G}).$$

Thus $N^{**} = \tilde{\rho}^*$ and by Lemma 2.6, for all $c > -\rho(0)$, $\overline{L_N^{\leq c}} = L_{\tilde{\rho}^*}^{\leq c}$. Using Lemma 2.7,

$$\begin{aligned}\rho \text{ is } \mathcal{G}\text{-law invariant} &\Leftrightarrow \tilde{\rho} \text{ is law invariant} \\ &\Leftrightarrow L_{\tilde{\rho}^*}^{\leq c} \text{ is } \mathcal{G}\text{-law invariant for all } c > -\rho(0) \\ &\Leftrightarrow \overline{L_N^{\leq c}} \text{ is } \mathcal{G}\text{-law invariant for all } c > -\rho(0) \\ &\Leftrightarrow L_N^{\leq c} \text{ is almost } \mathcal{G}\text{-law invariant for all } c > -\rho(0).\end{aligned}$$

This completes the proof. □

2.13 Additional results and proofs accompanying Section 2.6

2.13.1 Additional results and proofs accompanying Section 2.6.1

We call $\mathcal{X}$ *upward-directed* if for all $X_1, X_2 \in \mathcal{X}$, $\max\{X_1, X_2\} \in \mathcal{X}$.

Proposition 2.23 (Föllmer and Schied [2016, Theorem A.11])**.** *Let $\mathcal{X} \subseteq L^\infty$ be such that $\sup_{X\in\mathcal{X}} \|X\|_\infty$ is finite. Then* ess-sup $\mathcal{X}$ *exists. Moreover, if $\mathcal{X}$ is upward-directed, then it holds that* $\mathbb{E}\left[\text{ess-sup}\,\mathcal{X}\right] = \sup_{X\in\mathcal{X}} \mathbb{E}[X]$.

Lemma 2.8. *Let $\mathcal{Q} \subseteq \mathcal{E}(\mathbb{P})$ be $\mathcal{G}$-convex. For all $X \in L^\infty$,*

$$\sup_{\mu\in\mathcal{Q}} \mathbb{E}^\mu[X] = \mathbb{E}\left[\operatorname*{ess\text{-}sup}_{\mu\in\mathcal{Q}} \mathbb{E}^\mu[X|\mathcal{G}]\right].$$

Proof. Proof: Let $X \in L^\infty$, define $I(X) = \{\mathbb{E}^\mu[X|\mathcal{G}] : \mu \in \mathcal{Q}\}$. We claim that $I(X)$ is upward-directed. Given $\mu_1, \mu_2 \in \mathcal{Q}$, let

$$A = \{\omega : \mathbb{E}^{\mu_1}[X|\mathcal{G}](\omega) \geq \mathbb{E}^{\mu_2}[X|\mathcal{G}](\omega)\}.$$

Since $\{D_\mu : \mu \in \mathcal{Q}\}$ is $\mathcal{G}$-convex, $\mathbb{1}_A D_{\mu_1} + \mathbb{1}_{A^c} D_{\mu_2} \in \{D_\mu : \mu \in \mathcal{Q}\}$. Let ν be the probability measure associated with $\mathbb{1}_A D_{\mu_1} + \mathbb{1}_{A^c} D_{\mu_2}$. We have

$$\mathbb{E}^\nu[X|\mathcal{G}] = I_A \mathbb{E}^{\mu_1}[X|\mathcal{G}] + I_{A^c}\mathbb{E}^{\mu_2}[X|\mathcal{G}] = \max\left\{\mathbb{E}^{\mu_1}[X|\mathcal{G}], \mathbb{E}^{\mu_2}[X|\mathcal{G}]\right\}.$$

Since $\mathbb{E}^\nu[X|\mathcal{G}] \in I(X)$, $I(X)$ is upward directed. By Proposition 2.23,

$$\mathbb{E}[\text{ess-sup}\, I(X)] = \sup_{\mu\in\mathcal{Q}} \mathbb{E}[\mathbb{E}^\mu[X|\mathcal{G}]] = \sup_{\mu\in\mathcal{Q}} \mathbb{E}^\mu[\mathbb{E}^\mu[X|\mathcal{G}]] = \sup_{\mu\in\mathcal{Q}} \mathbb{E}^\mu[X].$$

Since $X \in L^\infty$ was arbitrary, the proof is complete. □

Proof. Proof of Proposition 2.7: Fix some $X \in L^\infty$, we have

$$\overline{\mathrm{ES}}^{\mathcal{Q}}_\alpha[X] = \sup_{\mu\in\mathcal{Q}} \mathrm{ES}^\mu_\alpha[X] = \sup_{\mu\in\mathcal{Q}} \min_{x\in\mathbb{R}} \left(x + \frac{1}{1-\alpha}\mathbb{E}^\mu\left[(X-x)^+\right] \right). \tag{2.22}$$

Define $f : \mathcal{Q} \times \mathbb{R} \to \mathbb{R} : (\mu, x) \mapsto x + \frac{1}{1-\alpha}\mathbb{E}^\mu[(X-x)^+]$. Since $\mathcal{Q}$ is $\mathcal{G}$-convex, it is convex. Clearly, $f(\cdot, x)$ is $w(\mathcal{M}, L^\infty)|_{\mathcal{Q}}$-continuous and affine for all $x \in \mathbb{R}$. Likewise, $f(\mu, \cdot)$ is continuous and convex for all $\mu \in \mathcal{Q}$. By the minimax Theorem from Mertens et al. [2015, Theorem I.1.1], we are able to switch the minimum and supremum in (2.22), thus

$$\begin{aligned}
\overline{\mathrm{ES}}^{\mathcal{Q}}_\alpha[X] &= \min_{x\in\mathbb{R}} \sup_{\mu\in\mathcal{Q}} \left(x + \frac{1}{1-\alpha}\mathbb{E}^\mu\left[(X-x)^+\right] \right) \\
&= \min_{x\in\mathbb{R}} \left(x + \frac{1}{1-\alpha} \sup_{\mu\in\mathcal{Q}} \mathbb{E}^\mu\left[(X-x)^+\right] \right) \\
&= \min_{x\in\mathbb{R}} \left(x + \frac{1}{1-\alpha}\mathbb{E}\left[\text{ess-sup}_{\mu\in\mathcal{Q}}\, \mathbb{E}^\mu\left[(X-x)^+|\mathcal{G}\right] \right] \right) \\
&= \min_{x\in\mathbb{R}} \left(x + \frac{1}{1-\alpha}\mathbb{E}\left[\rho^{\mathcal{Q}}_{\mathcal{G}}\left((X-x)^+\right)\right] \right),
\end{aligned}$$

where the third equality follows from Lemma 2.8.

Let $\mu_1, \mu_2 \in \mathcal{Q}_\beta$ and $\lambda \in L^\infty(\mathcal{G})$ with $0 \le \lambda \le 1$, clearly

$$0 \le \lambda D_{\mu_1} + (1-\lambda) D_{\mu_2} \le \frac{1}{1-\beta}.$$

Let $A \in \mathcal{G}$,

$$\begin{aligned}
\mathbb{E}[(\lambda D_{\mu_1} + (1-\lambda)D_{\mu_2})\mathbb{1}_A] &= \mathbb{E}[\lambda \mathbb{1}_A \mathbb{E}[D_{\mu_1}|\mathcal{G}] + (1-\lambda)\mathbb{1}_A \mathbb{E}[D_{\mu_1}|\mathcal{G}]] \\
&= \mathbb{E}[\lambda \mathbb{1}_A + (1-\lambda)\mathbb{1}_A] = \mathbb{P}(A).
\end{aligned}$$

Therefore, $\lambda D_{\mu_1} + (1-\lambda)D_{\mu_2} \in \widehat{\mathcal{Q}}_\beta$. □

2.13.2 Additional results and proofs accompanying Section 2.6.2

Proposition 2.24. *Let $\tilde{\rho}$ in (2.13) be monotonic (resp. coherent) and Fatou. Then, the resulting risk measure ρ in (2.13) is also monotonic (resp. coherent) and Fatou.*

Proof. Proof: It is straightforward that ρ is monotonic. Let $(X_n)_{n\in\mathbb{N}}$ be an increasing sequence in L^∞ that converges $\mathbb{P}$-a.s. to some $X \in L^\infty$. Since $(X_n)_{n\in\mathbb{N}}$ is uniformly bounded, it is uniformly integrable. Thus X_n converges to X in L^1. Therefore, $(\mathbb{E}[X_n|\mathcal{G}])_{n\in\mathbb{N}}$ converges in L^1 to $\mathbb{E}[X|\mathcal{G}]$ by Jensen's inequality. Thus, $(\mathbb{E}[X_n|\mathcal{G}])_{n\in\mathbb{N}}$ converges in probability to $\mathbb{E}[X|\mathcal{G}]$. Since $(\mathbb{E}[X_n|\mathcal{G}])_{n\in\mathbb{N}}$ is also increasing, it converges $\mathbb{P}$-a.s. to $\mathbb{E}[X|\mathcal{G}]$. Therefore, by Fatou continuity of $\tilde{\rho}$,

$$\lim_{n\to\infty} \rho(X_n) \leq \rho(X) = \tilde{\rho}(\mathbb{E}[X|\mathcal{G}]) \leq \liminf_{n\to\infty} \tilde{\rho}(\mathbb{E}[X_n|\mathcal{G}]) = \lim_{n\to\infty} \rho(X_n).$$

Hence, $\rho(X) = \lim_{n\to\infty} \rho(X_n)$. This means that ρ is continuous from below. It is well-known that continuity from below implies Fatou-continuity under the monotonicity assumption. The fact that ρ is coherent is straightforward. □

Proof. Proof of Proposition 2.8: (i) ⇒ (ii) and (ii) ⇒ (iii): These are clear. (iii) ⇒ (iv): We have

$$\rho(X) = \rho(\mathbb{E}[X|\mathcal{G}]) = \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}[D_\mu \mathbb{E}[X|\mathcal{G}]] = \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}[\mathbb{E}[D_\mu|\mathcal{G}]X].$$

Therefore the set $\{\mathbb{E}[D_\mu|\mathcal{G}] : \mu \in \mathcal{S}_\rho\} \subseteq \widehat{\mathcal{S}}_\rho$ is $w(L^1, L^\infty)$-dense by Proposition 2.10. Since $L^1(\mathcal{G})$ is $w(L^1, L^\infty)$-closed by Corollary 2.2, we have $\widehat{\mathcal{S}}_\rho \subseteq L^1(\mathcal{G})$. (iv) ⇒ (i): Define the map

$$\tilde{\rho} : L^\infty(\mathcal{G}) \to \mathbb{R} : X \mapsto \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}[D_\mu X].$$

The mapping $\tilde{\rho}$ is clearly coherent and Fatou. By the tower property of conditional expectation, we have

$$\rho(X) = \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}[D_\mu X] = \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}[D_\mu \mathbb{E}[X|\mathcal{G}]] = \tilde{\rho}(\mathbb{E}[X|\mathcal{G}]).$$

The final statement of the proposition follows by definition. □

Proof. Proof of Proposition 2.9: Define $\tilde{\rho} : L^\infty \to \mathbb{R}$ by $\tilde{\rho}(X) = \gamma(X_\#\mathbb{P})$, $X \in L^\infty$. By definition, $\tilde{\rho}$ is coherent and Fatou. For $\mu \in \mathcal{S}$, define $\rho^\mu : L^\infty \to \mathbb{R}$ by $\rho^\mu(X) = \gamma(X_\#\mu)$,

$X \in L^\infty$. It is straightforward to see that for $X, Y \in L^\infty$, $\rho^\mu(X) = \tilde{\rho}(Y)$ if $F_{\mu,X} = F_{\mathbb{P},Y}$. We will first prove the coherence of ρ^μ. Translation invariance and positive homogeneity of ρ^μ is apparent. Let $X, Y \in L^\infty$ with $X \leq Y$, we have that $F_{\mu,Y}(x) \leq F_{\mu,X}(x)$ for all $x \in \mathbb{R}$ or $F^{-1}_{\mu,X}(u) \leq F^{-1}_{\mu,Y}(u)$ for all $u \in [0,1]$. Since $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, there exists a uniform $(0,1)$ random variable U. Define $X_\mu = F^{-1}_{\mu,X}(U)$ and $Y_\mu = F^{-1}_{\mu,Y}(U)$. We have $F_{\mu,X} = F_{\mathbb{P},X_\mu}$, $F_{\mu,Y} = F_{\mathbb{P},Y_\mu}$ and $X_\mu \leq Y_\mu$. Thus, $\rho^\mu(X) = \tilde{\rho}(X_\mu) \leq \tilde{\rho}(Y_\mu) = \rho^\mu(Y)$. Therefore ρ^μ is monontonic.

Let $X, Y \in L^\infty$. By Dudley [2002, Theorem 11.7.5], there exists a Borel-measurable $f : (0,1) \to \mathbb{R}^2$ such that $f_\# \lambda = (X, Y)_\# \mu$, where λ is the Lebesgue measure. Since $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, there exists a uniform $(0,1)$ random variable $U \in L^\infty$. Define $(X_\mu, Y_\mu) = f \circ U$, we have $(X, Y)_\# \mu = (X_\mu, Y_\mu)_\# \mathbb{P}$. Therefore $F_{\mu,X} = F_{\mathbb{P},X_\mu}$, $F_{\mu,Y} = F_{\mathbb{P},Y_\mu}$ and $F_{\mu,X+Y} = F_{\mathbb{P},X_\mu+Y_\mu}$. Thus,

$$\rho^\mu(X+Y) = \tilde{\rho}(X_\mu + Y_\mu) \leq \tilde{\rho}(X_\mu) + \tilde{\rho}(Y_\mu) = \rho^\mu(X) + \rho^\mu(Y).$$

Therefore, ρ^μ is subadditive.

Let $(X_n)_{n\in\mathbb{N}}$ in L^∞ be uniformly bounded and converge $\mathbb{P}$-a.s. to $X \in L^\infty$. Therefore, $(X_n)_{n\in\mathbb{N}}$ converges μ-a.s. to X. Thus,

$$\lim_{n\to\infty} F^{-1}_{\mu,X_n}(q) = F^{-1}_{\mu,X}(q)$$

for all the continuity points of $F^{-1}_{X,\mu}$. Since $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, there exists a uniform $(0,1)$ random variable U. Define $X^n_\mu = F^{-1}_{\mu,X_n}(U)$, $X_\mu = F^{-1}_{\mu,X}(U)$. We have $F_{\mu,X_n} = F_{\mathbb{P},X^n_\mu}$, $F_{\mu,X} = F_{\mathbb{P},X_\mu}$. Since the set of discontinuity points of $F^{-1}_{X,\mu}$ is countable, X^n_μ converges $\mathbb{P}$-a.s. to X_μ. Thus,

$$\rho^\mu(X) = \tilde{\rho}(X_\mu) \leq \liminf_{n\to\infty} \tilde{\rho}(X^n_\mu) = \liminf_{n\to\infty} \rho^\mu(X_n),$$

proving Fatou continuity.

Finally, since

$$\rho(X) = \sup_{\mu\in\mathcal{S}} \gamma(X_\# \mu) = \sup_{\mu\in\mathcal{S}} \rho^\mu(X),$$

it follows that ρ is the supremum of coherent and Fatou risk measures. Thus, ρ is coherent and Fatou. The fact that ρ is $\mathcal{G}$-law invariant is a straightforward consequence of the fact that $\mathcal{S} \subseteq \mathcal{E}(\mathbb{P}_0)$. □

Proof. Proof of the claims in Example 2.7: **Claim 1:** The first equality is straightforward. To show the second equality, let $X \in L^\infty(\mathcal{H})$. From the dual representation of ES in

Example 2.9, we get

$$\rho(X) = \sup_{\mu \in \mathcal{S}} \mathbb{E}^{\mu}[X] \leq \sup_{\mu \in \mathcal{S}_{\beta,\mathbb{P}}} \mathbb{E}^{\mu}[X] = \mathrm{ES}_{\beta}(X).$$

For the other direction, let $A \in \mathcal{F}$ be an event independent of $\mathcal{G}$ satisfying $\mathbb{E}[X|A] = \mathrm{ES}_{\beta}(X)$ and $\mathbb{P}(A) = 1 - \beta$. The existence of such an event (called a β-tail event) is justified by Wang and Ziegel [2021, Lemmas A.4 and A.7]. Let $\nu \in \mathcal{M}_1$ be given by $D_\nu = \mathbb{1}_A/(1-\beta)$. Since $\mathbb{1}_A$ is independent of $\mathcal{G}$, we have $\nu \in \mathcal{E}(\mathbb{P})$, and therefore $\nu \in \mathcal{S}$, and $\rho(X) \geq \mathbb{E}^{\nu}[X] = \mathbb{E}[X|A] = \mathrm{ES}_{\beta}(X)$. **Claim 2:** Take $X_1 \in L^{\infty}(\mathcal{G})$, $X_2 \in L^{\infty}(\mathcal{H})$, $X_3 \in L^{\infty}$ with $X_1 \overset{\mathrm{d}}{=}_{\mathbb{P}} X_2 \overset{\mathrm{d}}{=}_{\mathbb{P}} X_3$. We have

$$\begin{aligned}
\rho(X_3) &= \sup_{\mu \in \mathcal{S}} \mathbb{E}^{\mu}[X_3] \geq \mathbb{E}[X_3] = \mathbb{E}[X_1] = \rho(X_1); \\
\rho(X_3) &= \sup_{\mu \in \mathcal{S}} \mathbb{E}^{\mu}[X_3] \leq \sup_{\mu \in \mathcal{S}_{\beta,\mathbb{P}}} \mathbb{E}^{\mu}[X_3] = \mathrm{ES}_{\beta}[X_3] = \mathrm{ES}_{\beta}[X_2] = \rho(X_2),
\end{aligned}$$

as desired. □

Proof. Proof of the claim in Example 2.8: This proof follows a very similar line of reasoning as Liu et al. [2022, Proposition 2]; we reproduce the proof here as there are some nuances in this setting that need to be considered.

The β-tail measure $\widehat{\rho}^{\beta} : L^{\infty} \to \mathbb{R}$ generated by $\widehat{\rho}$ is defined by

$$\widehat{\rho}^{\beta}(X) = \widehat{\rho}(F^{-1}_{\mathbb{P},X}\left(\beta + (1-\beta)U\right)),$$

where U is a uniform $(0,1)$ random variable. Since $\widehat{\rho}$ is law-invariant, this mapping is clearly well-defined. For $\mu \in \mathcal{M}_1$, define $\rho^{\mu} : L^{\infty} \to \mathbb{R}$ by $\rho^{\mu}(X) = \gamma(X_{\#}\mu)$, $X \in L^{\infty}$. It is straightforward to see that for $X, Y \in L^{\infty}$, $\rho^{\mu}(X) = \widehat{\rho}(Y)$ if $F_{\mu,X} = F_{\mathbb{P},Y}$. Recall that $\mathcal{S}_{\beta,\mathbb{P}} = \{\nu \in \mathcal{M}_1 \mid (1-\alpha)D_\nu \leq 1\}$. By Liu et al. [2022, Proposition 2], we have

$$\rho(X) \leq \sup_{\mu \in \mathcal{S}_{\beta,\mathbb{P}}} \rho^{\mu}(X) = \widehat{\rho}^{\beta}(X),$$

for all $X \in L^{\infty}$. When $X \in L^{\infty}(\mathcal{H})$, by the fact that $(\Omega, \mathcal{H}, \mathbb{P})$ is atomless and Lemma 2.3, we can find a uniform $(0,1)$ random variable $U_X \in L^{\infty}(\mathcal{H})$ such that $F^{-1}_{\mathbb{P},X}(U_X) = X$ $\mathbb{P}$-a.s. Let $\mu \in \mathcal{M}_1$ have the Radon–Nikodym derivative

$$D_\mu = \frac{1}{1-\beta}\mathbb{1}_{\{U_X > \beta\}}.$$

Since D_μ is independent of $\mathcal{G}$, $\mathbb{E}[D_\mu|\mathcal{G}] = 1$ and $\mu \in \mathcal{S}$. Letting $Z = F^{-1}_{\mathbb{P},X}(\beta + (1-\beta)U_X)$ and $x \in \mathbb{R}$, we have

$$\mu(X > x) = \mu(X > x, U_X > \beta) = \frac{1}{1-\beta}\mathbb{P}(X > x, U_X > \beta) = \min\left(\frac{1}{1-\beta}\mathbb{P}(X > x), 1\right),$$
$$F_{\mu,X}(x) = \left(\frac{1}{1-\beta}(F_{\mathbb{P},X}(x) - \beta)\right)^+ = F_{\mathbb{P},Z}(x).$$

As a consequence,

$$\widehat{\rho}^\beta(X) = \widehat{\rho}(Z) = \rho^\mu(X) \leq \rho(X),$$

and $\rho(X) = \widehat{\rho}^\beta(X)$. Let $X \in L^\infty(\mathcal{G})$ and $\mu \in \mathcal{S}$, since $\mu \in \mathcal{E}(\mathbb{P})$, $F_{\mu,X} = F_{\mathbb{P},X}$. Therefore, $\rho^\mu(X) = \widehat{\rho}(X)$ and $\rho(X) = \widehat{\rho}(X)$. □

2.13.3 Additional results and proofs accompanying Section 2.6.3

Proof. A computable formula for (2.16): Let Φ (resp. ϕ) denote the standard normal cdf (resp. pdf). For $X \sim \mathrm{N}(r, \sigma^2)$, we have

$$\mathrm{VaR}_\beta\left((X-x)^+\right) = \left(r - x + \sigma\Phi^{-1}(\beta)\right)^+,$$
$$\mathrm{ES}_\beta\left((X-x)^+\right) = \begin{cases} r - x + \sigma\frac{\phi(\Phi^{-1}(\beta))}{1-\beta} & x \leq r + \sigma\Phi^{-1}(\beta) \\ \frac{(r-x)(1-\Phi\left(\frac{x-r}{\sigma}\right))}{1-\beta} + \sigma\frac{\phi(\frac{x-r}{\sigma})}{1-\beta} & x > r + \sigma\Phi^{-1}(\beta). \end{cases}$$

For $l \in \mathbb{R}$, the conditional distribution

$$\pi_1 X_1 + \pi_2 X_2 | X_1 = l \sim N\left(\pi_1 l + \pi_2 m_2 + c\pi_2\frac{\sigma_2}{\sigma_1}(l - m_1), \pi_2^2(1-c^2)\sigma_2^2\right).$$

Define the functions $M(\pi_1, \pi_2, l) = \pi_1 l + \pi_2 m_2 + c\pi_2\frac{\sigma_2}{\sigma_1}(l - m_1)$, $\sigma(\pi_2) = \pi_2\sqrt{(1-c^2)}\sigma_2$, and

$$g_\beta(r, \sigma, x) = \begin{cases} r - x + \sigma\frac{\phi(\Phi^{-1}(\beta))}{1-\beta} & x \leq r + \sigma\Phi^{-1}(\beta) \\ \frac{(r-x)(1-\Phi\left(\frac{x-r}{\sigma}\right))}{1-\beta} + \sigma\frac{\phi(\frac{x-r}{\sigma})}{1-\beta} & x > r + \sigma\Phi^{-1}(\beta). \end{cases}$$

We can now verify the following identity

$$\mathrm{ES}_\beta((\pi_1 X_1 + \pi_2 X_2 - x)^+ | X_1) = g_\beta(M(\pi_1, \pi_2, X_1), \sigma(\pi_2), x). \tag{2.23}$$

Therefore,

$$f_\beta(\pi_1, \pi_2, x) = \int_{\mathbb{R}} g_\beta(M(\pi_1, \pi_2, z), \sigma(\pi_2), x)\ \frac{1}{\sigma_1}\phi\left(\frac{z - m_1}{\sigma_1}\right)\ \mathrm{d}z,$$

which is an integral formula for (2.16) that is easy to numerically compute. □

2.14 Analysis in a finite probability space

We will first show that law invariance is equivalent to permutation invariance in a finite uniform probability space. Afterward, we will study partially law-invariant coherent risk measures when our probability space has a natural structure.

2.14.1 Law-invariant coherent risk measures

In the literature, when proving results regarding law-invariant risk measures, one often assumes that the underlying probability space is atomless. Since finite probability spaces always have atoms, we instead will assume that the underlying probability measure is uniform.

Let $N \in \mathbb{N}$, denote by $[N] = \{1, \dots, N\}$. Assume that $\Omega = [N]$ and $\mathcal{F} = 2^\Omega$ and put the uniform probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, that is $\mathbb{P}(\{k\}) = 1/N$ for all $k \in [N]$. We can identify elements $X \in L^\infty$ with $\mathbb{R}^N$. For the rest of the section, we will refer only to $\mathbb{R}^N$, but this is to be regarded as L^∞.

Let $\Delta^N = \left\{\mu \in \mathbb{R}^N_+ : \mathbb{E}[\mu] = 1\right\}$, it is easy to see that we can identify $\mathcal{M}_1$ with Δ^N. Since we are working in $\mathbb{R}^N$, the $w\left(\mathbb{R}^N, \mathbb{R}^N\right)$ topology ($w\left(L^\infty, \mathcal{M}\right)$ and $w\left(\mathcal{M}, L^\infty\right)$) is the standard euclidean topology. Therefore, there will be no mention in this section as to which topology we are dealing with. Note that any closed set $\mathcal{R} \subseteq \Delta^N$ is compact. Finally, Fatou continuity is implied by any coherent risk measure $\rho : \mathbb{R}^N \to \mathbb{R}$. Thus, there will be no mention of continuity in this section.

To describe law invariance in this space define $S_N = \{\sigma : [N] \to [N] \mid \sigma \text{ is a bijection}\}$. For $\sigma \in S_N$ define the map $T_\sigma : \mathbb{R}^N \to \mathbb{R}^N$ where $(T_\sigma(X))_n = X_{\sigma(n)}$.

Definition 2.4. We say that a mapping $\rho : \mathbb{R}^N \to \mathbb{R}$ is *permutation invariant* if $\rho(T_\sigma(X)) = \rho(X)$ for all $\sigma \in S_N$ and $X \in \mathbb{R}^N$. We say a set $\mathcal{R} \subseteq \mathbb{R}^N$ is *permutation invariant* if $T_\sigma(\mathcal{R}) \subseteq \mathcal{R}$ for all $\sigma \in S_N$.

It is easy to see that a mapping is law invariant if and only if it is permutation invariant. If $\mathcal{R} \subseteq \mathbb{R}^N$ is permutation invariant then $T_\sigma(\mathcal{R}) \supseteq T_\sigma(T_{\sigma^{-1}}(\mathcal{R})) = \mathcal{R}$, so we actually have equality.

Proposition 2.25. *Let $\rho : \mathbb{R}^N \to \mathbb{R}$ be a coherent risk measure. Then ρ is law invariant if and only if $\mathcal{S}_\rho$ is permutation invariant.*

Proof. Proof: Let $\mathcal{S}_\rho$ be permutation invariant then we have

$$\begin{aligned}\rho(T_\sigma(X)) = \sup_{\mu \in \mathcal{S}_\rho} \mathbb{E}[\mu T_\sigma(X)] &= \frac{1}{N} \sup_{\mu \in \mathcal{S}_\rho} (\mu \cdot T_\sigma(X)) \\ &= \frac{1}{N} \sup_{\mu \in \mathcal{S}_\rho} (T_{\sigma^{-1}}(\mu) \cdot X) = \frac{1}{N} \sup_{\mu \in \mathcal{S}_\rho} (\mu \cdot X) = \rho(X),\end{aligned}$$

for all $\sigma \in S_n$. Conversely, assume that $\mathcal{S}_\rho$ is not permutation invariant. Find $R \in \mathcal{S}_\rho$ and $\sigma \in S_n$ such that $T_\sigma(R) \notin \mathcal{S}_\rho$. Since $\mathcal{S}_\rho$ is compact and convex, by the Hahn Banach separation theorem, we can find $X \in \mathbb{R}^N$ such that

$$\rho(X) = \frac{1}{N} \sup_{\mu \in \mathcal{S}_\rho} (\mu \cdot X) < \frac{1}{N} X \cdot T_\sigma(R).$$

Therefore we have

$$\rho(X) < \frac{1}{N} X \cdot T_\sigma(R) = \frac{1}{N} T_{\sigma^{-1}}(X) \cdot R \leq \rho(T_{\sigma^{-1}}(X)),$$

implying that ρ is not permutation invariant and thus not law invariant. □

2.14.2 Partially law-invariant coherent risk measures

We write $A \times B = \{(a, b) : a \in A,\ b \in B\}$ for two sets A and B.

Let $\Omega = [M] \times [N]$ for some $M, N \in \mathbb{N}$, $\mathcal{F} = 2^\Omega$, $\mathcal{G}$ is generated by the projection map onto the second coordinate, and $\mathbb{P}$ is positive (that is, $\mathbb{P} > 0$ on $\mathcal{F} \setminus \{\varnothing\}$) and uniform in the second coordinate (that is, $\mathbb{P}([M] \times \{n\}) = 1/N$ for all $n \in [N]$).

We can identify elements $X \in L^\infty$ with $\mathbb{R}^{M \times N}$. For the rest of the section, we will refer only to $\mathbb{R}^{M \times N}$ but this is to be regarded as L^∞. Let $\Delta^{M \times N} = \{\mu \in \mathbb{R}_+^{M \times N} : \mathbb{E}[\mu] = 1\}$, it is easy to see that $\mathcal{M}_1$ can be identified with $\Delta^{M \times N}$. The same comments about topology and Fatou continuity from the previous section also apply in this section.

For $k \in [N]$ define $E^k \in \mathbb{R}^{M\times N}$ by $E^k_{mn} = \delta_{kn}$, where δ_{kn} is the Kronecker delta. Let $\mathcal{E} = \left\{E^k : k \in [N]\right\}$, clearly $\mathcal{E}$ is a basis for $L^\infty(\mathcal{G})$. Define the map $\mathbb{L} : \mathbb{R}^{M\times N} \to \mathbb{R}^N : X \mapsto [\mathbb{E}[X|\mathcal{G}]]_\mathcal{E}$. Given some $X \in \mathbb{R}^N$ we will define $\tilde{X} \in \mathbb{R}^{M\times N}$ by $\tilde{X} = \sum_{k=1}^N X_k E^k$. A set $\mathcal{R} \subseteq \mathbb{R}^{M\times N}$ is $\mathcal{G}$-law invariant if $\mathbb{L}(\mathcal{R})$ is permutation invariant.

The following proposition gives necessary and sufficient conditions for the supporting set of an $\mathcal{G}$-law invariant coherent risk measure.

Proposition 2.26. *Let $\rho : \mathbb{R}^{M\times N} \to \mathbb{R}$ be a coherent risk measure. Then ρ is $\mathcal{G}$-law invariant if and only if $\mathcal{S}_\rho$ is $\mathcal{G}$-law invariant.*

Proof. Proof: Define the mapping

$$\tilde{\rho} : \mathbb{R}^N \to \mathbb{R} : X \mapsto \rho(\tilde{X}).$$

It is easy to verify that this mapping is a coherent risk measure. Let $\mathcal{S}_{\tilde{\rho}} \subseteq \Delta^N$ be the supporting set for $\tilde{\rho}$. It follows that ρ is $\mathcal{G}$-law invariant if and only if $\tilde{\rho}$ is law invariant. For $X \in \mathbb{R}^N$ we have that

$$\begin{aligned}\tilde{\rho}(X) = \rho(\tilde{X}) = \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}\left[\mu\tilde{X}\right] &= \sup_{\mu\in\mathcal{S}_\rho} \mathbb{E}\left[\mathbb{E}[\mu|\mathcal{G}]\tilde{X}\right] \\ &= \frac{1}{N}\sup_{\mu\in\mathcal{S}_\rho} \mathbb{L}(\mu)\cdot X = \frac{1}{N}\sup_{\mu\in\mathbb{L}(\mathcal{S}_\rho)} \mu\cdot X.\end{aligned}$$

Thus $\mathbb{L}(\mathcal{S}_\rho) \subseteq \mathcal{S}_{\tilde{\rho}}$ by Proposition 2.10. Since $\mathbb{L}(\mathcal{S}_\rho)$ is compact and convex (image of a compact and convex set under a continuous linear map), the same theorem states that $\mathbb{L}(\mathcal{S}_\rho) = \mathcal{S}_{\tilde{\rho}}$. We then have

$$\begin{aligned}\rho \text{ is } \mathcal{G}\text{-law invariant } &\Leftrightarrow \tilde{\rho} \text{ is law invariant} \Leftrightarrow \mathcal{S}_{\tilde{\rho}} \text{ is permutation invariant} \\ &\Leftrightarrow \mathbb{L}(\mathcal{S}_\rho) \text{ is permutation invariant} \\ &\Leftrightarrow \mathcal{S}_\rho \text{ is } \mathcal{G}\text{-law invariant.}\end{aligned}$$

This completes the proof. □

2.14.3 Representation results

This section provides representation results similar to Theorem 2.1 and Theorem 2.3. A natural corollary to Proposition 2.26 is the following Theorem.

Theorem 2.6. *The mapping $\rho : \mathbb{R}^{M\times N} \to \mathbb{R}$ is a $\mathcal{G}$-law-invariant coherent risk measure if and only if*

$$\rho(X) = \sup_{\mu\in\mathcal{S}} \mathbb{E}[\mu X], \quad X \in L^\infty$$

holds for some closed and convex $\mathcal{S} \subseteq \mathcal{M}_1$ such that $\mathbb{L}(\mathcal{S})$ is $\mathcal{G}$-law invariant.

Define the set $\Delta_{\mathcal{G}}^{M\times N} = \Delta^{M\times N} \cap L^\infty(\mathcal{G})$, which can be identified with $\mathcal{M}_1(\mathcal{G})$.

Theorem 2.7. *The mapping $\rho : \mathbb{R}^{M\times N} \to \mathbb{R}$ is a coherent risk measure if and only if there exists a unique compact and convex set $\mathcal{R} \subseteq \Delta_{\mathcal{G}}^{M\times N}$ and a CA-assignment $(\tau_\mu)_{\mu\in\mathcal{R}}$ such that*

$$\rho(X) = \sup_{\mu\in\mathcal{R}} \left(\mathbb{E}\left[\mu X\right] + \tau_\mu(X)\right), \; X \in L^\infty. \tag{2.24}$$

Moreover, in (2.24), ρ is $\mathcal{G}$-law invariant if and only if $\mathcal{R}$ is $\mathcal{G}$-law invariant.

Proof. Proof: Let ρ be a coherent risk measure. Define the set $\mathcal{R} = \{\mathbb{E}[\mu|\mathcal{G}] : \mu \in \mathcal{S}_\rho\} \subseteq \mathcal{R} \subseteq \Delta_{\mathcal{G}}^{M\times N}$. The set $\mathcal{R}$ is compact and convex since it is the image of a compact convex set under a (continuous) linear map. For $\mu \in \mathcal{R}$, define $\mathcal{S}(\mu) = \{\nu \in \mathcal{S}_\rho : \mathbb{E}[\nu|\mathcal{G}] = \mu\}$. It is clear that $\mathcal{S}_\rho = \bigcup_{\mu\in\mathcal{R}} \mathcal{S}(\mu)$. We have

$$\begin{aligned}\rho(X) = \sup_{\nu\in\mathcal{S}_\rho} \mathbb{E}[\mu X] = \sup_{\mu\in\mathcal{R}} \sup_{\nu\in\mathcal{S}(\mu)} \mathbb{E}[\nu X] &= \sup_{\mu\in\mathcal{R}} \sup_{\nu\in\mathcal{S}(\mu)} \left(\mathbb{E}[\mu X] + \mathbb{E}[(\nu-\mu)X]\right) \\ &= \sup_{\mu\in\mathcal{R}} \left(\mathbb{E}[\mu X] + \tau_\mu(X)\right)\end{aligned}$$

for all $X \in L^\infty$, where $\tau_\mu(X) = \sup_{\nu\in\mathcal{S}(\mu)} \mathbb{E}[(\nu-\mu)X]$. The fact that $\tau_\mu \in \mathrm{CA}(\mu)$ follows a similar argument as the proof of Proposition 2.15. Since $\mathcal{R}$ supports $\rho|_{L^\infty(\mathcal{G})}$ and is closed and convex, it is unique from Proposition 2.10. For the converse, note that (2.8) is the supremum of coherent risk measures (by Proposition 2.6; see (2.7)) and thus a coherent risk measure.

To prove the last claim, let $\tilde{\rho} : \mathbb{R}^N \to \mathbb{R} : X \mapsto \rho(\tilde{X})$ and $\mathcal{S}_{\tilde{\rho}} \subseteq \Delta^N$. We have

$$\tilde{\rho}(X) = \rho(\tilde{X}) = \sup_{\mu\in\mathcal{R}} \mathbb{E}[\mu\tilde{X}] = \frac{1}{N} \sup_{\mu\in\mathbb{L}(\mathcal{R})} \mu \cdot X$$

for all $X \in \mathbb{R}^N$. Therefore $\mathbb{L}(\mathcal{R}) = \mathcal{S}_{\tilde{\rho}}$, and

$$\begin{aligned}\rho \text{ is } \mathcal{G}\text{-law invariant} \Leftrightarrow \tilde{\rho} \text{ is law invariant} &\Leftrightarrow \mathcal{S}_{\tilde{\rho}} \text{ is permutation invariant} \\ &\Leftrightarrow \mathcal{R} \text{ is } \mathcal{G}\text{-law invariant.}\end{aligned}$$

This completes the proof. □

Chapter 3

Choquet rank-dependent utility with an exogenous unambiguous source

3.1 Introduction

The expected utility (EU) model, axiomatized by von Neumann and Morgenstern [1947], serves as the theoretical foundation for much of today's theory of decision-making. The EU model has two well-known limitations. First, the EU model is formulated within the pure-risk setting, where acts are given by lotteries over outcomes. This setting is unrealistic in many situations, as a decision maker (DM) is often not fully aware of the distributions of the acts they are comparing. Second, even in the pure-risk setting, the EU model fails to account for both conceptual and empirical considerations. For example, the EU model cannot account for the empirically observed Allais paradox [Allais, 1953].

To address concerns associated with the pure-risk setting, the seminal work of Savage [1954] redefined acts as functions from states of the world to outcomes, rather than as lotteries over outcomes. This setting, commonly referred to as the Savage setting, does not assume that the likelihoods of events are well-specified a priori, and thus, there is no objective way to view the acts as lotteries. By imposing a set of axioms on preferences over these acts, Savage [1954] was able to elicit a subjective probability, thereby converting from the Savage setting into the pure-risk setting. Furthermore, the Savage axioms yield the EU model in the elicited pure-risk setting. This model is now known as the subjective expected utility (SEU) model.

The SEU model remained the prevailing decision model in the Savage setting until the critique of Ellsberg [1961], which showed that if the DM has asymmetric information

regarding the likelihoods of the events, the sure-thing principle (an axiom used in Savage [1954] to derive the SEU model) can fail to be behaviorally feasible. In fact, this critique called into question the behavioral feasibility of probabilistically sophisticated [Machina and Schmeidler, 1992] decision models: models that start by representing uncertainty with a subjective probability and then evaluate acts solely based on the lotteries they induce over outcomes.This foundational critique led to the development of what are now known as ambiguity models. Some prominent ambiguity models include the Choquet expected utility (CEU) model of Schmeidler [1989], the maxmin expected utility (MEU) model of Gilboa and Schmeidler [1989], the smooth ambiguity model of Klibanoff et al. [2005], and the variational preferences model of Maccheroni et al. [2006]. All of these models share a common theme: in the absence of ambiguity, i.e., when a probability can describe uncertainty and the DM can operate in the pure-risk setting, they return to the SEU model.

To address the empirical violations of the EU model within the pure-risk setting, the work of Quiggin [1982] introduced the rank-dependent utility (RDU) model. This model was a generalization of the EU model that could account for the Allais paradox. Further advancements came with the prospect theory [Kahneman and Tversky, 1979] and the cumulative prospect theory [Tversky and Kahneman, 1992], a generalization of the RDU model. This model was more behaviorally realistic than the RDU model, as it accounted for framing effects and asymmetries in the valuation of gains and losses. For a comprehensive treatment of the prospect theory, see Wakker [2010].

We are now at the intersection of two research streams: ambiguity models, which reduce to the SEU model in the case of pure risk, and rank-dependent models, which are in the pure-risk setting and do not account for ambiguity. However, some work has been done to bridge this gap. For example, the framework of distributionally robust optimization extends the MEU model by allowing for the reduction to general pure-risk models in the case of no ambiguity; see Delage et al. [2019] and Fadina et al. [2024] for approaches along this line through the lens of risk measures. A different generalization, and the main focus of this chapter, is a conceptual extension of the CEU model that reduces to the RDU model under pure risk. This framework, originally introduced in Tversky and Fox [1995] as the two-stage approach, retains the functional form of the CEU model but incorporates an additional distortion outside the capacity to capture how the DM distorts probabilities in the case of pure risk. In this chapter, we will refer to this as the Choquet rank-dependent utility (CRDU) model.

As the CRDU model is an instance of the biseperable model (Ghirardato and Marinacci, 2001), we provide a biseperable axiomatization of the CRDU model. Afterwards, we present a novel analysis of the ambiguity attitudes associated with this model. Recently, Buchak [2025] proposed an axiomatization of the CRDU model. Our approach begins from

a fundamentally different modeling assumption. The motivation for this choice comes from the empirical literature validating the behavioral feasibility of the CRDU model; see, for example, Abdellaoui et al. [2005], Dimmock et al. [2016], and Abdellaoui et al. [2021]. The empirical literature frequently employs the concept of a matching probability (Baillon et al., 2018) to determine the value of the capacity under the distortion for different events. To properly define matching probabilities, we need to have an objective notion of pure risk. Furthermore, our axiomatization relies on regularity axioms that allow for the independent verification of the existence of matching probabilities.

In this chapter, we adopt the Savage setting, where acts are viewed as mappings from states of the world to real-valued outcomes.[1] To define matching probabilities, we will assume the DM has an exogenous source of unambiguous events. This source will be equipped with a probability measure identifying the likelihoods of these events. Our main regularity axiom is that the decision maker is probabilistically sophisticated regarding acts that depend on events from the exogenous source. Models following this axiom have been studied recently by Baillon et al. [2025] in general decision models and by Shen et al. [2026] in risk assessment.

We begin in Section 3.2.1 by defining notation and reviewing the mathematical tools that will be used throughout the chapter. In Section 3.2.2, we will define the Savage setting with pure risk and discuss how the CEU and CRDU models fit into this setting. In Section 3.3, we present the axioms for our axiomatization of the CRDU model and derive a representation in Theorem 3.1. To properly discuss how the ambiguity component of the preferences is incorporated into this representation, we relate an aspect of the representation to the concept of matching probabilities. Section 3.4 focuses on analyzing various ambiguity attitudes. We start in Section 3.4.1 by defining comparative ambiguity attitudes and deriving their representation through matching probabilities. Afterwards, we compare our definition to the comparative ambiguity attitudes of Epstein [2004]. We finish the section by defining absolute ambiguity attitudes once a normalization for ambiguity neutrality is fixed. In Section 3.4.2, we discuss the propensity for diversification. We compare the consequences of this preference in the CRDU model with those in the restricted CEU model. Finally, in Section 3.4.3, we show how some of the ideas from Section 3.4.2 give rise to a robust representation, analogous to the way that propensity for diversification connects the CEU and MEU models. In Section 3.5, we adopt the setting of Shen et al. [2026] and assume that the DM has a reference probability measure on the whole space whose restriction agrees with the probability measure on the unambiguous source. This allows us to define and characterize a natural form of ambiguity aversion that we call reference

[1] The setting of Savage [1954] is, in general, not limited to real-valued acts; see Abdellaoui and Wakker [2020]. We take the perspective of real-valued acts in this chapter.

ambiguity aversion. We finish the section by providing an alternative representation for the CRDU model using act-dependent distortion functions in Theorem 3.6, offering a different view on decision-making under ambiguity: both pure risks and ambiguous acts are distorted. We finish by characterizing an ambiguity model that reduces to the dual model of Yaari [1987] for pure risks, which has a concise axiomatic formulation. We conclude the chapter in Section 3.7.

3.2 Preliminaries

3.2.1 Definitions and notation

Throughout the chapter, we consider a measurable space $(\Omega, \mathcal{F})$ representing the future states of the world. A *capacity* is a function $\nu : \mathcal{F} \mapsto \mathbb{R}$ that satisfies $\nu(\varnothing) = 0$, $\nu(\Omega) = 1$, and $\nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{F}$ with $A \subseteq B$. In some literature (e.g., Marinacci and Montrucchio, 2004), a capacity is not assumed to satisfy $\nu(\Omega) = 1$, but this difference is not essential. A capacity ν is said to be: *upward continuous* if for all increasing sequences $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ it holds that $\lim_{n\to\infty} \nu(A_n) = \nu\left(\bigcup_{n=1}^{\infty} A\right)$, *downward continuous* if for all decreasing sequences $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ it holds that $\lim_{n\to\infty} \nu(A_n) = \nu\left(\bigcap_{n=1}^{\infty} A_n\right)$, and *continuous* if it is both upward continuous and downward continuous. A capacity is called a *probability measure* if it is countably additive. We denote the set of probability measures on $(\Omega, \mathcal{F})$ by $\mathcal{M}_1$.

As in the Savage setting, an *act* is defined as a bounded real-valued measurable function. We will denote the space of acts by $\mathcal{X}$. Constant acts in $\mathcal{X}$ are identified with constants in $\mathbb{R}$. Given $X \in \mathcal{X}$, we write $\|X\| = \sup_{\omega\in\Omega} |X(\omega)|$. A set $\mathcal{S} \subseteq \mathcal{X}$ is said to be *bounded* if $\sup_{X\in\mathcal{S}} \|X\| < \infty$ and *B-closed* if, for any bounded $(X_n)_{n\in N} \subseteq \mathcal{S}$ that pointwise converges to some $X \in \mathcal{X}$, it holds that $X \in \mathcal{S}$.

For $X, Y \in \mathcal{X}$, $X \geq Y$ means $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$. Given $\mu \in \mathcal{M}_1$ and $X, Y \in \mathcal{X}$, we say that $X \geq^{\text{as}}_{\mu} Y$ if $\mu(\{X \geq Y\}) = 1$, and $X \geq^{\text{sd}}_{\mu} Y$ if $\mu(X > x) \geq \mu(Y > x)$ for all $x \in \mathbb{R}$. Clearly, $X \geq Y \implies X \geq^{\text{as}}_{\mu} Y \implies X \geq^{\text{sd}}_{\mu} Y$.

For a continuous capacity ν, the *core of* ν, denoted by $\mathfrak{C}(\nu)$, is defined as

$$\mathfrak{C}(\nu) = \{\mu \in \mathcal{M}_1 : \mu(A) \geq \nu(A) \text{ for all } A \in \mathcal{F}\}.$$

The following are common properties of a continuous capacity ν.

Supermodularity: For all $A, B \in \mathcal{F}$, $\nu(A) + \nu(B) \leq \nu(A \cup B) + \nu(A \cap B)$.

Submodularity: For all $A, B \in \mathcal{F}$, $\nu(A) + \nu(B) \geq \nu(A \cup B) + \nu(A \cap B)$.

Balanced: $\mathfrak{C}(\nu) \neq \varnothing$.

Exactness: For all $A \in \mathcal{F}$, $\nu(A) = \min_{\mu \in \mathfrak{C}(\nu)} \mu(A)$.

For more on the core and the previous properties of capacities, see Denneberg [1994] and Marinacci and Montrucchio [2004].

Given a capacity ν, the *Choquet integral* with respect to ν is given by

$$\int X \mathrm{d}\nu = \int_0^\infty \nu(X > x)\mathrm{d}x + \int_{-\infty}^0 (\nu(X > x) - 1)\,\mathrm{d}x, \quad X \in \mathcal{X}.$$

We say that $X, Y \in \mathcal{X}$ are *comonotonic* if for all $\omega, \tilde{\omega} \in \Omega$, $(X(\omega) - X(\tilde{\omega}))(Y(\omega) - Y(\tilde{\omega})) \geq 0$. As shown by Schmeidler [1986], a function $V : \mathcal{X} \to \mathbb{R}$ can be represented as $V(X) = \int_\Omega X \mathrm{d}\nu$ for some capacity ν if and only if V is monotone ($V(X) \geq V(Y)$ whenever $X \geq Y$), comonotonic additive ($V(X+Y) = V(X) + V(Y)$ for comonotonic X, Y), and normalized ($V(1) = 1$). An increasing (always in the non-strict sense) function $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$ is called a *distortion function*. A *von Neumann–Morgenstern (vNM) utility function* is a non-constant increasing function $u : \mathbb{R} \to \mathbb{R}$.

3.2.2 The Savage setting with pure risk and the Choquet rank-dependent utility model

We will work in a refined Savage setting, as described below. Fix a sub-σ-algebra $\mathcal{G}$, which is to be interpreted as an exogenous source of unambiguous events. As the events in $\mathcal{G}$ are unambiguous, the DM is able to assign likelihoods to them consistent with a probability measure. Let $\mathbb{P} : \mathcal{G} \to [0,1]$ denote the probability measure assigning these likelihoods. As common in risk management, we will assume that $\mathbb{P}$ is an atomless; that is, $\{\mathbb{P}(B) : B \in \mathcal{G},\ B \subseteq A\} = [0, \mathbb{P}(A)]$ for every $A \in \mathcal{G}$. We refer to this setting as the *Savage setting with pure risk*. We call $\mathbb{P}$ the *exogenous probability measure*.

A preference relation $\succsim$ representing the DM's preferences is a total preorder on $\mathcal{X}$; that is, (i) for all $X, Y \in \mathcal{X}$, $X \succsim Y$ or $Y \succsim X$; (ii) for all $X, Y, Z \in \mathcal{X}$, if $X \succsim Y$ and $Y \succsim Z$ then $X \succsim Z$. We write $\mathcal{X}(\mathcal{G})$ for the subspace of $\mathcal{X}$ consisting of $\mathcal{G}$-measurable acts. Under our setting, acts $X \in \mathcal{X}(\mathcal{G})$ exhibit no distributional ambiguity, and the preference relation $\succsim$ restricted to $\mathcal{X}(\mathcal{G})$ should be determined solely by the distributions of these acts under the exogenous probability measure $\mathbb{P}$. This brings us to our first axiom of risk conformity.[2]

[2]Risk conformity was called differently by Shen et al. [2026] as partial law invariance.

(RC) Risk conformity: For all $X, Y \in \mathcal{X}(\mathcal{G})$ satisfying $X =^{\text{sd}}_{\mathbb{P}} Y$, $X \simeq Y$.

If the preference relation $\succsim$ satisfies Axiom (RC), then the restriction of $\succsim$ to $\mathcal{X}(\mathcal{G})$, called the *risk preferences* of $\succsim$, corresponds to a preference relation over lotteries. To evaluate general acts, i.e., acts not necessarily in $\mathcal{X}(\mathcal{G})$, the DM must consider both their risk preferences and their perception and attitudes regarding ambiguity. In the discussions later, the *risk component* of $\succsim$ refers to the influence of the risk preferences on the evaluation of general acts, and the *ambiguity component* of $\succsim$ refers to the influence of ambiguity on the evaluation of general acts.

A capacity ν is *risk conforming* if $\nu|_{\mathcal{G}} = \mathbb{P}$. Risk conformity means that, for unambiguous events, the capacity behaves like the exogenous probability measure $\mathbb{P}$. When ν describes the DM's subjective evaluation regarding the likelihoods of events, risk conformity is essential, as the DM believes that $\mathbb{P}$ correctly specifies the likelihood of unambiguous events.

The CEU model (Schmeidler, 1989) was originally formulated in the setting of Anscombe and Aumann [1963]. We now explain the relevant formulation of the CEU model in the Savage setting with pure risk.[3] The DM has a continuous strictly increasing vNM utility u which describes their evaluation of outcomes. To describe the ambiguity component of the preferences, the DM has a continuous risk-confirming capacity ν that can be interpreted as their subjective assessment of the likelihood of events. The preference relation $\succsim$ of the DM in the CEU model is given by

$$X \succsim Y \iff \int u(X)\mathrm{d}\nu \geq \int u(Y)\mathrm{d}\nu, \quad X, Y \in \mathcal{X}. \tag{3.1}$$

It is well known that, given a preference relation $\succsim$ satisfying (3.1), u is unique up to positive affine transformations and ν is unique. Since ν in (3.1) is risk conforming, it is straightforward to show that a CEU preference relation $\succsim$ satisfies Axiom (RC) and the risk preferences for $\succsim$ are given by the EU model. As noted in the introduction, the majority of ambiguity models yield the EU model for risk preferences. A key conceptual assumption of the CEU model is that if the DM believes a probability measure $\mu \in \mathcal{M}_1$ uniquely describes their subjective assessment of the likelihood of events, i.e., they perceive no ambiguity, then they take $\nu = \mu$ in (3.1) and their preference relation $\succsim$ is given by the SEU model. The work of Sarin and Wakker [1992] provides an axiomatization of CEU

[3]The choice below to take u to be continuous and strictly increasing, as well as ν to be continuous, is not required in general, but this choice is adopted for reasons that will become apparent in the next section.

preference relations, but rather than assuming that $\mathbb{P}$ is given exogenously they incorporate its derivation into the axiomatization.

A conceptual generalization of the CEU model is the CRDU model, first considered by Tversky and Fox [1995]. We now explain the relevant formulation of the CRDU model in the Savage setting with pure risk. Once again, the DM has a continuous strictly increasing vNM utility u and a continuous risk-confirming capacity ν, which serve the same purpose as the CEU model. However, unlike the CEU model, the DM has an additional element describing the risk component of her preferences. This element is given by a continuous strictly increasing distortion function g, which describes how the DM distorts probabilities. The preference relation $\succsim$ of the DM in the CRDU model is given by

$$X \succsim Y \iff \int u(X)\mathrm{d}(g \circ \nu) \geq \int u(Y)\mathrm{d}(g \circ \nu), \quad X, Y \in \mathcal{X}. \tag{3.2}$$

Since any preference relation satisfying (3.2) is biseparable (Ghirardato and Marinacci, 2001), and ν must be risk conforming, it is straightforward to show that for any such preference relation, the utility function u is unique up to positive affine transformations, g is unique, and ν is unique.

Given a CRDU preference relation $\succsim$ in (3.2), since ν is risk conforming, we have

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}(g \circ \mathbb{P}) \geq \int_\Omega u(Y)\mathrm{d}(g \circ \mathbb{P}), \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

Therefore, $\succsim$ satisfies Axiom (RC) and the risk preferences are given by the RDU model of Quiggin [1982]. This explains the terminology "Choquet rank-dependent utility", as the model has a similar form as the CEU model but allows for risk preferences given by the RDU model. Furthermore, if there exists $x \in [0, 1]$ such that $g(x) \neq x$, then the risk preferences are not an instance of the EU model.

An axiomatization for the CRDU model was recently proposed by Buchak [2025] in a setting similar to ours; however, they do not assume that $\mathbb{P}$ is given exogenously. In the next section, we will provide an axiomatization of the CRDU model fundamentally different from Buchak [2025]. This axiomatization will consist of the regularity Axiom (RC), three additional regularity axioms that are also natural in the Savage setting with pure risk, and two additional axioms for biseparable preferences. The primary motivation for this axiomatization is that the regularity axioms enable us to show the existence of matching probabilities (Buchak [2025] has this as an axiom) and define the necessary tools for establishing the two axioms for biseparable preferences. This highlights the suitability of the Savage setting with pure risk for deriving axiomatizations of non-traditional models

of ambiguity within a Savage-like setting. The proofs of the main results, along with several supplementary results, are provided in the Appendix.

3.3 Axioms and representation

We now state and discuss several additional axioms for preference relations $\succsim$. Afterwards, in Theorem 3.1, we will present the main representation.

(M) Monotonicity: For all $X, Y \in \mathcal{X}$ satisfying $X \geq Y$, $X \succsim Y$.

Axiom (M) is standard for preference relations in the Savage setting. If the act X pays more than the act Y for every state of the world, then it is natural that the DM would prefer X to Y.

(SRM) Strict risk monotonicity: For all $X, Y \in \mathcal{X}(\mathcal{G})$ satisfying $X >^{\text{as}}_{\mathbb{P}} Y$, $X \succ Y$.

Axiom (SRM) reflects the fact that the DM believes that the probability measure $\mathbb{P}$ is well-specified on $\mathcal{G}$. Thus, a stronger form of monotonicity, using distributions, holds for acts in $\mathcal{X}(\mathcal{G})$.

(C) Continuity: For all $X \in \mathcal{X}$, $\{Y \in \mathcal{X} : X \succsim Y\}$ and $\{Y \in \mathcal{X} : Y \succsim X\}$ are B-closed.

Axiom (C) appeared in Chew and Mao [1995] and Chateauneuf and Lakhnati [2007] in the pure-risk setting. The use of this axiom in a Savage setting is common in the literature on risk measures, where it is referred to as the Lebesgue property; see Föllmer and Schied [2016].

The axioms discussed thus far concern only certain notions of regularity and essentially impose no specific functional forms on the preference relations. The axioms governing specific forms of the risk preferences will be discussed later in the section. The following proposition confirms that a CRDU preference relation satisfies the above regularity axioms.

Proposition 3.1. *Let $\succsim$ be a CRDU preference relation. Then $\succsim$ satisfies Axioms (RC), (M), (SRM), and (C).*

An additional confirmation of the suitability of our regularity axioms is that they ensure the existence of matching probabilities, an object we now recall. For notational simplicity, we will write $A \succsim B$ for $A, B \in \mathcal{F}$ to mean $\mathbb{1}_A \succsim \mathbb{1}_B$, where $\mathbb{1}_A$ denotes the binary act which yields 1 if $\omega \in A$ and 0 otherwise.

Definition 3.1. Given a preference relation $\succsim$, a function $\nu : \mathcal{F} \to [0,1]$ is a $\succsim$-*matching probability* if for all $A \in \mathcal{F}$, there exists $R_A \in \mathcal{G}$ such that $\nu(A) = \mathbb{P}(R_A)$ and $A \simeq R_A$.

Intuitively, a $\succsim$-matching probability uses the restriction of the preference relation $\succsim$ on $\mathcal{F}$ to produce a description of ambiguity on the level of events. The definition formalizes the idea that, for each event $A \in \mathcal{F}$, the preference relation $\succsim$ is used to identify a unambiguous probability $\nu(A)$ such that a bet paying \$1 if A occurs (and \$0 otherwise) and a unambiguous bet that pays \$1 with probability $\nu(A)$ (and \$0 otherwise) are equally preferred. Our definition of a $\succsim$-matching probability only works in the Savage setting with pure risk, as $\mathcal{G}$ needs to contain unambiguous events to have the aforementioned interpretation.

Proposition 3.2. *For every preference relation $\succsim$ satisfying Axioms (RC), (M), (SRM), and (C), there exists a unique $\succsim$-matching probability ν. Furthermore, ν is a continuous risk-conforming capacity, and*

$$A \succsim B \iff \nu(A) \geq \nu(B), \quad A, B \in \mathcal{F}. \tag{3.3}$$

Let $\succsim$ satisfy Axioms (RC), (M), (SRM), and (C). By Proposition 3.2, the $\succsim$-matching probability is risk conforming. Therefore, the $\succsim$-matching probability correctly identifies the likelihood of unambiguous events specified by $\mathbb{P}$. The following shows that the $\succsim$-matching probability can be used to isolate the ambiguity component of a CRDU preference relation, which is conceptually similar to Dimmock et al. [2016, Theorem 3.1].

Proposition 3.3. *Let $\succsim$ be a CRDU preference relation in (3.2), then ν is the unique $\succsim$-matching probability.*

The remaining two axioms are motivated by the fact that a CRDU preference relation is biseperable. These axioms make reference to certain constructions guaranteed to exist from our regularity axioms. For example, both make reference to certainty equivalents. To show existence, as well as a continuity property, we have the following proposition.

Proposition 3.4. *Given a preference relation $\succsim$ satisfying Axioms (RC), (M), (SRM), and (C), for each $X \in \mathcal{X}$, there exists a unique $c_X \in \mathbb{R}$ such that $X \simeq c_X$. Moreover, if $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ is a bounded sequence converging pointwise to $X \in \mathcal{X}$, $\lim_{n\to\infty} c_{X_n} = c_X$.*

The following axiom, or variations of it, has often been used for the axiomatization of the RDU model. In what follows, we use $\mathcal{G}_{\text{cf}}$ to denote the set $\{R \in \mathcal{G} : \mathbb{P}(R) = 1/2\}$ of "coin flip" events. Given $A \in \mathcal{F}$ and $x, y \in \mathbb{R}$, xAy denotes the binary act which yields x if $\omega \in A$ and y otherwise.

(RS) Risk symmetry: For all $R \in \mathcal{G}_{\text{cf}}$ and $x, y, z, z' \in \mathbb{R}$ satisfying $x \geq y$ and $z, z' \in [y, x]$

$$c_{xRz} R c_{z'Ry} \simeq c_{xRz'} R c_{zRy}.$$

It is not hard to show that Axiom (RS) is a weaker version of the weak independence axiom from Quiggin [1982]. The formulation of Axiom (RS) is similar to Axiom (A6) in Ghirardato et al. [2001], where various interpretations of axioms of this type and their historical development are discussed.

To define the final axiom, we need to define a binary operation on $\mathcal{X}$ that can be viewed as a subjective mixture operation for acts. As a preliminary step, following the approach of Ghirardato et al. [2003], we first introduce a subjective mixture for outcomes. The following proposition demonstrates that, once a non-trivial risky event is fixed, our regularity axioms guarantee both the existence and uniqueness of a subjective mixture for outcomes associated with that event.

Proposition 3.5. *Let $\succsim$ satisfy Axioms (RC), (M), (SRM), and (C). Given $R \in \mathcal{G}$ satisfying $\mathbb{P}(A) \in (0, 1)$ and $x, y \in \mathbb{R}$ with $x \geq y$, there exists a unique $z \in [y, x]$ such that $xRy \simeq c_{xRz} R c_{zRy}$.*

If the preference relation $\succsim$ satisfies Axioms (RC), (M), (SRM), and (C), then, given $x, y \in \mathbb{R}$, as $x \geq y$ or $y \geq x$, we define $(1/2)x \oplus (1/2)y$ as the unique z given in Proposition 3.5 with $R \in \mathcal{G}_{\text{cf}}$, following Ghirardato et al. [2003]. Clearly, by Axiom (RC), this binary operation is independent of which specific set $R \in \mathcal{G}_{\text{cf}}$ we choose. For two acts $X, Y \in \mathcal{X}$, we define the subjective mixture of acts $(1/2)X \oplus (1/2)Y$ by

$$[(1/2)X \oplus (1/2)Y](\omega) = (1/2)X(\omega) \oplus (1/2)Y(\omega).$$

Proposition 3.11 in Appendix 3.8 confirms that $(1/2)X \oplus (1/2)Y \in \mathcal{X}$; that is, $(1/2)X \oplus (1/2)Y$ is bounded and measurable.

We are now ready to present the final axiom, which is the comonotonic independence axiom of Schmeidler [1989], but for the subjective mixture of acts. The following axiom assumes that the preference relation $\succsim$ already satisfies Axioms (RC), (M), (SRM), and (C), since the existence of the subjective mixture operation for acts relies on these axioms.

(SCI) Subjective comonotonic independence: For all pairwise comonotonic $X, Y, Z \in \mathcal{X}$,

$$X \simeq Y \Longrightarrow (1/2)X \oplus (1/2)Z \simeq (1/2)Y \oplus (1/2)Z.$$

The following result shows that the above axioms fully characterize CRDU preference relations.

Theorem 3.1. *The preference relation $\succsim$ satisfies Axioms (RC), (M), (SRM), (C), (RS), and (SCI) if and only if $\succsim$ is a CRDU preference relation, that is, there exist a continuous strictly increasing vNM utility function u, a strictly increasing continuous distortion function g, and a continuous risk-conforming capacity ν such that*

$$X \succsim Y \iff \int u(X)\mathrm{d}(g \circ \nu) \geq \int u(Y)\mathrm{d}(g \circ \nu), \quad X, Y \in \mathcal{X}.$$

Given a CRDU preference relation $\succsim$, as the strictly increasing continuous distortion function g in (3.2) is uniquely determined, we will henceforth refer to g as the $\succsim$-distortion. Likewise, the strictly increasing vNM utility function u in (3.2) is unique up to positive affine transformations, and we will henceforth refer to the version of u satisfying $u(0) = 0$ and $u(1) = 1$ as the $\succsim$-utility.

In the Savage setting with pure risk, Axiom (RC) ensures that the $\succsim$-distortion, which represents an aspect of the risk component of $\succsim$, can be identified directly from the preference relation $\succsim$. This, in turn, separates it from the $\succsim$-matching probability, which represents the ambiguity component of $\succsim$. In the general Savage setting, i.e., the setting where $\mathcal{G}$ is not assumed to be an exogenously given unambiguous source, such a separation of the risk and ambiguity components may not be possible. The next example illustrates this point with a variational preference model. A similar example can be constructed for RDU and omitted here.

Example 3.1. Consider the general Savage setting without specifying $\mathcal{G}$. Suppose that the DM's preference relation $\succsim$ satisfies

$$X \succsim Y \iff \inf_{\mu \in \mathcal{M}_1} \left\{ \int_\Omega X \mathrm{d}\mu + \beta H(\mu|\mathbb{Q}) \right\} \geq \inf_{\mu \in \mathcal{M}_1} \left\{ \int_\Omega Y \mathrm{d}\mu + \beta H(\mu|\mathbb{Q}) \right\}, \tag{3.4}$$

where $\mathbb{Q} \in \mathcal{M}_1$ is a reference probability measure fixed by the DM, $\beta > 0$, and $H(\mu|\mathbb{Q})$ is the Kullback–Leibler divergence (relative entropy) of μ from $\mathbb{Q}$.[4] The model (3.4) is a special case of both the multiplier preferences of Hansen and Sargent [2001] and the variational preferences of Maccheroni et al. [2006]. Via the Donsker–Varadhan variational

[4] If μ is absolutely continuous with respect to $\mathbb{Q}$, then $H(\mu|\mathbb{Q}) = \mathbb{E}^\mu[\log(\mathrm{d}\mu/\mathrm{d}\mathbb{Q})]$, where $\mathrm{d}\mu/\mathrm{d}\mathbb{Q}$ is the Radon–Nikodym derivative; otherwise $H(\mu|\mathbb{Q}) = \infty$.

formula, (3.4) can be alternatively expressed by

$$X \succsim Y \iff -\beta \log\left(\int_\Omega \exp(-X/\beta)\mathrm{d}\mathbb{Q}\right) \geq -\beta \log\left(\int_\Omega -\exp(-Y/\beta)\mathrm{d}\mathbb{Q}\right). \tag{3.5}$$

The model (3.4)–(3.5) can be explained by two distinct behavioral considerations.

(a) The DM has the expected-value risk preferences, but she is unsure about the accuracy of the reference probability measure $\mathbb{Q}$. As alternative probability measures $\mu \in \mathcal{M}_1$ are plausible, the DM uses the relative entropy to penalize the deviation of μ from $\mathbb{Q}$ and then aggregates these evaluations.

(b) The DM is confident about the accuracy of the reference probability measure $\mathbb{Q}$, and she has EU risk preferences with an exponential utility function $u : x \mapsto -\exp(-x/\beta)$.

Since both cases lead to the same preference relation, the risk and ambiguity components of $\succsim$ cannot be uniquely identified from the DM's choices over $\mathcal{X}$. In case (a), β is a parameter of ambiguity aversion and in case (b), it is a parameter of risk aversion. With the help of the set $\mathcal{X}(\mathcal{G})$ in our Savage setting with pure risk, we are able to correctly separate these two components.

3.4 Ambiguity attitudes

As discussed in the previous section, given a CRDU preference relation $\succsim$, the ambiguity component of $\succsim$ is described by the $\succsim$-matching probability. Therefore, in this section, we study ambiguity attitudes of CRDU preferences through properties of the matching probabilities.

3.4.1 Comparative and absolute ambiguity attitudes

We will first define comparative ambiguity attitudes for CRDU preference relations. This will then allow us to define absolute ambiguity attitudes once a normalization for ambiguity neutrality is provided. Many of the notions presented here will be similar to those in Epstein [2004], but one key difference exists: Epstein's (2004) definition of comparative ambiguity attitudes between two preference relations $\succsim_1$ and $\succsim_2$[5] implies the equivalence of risk preferences between $\succsim_1$ and $\succsim_2$, whereas our notion is independent of risk preferences.

[5] Epstein [2004] used the term "uncertainty" in place of our term "ambiguity".

Given the CRDU preference relations $\succsim_1$ and $\succsim_2$, we say that $\succsim_2$ is *more ambiguity averse than* $\succsim_1$ if

$$R \succsim_1 A \implies R \succsim_2 A, \qquad R \in \mathcal{G} \text{ and } A \in \mathcal{F}.$$

The interpretation of this definition is quite natural: As events $R \in \mathcal{G}$ have no ambiguity, they can serve as the benchmark for comparing ambiguous events. This is similar to the classic notion of comparative risk aversion (Pratt, 1964 and Yaari, 1969), where constant acts serve as the benchmark.

We have the following proposition characterizing comparative ambiguity attitudes.

Theorem 3.2. *Let $\succsim_1$ and $\succsim_2$ be CRDU preference relations with matching probabilities ν_1 and ν_2 respectively. Then, $\succsim_2$ is more ambiguity averse than $\succsim_1$ if and only if $\nu_1(A) \geq \nu_2(A)$ for all $A \in \mathcal{F}$.*

In Theorem 3.2, the $\succsim_1$-distortion and the $\succsim_2$-distortion may not be the same, and similarly, the two utility functions may not be the same. This is consistent with the fact that the ambiguity component in the CRDU model is separated from the risk component. To get a full comparative implication on all acts, instead of only binary ones, we need to force the risk components of both preference relations to be identical, as shown in the next result.

Proposition 3.6. *Let $\succsim_1$ and $\succsim_2$ be CRDU preference relations. Then, the comparative implication*

$$X \succsim_1 Y \ (X \succ_1 Y) \implies X \succsim_2 Y \ (X \succ_2 Y), \quad \text{for all } X \in \mathcal{X}(\mathcal{G}) \text{ and } Y \in \mathcal{X} \tag{3.6}$$

holds if and only if the $\succsim_1$-utility is equal to the $\succsim_2$-utility, the $\succsim_1$-distortion is equal to the $\succsim_2$-distortion, and $\succsim_2$ is more ambiguity averse than $\succsim_1$.

As Proposition 3.6 demonstrates, our notion of comparative ambiguity attitudes for CRDU preference relations generalizes the comparative ambiguity attitudes of Epstein [2004] given by (3.6).

Our next goal is to define a normalization for ambiguity neutrality. Similar to the definition of comparative ambiguity attitudes, we define ambiguity neutrality as a property of the preference relation restricted to events.

(AN) Ambiguity neutrality: For all $A, B, C \in \mathcal{F}$ satisfying $(A \cup B) \cap C = \varnothing$,

$$A \succsim B \iff A \cup C \succsim B \cup C. \tag{3.7}$$

We have the following theorem characterizing Property (AN). However, this characterization only requires the four regularity axioms, which, by Proposition 3.2, guarantee the existence and uniqueness of a matching probability.

Theorem 3.3. *Let $\succsim$ satisfy Axioms (RC), (M), (SRM), and (C). Then $\succsim$ satisfies Property (AN) if and only if the $\succsim$-matching probability ν is a probability measure.*

In Epstein [2004], the normalization for ambiguity neutrality was given by preference relations consistent with respect to first-order stochastic dominance for some $\mu \in \mathcal{M}_1$. That is, there exists $\mu \in \mathcal{M}_1$ such that

$$X \geq_{\mu}^{\text{sd}} Y \implies X \succsim Y, \quad X, Y \in \mathcal{X}. \tag{3.8}$$

If $\succsim$ satisfies (3.8), then $\succsim$ is probabilistically sophisticated with respect to μ, i.e., for all $X, Y \in \mathcal{X}$ satisfying $X =_{\mu}^{\text{sd}} Y$, we have $X \simeq Y$; see Machina and Schmeidler [1992]. The following proposition demonstrates that our normalization for ambiguity neutrality aligns with the one presented in Epstein [2004].

Proposition 3.7. *Let $\succsim$ be a CRDU preference relation. Then $\succsim$ satisfies Property (AN) if and only if there exists $\mu \in \mathcal{M}_1$ such that (3.8) holds.*

Consistency with respect to first-order stochastic dominance, i.e., (3.8), does not represent ambiguity neutrality in the general Savage setting. For instance, in the MEU model of Gilboa and Schmeidler [1989], there are preference relations that are consistent with respect to first-order stochastic dominance, but do not have a prior set given by a singleton. Consider, for example, when $Q_\epsilon = \{\mu \in \mathcal{M}_1 : H(\mu|\mathbb{Q}) \leq \epsilon\}$, where $\mathbb{Q} \in \mathcal{M}_1$ is a reference probability measure and $\epsilon > 0$. The preference relation $\succsim$ given by

$$X \succsim Y \iff \min_{\mu \in \mathcal{Q}_\epsilon} \int_\Omega u(X) \mathrm{d}\mu \geq \min_{\mu \in \mathcal{Q}_\epsilon} \int_\Omega u(Y) \mathrm{d}\mu, \quad X, Y \in \mathcal{X}$$

satisfies (3.8) with $\mu = \mathbb{Q}$; see Shapiro [2013, Corollary 1]. As sets of multiple priors in the MEU model represent both the perception of ambiguity and a strong form of ambiguity aversion, as discussed by Ghirardato et al. [2004], consistency with respect to first-order stochastic dominance is not sufficient to claim ambiguity neutrality. Example 3.1 gives a similar example in the framework of variational preferences.

In our Savage setting with pure risk, Axiom (RC) enables us to pin down the DM's risk preferences. Furthermore, by Theorem 3.3, Property (AN) implies that the DM transforms reduces all acts to lotteries via the matching probability in $\mathcal{M}_1$, and then uses their risk preferences on these lotteries to make decisions. This is why the example for the MEU

model presented above does not capture ambiguity neutrality. In the MEU model, the risk preferences are given by the EU model. Under the situation where the set of priors is not a singleton, but the preferences are consistent with respect to first-order stochastic dominance, the preference relation induced on lotteries is not EU. In the example above, the induced preference relation on lotteries coincides with the so-called coherent entropic risk measure (Föllmer and Knispel, 2011 and Ahmadi-Javid, 2012) composed with the utility transformation. Therefore, ambiguity neutrality, in general, should be equivalent to consistency with respect to first-order stochastic dominance and the induced preference relation on lotteries coinciding with the risk preferences. This definition is consistent with Ghirardato and Marinacci [2002], where the authors' normalization for ambiguity neutrality corresponds to preference relations from the SEU model.

Finally, we say that a CRDU preference relation $\succsim$ is *ambiguity averse (AA)* if it is more ambiguity averse than a CRDU preference relation satisfying Property (AN). A straightforward consequence of Theorem 3.2 and Theorem 3.3 is that given a CRDU preference relation $\succsim$, Property (AA) is equivalent to the $\succsim$-matching probability being balanced. The idea of using a balanced matching probability in the CEU model to capture a form of ambiguity aversion was discussed in Chateauneuf and Tallon [2002]. For alternative forms of ambiguity aversion in the CEU model and their connection to matching probabilities, see Hartmann and Kauffeldt [2026].

3.4.2 Propensity for diversification

The idea of propensity for diversification in finance has appeared in the setting of risk since the foundational work of Markowitz [1952]. For decision models that incorporate ambiguity, such a property has been formalized in the classic papers by Gilboa and Schmeidler [1989] and Schmeidler [1989].[6] We present the definition of this property below, which is also sometimes called the convexity of the preferences.

(DS) Diversification seeking: For all $X, Y \in \mathcal{X}$ satisfying $X \simeq Y$ and $\lambda \in [0,1]$,

$$\lambda X + (1-\lambda)Y \succsim X.$$

If the preference relation of the DM satisfies Property (DS), then the DM prefers mixtures of equally preferred acts. This aligns with the intuitive idea of diversification. We immediately get the following proposition based on the work of Wakker and Yang [2019].

[6] In Gilboa and Schmeidler [1989] and Schmeidler [1989], diversification seeking is called uncertainty aversion.

Proposition 3.8. *Let $\succsim$ be a CRDU preference relation in (3.2). Then $\succsim$ satisfies Property (DS) if and only if the u is concave and the capacity $g \circ \nu$ is supermodular.*

By Proposition 3.8, for a CRDU preference relation $\succsim$, Property (DS) depends on both the risk component of $\succsim$, i.e., the $\succsim$-utility and the $\succsim$-distortion, and the ambiguity component of $\succsim$, i.e., the $\succsim$-matching probability. Furthermore, since the $\succsim$-matching probability ν is risk conforming and $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, it follows from the supermodularity of $g \circ \nu$, where g is the $\succsim$-distortion, that g is convex, see Föllmer and Schied [2016, Proposition 4.75]. Consequently, the Property (DS) implies a well-known risk attitude, which we now recall.

Given $\mu \in \mathcal{M}_1$ and the acts $X, Y \in \mathcal{X}$, we say that X *dominates* Y *in second-order stochastic dominance (SSD) under* μ, written $X \geq_\mu^{\text{ssd}} Y$, if for all increasing concave functions $f : \mathbb{R} \to \mathbb{R}$, it holds that $\int_\Omega f(X) \mathrm{d}\mu \geq \int_\Omega f(Y) \mathrm{d}\mu$. SSD is closely related to strong risk aversion as studied by Rothschild and Stiglitz [1970]. Note that the following property is defined only for pure-risk acts, that is, acts in $\mathcal{X}(\mathcal{G})$.

(SRA) Strong risk aversion: For all $X, Y \in \mathcal{X}(\mathcal{G})$ satisfying $X \geq_{\mathbb{P}}^{\text{ssd}} Y$, $X \succsim Y$.

A direct consequence of Schmidt and Zank [2008] is that a CRDU preference relation $\succsim$ satisfies Property (SRA) if and only if the $\succsim$-utility is concave and the $\succsim$-distortion is convex. Therefore, Property (DS) implies Property (SRA).

Property (SRA) is strictly weaker than Property (DS), as Property (DS) is a consequence of both the risk and ambiguity components of the preferences, but Property (SRA) is only a consequence of the risk component of the preferences. Ideally, one would like to identify an ambiguity attitude that, together with Property (SRA), implies Property (DS). The following example shows that this can be done for the specific class of CEU preference relations.

Example 3.2. Let $\succsim$ be a CEU preference relation; that is,

$$X \succsim Y \iff \int_\Omega u(X) \mathrm{d}\nu \geq \int_\Omega u(Y) \mathrm{d}\nu, \quad X, Y \in \mathcal{X},$$

where u is the $\succsim$-utility and ν is the $\succsim$-matching probability. By Proposition 3.8, we know that $\succsim$ satisfies Property (DS) if and only if u is concave and ν is supermodular. Since the supermodularity of ν implies that ν is balanced, we have $\succsim$ satisfies Property (AA). Therefore, the supermodularity of the $\succsim$-matching probability is a stronger form of ambiguity aversion. Furthermore, this stronger form of ambiguity aversion, combined with Property (SRA) guaranteeing the concavity of u, yields Property (DS).

For general CRDU preferences, the forward direction of the equivalence established in Example 3.2 for CEU preferences still holds, as shown in the following theorem.

Theorem 3.4. *Let $\succsim$ be a CRDU preference relation. If $\succsim$ satisfies Property (SRA) and the $\succsim$-matching probability is supermodular, then $\succsim$ satisfies Property (DS).*

Therefore, a combination of both strong risk aversion (Property (SRA)) and a stronger version of ambiguity aversion (Property (AA)) will ensure that the DM prefers diversification (Property (DS)). However, the next example demonstrates that the reverse direction of the equivalence established in Example 3.2 for CEU preferences does not generally extend to CRDU preferences. In particular, we show the existence of a CRDU preference relation $\succsim$ that satisfies Property (DS) but not Property (AA), which implies that the $\succsim$-matching probability is not supermodular.

Example 3.3. Assume there exists a risk-conforming $\mathbb{Q} \in \mathcal{M}_1$ and a sub-σ-algebra $\mathcal{H}$, independent of $\mathcal{G}$ under $\mathbb{Q}$,[7] such that there exists $A_0 \in \mathcal{H}$ satisfying $\mathbb{Q}(A_0) \in (0,1)$. Given a strictly concave continuous distortion function h, we show in Appendix 3.9 the existence of a continuous risk-conforming capacity ν for which

$$h(\mathbb{Q}(A)) \geq \nu(A) \geq \mathbb{Q}(A) \text{ for all } A \in \mathcal{F}, \qquad \nu(A) = h(\mathbb{Q}(A)) \text{ for all } A \in \mathcal{H},$$

and $g \circ \nu$ is supermodular, where g is the inverse distortion function of h. Fix the strictly increasing concave vNM utility u and define the CRDU preference relation $\succsim$ by

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}(g \circ \nu) \geq \int_\Omega u(Y)\mathrm{d}(g \circ \nu), \quad X, Y \in \mathcal{X}.$$

As $g \circ \nu$ is supermodular and u is concave, by Proposition 3.8, $\succsim$ satisfies Property (DS). By the strict concavity of h, $\nu(A_0) = h(\mathbb{Q}(A_0)) > \mathbb{Q}(A_0)h(1) + (1 - \mathbb{Q}(A_0))h(0) = \mathbb{Q}(A_0)$. Similarly, as $\mathbb{Q}(A_0^c) \in (0,1)$, $\nu(A_0^c) > \mathbb{Q}(A_0^c)$. Thus, $\nu(A) + \nu(A^c) > 1$ and $\succsim$ does not satisfy Property (AA).

The preference relation $\succsim$ constructed in Example 3.3 further illustrates the interplay between the risk and ambiguity components of CRDU preferences. Suppose that we fix the generalized probabilistic beliefs of $\succsim$ from Example 3.3 (we fix the $\succsim$-matching probability ν), but replace the risk preferences with the expected-value risk preferences, namely,

$$X \succsim' Y \iff \int_\Omega X\mathrm{d}\nu \geq \int_\Omega Y\mathrm{d}\nu, \quad X, Y \in \mathcal{X}.$$

[7] Independence between $\mathcal{H}$ and $\mathcal{G}$ means $\mathbb{Q}(A \cap B) = \mathbb{Q}(A)\mathbb{Q}(B)$ for all $A \in \mathcal{G}$ and $B \in \mathcal{H}$.

Therefore, it holds that

$$X \succsim' Y \iff \int_\Omega X \mathrm{d}(h \circ \mathbb{Q}) \geq \int_\Omega Y \mathrm{d}(h \circ \mathbb{Q}), \quad X, Y \in \mathcal{X}(\mathcal{H}).$$

Since $h \circ \mathbb{P}$ is submodular (Föllmer and Schied, 2016, Proposition 4.75), for all $X, Y \in \mathcal{X}(\mathcal{H})$, if $X \simeq' Y$, then for all $\lambda \in (0,1)$, $X \succsim' \lambda X + (1-\lambda)Y$. Therefore, for the preference relation $\succsim'$, there is an aversion to diversification locally on $\mathcal{X}(\mathcal{H})$, contrasting the global propensity for diversification of $\succsim$, although both preference relations satisfy Property (SRA).

3.4.3 The robust interpretation

A preference relation $\succsim$ is called a *maxmin expected utility (MEU)* preference relation (Gilboa and Schmeidler, 1989) if there exists a collection of risk-conforming probability measures $\mathcal{S}$[8] and a vNM utility u such that

$$X \succsim Y \iff \min_{\mu \in \mathcal{S}} \int_\Omega u(X) \mathrm{d}\mu \geq \min_{\mu \in \mathcal{S}} \int_\Omega u(Y) \mathrm{d}\mu, \quad X, Y \in \mathcal{X}. \tag{3.9}$$

Intuitively, a DM with an MEU preference relation chooses the set $\mathcal{S}$ to represent their ambiguity perception and then aggregates their risk preferences, given by the EU model, over this set. This form of aggregation is known as the robust aggregation of risk preferences, and it corresponds to a very strong form of ambiguity aversion; see Ghirardato et al. [2004].

For the CRDU preference relation $\succsim$ in (3.2), the above idea of robust aggregation of risk preferences means the following property:

$$X \succsim Y \iff \min_{\mu \in \mathfrak{C}(\nu)} \int_\Omega u(X) \mathrm{d}(g \circ \mu) \geq \min_{\mu \in \mathfrak{C}(\nu)} \int_\Omega u(Y) \mathrm{d}(g \circ \mu), \quad X, Y \in \mathcal{X}, \tag{3.10}$$

where the set of probability measures $\mathfrak{C}(\nu)$ is understood as reflecting the DM's ambiguity perception. As the following theorem demonstrates, the supermodularity of the $\succsim$-matching probability enables the robust aggregation of risk preferences in the case of CRDU preference relations.

Theorem 3.5. *Let $\succsim$ be a CRDU preference relation in (3.2). Then the robust aggregation (3.10) holds if and only if ν is supermodular.*

[8] Here, we suppose each $\mu \in \mathcal{S}$ is risk conforming to ensure we remain in the Savage setting with pure risk. That is, since $\mathbb{P}$ is well specified on $\mathcal{G}$, it should hold that $\mu|_{\mathcal{G}} = \mathbb{P}$ for all $\mu \in \mathcal{S}$.

It is well-known that Property (DS) connects CEU preference relations to MEU preference relations (Schmeidler, 1989). In the case of CRDU preference relations, Property (DS) is not sufficient to guarantee the robust aggregation of risk preferences. To see this, let $\succsim$ be a CRDU preference relation in (3.2) satisfying Property (DS). Then, by Proposition 3.8 and Marinacci and Montrucchio [2004, Theorem 4.6], it holds that

$$X \succsim Y \iff \min_{\mu \in \mathfrak{C}(g \circ \nu)} \int_{\Omega} u(X) \mathrm{d}\mu \geq \min_{\mu \in \mathfrak{C}(g \circ \nu)} \int_{\Omega} u(Y) \mathrm{d}\mu, \quad X, Y \in \mathcal{X}. \tag{3.11}$$

The expression in (3.11) represents an aggregation, but it does not correspond to the robust aggregation of the DM's risk preferences. This follows since (3.11) corresponds to an aggregation of the EU risk preferences. However, as $\succsim$ is a CRDU preference relation, the risk preferences should be given by the RDU model with distortion g. Furthermore, the set $\mathfrak{C}(g \circ \nu)$ in (3.11) should not be interpreted as the DM's ambiguity set as the capacity $g \circ \nu$ does not represent the ambiguity component of $\succsim$; rather, the $\succsim$-matching probability ν represents the ambiguity component of $\succsim$.

3.5 An exogenous global probability measure

In this section, we will assume that the DM has, through some statistical procedure, estimated a reference probability measure $\mathbb{Q} \in \mathcal{M}_1$ to initially describe their uncertainty. We will assume that the reference probability measure agrees with the likelihood evaluations on the unambiguous events; that is, $\mathbb{Q}$ is risk conforming. There are two ways to interpret the relationship between $\mathbb{Q}$ and $\mathbb{P}$. The first is that the DM begins with the exogenous unambiguous events and the probability measure $\mathbb{P}$. From here, when estimating the model $\mathbb{Q}$, the DM ensures that the statistical procedure produces a risk-conforming $\mathbb{Q}$. Example 3.4 below illustrates this interpretation in a setup used by Klibanoff et al. [2005]. For the second way of interpreting this assumption, the DM can first estimate $\mathbb{Q}$ and then, through introspection or some quantitative statistical procedure, designate $\mathcal{G}$ as being unambiguous. Then, the DM takes $\mathbb{P} = \mathbb{Q}|_{\mathcal{G}}$. This modeling approach was discussed in Shen et al. [2026].

Regardless of the interpretation, the key motivating takeaway is that the DM has a global probability measure $\mathbb{Q}$, but a priori, only fully trusts the model on $\mathcal{G}$.

Example 3.4. Let $(\Omega_1, \mathcal{F}_1)$ denote a measurable space where the DM lacks any prior probabilistic information. Let $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ denote a probability space where the DM knows that observations from this space happen according to $\mathbb{P}_2$. For example, the probability

space $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ could represent a lottery generating device. Define $(\Omega, \mathcal{F}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. Define $\mathcal{G} = \{\Omega_1 \times A : A \in \mathcal{F}_2\}$ and $\mathbb{P} : \mathcal{G} \to [0,1]$ by

$$\mathbb{P}(\Omega_1 \times A) = \mathbb{P}_2(A), \quad A \in \mathcal{F}_2.$$

Assume that the DM, through some statistical estimation procedure, has estimated the probability measure $\mathbb{P}_1$ for $(\Omega_1, \mathcal{F}_1)$. Additionally, assume the DM believes that these two sources of randomness are independent of each other, which is consistent with the interpretation that $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ represents a lottery generating device. Therefore, the DM takes as their reference probability measure $\mathbb{Q} \in \mathcal{M}_1$ the unique probability measure satisfying

$$\mathbb{Q}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2), \quad A_1 \in \mathcal{F}_1 \text{ and } A_2 \in \mathcal{F}_2.$$

It is straightforward to verify that $\mathbb{Q}$ is risk conforming.

Consider the case where the DM estimated the model $\mathbb{Q}$ to inform investment decisions for their financial portfolio, which they plan to liquidate within a week. Under the reference model $\mathbb{Q}$, assume that the weekly losses of a well-established equity X and the weekly losses of a recently launched cryptocurrency Y share the same distribution. Because the equity has been observed for an extended period, the DM perceives the events related to X as unambiguous, i.e., $X \in \mathcal{X}(\mathcal{G})$. When deciding between an investment in X versus an investment in Y, the DM is automatically drawn to X because it has the same estimated distribution as Y, without any ambiguity. Therefore, the assumption $X \succsim Y$ is reasonable. This discussion motivates the following property.

(RAA) Reference ambiguity aversion: For all $X \in \mathcal{X}(\mathcal{G})$ and $Y \in \mathcal{X}$ satisfying $X =^{\text{sd}}_{\mathbb{Q}} Y$, $X \succsim Y$.

As $\mathbb{Q}$ is risk conforming, it is straightforward to verify that Property (RAA) is a stronger condition than Axiom (RC). The terminology "reference ambiguity aversion" comes from the following proposition.

Proposition 3.9. *Let $\succsim$ be a CRDU preference. Then $\succsim$ satisfies Property (RAA) if and only if the $\succsim$-matching probability ν is set-wise dominated by the reference probability measure $\mathbb{Q}$, i.e., $\mathbb{Q} \in \mathfrak{C}(\nu)$.*

If the CRDU preference relation $\succsim$ satisfies Property (RAA), then, by Proposition 3.9, the $\succsim$-matching probability must be balanced. In general, we have

$$\text{Property (RAA)} \implies \text{Property (AA) and Axiom (RC)}$$

To see that the converse implication does not hold, we can take ν such that $\mathfrak{C}(\nu)$ does not contain $\mathbb{Q}$ but contains some risk conforming $\mu \in \mathcal{M}_1$ that differs from $\mathbb{Q}$ outside $\mathcal{G}$.

We now derive an alternative representation for CRDU preference relations when the following additional property is satisfied.

(NSC) Null set consistency: For all $X, Y \in \mathcal{X}$ satisfying $X =^{\text{as}}_{\mathbb{Q}} Y$, $X \simeq Y$.

Property (NSC) encodes the idea that the DM is confident that the reference probability measure $\mathbb{Q}$ correctly identifies sets of probability zero and is indifferent to acts that only differ on a set of probability zero. Property (NSC) is not implied by Axiom (RC), which does not restrict $\succsim$ outside $\mathcal{G}$. For $A, B \in \mathcal{F}$, denote by $A \triangle B$ the symmetric set difference, that is, $A \triangle B = (A \setminus B) \cup (B \setminus A)$. For a capacity ν, we say that ν is $\mathbb{Q}$-*consistent* if for all $A, B \in \mathcal{F}$ with $\mathbb{Q}(A \triangle B) = 0$, $\nu(A) = \nu(B)$.

Proposition 3.10. *Let $\succsim$ be a CRDU preference relation. Then $\succsim$ satisfies Property (NSC) if and only if the $\succsim$-matching probability is $\mathbb{Q}$-consistent.*

We now provide an alternative representation of CRDU when Property (NSC) is also assumed. For this purpose, we need the following definitions of distortion families. In the sequel, we write $\mathcal{D}(X) = \{\mathbb{Q}(X > x) : x \in \mathbb{R}\} \subseteq [0, 1]$.

Definition 3.2. (i) An indexed family $(g_X)_{X \in \mathcal{X}}$ is called a *distortion family* if g_X is distortion function for all $X \in \mathcal{X}$ and the following three properties are satisfied: (a) For all $A, B \in \mathcal{F}$ satisfying $A \subseteq B$, $g_{\mathbb{1}_A}(\alpha) \leq g_{\mathbb{1}_B}(\alpha)$ for all $\alpha \in [0, 1]$; (b) For all $X \in \mathcal{X}$, g_X is continuous when restricted to $\mathcal{D}(X)$; (c) For all comonotonic $X, Y \in \mathcal{X}$, $g_X(\alpha) = g_Y(\alpha)$ for all $\alpha \in \mathcal{D}(X) \cap \mathcal{D}(Y)$.

(ii) For a distortion function g, a distortion family $(g_X)_{X \in \mathcal{X}}$ is called a *g-distortion family* if for all $X \in \mathcal{X}(\mathcal{G})$, $g_X(\alpha) = g(\alpha)$ for all $\alpha \in \mathcal{D}(X)$, A g-distortion family $(g_X)_{X \in \mathcal{X}}$ is *ambiguity averse* if for all $X \in \mathcal{X}$, $g_X(\alpha) \leq g(\alpha)$ for all $\alpha \in \mathcal{D}(X)$.

Theorem 3.6. *The CRDU preference $\succsim$ in (3.2) with Property (NSC) has the alternative representation*

$$X \succsim Y \iff \int_\Omega u(X) \mathrm{d}(g_X \circ \mathbb{Q}) \geq \int_\Omega u(Y) \mathrm{d}(g_Y \circ \mathbb{Q}), \quad X, Y \in \mathcal{X}. \tag{3.12}$$

where $(g_X)_{X \in \mathcal{X}}$ is a g-distortion family that is uniquely determined on $\mathcal{D}(X)$ for each X. Furthermore, $\succsim$ satisfies (RAA) if and only if $(g_X)_{X \in \mathcal{X}}$ is ambiguity averse.

The representation in Theorem 3.6 gives an alternative view on how the DM incorporates the ambiguity component of their preferences. Intuitively, the DM uses the reference measure $\mathbb{Q}$ to make decisions; however, to account for ambiguity, how the reference measure is distorted depends on the act being evaluated. Since there is no distributional ambiguity for the acts $\mathcal{X}(\mathcal{G})$, the distortion functions associated with acts in $\mathcal{X}(\mathcal{G})$ coincide. The common distortion function fixes how the DM distorts unambiguous probabilities and is an aspect of the risk component of the preferences. Given some $X \in \mathcal{X} \setminus \mathcal{X}(\mathcal{G})$, how g_X differs from g encodes the DM's perception of and attitude toward the ambiguity in X.

Remark 3.1. We show in Appendix 3.10 that representation (3.12) is indeed a sufficient condition for $\succsim$ to be a CRDU preference relation. The representation (3.12) is primarily intended to motivate an alternative view on how the DM incorporates the ambiguity component of their preferences, rather than as a reference point for modeling such preference relations.

Assume that $\succsim$ satisfies representation (3.12). Given some $X \in \mathcal{X} \setminus \mathcal{X}(\mathcal{G})$, if it holds that $g_X = g$, then the DM perceives no ambiguity in X or is neutral towards the ambiguity perceived in X. For example, if $g_X = g$ holds for every $X \in \mathcal{X}$, then $\succsim$ satisfies

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}(g \circ \mathbb{Q}) \geq \int_\Omega u(Y)\mathrm{d}(g \circ \mathbb{Q}), \quad X, Y \in \mathcal{X}.$$

That is, the DM uses the RDU model globally under the reference probability measure $\mathbb{Q}$ and, thereby, satisfies Property (AN). If $\succsim$ satisfies Property (RAA), then, by Theorem 3.6, for all $X \in \mathcal{X}$,

$$g_X(\alpha) \leq g(\alpha), \quad \alpha \in \mathcal{D}(X)$$

Therefore, for all acts $X \in \mathcal{X} \setminus \mathcal{X}(\mathcal{G})$, the distortion g_X will be more risk averse than the DM's risk distortion g, in the classic sense of Pratt [1964] and Yaari [1969].

3.6 Risk preferences given by the dual utility model

The dual utility model of Yaari [1987] is the special case of the RDU model when the vNM utility is linear. The goal of this section is to axiomatize the subclass of CRDU preferences that have risk preferences given by the dual utility model. For this purpose, we introduce the following axiom, originally formulated by Yaari [1987] in the risk setting and by Schmeidler [1989] in the setting of ambiguity.

(CI) Comonotonic independence: For all pairwise comonotonic $X, Y, Z \in \mathcal{X}$,

$$X \simeq Y \Rightarrow (1/2)X + (1/2)Z \simeq (1/2)Y + (1/2)Z.$$

Intuitively, the Axiom (CI) says that an act, mutually comonotonic to two indifferent comonotonic acts, cannot be used as a way to hedge, i.e., as a way to change the indifference. Axiom (CI) is structurally similar to Axiom (SCI) presented in Section 3.3. However, Axiom (CI) is much simpler to define as it does not need the subjective mixture operation for acts. It turns out that Axiom (CI), together with the earlier regularity axioms, suffices to characterize all CRDU preference relations whose risk preferences are given by the dual utility model.

Theorem 3.7. *The preference relation $\succsim$ satisfies Axioms (RC), (M), (SRM), (C), and (CI) if and only if there exists a strictly increasing continuous distortion function g and a continuous risk-conforming capacity ν such that*

$$X \succsim Y \iff \int_\Omega X \mathrm{d}(g \circ \nu) \geq \int_\Omega Y \mathrm{d}(g \circ \nu), \quad X, Y \in \mathcal{X}. \tag{3.13}$$

Moreover, in the representation (3.13), g is unique and ν is unique and coincides with the $\succsim$-matching probability.

Any preference relation satisfying representation (3.13) is evidently a CRDU preference relation. This consequence is further clarified by the following observation. If a preference relation $\succsim$ satisfies the Axioms (RC), (M), (SRM), and (C), then, given $x, y \in \mathbb{R}$, we can define the subjective mixture of these outcomes $(1/2)x \oplus (1/2)y$; see Section 3.3. If we further assume that $\succsim$ satisfies Axiom (CI), then it holds that $(1/2)x \oplus (1/2)y = (1/2)x + (1/2)y$, that is, the subjective mixture is given by the usual arithmetic average. Therefore, Axiom (CI) implies Axiom (SCI).

3.7 Conclusion

We provided a new axiomatization of the Choquet rank-dependent utility model, an ambiguity model that is structurally similar to the Choquet expected utility model, but when confined to pure risks, it reduces to the rank-dependent utility model, rather than to the expected utility model. Our model is formulated in the Savage setting, instead of the Anscombe–Aumann setting.

The main result, Theorem 3.1, axiomatizes this model in the Savage setting with pure risk and, notably, requires only two axioms apart from some standard regularity axioms. The ambiguity component of the preferences is fully captured by the notion of a matching probability. To study ambiguity attitudes, we define comparative ambiguity attitudes, which involve only preference comparisons between binary acts rather than all acts. Specifically, Theorem 3.2 shows that, for CRDU preference relations, comparative ambiguity attitudes are fully characterized by their matching probabilities. After fixing a normalization for ambiguity neutrality, characterized in Theorem 3.3, we define and characterize absolute ambiguity attitudes. On the interaction between risk and ambiguity attitudes shapes the DM's propensity for diversification, Theorem 3.4 shows that strong risk aversion and the supermodularity of the matching probability are sufficient for propensity for diversification. However, unlike the Choquet expected utility model, these conditions are not necessary. This result allows a model to maintain the propensity for diversification by adopting flexible risk and ambiguity attitudes. In addition, Theorem 3.5 shows that the supermodularity of the matching probability is equivalent to the preference relation admitting a robust aggregation interpretation of the risk preferences. To give a complementary perspective on the ambiguity component of CRDU preference relations, Theorem 3.6 provides a representation derived under an additional property. This representation states that the DM distorts a reference probability measure, but allows this distortion to depend on the act being evaluated. The extent to which the distortion function associated with an act differs from the risk distortion determines the DM's perception of and attitude toward the ambiguity in that act. Finally, Theorem 3.7 provides an axiomatization for the subclass of CRDU preference relations whose risk preferences are given by the dual utility model. This axiomatization is concise as, apart from the regularity axioms from Theorem 3.1, only the single axiom of comonotonic independence is needed.

3.8 Results and proofs accompanying Section 3.3

Proof of Proposition 3.1. Since $\succsim$ is a CRDU preference relation, there exists a continuous strictly increasing vNM utility u, a continuous strictly increasing distortion function g, and a continuous risk-conforming capacity ν such that (3.2) holds true. Since ν is risk conforming, it is clear that $\succsim$ satisfies Axiom (RC). The fact that $\succsim$ satisfies Axiom (M) follows from the monotonicity of the Choquet integral and the fact that u, by virtue of being a vNM utility, is increasing.

Let $X, Y \in \mathcal{X}(\mathcal{G})$ satisfy $X >^{\text{as}}_{\mathbb{P}} Y$. Since u is strictly increasing, $u(X) >^{\text{as}}_{\mathbb{P}} u(Y)$. Therefore, for all $x \in \mathbb{R}$, $\mathbb{P}(u(X) > x) \geq \mathbb{P}(u(Y) > x)$, and there exists $x_0 \in \mathbb{R}$ such that

$\mathbb{P}(u(X) > x_0) > \mathbb{P}(u(Y) > x_0)$. Since g is strictly increasing, for all $x \in \mathbb{R}$, $g(\mathbb{P}(u(X) > x)) \geq g(\mathbb{P}(u(Y) > x))$, and $g(\mathbb{P}(u(X) > x_0)) > g(\mathbb{P}(u(Y) > x_0))$. By the right continuity of the mappings

$$x \mapsto g(\mathbb{P}(u(X) > x)) \quad \text{and} \quad x \mapsto g(\mathbb{P}(u(Y) > x)),$$

there exists $\epsilon > 0$ such that for all $x \in [x, x + \epsilon)$, $g(\mathbb{P}(u(X) > x)) > g(\mathbb{P}(u(Y) > x))$. Therefore,

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu) = \int_\Omega u(X)\mathrm{d}(g \circ \mathbb{P}) > \int_\Omega u(Y)\mathrm{d}(g \circ \mathbb{P}) = \int_\Omega u(Y)\mathrm{d}(g \circ \nu),$$

$X \succ Y$, and $\succsim$ satisfies Axiom (SRM).

Let $Y \in \mathcal{X}$ and $(X_n)_{n \in N} \subseteq \mathcal{X}$ be a bounded sequence converging pointwise to $X \in \mathcal{X}$ satisfying $Y \succsim X_n$ for all $n \in \mathbb{N}$. Since u is continuous and increasing, $(u(X_n))_{n \in \mathbb{N}}$ is bounded and converges pointwise to $u(X)$. As the capacity $g \circ \nu$ is continuous, by Wang and Klir [2010, Theorem 11.9], we have

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu) = \lim_{n \to \infty} \int_\Omega u(X_n)\mathrm{d}(g \circ \nu) \leq \int_\Omega u(Y)\mathrm{d}(g \circ \nu).$$

Therefore, $Y \succsim X$. If we instead assumed that $X_n \succsim Y$ for all $n \in \mathbb{N}$, a similar argument would show that $X \succsim Y$. Therefore, $\succsim$ satisfies Axiom (C). □

Proof of Proposition 3.2. Given $A \in \mathcal{F}$, define $\mathcal{L}(A) = \{\mathbb{P}(R) : A \succsim R \text{ and } R \in \mathcal{G}\}$ and $\mathcal{U}(A) = \{\mathbb{P}(R) : R \succsim A \text{ and } R \in \mathcal{G}\}$. Since, by Axiom (M), $\Omega \succsim A \succsim \varnothing$, it holds that $0 \in \mathcal{L}(A) \subseteq [0, 1]$ and $1 \in \mathcal{U}(A) \subseteq [0, 1]$. Thus, both $\mathcal{L}(A)$ and $\mathcal{U}(A)$ are non-empty. Let $a_A = \sup(\mathcal{L}(A))$ and $b_A = \inf(\mathcal{U}(A))$, we claim that $a_A \in \mathcal{L}(A)$. To see this, let $(R_n)_{n \in \mathbb{N}} \subseteq \mathcal{G}$ with $A \succsim R_n$ such that $\mathbb{P}(R_n) \uparrow a_A$. If $\mathbb{P}(R_n) = a_A$ for some $n \in \mathbb{N}$, then $a_A \in \mathcal{L}(A)$. Therefore, we will assume that $\mathbb{P}(R_n) < a_A$ for all $n \in \mathbb{N}$. Since $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, there exists a $\mathcal{G}$-measurable $U : \Omega \to (0, 1)$ with $\mathbb{P}(U \leq x) = x$ for all $x \in (0, 1)$, i.e., U has a uniform distribution under $\mathbb{P}$. Define $R'_n = \{U \leq \mathbb{P}(R_n)\}$. By Axiom (RC), $A \succsim R'_n$. Since $\mathbb{1}_{R'_n} \to \mathbb{1}_{\{U < a_A\}}$ pointwise, by Axiom (C), $A \succsim \{U < a_A\}$. Therefore $a_A = \mathbb{P}(U < a_A) \in \mathcal{L}(A)$. Similarly, one can show that $b_A \in \mathcal{U}(A)$. Therefore, as $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, by Axiom (M), $\mathcal{L}(A) = [0, a_A]$ and $\mathcal{U}(A) = [b_A, 1]$. Furthermore, as $\mathcal{U}(A) \cup \mathcal{L}(A) = [0, 1]$, $a_A \geq b_A$. For the sake of contradiction, assume that $a_A > b_A$. We can find $R_1, R_2 \in \mathcal{G}$ such that $R_1 \subseteq R_2$ and $\mathbb{P}(R_1) = b_A$ and $\mathbb{P}(R_2) = a_A$. By Axiom (SRM), $R_2 \succ R_1$. On the other hand, by Axiom (RC), $R_1 \succsim A \succsim R_2$, a contradiction. Therefore $a_A = b_A$. Since $A \in \mathcal{F}$ was general, we can define the set function $\nu : \mathcal{F} \to [0, 1]$ by

$\nu(A) = a_A = b_A$. By Axiom (RC) and the definition of $\mathcal{L}(U)$ and $\mathcal{U}(A)$, given any $R \in \mathcal{G}$ with $\mathbb{P}(R) = \nu(A)$, $A \simeq R$. Therefore, ν is a $\succsim$-matching probability. The uniqueness of ν is a consequence of the fact that $a_A = b_A$ for all $A \in \mathcal{F}$. Furthermore, the uniqueness of ν forces $\nu|_{\mathcal{G}} = \mathbb{P}$.

We will show that (3.3) is true. Let $A, B \in \mathcal{F}$ satisfy $A \succsim B$. Since ν is a matching probability, there exists $R_A, R_B \in \mathcal{G}$ satisfying $A \simeq R_A$ and $B \simeq R_B$ such that $\nu(A) = \mathbb{P}(R_A)$ and $\nu(B) = \mathbb{P}(R_B)$. Therefore, $R_A \simeq A \succsim B \simeq R_B$. For the sake of contradiction, assume that $\mathbb{P}(R_B) > \mathbb{P}(R_A)$. Under this assumption, as $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, we can find R_A^* such that $R_A^* \subseteq R_B$ and $\mathbb{P}(R_A^*) = \mathbb{P}(R_A)$. By Axioms (RC) and (SRM), $R_B \succ R_A^* \simeq R_A$, a contradiction. Thus, $\mathbb{P}(R_A) \geq \mathbb{P}(R_B)$ and $\nu(A) \geq \nu(B)$. Let $A, B \in \mathcal{F}$ satisfy $\nu(A) \geq \nu(B)$. Since ν is a matching probability, there exists $R_A, R_B \in \mathcal{G}$ satisfying $A \simeq R_A$ and $B \simeq R_B$ such that $\nu(A) = \mathbb{P}(R_A)$ and $\nu(B) = \mathbb{P}(R_B)$. As $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, we can R_B^* such that $R_B^* \subseteq R_A$ and $\mathbb{P}(R_B^*) = \mathbb{P}(R_B)$. Therefore, by Axioms (RC) and (M), $A \simeq R_A \succsim R_B^* \simeq R_B \simeq B$. Therefore, (3.3) holds and, by Axiom (M), ν is a capacity.

We will show that ν is upward continuous. Let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ be an increasing sequence with $A = \bigcup_{n\in\mathbb{N}} A_n$. For each $n \in \mathbb{N}$, define $R_n = \{U \leq \nu(A_n)\}$. By Axiom (RC) and the fact that ν is a matching probability, $R_n \simeq A_n$ for all $n \in N$. Let $R = \bigcup_{n\in\mathbb{N}} R_n$. By Axiom (M), $R \succsim R_n \simeq A_n$. As $\mathbb{1}_{A_n} \to \mathbb{1}_A$ pointwise, by Axiom (C), $R \succsim A$. By Axiom (M), $A \succsim A_n \simeq R_n$. As $\mathbb{1}_{R_n} \to \mathbb{1}_R$ pointwise, by Axiom (C), $A \succsim R$. Thus $A \simeq R$. Therefore,

$$\nu(A) = \mathbb{P}(R) = \lim_{n\to\infty} \mathbb{P}(R_n) = \lim_{n\to\infty} \nu(A_n).$$

A similar argument will show that ν is downward continuous. □

Proof of Proposition 3.3. Let $A \in \mathcal{F}$ and $R_A \in \mathcal{G}$ such that $\nu(A) = \mathbb{P}(R_A)$, which is guaranteed to exist since $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless. As

$$\begin{aligned}\int_\Omega u(\mathbb{1}_A)\mathrm{d}(g \circ \nu) &= u(1)g(\nu(A)) + u(0)(1 - g(\nu(A))) \\ &= u(1)g(\mathbb{P}(R_A)) + u(0)(1 - g(\mathbb{P}(R_A))) = \int_\Omega u(\mathbb{1}_{R_A})\mathrm{d}(g \circ \nu),\end{aligned}$$

$A \simeq R_A$. Therefore, ν is the $\succsim$-matching probability. □

Proof of Proposition 3.4. Let $X \in \mathcal{X}$, find $a, b \in \mathbb{R}$ such $a \geq X \geq b$. Therefore, by Axiom (M), the sets $\mathcal{L}_X = \{c \in \mathbb{R} : X \succsim c\}$ and $\mathcal{U}_X = \{c \in \mathbb{R} : c \succsim X\}$ are non-empty. It must hold that $\mathcal{L}_X \subseteq (-\infty, a]$, because if $X \succsim c$ for $c \in \mathbb{R}$ with $c > a$, then by Axiom (SRM), it would follow that $X \succ a \succsim X$. Similarly, we have that $\mathcal{U}_X \subseteq [b, \infty)$. Let

$a^* = \sup(\mathcal{L}_X) \in \mathbb{R}$ and $b^* = \inf(\mathcal{U}_X) \in \mathbb{R}$. By Axioms (M) and (C), it is clear that $\mathcal{L}_X = (-\infty, a^*]$ and $\mathcal{U}_X = [b^*, \infty)$. As $\mathcal{L}_X \cup \mathcal{U}_X = \mathbb{R}$, $b^* \leq a^*$. Take $c_X \in \mathcal{L}_X \cap \mathcal{U}_X$. Uniqueness of c_X follows from Axiom (SRM).

Let $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ be a bounded sequence converging pointwise to $X \in \mathcal{X}$. For $n \in \mathbb{N}$, define $Z_n = \sup_{m\geq n} X_m$. The sequence $(Z_n)_{n\in\mathbb{N}}$ is decreasing and converges pointwise to X. By Axiom (M), $Z_n \succsim X_m$ for all $n \in \mathbb{N}$ and $m \geq n$. Therefore, $Z_n \succsim \limsup_{m\to\infty} c_{X_m}$ for all $n \in \mathbb{N}$. By Axiom (C), $X \succsim \limsup_{m\to\infty} c_{X_m}$ and $c_X \geq \limsup_{m\to\infty} c_{X_m}$. A similar proof will show that $\liminf_{m\to\infty} c_{X_m} \geq c_X$. Thus, $\lim_{n\to\infty} c_{X_n} = c_X$. □

Proof of Proposition 3.5. If $x = y$, this is true; therefore, we can assume that $x > y$. Since $R \in \mathcal{G}$ satisfies $\mathbb{P}(A) \in (0, 1)$, by Axiom (SRM), $x \succ xRy \succ y$. Therefore, $x > c_{xRy} > y$. By Axiom (SRM),

$$c_{xRx}Rc_{xRy} = xRc_{xRy} \succ xRy \succ c_{xRy}Ry = c_{xRy}Rc_{yRy}. \tag{3.14}$$

By Proposition 3.4, the mappings $\phi_1, \phi_2 : [y, x] \to [y, x]$ defined by $\phi_1(z) = c_{xRz}$ and $\phi_2(z) = c_{zRy}$ are continuous. Furthermore, by Axiom (SRM), ϕ_1 and ϕ_2 are strictly increasing. Therefore, by Proposition 3.4 and Axiom (SRM), the mapping $\phi : [y, x] \to [y, x]$ defined by $\phi(z) = c_{\phi_1(z)R\phi_2(z)}$ is continuous and strictly increasing. By (3.14), $\phi(x) > c_{xRy} > \phi(y)$. Therefore, by the intermediate value theorem, there exists $z' \in [x, y]$ such that $\phi(z') = c_{xRy}$ or $c_{xRz'}Rc_{z'Ry} \simeq xRy$. The uniqueness of z' follows from the fact that ϕ is strictly increasing. □

Proposition 3.11. *Let $\succsim$ satisfy Axioms (RC), (M), (SRM), (C). Then for all $X, Y \in \mathcal{X}$,*

$$(1/2)X \oplus (1/2)Y \in \mathcal{X}.$$

Proof. Since there exists $M > 0$ such that $-M \leq X, Y \leq M$, we have that

$$-M \leq (1/2)X(\omega) \oplus (1/2)Y(\omega) \leq M, \quad \omega \in \Omega.$$

Therefore, $(1/2)X \oplus (1/2)Y$ is bounded.

If the binary operation $\oplus$ is measurable, then the $(1/2)X \oplus (1/2)Y$ is measurable as it is then the composition of measurable mappings. To show measurability, we will show that the binary operation $\oplus$ is continuous. Let $((x_n, y_n))_{n\in\mathbb{N}} \subseteq \mathbb{R}^2$ be a convergent sequence converging to $(x, y) \in \mathbb{R}^2$. We claim that there exists a subsequence $((x_{n_k}, y_{n_k}))_{k\in\mathbb{N}}$ such that $(x_{n_k} \oplus y_{n_k})_{k\in\mathbb{N}}$ converges to $x \oplus y$.

If $x = y$, we claim that we can take the original sequence $((x_n, y_n))_{n\in\mathbb{N}}$. Let $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $x - \epsilon < x_n < x + \epsilon$ and $x - \epsilon < y_n < x + \epsilon$. Therefore, $x - \epsilon < x_n \oplus y_n < x + \epsilon$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty}(x_n \oplus y_n) = x = x \oplus y$.

If $x \neq y$, without loss of generality, we may assume $x > y$ and $x_n > y_n$ for all $n \in \mathbb{N}$. Since the sequence $((x_n, y_n))_{n\in\mathbb{N}}$ is convergent, it is bounded. Therefore, the sequence $(x_n \oplus y_n)_{n\in\mathbb{N}}$ is bounded. By the Bolzano–Weierstrass theorem, there exists a subsequence $(x_{n_k} \oplus y_{n_k})_{k\in\mathbb{N}}$ that converges to some $z \in \mathbb{R}$. Fix $R \in \mathcal{G}_{\text{cf}}$, to ease notation, for $k \in \mathbb{N}$, let $a_k, b_k, d_k \in \mathbb{R}$ satisfy

$$a_k \simeq x_{n_k} R y_{n_k}, \quad b_k \simeq x_{n_k} R z_{n_k}, \quad \text{and} \quad d_k \simeq z_{n_k} R y_{n_k},$$

where $z_{n_k} = x_{n_k} \oplus y_{n_k}$. Furthermore, let $b, d \in \mathbb{R}$ satisfy $b \simeq xRz$ and $d \simeq zRy$. By the definition of $\oplus$, we know that $a_k \simeq b_k R d_k$ for all $k \in \mathbb{N}$. Since $(x_{n_k} R y_{n_k})_{k\in\mathbb{N}}$, $(x_{n_k} R z_{n_k})_{k\in\mathbb{N}}$, and $(z_{n_k} R y_{n_k})_{k\in\mathbb{N}}$ are bounded, by Proposition 3.4, $\lim_{k\to\infty} a_k = c_{xRy}$, $\lim_{k\to\infty} b_k = b$, and $\lim_{k\to\infty} d_k = d$. Furthermore, as $(b_k R d_k)_{k\in\mathbb{N}}$ is bounded, by Proposition 3.4, $\lim_{k\to\infty} c_{b_k R d_k} = c_{bRd}$. Therefore,

$$xRy \simeq c_{xRy} = \lim_{k\to\infty} a_k = \lim_{k\to\infty} c_{b_k R d_k} = c_{bRd} \simeq bRd = c_{xRz} R c_{zRy}.$$

Thus, $\lim_{k\to\infty}(x_{n_k} \oplus y_{n_k}) = x \oplus y$.

Assume for the sake of contradiction that $\oplus$ was not continuous. Then there would exist a convergent sequence $((x_n, y_n))_{n\in\mathbb{N}} \subseteq \mathbb{R}^2$ converging to $(x, y) \in \mathbb{R}^2$ and $\epsilon > 0$, such that for all $n \in \mathbb{N}$, $|(x_n \oplus y_n) - (x \oplus y)| \geq \epsilon$. By moving to a subsequence as described above, we get a contradiction. Therefore $\oplus$ is continuous. □

In the following proof, $\mathcal{X}_0$ denotes the collection of simple acts.

Proof of Theorem 3.1. Assume that $\succsim$ satisfies Axioms (RC), (M), (SRM), (C), (RS), and (SCI). Consider Axioms (A1)-(A6) and (A9) in Ghirardato et al. [2001]. It is clear that $\succsim$ satisfies Axioms (A1) and (A2). Fix $R \in \mathcal{G}_{\text{cf}}$. By Axiom (SRM), for all $x, y \in \mathbb{R}$ satisfying $x > y$, it holds that $x \succ xRy \succ y$. Therefore, R is essential as defined in Ghirardato et al. [2001], and $\succsim$ satisfies Axiom (A3). Furthermore, it is clear that Axiom (SRM) also implies that $\succsim$ satisfies Axiom (A4). Let $(a_\upsilon)_{\upsilon\in\Upsilon} \subseteq \mathbb{R}$ and $(b_\upsilon)_{\upsilon\in\Upsilon} \subseteq \mathbb{R}$ be nets converging to $a, b \in \mathbb{R}$ respectively. Assume that $X \in \mathcal{X}$ satisfies $a_\upsilon R b_\upsilon \succsim X$ for all $\upsilon \in \Upsilon$. It is straightforward to show that existence of a mapping $\phi : \mathbb{N} \to \Upsilon$ such that $\phi(n) \leq \phi(n+1)$ for all $n \in \mathbb{N}$, $\lim_{n\to\infty} a_{\phi(n)} = a$, and $\lim_{n\to\infty} b_{\phi(n)} = b$. Therefore, as the sequence $\left(a_{\phi(n)} R b_{\phi(n)}\right)_{n\in\mathbb{N}}$ is bounded, by Axiom (C), it holds that $aRb \succsim X$. A similar argument will shows that $X \succsim a_\upsilon R b_\upsilon$ for all $\upsilon \in \Upsilon$ implies $X \succsim aRb$. Therefore $\succsim$ satisfies

Axiom (A5). Let $x, y, z, z' \in \mathbb{R}$ satisfy $x \geq y$ and $z, z' \in [y, x]$. Then $\succsim$ satisfies Axiom (A6) if, for all $A, B \in \{R, R^c, \varnothing, \Omega\}$, it holds that

$$c_{xAz}Bc_{z'Ay} \simeq c_{xBz'}Ac_{zBy}. \tag{3.15}$$

It is straightforward to show that (3.15) holds true if $A \in \{\varnothing, \Omega\}$ or $B \in \{\varnothing, \Omega\}$. If $A = B = R$ or $A = B = R^c$, then (3.15) is the same statement as Axiom (RS). By Axiom (RC), we have $c_{xRz}Rc_{z'Ry} \simeq c_{xRz}R^c c_{z'Ry}$, $c_{xRz'} = c_{xR^c z'}$, and $c_{zRy} = c_{zR^c y}$. Therefore, by Axiom (RS),

$$c_{xRz}R^c c_{z'Ry} \simeq c_{xRz}Rc_{z'Ry} \simeq c_{xRz'}Rc_{zRy} = c_{xR^c z'}Rc_{zR^c y}.$$

Thus, for $A = R$ and $B = R^c$, (3.15) holds. The case of $A = R^c$ and $B = R$ is identical. Therefore, $\succsim$ satisfies Axiom (A6). Finally, the subjective mixture operation for acts $\oplus$, as defined in this chapter, coincides with the definition from Ghirardato et al. [2001]. Therefore, Axiom (SCI) is equivalent to Axiom (A9).

Since $\succsim$ satisfies Axioms (A1)-(A6) and (A9), by Ghirardato et al. [2001, Proposition 9], there exists a continuous vNM utility u and a capacity $\tilde{\nu}$ such that

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}\tilde{\nu} \geq \int_\Omega u(Y)\mathrm{d}\tilde{\nu}, \quad X, Y \in \mathcal{X}_0.$$

By Axiom (SRM), the function u must be strictly increasing. Therefore, we may assume without loss of generality that $u(0) = 0$ and $u(1) = 1$. Fix the distortion function φ such that $\varphi(u(x)) = u(\varphi(x)) = x$ for all $x \in [0, 1]$. It is clear that φ is continuous, strictly increasing, and that $c_A = \varphi(\tilde{\nu}(A))$ for all $A \in \mathcal{F}$, where $c_A = c_{\mathbb{1}_A}$. Let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ be an increasing sequence and define $A = \bigcup_{n\in\mathbb{N}} A_n$. By Proposition 3.4,

$$\varphi(\tilde{\nu}(A)) = c_A = \lim_{n\to\infty} c_{A_n} = \lim_{n\to\infty} \varphi(\tilde{\nu}(A_n)) = \varphi\left(\lim_{n\to\infty} \tilde{\nu}(A_n)\right).$$

Therefore, $\tilde{\nu}$ is upward continuous. A similar proof will show that $\tilde{\nu}$ is downward continuous and, therefore, $\tilde{\nu}$ is continuous.

Let $R_1, R_2 \in \mathcal{G}$ satisfy $\mathbb{P}(R_1) = \mathbb{P}(R_2)$. By Axiom (RC), $R_1 \simeq R_2$ and $\tilde{\nu}(R_1) = \tilde{\nu}(R_2)$. Therefore, there exists a distortion function g such that for all $R \in \mathcal{G}$, $\tilde{\nu}(R) = g(\mathbb{P}(R))$. Since $\tilde{\nu}$ is continuous, g is continuous. Furthermore, by Axiom (SRM), g is strictly increasing. Denote by h the continuous and strictly increasing distortion function satisfying $g(h(x)) = h(g(x)) = x$ for all $x \in [0, 1]$. Define the capacity $\nu = h \circ \tilde{\nu}$. Clearly ν is a continuous risk-conforming capacity. Therefore (3.2) holds for $\mathcal{X}_0$.

Finally, we need to show that (3.2) holds for $\mathcal{X}$. Given $X, Y \in \mathcal{X}$ satisfying $X \succsim Y$, fix

the bounded sequences $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}_0$ and $(Y_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}_0$ satisfying $X_n \geq X$ for all $n \in \mathbb{N}$, $Y \geq Y_n$ for all $n \in \mathbb{N}$, $X_n \to X$ pointwise, and $Y_n \to Y$ pointwise. Since X and Y are bounded, the sequences $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ exist. By Axiom (M), it holds that for all $n \in \mathbb{N}$, $X_n \succsim Y_n$. Therefore, for all $n \in \mathbb{N}$,

$$\int_\Omega u(X_n)\mathrm{d}(g \circ \nu) \geq \int_\Omega u(Y_n)\mathrm{d}(g \circ \nu).$$

Because $u(X_n) \to u(X)$ pointwise, $u(Y_n) \to u(Y)$ pointwise, $(u(X_n))_{n\in\mathbb{N}}$ is bounded, $(u(Y_n))_{n\in\mathbb{N}}$ is bounded, and $g \circ \nu$ is continuous, by using Wang and Klir [2010, Theorem 11.9], we have

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu) = \lim_{n\to\infty} \int_\Omega u(X_n)\mathrm{d}(g \circ \nu) \geq \lim_{n\to\infty} \int_\Omega u(Y_n)\mathrm{d}(g \circ \nu) = \int_\Omega u(Y)\mathrm{d}(g \circ \nu).$$

Given $X, Y \in \mathcal{X}$ satisfying

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu) \geq \int_\Omega u(Y)\mathrm{d}(g \circ \nu),$$

fix the same bounded sequences $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}_0$ and $(Y_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}_0$ as before. By the monotonicity of the Choquet integral, we have that for all $m, n \in \mathbb{N}$,

$$\int_\Omega u(X_m)\mathrm{d}(g \circ \nu) \geq \int_\Omega u(Y_n)\mathrm{d}(g \circ \nu).$$

Therefore, for all $m, n \in \mathbb{N}$, $X_m \succsim Y_n$. Since $(Y_n)_{n\in\mathbb{N}}$ is bounded and $Y_n \to Y$ pointwise, by Axiom (C), $X_m \succsim Y$ for all $m \in \mathbb{N}$. Since $(X_m)_{m\in\mathbb{N}}$ is bounded and $X_m \to X$ pointwise, by Axiom (C), $X \succsim Y$. Thus, (3.2) holds.

Conversely, assume that $\succsim$ is a CRDU preference relation. The fact that $\succsim$ satisfies Axioms (RC), (M), (SRM), and (C) follows from Proposition 3.1.

Let $R \in \mathcal{G}_{\mathrm{cf}}$ and $x, y, z, z' \in \mathbb{R}$ be chosen such that $x \geq y$ and $z, z' \in [y, x]$. We have

$$u(c_{xRz}) = u(x)g(1/2) + u(z)(1 - g(1/2)) \geq u(z')g(1/2) + u(y)(1 - g(1/2)) = u(c_{z'Ay}).$$

Therefore, $c_{xAz} \geq c_{z'Ay}$. Using a similar argument, one can show that $c_{xRz'} \geq c_{zRy}$. Thus,

$$\begin{aligned}
&\int_\Omega u(c_{xRz}Rc_{z'Ry})\mathrm{d}(g\circ\nu)\\
&= u(c_{xRz})g(1/2) + u(c_{z'Ry})(1-g(1/2))\\
&= [u(x)g(1/2) + u(z)(1-g(1/2))]\, g(1/2) + [u(z')g(1/2) + u(y)(1-g(1/2))]\,(1-g(1/2))\\
&= [u(x)g(1/2) + u(z')(1-g(1/2))]\, g(1/2) + [u(z)g(1/2) + u(y)(1-g(1/2))]\,(1-g(1/2))\\
&= u(c_{xRz'})g(1/2) + u(c_{zRy})(1-g(1/2)) = \int_\Omega u(c_{xRz'}Rc_{zRy})\mathrm{d}(g\circ\nu),
\end{aligned}$$

that is, $c_{xRz}Rc_{z'Ry} \simeq c_{xRz'}Rc_{zRy}$, and hence $\succsim$ satisfies Axiom (RS).

Let $R \in \mathcal{G}_{\mathrm{cf}}$, $x, y \in \mathbb{R}$ with $x \geq y$, and $z = x \oplus y$. It holds that

$$\begin{aligned}
&[u(x)g(1/2) + u(z)(1-g(1/2))]\, g(1/2) + [u(z)g(1/2) + u(y)(1-g(1/2))]\,(1-g(1/2))\\
&= u(c_{xRz})g(1/2) + u(c_{zRy})(1-g(1/2))\\
&= u(c_{c_{xRz}Rc_{zRy}})\\
&= u(c_{xRy}) = u(x)g(1/2) + u(y)(1-g(1/2)).
\end{aligned}$$

Therefore,

$$u(x)g(1/2)(1-g(1/2)) + u(y)g(1/2)(1-g(1/2)) = 2u(z)g(1/2)(1-g(1/2)).$$

As $1 > g(1/2) > 0$, it holds that $u(z) = (1/2)u(x) + (1/2)u(y)$. Therefore, $u((1/2)X \oplus_R (1/2)Y) = (1/2)u(X) + (1/2)u(Y)$ for all $X, Y \in \mathcal{X}$. Using the comonotonic additivity of the Choquet integral, it is clear that $\succsim$ satisfies Axiom (SCI). □

3.9 Proofs accompanying Section 3.4

Proof of Theorem 3.2. Assume that $\succsim_2$ is more ambiguity averse than $\succsim_1$. Let $A \in \mathcal{F}$ and $R_1 \in \mathcal{G}$ such that $A \simeq_1 R_1$. By the definition of comparative ambiguity attitudes, $R_1 \succsim_2 A$. Find $R_2 \in \mathcal{G}$ such that $A \simeq_2 R_2$. As $R_1 \succsim_2 R_2$, it holds that $\mathbb{P}(R_1) \geq \mathbb{P}(R_2)$ and $\nu_1(A) \geq \nu_2(A)$. To show the converse, let $R \in \mathcal{G}$ and $A \in \mathcal{F}$ satisfy $R \succsim_1 A$. Furthermore, let $R_1, R_2 \in \mathcal{G}$ satisfy $A \simeq_1 R_1$ and $A \simeq_2 R_2$. Since $\nu_1(A) \geq \nu_2(A)$ and $R \succsim_1 A$, $\mathbb{P}(R) \geq \mathbb{P}(R_1) \geq \mathbb{P}(R_2)$. Therefore, $R \succsim_2 R_2 \simeq_2 A$. □

Proof of Proposition 3.6. Assume that the $\succsim_1$-utility is equal to the $\succsim_2$-utility, the $\succsim_1$-distortion is equal to the $\succsim_2$-distortion, and $\succsim_2$ is more ambiguity averse than $\succsim_1$. Denote

the shared utility by u and the shared distortion by g. Denote the $\succsim_1$-matching probability by ν_1 and the $\succsim_2$-matching probability by ν_2. By Theorem 3.2, for all $A \in \mathcal{F}$, $g(\nu_1(A)) \geq g(\nu_2(A))$. Let $X \in \mathcal{X}(\mathcal{G})$ and $Y \in \mathcal{X}$ satisfy $X \succsim_1 Y$ ($X \succ_1 Y$). We have

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu_2) = \int_\Omega u(X)\mathrm{d}(g \circ \mathbb{P}) = \int_\Omega u(X)\mathrm{d}(g \circ \nu_1) \overset{(>)}{\geq} \int_\Omega Y\mathrm{d}(g \circ \nu_1) \geq \int_\Omega Y\mathrm{d}(g \circ \nu_2).$$

Therefore, $X \succsim_2 Y$ ($X \succ_2 Y$).

For the converse, assume that (3.6) holds true. Since (3.6) is stronger than the definition of comparative ambiguity aversion, $\succsim_2$ is more ambiguity averse than $\succsim_1$. Let $X, Y \in \mathcal{X}(\mathcal{G})$. If $X \succsim_1 Y$, then $X \succsim_2 Y$ by (3.6). If $X \succsim_2 Y$, then $X \succsim_1 Y$ or else there would be a contradiction with (3.6). Thus,

$$X \succsim_1 Y \iff X \succsim_2 Y, \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

Denote the $\succsim_1$-utility by u_1, the $\succsim_1$-distortion by g_1, the $\succsim_2$-utility by u_2, and the $\succsim_2$-distortion by g_2. Therefore, the functions

$$I_1(X) = \int_\Omega u_1(X)\mathrm{d}(g_1 \circ \mathbb{P}) \text{ and } I_2(X) = \int_\Omega u_2(X)\mathrm{d}(g_2 \circ \mathbb{P}), \quad X \in \mathcal{X}(\mathcal{G}),$$

both represent the preference relation on $\mathcal{X}(\mathcal{G})$. By Ghirardato and Marinacci [2001, Theorem 11], as $\succsim$ is a biseperable preference relation, $u_1 = au_2 + b$ for some $a > 0$ and $b \in \mathbb{R}$, and $g_1 \circ \mathbb{P} = g_2 \circ \mathbb{P}$. Clearly, this means that $g_1 = g_2$. As $u_1(0) = u_2(0) = 0$ and $u_1(1) = u_2(1) = 1$, we have that $u_1 = u_2$. □

The next lemma is a variant of a well-known result; see Lemma 6 of Abdellaoui and Wakker [2020]. Although their result assumes some additional properties, Property (AN) is sufficient for the proof of their part (ii) to go through, which is what the following lemma states.

Lemma 3.1. *Let $\succsim$ satisfy Property (AN). Then for all $A_1, A_2, B_1, B_2 \in \mathcal{F}$ satisfying $A_1 \cap A_2 = \varnothing$, $B_1 \cap B_2 = \varnothing$, and $A_1 \succsim B_1$, it holds that*

$$A_2 \succsim B_2 \text{ (resp. } A_2 \succ B_2) \implies A_1 \cup A_2 \succsim B_1 \cup B_2 \text{ (resp. } A_1 \cup A_2 \succ B_1 \cup B_2).$$

Proof of Theorem 3.3. Let ν denote the $\succsim$-matching probability, whose existence is given by Proposition 3.2. Assume that $\succsim$ satisfies Property (AN). Let $A, B \in \mathcal{F}$ be disjoint. For the sake of contradiction, assume that $\nu(A) + \nu(B) > 1$. Let $R_A \in \mathcal{G}$ satisfy $A \sim R_A$. Therefore, $\nu(A) = \mathbb{P}(R_A)$. Since $\nu(B) > 1 - \nu(A) = \mathbb{P}(R_A^c)$, we have, by Proposition 3.2,

$B \succ R_A^c$. Therefore, by Lemma 3.1 and Axiom (MO), $\Omega \succsim A \cup B \succ R_A \cup R_A^c = \Omega$, which is a contradiction.

Since $\nu(A)+\nu(B) \leq 1$, we can find disjoint $R_A, R_B \in \mathcal{G}$ such that $A \sim R_A$ and $B \sim R_B$. Therefore, by Lemma 3.1, we have $A \cup B \sim R_A \cup R_B$. Thus,

$$\nu(A \cup B) = \mathbb{P}(R_A \cup R_B) = \mathbb{P}(R_A) + \mathbb{P}(R_B) = \nu(A) + \nu(B).$$

Since ν is continuous (Proposition 3.2), we have that $\nu \in \mathcal{M}_1$.

Conversely, assume that $\nu \in \mathcal{M}_1$. Let $A, B, C \in \mathcal{F}$ satisfy $(A \cup B) \cap C = \varnothing$. By Proposition 3.2,

$$A \succsim B \iff \nu(A) \geq \nu(B) \iff \nu(A \cup C) \geq \nu(B \cup C) \iff A \cup C \succsim B \cup C.$$

Thus, $\succsim$ satisfies (AN). □

In the following proof, $\mathcal{X}_0(\mathcal{G}) = \mathcal{X}_0 \cap \mathcal{X}(\mathcal{G})$, where $\mathcal{X}_0$ denotes the collection of simple acts.

Proof of Proposition 3.7. Let ν denote the $\succsim$-matching probability. Assume that Property (AN) holds. Therefore, by Theorem 3.3, $\nu \in \mathcal{M}_1$. Let $X, Y \in \mathcal{X}$ satisfy $X \geq_\nu^{\text{sd}} Y$. We have that

$$g(\nu(u(X) > x)) \geq g(\nu(u(Y) > x)), \quad x \in \mathbb{R},$$

where u is the $\succsim$-utility and g is the $\succsim$-distortion. Therefore, $\int_\Omega u(X)\mathrm{d}(g\circ\nu) \geq \int_\Omega u(Y)\mathrm{d}(g\circ\nu)$ and $X \succsim Y$. Thus $\succsim$ satisfies (3.8) with $\mu = \nu$.

Assume that (3.8) holds with $\mu \in \mathcal{M}_1$. We claim that

$$\mathbb{P}(A) = 0 \iff \mu(A) = 0, \quad A \in \mathcal{G}.$$

Define $\epsilon_\mu = \inf\{\mu(B) : B \in \mathcal{G} \text{ and } B \succ \varnothing\}$. If $B \in \mathcal{G}$ satisfies $\mu(B) > \epsilon_\mu$, then there exists $B' \in \mathcal{G}$ such that $\mu(B) > \mu(B')$ and $B \succ \varnothing$. As $\succsim$ satisfies Property (3.8), this implies that $B \succsim B' \succ \varnothing$. Furthermore, if $B \in \mathcal{G}$ satisfies $B \succ \varnothing$, then $\mu(B) \geq \epsilon_\mu$. For the sake of contradiction, assume that $\epsilon_\mu > 0$ and fix $n \in \mathbb{N}$ satisfying $n > 1/\epsilon_\mu$. Since $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, we can find a partition of Ω, $\{A_1, \ldots, A_n\} \subseteq \mathcal{G}$, such that $\mathbb{P}(A_k) = 1/n$ for each $k \in \{1, \ldots, n\}$. By Axiom (SRM), $A_k \succ \varnothing$ for all $k \in \{1, \ldots, n\}$. Therefore $\mu(A_k) \geq \epsilon_\mu$ for all $k \in \{1, \ldots, n\}$. Thus

$$1 = \mu(\Omega) = \sum_{k=1}^{n} \mu(A_k) \geq n\epsilon_\mu > 1,$$

a contradiction. Therefore, $\epsilon_\mu = 0$. For all $A \in \mathcal{G}$, by Axiom (SRM), we have

$$\mathbb{P}(A) = 0 \iff A \simeq \varnothing \iff \mu(A) \leq \epsilon_\mu \iff \mu(A) = 0.$$

Since $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, $(\Omega, \mathcal{G}, \mu|_{\mathcal{G}})$ is atomless. Define the function $\psi : \mathcal{X}_0(\mathcal{G}) \to \mathbb{R}$ by

$$\psi(X) = \int_\Omega u(X) d(g \circ \mathbb{P}),$$

where u is the $\succsim$-utility and g is the $\succsim$-distortion. For the sake of contradiction, assume that $\mathbb{P} \neq \mu|_{\mathcal{G}}$. Then, by [Liebrich, 2024, Theorem 6], ψ must satisfy

$$\psi(X) = T\left(\max\{X(\omega) : \omega \in \Omega\}, \min\{X(\omega) : \omega \in \Omega\}\right),$$

where $T : \mathbb{R}^2 \to \mathbb{R}$. Let $\{R_1, R_2, R_3\} \subseteq \mathcal{G}$ be a partition of Ω satisfying $\mathbb{P}(R_1) = \mathbb{P}(R_2) = \mathbb{P}(R_3) = 1/3$ and $x_1, x_2, x_3, x_4 \in \mathbb{R}$ satisfy $x_1 > x_2 > x_3 > x_4$. Therefore, we have

$$\begin{aligned} T(x_1, x_4) = \psi(x_1 R_1 x_2 R_2 x_4) &= \int_\Omega u(x_1 R_1 x_2 R_2 x_4) \mathrm{d}(g \circ \mathbb{P}) \\ &> \int_\Omega u(x_1 R_1 x_3 R_2 x_4) \mathrm{d}(g \circ \mathbb{P}) = \psi(x_1 R_1 x_3 R_2 x_4) = T(x_1, x_4), \end{aligned}$$

which is a contradiction. Thus, $\mathbb{P} = \mu|_{\mathcal{G}}$. Let $A \in \mathcal{F}$ and $R_A \in \mathcal{G}$ such that $\mu(A) = \mathbb{P}(R_A)$. Since $\succsim$ satisfies (3.8), we have $A \simeq R_A$. Therefore $\nu(A) = \mathbb{P}(R_A) = \mu(A)$. Thus $\nu = \mu$ and $\nu \in \mathcal{M}_1$. By Theorem 3.3, $\succsim$ satisfies Property (AN). □

Proof of Proposition 3.8. Assume that $\succsim$ satisfies Property (DS). Let $X, Y \in \mathcal{X}$ satisfy $X \succsim Y$. We claim that there exists $c_0 \geq 0$ such that $X - c_0 \simeq Y$. Define $\phi : [0, \infty) \to \mathbb{R}$ by

$$\phi(c) = \int_\Omega u(X - c) \mathrm{d}(g \circ \nu).$$

By Wang and Klir [2010, Theorem 11.9], the function ϕ is continuous. Define the constant $c_1 = \|X\| + \|Y\|$. It holds that $X - c_1 \leq -\|Y\| \leq Y$. By Axiom (M), $Y \succsim X - c_1$ and

$$\phi(0) = \int_\Omega u(X) \mathrm{d}(g \circ \nu) = \int_\Omega u(Y) \mathrm{d}(g \circ \nu) \geq \int_\Omega u(X - c_1) \mathrm{d}(g \circ \nu) = \phi(c_1).$$

Therefore, by the intermediate value theorem, there exists $c_0 \geq 0$ such that $X - c_0 \simeq Y$.

By Property (DS) and Axiom (M), for all $\lambda \in [0,1]$,

$$\lambda X + (1-\lambda)Y \succsim \lambda(X - c_0) + (1-\lambda)Y \succsim Y.$$

Therefore, $\succsim$ is convex as defined in Wakker and Yang [2019, Definition 1]. By Wakker and Yang [2019, Corollary 7], as u is strictly increasing, u is concave and $g \circ \nu$ is supermodular. The proof of the converse is straightforward. □

Proof of Theorem 3.4. This is a consequence of Proposition 3.13 in Appendix 3.12, Proposition 3.8, and the remark regarding the work of Schmidt and Zank [2008]. □

In the following proof, given $A \in \mathcal{F}$, we use $\mathbb{Q}(A|\mathcal{H})$ to denote the conditional probability of A given the sub-σ-algebra $\mathcal{H}$.

Proof of the claim in Example 3.3. Define the capacity

$$\tilde{\nu}(A) = \int_\Omega g(\mathbb{Q}(A|\mathcal{H}))\mathrm{d}\mathbb{Q}, \quad A \in \mathcal{F}.$$

An application of both the monotone convergence theorem and the conditional monotone convergence theorem will verify that $\tilde{\nu}$ is continuous. As $\mathcal{G}$ and $\mathcal{H}$ are independent, it is straightforward to show that for all $A \in \mathcal{G}$, $\tilde{\nu}(A) = g(\mathbb{Q}(A)) = g(\mathbb{P}(A))$. Also, for all $A \in \mathcal{H}$, $\tilde{\nu}(A) = \mathbb{Q}(A)$. Since g is convex (as g is the inverse of h), $g(x) \leq x$ for all $x \in [0,1]$. Therefore, $\tilde{\nu}(A) \leq \mathbb{Q}(A)$ for all $A \in \mathcal{F}$. Finally, by Jensen's inequality, it holds that

$$\tilde{\nu}(A) = \int_\Omega g(\mathbb{Q}(A|\mathcal{H}))\mathrm{d}\mathbb{Q} \geq g\left(\int_\Omega \mathbb{Q}(A|\mathcal{H})\mathrm{d}\mathbb{Q}\right) = g(\mathbb{Q}(A)), \quad A \in \mathcal{F}.$$

Given $A, B \in \mathcal{F}$, as $\mathbb{1}_{A\cup B} + \mathbb{1}_{A\cap B} = \mathbb{1}_A + \mathbb{1}_B$, $\mathbb{Q}(A \cup B|\mathcal{H}) + \mathbb{Q}(A \cap B|\mathcal{H}) =^{\text{as}}_{\mathbb{Q}} \mathbb{Q}(A|\mathcal{H}) + \mathbb{Q}(B|\mathcal{H})$. Using a similar argument as Föllmer and Schied [2016, Proposition 4.75], it holds that

$$g(\mathbb{Q}(A \cup B|\mathcal{H})) + g(\mathbb{Q}(A \cap B|\mathcal{H})) \geq^{\text{as}}_{\mathbb{Q}} g(\mathbb{Q}(A|\mathcal{H})) + g(\mathbb{Q}(B|\mathcal{H})). \tag{3.16}$$

Applying the Choquet integral with respect to $\mathbb{Q}$ to (3.16) shows that $\tilde{\nu}$ is supermodular.

Define the capacity ν by $\nu(A) = h(\tilde{\nu}(A))$ for all $A \in \mathcal{F}$. It is clear that ν is risk conforming, $\nu(A) = h(\mathbb{Q}(A))$ for all $A \in \mathcal{H}$, $h(\mathbb{Q}(A)) \geq \nu(A) \geq \mathbb{Q}(A)$ for all $A \in \mathcal{F}$. Since $g \circ \nu = \tilde{\nu}$, $g \circ \nu$ is supermodular. As h is continuous, ν is continuous. □

Proof of Theorem 3.5. Assume that ν is supermodular. Let $X \in \mathcal{X}$, by Marinacci and Montrucchio [2004, Theorem 4.7], as $\{u(X) > x\}_{x\in\mathbb{R}}$ is a chain, there exists $\tilde{\mu} \in \mathfrak{C}(\nu)$ such

that $\tilde{\mu}(u(X) > x) = \nu(u(X) > x) \leq \mu(u(X) > x)$ for all $x \in \mathbb{R}$ and $\mu \in \mathfrak{C}(\nu)$. Therefore, as g is increasing,

$$\int_\Omega u(X)\mathrm{d}(g \circ \tilde{\mu}) = \int_\Omega u(X)\mathrm{d}(g \circ \nu) \leq \int_\Omega u(X)\mathrm{d}(g \circ \mu), \quad \mu \in \mathfrak{C}(\nu).$$

Thus, $\min_{\mu \in \mathfrak{C}(\nu)} \int_\Omega u(X)\mathrm{d}(g \circ \mu) = \int_\Omega u(X)\mathrm{d}(g \circ \tilde{\mu}) = \int_\Omega u(X)\mathrm{d}(g \circ \nu)$. Therefore, for all $X, Y \in \mathcal{X}$,

$$\begin{aligned} X \gtrsim Y &\iff \int_\Omega u(X)\mathrm{d}(g \circ \nu) \geq \int_\Omega u(Y)\mathrm{d}(g \circ \nu) \\ &\iff \min_{\mu \in \mathfrak{C}(\nu)} \int_\Omega u(X)\mathrm{d}(g \circ \mu) \geq \min_{\mu \in \mathfrak{C}(\nu)} \int_\Omega u(Y)\mathrm{d}(g \circ \mu). \end{aligned}$$

Conversely, assume that (3.10) holds. Let $X \in \mathcal{X}$, we know that $u(c_X) = \int_\Omega u(X)\mathrm{d}(g \circ \nu)$. Therefore, by (3.10),

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu) = \min_{\mu \in \mathfrak{C}(\nu)} \int_\Omega u(X)\mathrm{d}(g \circ \mu).$$

Let $A, B \in \mathcal{F}$ satisfy $A \subseteq B$ and $c \in \mathbb{R}$ satisfy $u(c) = 2$, by the above result, there exists $\tilde{\mu} \in \mathfrak{C}(\nu)$ such that

$$\begin{aligned} g(\nu(A)) + g(\nu(B)) &= \int_\Omega u(\mathbb{1}_A)\mathrm{d}(g \circ \nu) + \int_\Omega u(\mathbb{1}_B)\mathrm{d}(g \circ \nu) \\ &= \int_\Omega u(\mathbb{1}_A) + u(\mathbb{1}_B)\mathrm{d}(g \circ \nu) \\ &= \int_\Omega u(c\mathbb{1}_A + \mathbb{1}_{B \setminus A})\mathrm{d}(g \circ \nu) \\ &= \int_\Omega u(c\mathbb{1}_A + \mathbb{1}_{B \setminus A})\mathrm{d}(g \circ \tilde{\mu}) = g(\tilde{\mu}(A)) + g(\tilde{\mu}(B)). \end{aligned}$$

As $g(\nu(A)) \leq g(\tilde{\mu}(A))$ and $g(\nu(B)) \leq g(\tilde{\mu}(B))$, $g(\nu(A)) = g(\tilde{\mu}(A))$ and $g(\nu(B)) = g(\tilde{\mu}(B))$. Since g is bijective, we have that $\nu(A) = \tilde{\mu}(A)$ and $\nu(B) = \tilde{\mu}(B)$. Therefore, by Marinacci and Montrucchio [2004, Theorem 4.7], ν is supermodular. □

3.10 Results and proofs accompanying Section 3.5

Proof of Proposition 3.9. Let ν denote the $\succsim$-matching probability. Assume that $\succsim$ satisfies Property (RAA). Let $A \in \mathcal{F}$ and $R_A \in \mathcal{G}$ such that $A \simeq R_A$. Therefore, $\nu(A) = \mathbb{P}(R_A)$. Furthermore, let $R \in \mathcal{G}$ satisfy $\mathbb{Q}(R) = \mathbb{Q}(A)$. By Property (RAA), $R \succsim A \simeq R_A$. Therefore, by Proposition 3.2 and the fact that $\mathbb{Q}$ is risk conforming,

$$\mathbb{Q}(A) = \mathbb{Q}(R) = \mathbb{P}(R) \geq \mathbb{P}(R_A) = \nu(A).$$

Since $A \in \mathcal{F}$ was arbitrary, $\mathbb{Q} \in \mathfrak{C}(\nu)$.

Conversely, assume that $\mathbb{Q} \in \mathfrak{C}(\nu)$. Denote by u the $\succsim$-utility and by g the $\succsim$-distortion. Let $X \in \mathcal{X}(\mathcal{G})$ and $Y \in \mathcal{X}$ satisfy $X =^{\mathrm{sd}}_{\mathbb{Q}} Y$. Since ν and $\mathbb{Q}$ are risk conforming, for all $x \in \mathbb{R}$, we have

$$\begin{aligned} g(\nu(u(X) > x)) = g(\mathbb{P}(u(X) > x)) = g(\mathbb{Q}(u(X) > x)) = g(\mathbb{Q}(u(Y) > x)) \\ \geq g(\nu(u(Y) > x)). \end{aligned}$$

Therefore, $\int_\Omega u(X)\mathrm{d}(g \circ \nu) \geq \int_\Omega u(Y)\mathrm{d}(g \circ \nu)$ and $X \succsim Y$. Therefore, $\succsim$ satisfies Property (RAA). □

Proof of Proposition 3.10. Let ν denote the $\succsim$-matching probability. Assume that $\succsim$ satisfies Property (NSC). If $A, B \in \mathcal{F}$ satisfy $\mathbb{Q}(A \triangle B) = 0$, then $\mathbb{1}_A =^{\mathrm{as}}_{\mathbb{Q}} \mathbb{1}_B$. By Property (NSC), $A \simeq B$. Therefore, $\nu(A) = \nu(B)$ and ν is $\mathbb{Q}$-consistent.

For the converse, assume that ν is $\mathbb{Q}$-consistent. Denote by u the $\succsim$-utility and by g the $\succsim$-distortion. Fix $X, Y \in \mathcal{X}$ satisfying $X =^{\mathrm{as}}_{\mathbb{Q}} Y$. As $\mathbb{Q}(u(X) \neq u(Y)) = 0$ and

$$\{u(X) > x\} \triangle \{u(Y) > x\} \subseteq \{u(X) \neq u(Y)\},$$

we obtain $\mathbb{Q}(\{u(X) > x\} \triangle \{u(Y) > x\}) = 0$. Therefore, since ν is $\mathbb{Q}$-consistent, we have that $g(\nu(u(X) > x)) = g(\nu(u(Y) > x))$ for all $x \in \mathbb{R}$. Thus,

$$\int_\Omega u(X)\mathrm{d}(g \circ \nu) = \int_\Omega u(Y)\mathrm{d}(g \circ \nu),$$

and we can conclude $X \simeq Y$. □

To ease notation, given a distortion family $(g_X)_{X \in \mathcal{X}}$ and $A \in \mathcal{F}$, we use g_A to denote $g_{\mathbb{1}_A}$.

Proof of Theorem 3.6. Assume that $\succsim$ is a CRDU preference relation satisfying (3.2) and Property (NSC). Denote by ν the $\succsim$-matching probability and by g the $\succsim$-distortion. Fix $X \in \mathcal{X}$ and define $\tilde{g}_X : \mathcal{D}(X) \to [0,1]$ by

$$\tilde{g}_X(\alpha) = g(\nu(X > x)), \quad \text{where } x \in \mathbb{R} \text{ satisfies } \alpha = \mathbb{Q}(X > x). \tag{3.17}$$

To see that $\tilde{g}_X$ is well-defined, let $x, y \in \mathbb{R}$ satisfy $\mathbb{Q}(X > x) = \mathbb{Q}(X > y)$. It is straightforward to verify that $\mathbb{Q}(\{X > x\} \triangle \{X > y\}) = 0$. Therefore, by Proposition 3.10, we have that $\nu(X > x) = \nu(X > y)$ and $\tilde{g}_X$ is well-defined. Furthermore, if $X \in \mathcal{X}(\mathcal{G})$, as ν and $\mathbb{Q}$ are risk conforming, it holds that

$$\tilde{g}_X(\mathbb{Q}(X > x)) = g(\nu(X > x)) = g(\mathbb{P}(X > x)) = g(\mathbb{Q}(X > x)), \quad x \in \mathbb{R}. \tag{3.18}$$

Thus, if $X \in \mathcal{X}(\mathcal{G})$, $\tilde{g}_X(\alpha) = g(\alpha)$ for all $\alpha \in \mathcal{D}(X)$.

We claim that $\tilde{g}_X$ is increasing. Let $\alpha, \beta \in \mathcal{D}(X)$ with $\alpha < \beta$. By the definition of $\mathcal{D}(X)$, there exists $x, y \in \mathbb{R}$ such that $\alpha = \mathbb{Q}(X > x)$ and $\beta = \mathbb{Q}(X > y)$. As $\alpha < \beta$, it must hold that $y < x$. Therefore, $\tilde{g}_X(\alpha) = g(\nu(X > x)) \leq g(\nu(X > y)) = \tilde{g}_X(\beta)$.

Next, we claim that $\tilde{g}_X$ is continuous. Let $(\alpha_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}(X)$ be a strictly increasing sequence converging to $\alpha \in \mathcal{D}(X)$. There exists a strictly decreasing sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ such that $\alpha_n = \mathbb{Q}(X > x_n)$. Define $x = \lim_{n\to\infty} x_n = \inf_{n\in\mathbb{N}} x_n$. It is clear that $\alpha = \mathbb{Q}(X > x)$. We have that

$$\lim_{n\to\infty} \tilde{g}_X(\alpha_n) = \lim_{n\to\infty} g(\nu(X > x_n)) = g(\nu(X > x)) = \tilde{g}_X(\alpha).$$

Let $(\alpha_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}(X)$ be a strictly decreasing sequence converging to $\alpha \in \mathcal{D}(X)$. There exists a strictly increasing sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ such that $\alpha_n = \mathbb{Q}(X > x_n)$. As $\alpha \in \mathcal{D}(X)$, there exists $z \in \mathbb{R}$ such that $\mathbb{Q}(X > z) = \alpha$. Define $x = \lim_{n\to\infty} x_n = \sup_{n\in\mathbb{N}} x_n$, as $z > x_n$ for all $n \in \mathbb{N}$, $z \geq x$. As

$$\alpha = \lim_{n\to\infty} \mathbb{Q}(X > x_n) = \mathbb{Q}(X \geq x) \geq \mathbb{Q}(X > x) \geq \mathbb{Q}(X > z) = \alpha,$$

we have $\mathbb{Q}(X = x) = 0$. Therefore, $\mathbb{Q}(\{X > x\} \triangle \{X \geq x\}) = 0$. By Proposition 3.10, it holds that $\nu(X > x) = \nu(X \geq x)$. Since $\mathbb{Q}(X > x) = \alpha$, we have

$$\lim_{n\to\infty} \tilde{g}_X(\alpha_n) = \lim_{n\to\infty} g(\nu(X > x_n)) = g(\nu(X \geq x)) = g(\nu(X > x)) = \tilde{g}_X(\alpha).$$

Therefore, $\tilde{g}_X$ is continuous.

If $X = \mathbb{1}_A$ for some $A \in \mathcal{F}$, define $g_A : [0,1] \to [0,1]$ by $g_A(\alpha) = \tilde{g}_A(\mathbb{Q}(A))$ for $\alpha \in (0,1)$,

$g_A(0) = 0$, and $g_A(1) = 1$. It is clear that g_A is a distortion function. Furthermore, as $\mathcal{D}(\mathbb{1}_A) = \{0, \mathbb{Q}(A), 1\}$, $g_A(\alpha) = \tilde{g}_A(\alpha)$ for all $\alpha \in \mathcal{D}(\mathbb{1}_A)$. If $X \neq \mathbb{1}_A$ for some $A \in \mathcal{F}$, define $g_X : [0,1] \to [0,1]$ by

$$g_X(\alpha) = \sup\{\tilde{g}_X(\beta) : \beta \in \mathcal{D}(X) \text{ and } \beta \leq \alpha\}.$$

It is straightforward to verify that g_X is a distortion function. Since $\tilde{g}_X$ is increasing, $g_X(\alpha) = \tilde{g}_X(\alpha)$ for all $\alpha \in \mathcal{D}(X)$.

Since for all $X \in \mathcal{X}$, $g_X(\alpha) = \tilde{g}_X(\alpha)$ for all $\alpha \in \mathcal{D}(X)$, $(g_X)_{X\in\mathcal{X}}$ satisfies (b). For $A, B \in \mathcal{F}$ satisfying $A \subseteq B$ and $\alpha \in (0,1)$, we have

$$g_A(\alpha) = \tilde{g}_A(\mathbb{Q}(A)) = g(\nu(A)) \leq g(\nu(B)) = g_B(\alpha).$$

Therefore, $(g_X)_{X\in\mathcal{X}}$ is satisfies (a). Let $X, Y \in \mathcal{X}$ be comonotonic and $\alpha \in \mathcal{D}(X) \cap \mathcal{D}(Y)$. By definition, there exists $x, y \in \mathbb{R}$ such that $\alpha = \mathbb{Q}(X > x) = \mathbb{Q}(Y > y)$. As X and Y are comonotonic, by Denneberg [1994, Proposition 4.5], either $\{X > x\} \subseteq \{Y > y\}$ or $\{Y > y\} \subseteq \{X > x\}$. Since $\mathbb{Q}(X > x) = \mathbb{Q}(Y > y)$, it must hold that $\mathbb{Q}(\{Y > y\} \triangle \{X > x\}) = 0$. Therefore, by Proposition 3.10,

$$g_X(\alpha) = \tilde{g}_X(\alpha) = g(\nu(X > x)) = g(\nu(Y > y)) = \tilde{g}_Y(\alpha) = g_Y(\alpha),$$

and $(g_X)_{X\in\mathcal{X}}$ satisfies (c). Thus $(g_X)_{X\in\mathcal{X}}$ is a distortion family. Furthermore, since $g_X(\alpha) = \tilde{g}_X(\alpha)$ holds for all $\alpha \in \mathcal{D}(X)$ and all $X \in \mathcal{X}$, it follows from (3.18) that for all $X \in \mathcal{X}(\mathcal{G})$, $g_X(\alpha) = g(\alpha)$ for all $\alpha \in \mathcal{D}(X)$. Therefore, $(g_X)_{X\in\mathcal{X}}$ is a g-distortion family.

Denote by u the $\gtrsim$-utility. Let $u' : u(\mathbb{R}) \to \mathbb{R}$ be the inverse of u. Given $X \in \mathcal{X}$ and $x \in u(\mathbb{R})$, by the definition of $\tilde{g}_X$,

$$\begin{aligned} g_X(\mathbb{Q}(u(X) > x)) = g_X(\mathbb{Q}(X > u'(x))) = \tilde{g}_X(\mathbb{Q}(X > u'(x))) &= g(\nu(X > u'(x))) \\ &= g(\nu(u(X) > x)). \end{aligned}$$

Therefore, $\int_\Omega u(X)\mathrm{d}(g \circ \nu) = \int_\Omega u(X)\mathrm{d}(g_X \circ \mathbb{Q})$ and $\gtrsim$ satisfies (3.12).

Let $(g_X)_{X\in\mathcal{X}}$ and $(\tilde{g}_X)_{X\in\mathcal{X}}$ be g-distortion families satisfying (3.12). Fix $X \in \mathcal{X}$ and $\alpha \in \mathcal{D}(X)$. By definition, there exists $x \in \mathbb{R}$ such that $\alpha = \mathbb{Q}(X > x)$. Denote by $A = \{X > x\}$. Since X and $\mathbb{1}_A$ are comonotonic and $\alpha \in \mathcal{D}(X) \cap \mathcal{D}(\mathbb{1}_A)$, $g_X(\alpha) = g_A(\alpha)$ and $\tilde{g}_X(\alpha) = \tilde{g}_A(\alpha)$. Let $c_A \in [0,1]$ satisfy $A \simeq c_A$, by (3.12),

$$g_A(\alpha) = \int_\Omega u(\mathbb{1}_A)\mathrm{d}(g_A \circ \mathbb{Q}) = u(c_A) = \int_\Omega u(\mathbb{1}_A)\mathrm{d}(\tilde{g}_A \circ \mathbb{Q}) = \tilde{g}_A(\alpha).$$

Therefore, $g_A(\alpha) = \tilde{g}_A(\alpha)$ and $g_X(\alpha) = \tilde{g}_X(\alpha)$.

Assume that $\succsim$ satisfies Property (RAA). By Proposition 3.9, $\mathbb{Q} \in \mathfrak{C}(\nu)$. Let $X \in \mathcal{X}$ and $\alpha \in \mathcal{D}(X)$. Find $x \in \mathbb{R}$ such that $\alpha = \mathbb{Q}(X > x)$. By (3.17), we have

$$g_X(\alpha) = \tilde{g}_X(\alpha) = g_X(\mathbb{Q}(X > x)) = g(\nu(X > x)) \leq g(\mathbb{Q}(X > x)) = g(\alpha).$$

Therefore, $(g_X)_{X\in\mathcal{X}}$ is ambiguity averse.

For the converse, assume that $(g_X)_{X\in\mathcal{X}}$ is ambiguity averse. Let $X \in \mathcal{X}(\mathcal{G})$ and $Y \in \mathcal{X}$ satisfy $X =^{\mathrm{sd}}_{\mathbb{Q}} Y$, meaning, $\mathbb{Q}(X > x) = \mathbb{Q}(Y > x)$ for all $x \in \mathbb{R}$. Since $g_X(\alpha) = g(\alpha)$ for all $\alpha \in \mathcal{D}(X)$, and $g_Y(\alpha) \leq g(\alpha)$ for all $\alpha \in \mathcal{D}(Y)$, we have, for $x \in u(\mathbb{R})$,

$$\begin{aligned} g_X(\mathbb{Q}(u(X) > x)) &= g_X(\mathbb{Q}(X > u'(x))) \\ &= g(\mathbb{Q}(X > u'(x))) \\ &= g(\mathbb{Q}(Y > u'(x))) \\ &\geq g_Y(\mathbb{Q}(Y > u'(x))) = g_Y(\mathbb{Q}(u(Y) > x)), \end{aligned}$$

where $u' : u(\mathbb{R}) \to \mathbb{R}$ is the inverse of u. Thus, $\int_\Omega u(X)\mathrm{d}(g_X \circ \mathbb{Q}) \geq \int_\Omega u(Y)\mathrm{d}(g_X \circ \mathbb{Q})$, and $X \succsim Y$. Therefore, $\succsim$ satisfies (RAA). □

The following theorem confirms that (3.12) is also sufficient for $\succsim$ to be a CRDU preference relation.

Theorem 3.8. *Let $\succsim$ be a preference relation. Then $\succsim$ is a CRDU preference relation satisfying Property (NSC) if and only if there exist a continuous strictly increasing vNM utility function u, a continuous strictly increasing distortion function g, and a g-distortion family $(g_X)_{X\in\mathcal{X}}$ such that*

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}(g_X \circ \mathbb{Q}) \geq \int_\Omega u(Y)\mathrm{d}(g_Y \circ \mathbb{Q}), \quad X, Y \in \mathcal{X}. \tag{3.19}$$

Moreover, u is unique up to positive affine transformations and can be taken as the $\succsim$-utility, g is unique and given by the $\succsim$-risk distortion, and for each $X \in \mathcal{X}$, g_X is uniquely defined on $\mathcal{D}(X)$.

Proof. The forward direction follows from Theorem 3.6. Conversely, assume that there exist a continuous strictly increasing vNM utility u, a continuous strictly increasing distortion function g, and a g-distortion family $(g_X)_{X\in\mathcal{X}}$ such that

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}(g_X \circ \mathbb{Q}) \geq \int_\Omega u(Y)\mathrm{d}(g_Y \circ \mathbb{Q}), \quad X, Y \in \mathcal{X}.$$

Given $A \in \mathcal{F}$, define $\tilde{\nu}(A) = g_A(\mathbb{Q}(A))$. Clearly, $\tilde{\nu}(\varnothing) = 0$ and $\tilde{\nu}(\Omega) = 1$. Let $A, B \in \mathcal{F}$ satisfy $A \subseteq B$, as $(g_X)_{X\in\mathcal{X}}$ satisfies (a),

$$\tilde{\nu}(A) = g_A(\mathbb{Q}(A)) \leq g_A(\mathbb{Q}(B)) \leq g_B(\mathbb{Q}(B)) = \tilde{\nu}(B).$$

Therefore, $\tilde{\nu}(A) \leq \tilde{\nu}(B)$. We claim that $\tilde{\nu}$ is continuous. Let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ be increasing and define $A = \bigcup_{n=1}^{\infty} A_n$ and $A_0 = \varnothing$. Define $X = \sum_{n=1}^{\infty} 2^{-(n-1)} \mathbb{1}_{A_n \setminus A_{n-1}}$, it is straightforward to show

$$A_n = \left\{ X > \frac{3}{4} \times 2^{-(n-1)} \right\}, \quad n \in \mathbb{N}, \quad \text{and} \quad A = \{X > 0\}.$$

Since X and $\mathbb{1}_{A_n}$ are comonotonic for all $n \in \mathbb{N}$, $(g_X)_{X\in\mathcal{X}}$ satisfies (c), and $\mathbb{Q}(A_n) \in \mathcal{D}(X) \cap \mathcal{D}(\mathbb{1}_{A_n})$ for all $n \in \mathbb{N}$, it follows that $g_X(\mathbb{Q}(A_n)) = g_{A_n}(\mathbb{Q}(A_n)) = \tilde{\nu}(A_n)$ for all $n \in \mathbb{N}$. A similar argument will show that $g_X(\mathbb{Q}(A)) = \tilde{\nu}(A)$. Therefore, as $(\mathbb{Q}(A_n))_{n\in\mathbb{N}} \subseteq \mathcal{D}(X)$ satisfies $\lim_{n\to\infty} \mathbb{Q}(A_n) = \mathbb{Q}(A) \in \mathcal{D}(X)$ and $(g_X)_{X\in\mathcal{X}}$ satisfies (b),

$$\lim_{n\to\infty} \tilde{\nu}(A_n) = \lim_{n\to\infty} g_X(\mathbb{Q}(A_n)) = g_X(\mathbb{Q}(A)) = \tilde{\nu}(A).$$

Therefore, $\tilde{\nu}$ is upward continuous. A similar argument will show that $\tilde{\nu}$ is downward continuous. Therefore, $\tilde{\nu}$ is continuous.

Since g is continuous and strictly increasing, there exists a continuous strictly increasing distortion function h such that $h(g(x)) = g(h(x)) = x$ for all $x \in [0, 1]$. Define the capacity $\nu = h \circ \tilde{\nu}$. Clearly, ν is a continuous capacity. For all $A \in \mathcal{G}$, $g_A(\mathbb{Q}(A)) = g(\mathbb{Q}(A))$ as $\mathbb{Q}(A) \in \mathcal{D}(\mathbb{1}_A)$. Therefore, as $\mathbb{Q}$ is risk conforming, we have

$$\nu(A) = h(\tilde{\nu}(A)) = h(g_A(\mathbb{Q}(A))) = h(g(\mathbb{Q}(A))) = \mathbb{Q}(A) = \mathbb{P}(A), \quad A \in \mathcal{G}.$$

Thus, ν is risk conforming.

Let $u' : u(\mathbb{R}) \to \mathbb{R}$ be the inverse of u. Given $X \in \mathcal{X}$ and $x \in u(\mathbb{R})$, it holds that X and $\mathbb{1}_{\{X>u'(x)\}}$ are comonotonic and $\mathbb{Q}(X > u'(x)) \in \mathcal{D}(X) \cap \mathcal{D}(\mathbb{1}_{\{X>u'(x)\}})$. Since $(g_X)_{X\in\mathcal{X}}$ satisfies (c), we have

$$\begin{aligned} g_X(\mathbb{Q}(u(X) > x)) = g_X(\mathbb{Q}(X > u'(x))) &= g_{\{X>u'(x)\}}(\mathbb{Q}(X > u'(x))) \\ &= \tilde{\nu}(X > u'(x)) \\ &= g(\nu(X > u'(x))) = g(\nu(u(X) > x)). \end{aligned}$$

Therefore,

$$X \succsim Y \iff \int_\Omega u(X)\mathrm{d}(g\circ\nu) \geq \int_\Omega u(Y)\mathrm{d}(g\circ\nu), \quad X, Y \in \mathcal{X}.$$

Thus, $\succsim$ is a CRDU preference relation.

To show that $\succsim$ satisfies Property (NSC), fix $A, B \in \mathcal{F}$ satisfying $\mathbb{Q}(A \triangle B) = 0$. Therefore, $\mathbb{Q}(A \cap B) = \mathbb{Q}(A)$. Furthermore, we have that $\mathbb{1}_{A\cap B}$ and $\mathbb{1}_A$ are comonotonic and $\mathcal{D}(\mathbb{1}_{A\cap B}) = \mathcal{D}(\mathbb{1}_A) = \{0, \alpha, 1\}$, where $\alpha = \mathbb{Q}(A \cap B) = \mathbb{Q}(A)$. Therefore, since the distortion family $(g_X)_{X\in\mathcal{X}}$ satisfies (c),

$$\nu(A \cap B) = g_{A\cap B}(\mathbb{Q}(A \cap B)) = g_{A\cap B}(\alpha) = g_A(\alpha) = g_A(\mathbb{Q}(A)) = \nu(A).$$

Similary, one can show that $\nu(A \cap B) = \nu(B)$. Therefore, ν is $\mathbb{Q}$-consistent and, by Proposition 3.10, $\succsim$ satisfies Property (NSC).

The uniqueness claims are straightforward to show. □

It is not immediately clear why (3.19) implies that $\succsim$ satisfies Property (NSC). Therefore, we have the following proposition to assist in the sanity check of Theorem 3.8.

Proposition 3.12. *Assume that $(g_X)_{X\in\mathcal{X}}$ is a distortion family. For all $X, Y \in \mathcal{X}$ satisfying $X =^{\mathrm{as}}_{\mathbb{Q}} Y$, it holds that $\mathcal{D}(X) = \mathcal{D}(Y)$ and $g_X(\alpha) = g_Y(\alpha)$ for all $\alpha \in \mathcal{D}(X)$.*

Proof. It is clear that $\mathcal{D}(X) = \mathcal{D}(Y)$. Let $\alpha \in \mathcal{D}(X)$, there exists $x \in \mathbb{R}$ such that $\alpha = \mathbb{Q}(X > x)$. Since $X =^{\mathrm{as}}_{\mathbb{Q}} Y$, it holds that $\alpha = \mathbb{Q}(Y > x)$. Let $A = \{X > x\}$ and $B = \{Y > x\}$. As $A \triangle B \subseteq \{X \neq Y\}$, $\mathbb{Q}(A \triangle B) = 0$. Therefore, $\mathbb{Q}(A) = \mathbb{Q}(B) = \mathbb{Q}(A \cap B) = \alpha$. As $\mathbb{1}_{A\cap B}$ and $\mathbb{1}_A$ are comonotonic and $\mathcal{D}(\mathbb{1}_{A\cap B}) = \mathcal{D}(\mathbb{1}_A) = \{0, \alpha, 1\}$, by (c), we have $g_{A\cap B}(\alpha) = g_A(\alpha)$. Similarly, one can show that $g_{A\cap B}(\alpha) = g_B(\alpha)$. As $\mathbb{1}_A$ and X are comonotonic and $\alpha \in \mathcal{D}(X) \cap \mathcal{D}(\mathbb{1}_A)$, by (c), we have $g_A(\alpha) = g_X(\alpha)$. Similarly, one can show that $g_B(\alpha) = g_Y(\alpha)$. Thus, by a chain of the above equalities, $g_X(\alpha) = g_Y(\alpha)$. □

3.11 Proofs accompanying Section 3.6

Proof of Theorem 3.7. Let $\succsim$ be a preference relation satisfying Axioms (RC), (M), (SRM), (C), and (CI). By Proposition 3.4, for each $X \in \mathcal{X}$, there exists a unique $c_X \in \mathbb{R}$ satisfying $X \simeq c_X$. Define $V : \mathcal{X} \to \mathbb{R}$ by $V(X) = c_X$. Therefore,

$$X \succsim Y \iff V(X) \geq V(Y), \quad X, Y \in \mathcal{X}.$$

Let $X, Y \in \mathcal{X}$ be comonotonic. As $X, Y, V(X)$ are pairwise comonotonic, $Y, V(Y), V(X)$ are pairwise comonotonic, $X \simeq V(X)$, and $Y \simeq V(Y)$, by Axiom (CI), we have

$$(1/2)X + (1/2)Y \simeq (1/2)V(X) + (1/2)Y \simeq (1/2)V(X) + (1/2)V(Y).$$

Therefore, $V((1/2)X + (1/2)Y) = (1/2)V(X) + (1/2)V(Y)$. Taking $X = 2Z$ (for $Z \in \mathcal{X}$) and $Y = 0$, we see that $V(2Z) = 2V(Z)$. Putting this all together, we get

$$V(X+Y) = V((1/2)2X + (1/2)2Y) = (1/2)V(2X) + (1/2)V(2Y) = V(X) + V(Y).$$

Moreover, by Axiom (M), for all $X, Y \in \mathcal{X}$ satisfying $X \geq Y$, $V(X) \geq V(Y)$. As discussed in Section 3.2.1, there exists a capacity ν such that $V(X) = \int_\Omega X \mathrm{d}\tilde{\nu}$ for all $X \in \mathcal{X}$.

We claim that $\tilde{\nu}$ is continuous. Let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ be an increasing sequence and define $A = \bigcup_{n\in\mathbb{N}} A_n$. Since $\mathbb{1}_{A_n} \to \mathbb{1}_A$ pointwise, by Proposition 3.4,

$$\lim_{n\to\infty} \nu(A_n) = \lim_{n\to\infty} \int_\Omega \mathbb{1}_{A_n} \mathrm{d}\tilde{\nu} = \lim_{n\to\infty} c_{\mathbb{1}_{A_n}} = c_{\mathbb{1}_A} = \nu(A).$$

Therefore, $\tilde{\nu}$ is upward continuous. A similar argument will show that $\tilde{\nu}$ is downward continuous. Therefore $\tilde{\nu}$ is continuous.

Since $\succsim$ satisfies Axiom (RC), if $A, B \in \mathcal{G}$ satisfy $\mathbb{P}(A) = \mathbb{P}(B)$, then

$$A \simeq B \implies V(\mathbb{1}_A) = V(\mathbb{1}_B) \implies \tilde{\nu}(A) = \tilde{\nu}(B).$$

Thus, there exists a distortion function g such that $\tilde{\nu}(A) = g(\mathbb{P}(A))$ for all $A \in \mathcal{G}$. Since $\succsim$ satisfies Axiom (SRM), g is strictly increasing. Since $\tilde{\nu}$ is continuous and $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, g is continuous. Let h be the strictly increasing continuous distortion function that is the inverse of g. Define the capacity $\nu = h \circ \tilde{\nu}$. It is straightforward to show that ν is a continuous risk-conforming capacity satisfying $g \circ \nu = \tilde{\nu}$.

The converse is straightforward, and the uniqueness claims follow from Theorem 3.1. □

3.12 Modularity results for lattices

This section presents a technical result showing that on a general lattice, a distorted supermodular function is supermodular when the distortion function is increasing and convex. This result is used in the proof of Theorem 3.4.

A lattice is a partially ordered set $(\mathfrak{X}, \leq)$ in which every pair $x, y \in \mathfrak{X}$ admits a least upper bound $x \vee y$ and a greatest lower bound $x \wedge y$. A function $\varphi : \mathfrak{X} \to \mathbb{R}$ is called *supermodular* if for all $x, y \in \mathfrak{X}$, $\varphi(x \vee y) + \varphi(x \wedge y) \geq \varphi(x) + \varphi(y)$, *submodular* if for all $x, y \in \mathfrak{X}$, $\varphi(x \vee y) + \varphi(x \wedge y) \leq \varphi(x) + \varphi(y)$, *increasing* if for all $x, y \in \mathfrak{X}$ with $x \geq y$, $\varphi(x) \geq \varphi(y)$, and *decreasing* if for all $x, y \in \mathfrak{X}$ with $x \geq y$, $\varphi(x) \leq \varphi(y)$.

Proposition 3.13. *Let $(\mathfrak{X}, \leq)$ be a lattice, $a, b \in \mathbb{R}$ with $a < b$, $\varphi : \mathfrak{X} \to [a, b]$ be monotonic and supermodular (submodular), and $f : [a, b] \to \mathbb{R}$ be increasing and convex (concave). It holds that $g \circ \nu$ is supermodular (submodular).*

Proof. Assume that φ is supermodular and f is convex. For $u, v \in \mathfrak{X}$, define $z = \varphi(u \wedge v) + \varphi(u \vee v) - \varphi(u) - \varphi(v) \geq 0$. Note that

$$a \leq \varphi(v) + z = \varphi(u \wedge v) + \varphi(u \vee v) - \varphi(u) \leq \begin{cases} \varphi(u \vee v) & \text{if } \varphi \text{ is increasing} \\ \varphi(u \wedge v) & \text{if } \varphi \text{ is decreasing} \end{cases} \leq b.$$

As φ is monotonic, there exists $\lambda \in [0, 1]$ such that

$$\varphi(u) = \lambda\varphi(u \wedge v) + (1 - \lambda)\varphi(u \vee v).$$

Therefore,

$$\begin{aligned} f(\varphi(u \wedge v)) + f(\varphi(u \vee v)) &\geq f(\lambda\varphi(u \wedge v) + (1 - \lambda)\varphi(u \vee v)) \\ &\quad + f((1 - \lambda)\varphi(u \wedge v) + \lambda\varphi(u \vee v)) \\ &= f(\varphi(u)) + f(\varphi(v) + z) \\ &\geq f(\varphi(u)) + f(\varphi(v)). \end{aligned}$$

A similar proof will show the claim when φ is submodular and f is concave. □

Chapter 4

Aggregate then evaluate

4.1 Introduction

The expected utility (EU) model, originally axiomatized by von Neumann and Morgenstern [1947], has long served as the cornerstone of modern decision theory. The EU model is formulated in the pure-risk setting, where acts are represented by distributions over outcomes with objectively specified probabilities. However, the pure-risk setting is often unrealistic in many applications, as decision makers (DMs) rarely have complete knowledge of the true distributions underlying the alternatives they face.

To move beyond the pure-risk setting, Savage [1954] reformulated acts as functions from states of the world to consequences, rather than as distributions over consequences. This framework is now referred to as the Savage setting. Additionally, by imposing reasonably motivated axioms on a DM's preferences over such acts, Savage showed that (a) a subjective probabilistic model could be elicited for the DM, transforming the problem back into one of the pure-risk setting and (b) within this elicited pure-risk setting, the DM's preferences are consistent with the EU model. This decision model is known as the subjective expected utility (SEU) model.

The SEU model is particularly attractive because it grounds probability theory in behavioral principles. However, Ellsberg [1961] raised a critical challenge: when agents possess asymmetric information about the likelihoods of events, Savage's sure-thing principle can fail to describe plausible behavior. This insight undermined the behavioral appeal of the SEU model and, more generally, probabilistically sophisticated models [Machina and Schmeidler, 1992], which assume that uncertainty can be captured by a single subjective probabilistic model and that acts are evaluated solely through their induced distributions

over outcomes. Ellsberg's critique was pivotal in launching the formal study of decision-making under ambiguity. Among the most influential decision models under ambiguity are the Choquet expected utility model of Schmeidler [1989], the maxmin expected utility model of Gilboa and Schmeidler [1989], the α-maxmin model of Ghirardato et al. [2004], the smooth ambiguity model of Klibanoff et al. [2005], and the variational preferences model of Maccheroni et al. [2006]. These decision models under ambiguity share the important feature that, when ambiguity is absent, i.e., when uncertainty can be described by a single probabilistic model, they collapse back to the SEU model.

While reverting to the SEU model in the absence of ambiguity is consistent with the historical development of decision theory, this restriction limits the scope of these decision models. This follows because numerous alternative pure-risk decision models, which are either empirically or normatively appealing in certain contexts, have been proposed and widely adopted. Examples include the mean-variance model of Markowitz [1952], used in finance; the rank-dependent utility model of Quiggin [1982], the prospect theory of Kahneman and Tversky [1979], and the cumulative prospect theory of Tversky and Kahneman [1992], used in behavioral decision theory; the expected shortfall model of Acerbi and Tasche [2002], used in quantitative risk management. The aim of this chapter is to analyze general decision models under ambiguity that reduce to probabilistically sophisticated models in the absence of ambiguity, rather than relying solely on the SEU model. To do this, we consider two frameworks that both provide a clear separation between pure risk and ambiguity. A main difference between these two frameworks is that whether pure risk or ambiguity is processed first.

The first framework we discuss is called *evaluate-then-aggregate (ETA)*. In these decision models, given an act, the DM first examines the distributions it produces under various probabilistic models. For each distribution, the DM evaluates it using a statistic consistent with their pure-risk preferences, e.g., the expected-utility statistic in the case of the EU model. These model-specific evaluations are then aggregated across all the relevant probabilistic models using some aggregator. This framework provides a clear separation between a *risk component* (the evaluation at each probabilistic model) and an *ambiguity component* (the aggregation across probabilistic models). This framework includes most of the decision models under ambiguity discussed above.

The second framework is called *aggregate-then-evaluate (ATE)*, which we introduce in this chapter. In this framework, given an act, instead of evaluating each model separately and then aggregating, the DM first aggregates their ambiguity regarding the act into a single distribution. Afterwards, the DM evaluates this distribution using a statistic consistent with their pure-risk preferences. This framework provides a clear separation between an ambiguity component (the aggregation into a single distribution) and a risk

component (the evaluation of this distribution). For example, the Choquet expected utility model with respect to a continuous capacity ν can be viewed as first mapping each act X to a distribution Q_X whose survival function S_{Q_X} is given by

$$S_{Q_X}(x) = \nu(X > x), \quad x \in \mathbb{R}, \tag{4.1}$$

and then applying the expected-utility statistic to these distributions. Allowing for general statistics in the second step is what we refer to as the *Choquet ATE* model.

We illustrate the conceptual differences between the two frameworks by an example adapted from Ellsberg [1961]. An urn contains 90 balls; 30 are red, and the remaining 60 consist of blue and yellow balls in unknown proportions.[1] The possible distributions of the outcome of randomly picking a ball from the urn are represented by μ_x for $x \in \{0, 1, \dots, 60\}$ that represents the number of blue balls in the urn. The DM needs to rank four acts: R pays \$4 when a red ball is drawn, B pays \$6 when a blue ball is drawn, Y pays \$5 when a yellow ball is drawn, and C pays \$1 in all states. Suppose that the DM has a statistic γ defined on distributions, which computes the (numerical) utility of pure risks (represented by distributions). Note that R and C have no ambiguity, and hence their distributions Q_R and Q_C are known, and the utilities $\gamma(Q_R)$ and $\gamma(Q_C)$ can be directly computed. On the other hand, both B and Y have 61 possible distributions, one under each μ_x, which are denoted by $B_{\#}\mu_x$ and $Y_{\#}\mu_x$. Below, we describe how the ETA and ATE approaches assess B and Y.

1. The ETA approach first evaluates the utility of B, denoted by $U_x = \gamma(B_{\#}\mu_x)$ under μ_x for every $x \in \{0, 1, \dots, 60\}$, and then aggregates the 61 utilities by a mapping ρ representing the DM's ambiguity attitude into a utility $\rho(U_0, U_1, \dots, U_{60})$. For instance, the utilities may be aggregated using the worst-case method, a weighted average, or a smooth ambiguity model. The same approach is taken to evaluate the utility of Y. The resulting aggregate utilities will be compared with the utilities $\gamma(Q_R)$ and $\gamma(Q_C)$ to make a comparison.

2. The ATE approach begins by assigning distributions Q_B and Q_Y to the acts B and Y, and then the utilities $\gamma(Q_B)$ and $\gamma(Q_Y)$ are computed and compared with $\gamma(Q_R)$ and $\gamma(Q_C)$. The distributions Q_B and Q_Y can be summarized by the two numbers $q_B = Q_B(\{6\})$ and $q_Y = Q_Y(\{5\})$. This approach does not assume Q_B and Q_Y are consistent with a probabilistic model. For example, it may hold that $q_B + q_Y < 2/3$ (e.g., when the DM is ambiguity averse) or $q_B + q_Y > 2/3$ (e.g., when the DM is

[1] Ellsberg [1961] used black balls rather than blue ones, but we use blue for better visual clarity in Figure 4.1.

ambiguity seeking). The distributions Q_B and Q_Y are constructed by aggregating the beliefs and attitudes toward ambiguity.

Both approaches are conceptually feasible and may yield the same preferences over $\{B, Y, R, C\}$ for some γ and ambiguity attitudes. Nevertheless, the ETA approach requires the (mental) computation of 61 models for each of B and Y, whereas the ATE approach only requires determining the values q_B and q_Y. For instance, suppose $q_B = 0.25$ and $q_Y = 0.28$ in the ATE approach. If γ is the expected value (i.e., risk neutral) then we have $B \succ Y \succ R \succ C$; if γ is the expected utility with the utility function $x \mapsto x^{1/2}$ on $\mathbb{R}_+$ then we have $C \succ R \succ Y \succ B$. In this example, the ATE approach seems conceptually simpler and more natural, though we cannot assert with certainty which approach better describes a real individual's thought process.

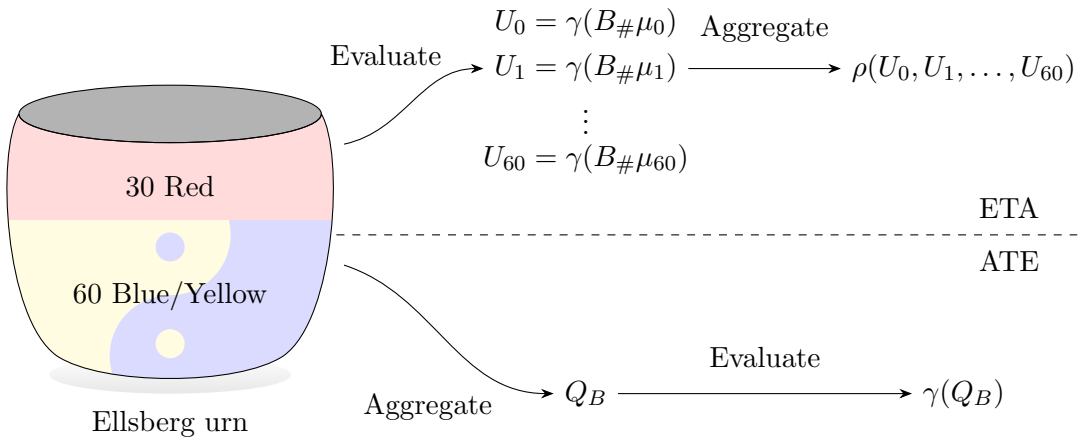


Figure 4.1: The ETA approach vs the ATE approach for the utility of act B

The aim of this chapter is to make explicit these two different frameworks in the Savage setting, where acts are viewed as mappings from states of the world to real-valued outcomes. Since the ATE framework is new, we will focus on it exclusively in the second half of the chapter. The main axiomatic work of this chapter will be to characterize the Choquet ATE model. To provide sufficient structure for this axiomatization, we will work in a refined Savage setting with an exogenously given source of unambiguous events. This setting is beneficial because it provides the proper conceptual structure for defining matching probabilities; see Dimmock et al. [2016] and Baillon et al. [2018], which are a major component of the axiomatization, and it allows us to explicitly study the interplay between risk attitudes and ambiguity attitudes.

We begin in Section 4.2.1 by reviewing the mathematical tools that will be used throughout the chapter. In Section 4.2.2, we formally present the Savage setting. In Section 4.2.3, we formulate the pure-risk setting and provide the mathematical tools for distributions that will support all later discussions of the risk component of preferences. In Section 4.3.1, we present the ETA model in the Savage setting. In Section 4.3.2, we do the same with the ATE model. Furthermore, Theorem 4.1 provides a novel result characterizing when a mapping assigning acts to distributions is given by (4.1). In Section 4.4, we introduce a refined Savage setting and provide the axioms and the characterization of the Choquet ATE model in Theorem 4.2. Notably, we find that the ambiguity component of the preferences is fully determined by the matching probability. Section 4.5 analyzes various ambiguity attitudes for the Choquet ATE model and their interplay with risk attitudes. We start in Section 4.5.1 by studying various risk attitudes. In Theorem 4.3, we fully characterize the propensity for risk diversification, which will help study its interplay with ambiguity. In Section 4.5.2, we discuss comparative and absolute ambiguity attitudes and confirm that we get a characterization of these properties through matching probabilities. In Section 4.5.3, we analyze the overall propensity for diversification. We first demonstrate that the propensity for risk diversification alone does not necessarily imply the overall propensity for diversification. Based on a comonotonic sum inequality, we establish a property for mappings that assign acts to distributions, given in (4.17), that guarantees (in the case of the ATE model) the propensity for risk diversification implies the overall propensity for diversification. Finally, Theorem 4.5 establishes that the mapping defined in (4.1) satisfies condition (4.17) if and only if ν is supermodular. In Section 4.5.4, we show how some of the ideas from Section 4.5.3 give rise to a distributionally robust representation in Theorem 4.6. We conclude the chapter in Section 4.6.

4.2 Preliminaries

4.2.1 Standard definitions and notation

For a given measurable space $(\Theta, \mathcal{E})$, a function $\nu : \mathcal{E} \mapsto [0,1]$ is a *capacity* if $\nu(\varnothing) = 0$, $\nu(\Theta) = 1$, and $\nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{E}$ with $A \subseteq B$. A capacity ν is said to be: *upward continuous* if for all increasing sequences $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ it holds that $\lim_{n\to\infty} \nu(A_n) = \nu\left(\bigcup_{n=1}^{\infty} A\right)$, *downward continuous* if for all decreasing sequences $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ it holds that $\lim_{n\to\infty} \nu(A_n) = \nu\left(\bigcap_{n=1}^{\infty} A_n\right)$, and *continuous* if it is upward and downward continuous. A capacity is called a *probability measure* if for all pairwise disjoint sequences $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ it holds that $\nu(\bigcup_{n=1}^{\infty} A_n) = \sum_{k=1}^{\infty} \nu(A_n)$. We denote the set of probability measures as $\mathcal{M}_1(\Theta, \mathcal{E})$.

For a continuous capacity ν , the *core of* ν, denoted by $\mathfrak{C}(\nu)$, is defined as

$$\mathfrak{C}(\nu) = \{\mu \in \mathcal{M}_1(\Theta, \mathcal{E}) : \mu(A) \geq \nu(A) \text{ for all } A \in \mathcal{E}\}.$$

The following are common properties of a continuous capacity ν.

Supermodularity: For all $A, B \in \mathcal{E}$, $\nu(A) + \nu(B) \leq \nu(A \cup B) + \nu(A \cap B)$.

Submodularity: For all $A, B \in \mathcal{E}$, $\nu(A) + \nu(B) \geq \nu(A \cup B) + \nu(A \cap B)$.

Balanced: It holds that $\mathfrak{C}(\nu) \neq \varnothing$.

Exactness: For all $A \in \mathcal{E}$, $\nu(A) = \min_{\mu \in \mathfrak{C}(\nu)} \mu(A)$.

For more on the core and the previous properties of capacities, see Denneberg [1994] and Marinacci and Montrucchio [2004].

Denote the set of real-valued bounded measurable functions on $(\Theta, \mathcal{E})$ by $\mathcal{X}(\Theta, \mathcal{E})$. Constant acts in $\mathcal{X}(\Theta, \mathcal{E})$ are identified with constants in $\mathbb{R}$. For $\eta, \zeta \in \mathcal{X}(\Theta, \mathcal{E})$, we write $\eta \geq \zeta$ when $\eta(s) \geq \zeta(s)$ for all $s \in \Theta$. Given a capacity ν, the *Choquet integral* with respect to ν is given by

$$\int \eta \mathrm{d}\nu = \int_0^\infty \nu(\eta > x)\mathrm{d}x + \int_{-\infty}^0 (\nu(\eta > x) - 1)\,\mathrm{d}x, \quad \eta \in \mathcal{X}(\Theta, \mathcal{E}).$$

We say that $\eta, \zeta \in \mathcal{X}$ are *comonotonic* if for all $s, s' \in \Theta$, $(\eta(s) - \eta(s'))(\zeta(s) - \zeta(s')) \geq 0$. As shown by Schmeidler [1986], a function $I : \mathcal{X}(\Theta, \mathcal{E}) \to \mathbb{R}$ can be represented as $I(\eta) = \int_\Omega \eta \mathrm{d}\nu$ for some capacity ν if and only if I is monotone ($I(\eta) \geq I(\zeta)$ whenever $\eta \geq \zeta$), comonotonic additive ($I(\eta+\zeta) = I(\eta)+I(\zeta)$ for comonotonic η, ζ), and normalized ($I(1) = 1$).

4.2.2 The Savage setting

Throughout the chapter, we consider a measurable space $(\Omega, \mathcal{F})$ representing the set of possible future states of the world. An element of the set $\mathcal{X}(\Omega, \mathcal{F})$ is called an *act*, and we write $\mathcal{X} = \mathcal{X}(\Omega, \mathcal{F})$ for brevity. Also, we will denote the set $\mathcal{M}_1(\Omega, \mathcal{F})$, also called the set of models, by Δ. Given $X \in \mathcal{X}$, we write $\|X\| = \sup_{\omega \in \Omega} |X(\omega)|$. A set $\mathcal{Y} \subseteq \mathcal{X}$ is said to be *bounded* if $\sup_{X \in \mathcal{Y}} \|X\| < \infty$ and *B-closed* if, for all bounded $(X_n)_{n \in N} \subseteq \mathcal{Y}$ satisfying $X_n \to X$ pointwise, it follows that $X \in \mathcal{Y}$.

Given any set $\mathcal{S}$, a *preference relation on* $\mathcal{S}$ is a total preorder $\succsim$; that is, (i) for all $\zeta, \eta \in \mathcal{S}$, $\zeta \succsim \eta$ or $\eta \succsim \zeta$; (ii) for all $\zeta, \eta, \xi \in \mathcal{X}$, if $\zeta \succsim \eta$ and $\eta \succsim \xi$ then $\zeta \succsim \eta$. The

Savage setting concerns preference relations on $\mathcal{X}$, where $\mathcal{X}$ represents the DM's set of choice objects, and the preference relation describes how the DM ranks them. Throughout the chapter, when we mention the preference relation without mentioning the underlying set $\mathcal{S}$, it should be understood as $\mathcal{S} = \mathcal{X}$.

In what follows, we introduce an axiom that will later be used in the axiomatization of a decision model under ambiguity in Section 4.4. We present it now because it will also be referenced in Section 4.3.

(M) For all $X, Y \in \mathcal{X}$, if $X \geq Y$ then $X \succsim Y$.

Axiom (M) is standard for preference relations in the Savage setting. If the act X pays more than the act Y for every state of the world, then it is natural that the DM would prefer X to Y.

4.2.3 The pure-risk setting, distributions, and statistics

Given a Borel probability measure Q on $\mathbb{R}$, the *support of* Q, denoted by $\text{supp}(Q)$, is the smallest closed set $C \subseteq \mathbb{R}$ satisfying $Q(C) = 1$. Denote by $\mathcal{D}$ the set of all Borel probability measures on $\mathbb{R}$ with compact support. We will refer to $\mathcal{D}$ as the set of distributions. In decision theory, the study of preference relations on distributions is known as the pure-risk setting.

Many decision models in the Savage setting begin by reducing the problem to the pure-risk setting, e.g., the SEU model. This reduction is carried out as follows. Suppose the DM does not perceive any ambiguity and describes their uncertainty with the unique probability measure $\mu \in \Delta$. Given any act $X \in \mathcal{X}$, the DM may represent this act with the distribution given by the pushforward probability measure $X_{\#}\mu$ defined by

$$X_{\#}\mu(B) = \mu(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel-σ-algebra on $\mathbb{R}$ and X^{-1} denote the preimage of X. This process allows the DM to move from the Savage setting to the pure-risk setting. Furthermore, if the DM has a preference relation on distributions $\succsim^{\text{d}}$, the DM is able to induce a preference relation $\succsim$ by

$$X \succsim Y \iff X_{\#}\mu \succsim^{\text{d}} Y_{\#}\mu, \quad X, Y \in \mathcal{X}. \tag{4.2}$$

We call preference relations satisfying (4.2) *probabilistically sophisticated*, slightly relaxing the original definition of Machina and Schmeidler [1992] via stochastic dominance.

A collection $\mathcal{C} \subseteq \mathcal{D}$ is said to have *uniformly bounded support* if there exists a compact set that contains the support of each $Q \in \mathcal{C}$. Given $Q \in \mathcal{D}$, we define the *survival function of* Q by $S_Q(x) = Q((x, \infty))$ for all $x \in \mathbb{R}$ and the *quantile function of* Q by

$$q_Q(\alpha) = \inf\{x \in \mathbb{R} : S_Q(x) < 1 - \alpha\}, \quad \alpha \in (0, 1);$$

the mapping $\mathrm{VaR}_\alpha : Q \mapsto q_Q(\alpha)$ is called the *Value-at-Risk of* Q at level α in finance. Given an increasing (always in the non-strict sense) function $f : \mathbb{R} \to \mathbb{R}$ and $Q \in \mathcal{D}$, we define the distribution $f_\# Q$ by

$$f_\# Q(B) = Q(f^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}),$$

where f^{-1} denote the preimage of f. It holds that $f_\# Q \in \mathcal{D}$ since f is increasing. Given $(Q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ and $Q \in \mathcal{D}$, we say $Q_n \to Q$ *in distribution* if $\lim_{n\to\infty} \int_{\mathbb{R}} f \mathrm{d}Q_n = \int_{\mathbb{R}} f \mathrm{d}Q$ for all continuous and bounded $f : \mathbb{R} \to \mathbb{R}$.

For $Q, P \in \mathcal{D}$, we write $Q \geq_{\mathrm{fsd}} P$ if $q_Q(\alpha) \geq q_P(\alpha)$ for all $\alpha \in (0, 1)$, $Q \geq_{\mathrm{ssd}} P$ if

$$\int_{\mathbb{R}} f \mathrm{d}Q \geq \int_{\mathbb{R}} f \mathrm{d}P \tag{4.3}$$

for all increasing concave functions $f : \mathbb{R} \to \mathbb{R}$, and $Q \geq_{\mathrm{cv}} P$ if (4.3) holds for all concave functions $f : \mathbb{R} \to \mathbb{R}$. Each of $\geq_{\mathrm{fsd}}$, $\geq_{\mathrm{ssd}}$, and $\geq_{\mathrm{cv}}$ defines a partial order on $\mathcal{D}$. For a partial order $\geq$ on $\mathcal{D}$, $\mathcal{C} \subseteq \mathcal{D}$, and $Q \in \mathcal{D}$, we say $\mathcal{C} \geq Q$ if for all $P \in \mathcal{C}$, $P \geq Q$. For $\mathcal{C} \subseteq \mathcal{D}$, the *first-order infimum of* $\mathcal{C}$, denoted by $\bigwedge_{\mathrm{fsd}} \mathcal{C}$, is defined as the distribution Q satisfying $P \geq_{\mathrm{fsd}} Q \geq_{\mathrm{fsd}} P'$ for all $P \in \mathcal{C}$ and $P' \in \mathcal{D}$ satisfying $\mathcal{C} \geq_{\mathrm{fsd}} P'$. Remark that if Q exists, it is unique. If $\mathcal{C} \subseteq \mathcal{D}$ has uniformly bounded support, then $\bigwedge_{\mathrm{fsd}} \mathcal{C}$ exists (see e.g., Mao et al., 2025).

In this chapter, a *statistic* is a mapping $\gamma : \mathcal{D} \to \mathbb{R}$ such that for all $Q, P \in \mathcal{D}$,

$$Q >_{\mathrm{fsd}} P \implies \gamma(Q) > \gamma(P),$$

and for all sequences $(Q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ with uniformly bounded support converging in distribution to some $Q \in \mathcal{D}$, it holds that $\gamma(Q_n) \to \gamma(Q)$. A statistic γ is said to be a *certainty-equivalent statistic* if $\gamma(\delta_c) = c$ for all $c \in \mathbb{R}$, where δ_c denotes the Dirac measure at c. The two defining properties of a statistic are natural, being nothing more than monotonicity and a mild continuity property, called compact continuity, studied by Chew and Mao [1995]. Statistics can be used to represent preference relations over distributions.

That is, given the statistic γ, we can define a preference relation on distributions $\succsim^{\mathrm{d}}$ by

$$Q \succsim^{\mathrm{d}} P \iff \gamma(Q) \geq \gamma(P), \quad Q, P \in \mathcal{D}. \tag{4.4}$$

Proposition 4.1. *For a given $\succsim^{\mathrm{d}}$ in (4.4), the statistic γ is ordinally unique, and there exists a unique certainty-equivalent statistic γ satisfying (4.4).*

The following are examples of statistics commonly used in decision theory. (a) The expected value statistic

$$\mathbb{E}(Q) = \int_{\mathbb{R}} x Q(\mathrm{d}x), \quad Q \in \mathcal{D}.$$

(b) The expected utility model of von Neumann and Morgenstern [1947] corresponds to the statistic

$$\mathbb{E}_u(Q) = \int_{\mathbb{R}} u \mathrm{d}Q = \mathbb{E}(u_{\#}Q), \quad Q \in \mathcal{D},$$

where $u : \mathbb{R} \to \mathbb{R}$ is a strictly increasing continuous function, which we call a *von Neumann–Morgenstern (vNM) utility function*. (c) The dual utility model of Yaari [1987] corresponds to the statistic

$$\mathbb{E}^g(Q) = \int_0^\infty g(S_Q(x))\mathrm{d}x + \int_{-\infty}^0 (g(S_Q(x)) - 1)\mathrm{d}x, \quad Q \in \mathcal{D},$$

where $g : [0,1] \to [0,1]$ is a strictly increasing continuous function, which we call a *distortion function*. (d) The rank-dependent utility model of Quiggin [1982] corresponds to the statistic

$$\mathbb{E}^g_u(Q) = \mathbb{E}^g\left(u_{\#}Q\right), \quad Q \in \mathcal{D}, \tag{4.5}$$

where u is a vNM utility function and g is a distortion function. Note that for $\mu \in \Delta$ and $X \in \mathcal{X}$,

$$\mathbb{E}^g_u(X_{\#}\mu) = \int_0^\infty g(\mu(u(X) > x))\mathrm{d}x + \int_{-\infty}^0 (g(\mu(u(X) > x)) - 1)\mathrm{d}x = \int_\Omega u(X)\mathrm{d}(g \circ \mu),$$

which is how rank-dependent utility models are usually formulated.

Remark 4.1. Since, in general, the above statistics are not continuous with respect to convergence in distribution, the weaker notion of compact continuity is essential to our study.

4.3 Two frameworks for separating pure risk and ambiguity

Both the ATE and ETA frameworks provide preference relations that clearly distinguish between pure risk and ambiguity. In both frameworks, we will assume that the DM represents the risk component of their preferences with a preference relation $\succsim^{\mathrm{d}}$ on $\mathcal{D}$. Additionally, we will assume that $\succsim^{\mathrm{d}}$ can be represented by a statistic as in (4.4). Where the frameworks differ is how they define and incorporate the ambiguity component of the preferences. We do not suggest that these two frameworks are the only ones from which ambiguity models can be constructed. We nevertheless find their juxtaposition conceptually compelling, which motivates the organization of this section.

4.3.1 The evaluate-then-aggregate framework

We equip the set of models Δ with the weak* topology, defined as the coarsest topology under which every map $\mu \mapsto \mu(A)$ is continuous for all $A \in \mathcal{F}$. It is common in decision theory to equip the space of models with the weak* topology, and no familiarity with this topology is needed for the discussion that follows. We denote by Σ the corresponding Borel σ-algebra.

In this framework, to describe the ambiguity component of the preferences, we will use the formulation proposed in Klibanoff et al. [2005] to axiomatize the smooth ambiguity model. In this framework, the DM also has a preference relation $\succsim^{\mathrm{m}}$ on $\mathcal{X}(\Delta, \Sigma)$, which is hereby called a preference relation on model-based acts. We will assume that $\succsim^{\mathrm{m}}$ satisfies the following regularity properties.

(M1) For all $\eta, \zeta \in \mathcal{X}(\Delta, \Sigma)$, if $\eta \geq \zeta$ then $\eta \succsim^{\mathrm{m}} \zeta$.

(M2) For all $\eta \in \mathcal{X}(\Delta, \Sigma)$, $\{c \in \mathbb{R} : \eta \succsim^{\mathrm{m}} c\}$ and $\{c \in \mathbb{R} : c \succsim^{\mathrm{m}} \eta\}$ are closed.

(M3) For $a, b \in \mathbb{R}$, if $a > b$ then $a \succ^{\mathrm{m}} b$.

Proposition 4.2. *Under Properties (M1)–(M3), there exists $\rho : \mathcal{X}(\Delta, \Sigma) \to \mathbb{R}$ with $\rho(\eta) \geq \rho(\zeta)$ for all $\eta, \zeta \in \mathcal{X}(\Delta, \Sigma)$ satisfying $\eta \geq \zeta$, and $\rho(a) > \rho(b)$ for all $a, b \in \mathbb{R}$ satisfying $a > b$, such that*

$$\eta \succsim^{\mathrm{m}} \zeta \iff \rho(\eta) \geq \rho(\zeta), \quad \eta, \zeta \in \mathcal{X}(\Delta, \Sigma). \tag{4.6}$$

Using the preference relation on distributions $\succsim^{\mathrm{d}}$, acts can be naturally transformed into model-based acts, as described below. Start by fixing the unique certainty-equivalent statistic γ that represents $\succsim^{\mathrm{d}}$ (Proposition 4.1). Given $X \in \mathcal{X}$, consider the mapping $\mathfrak{R}_X : \Delta \to \mathcal{D}$ defined by

$$\mathfrak{R}_X(\mu) = X_{\#}\mu, \quad \mu \in \Delta.$$

The mapping $\mathfrak{R}_X$ allows us to define the model-based act $\gamma \circ \mathfrak{R}_X$. We show in Section 4.9 that $\gamma \circ \mathfrak{R}_X \in \mathcal{X}(\Delta, \Sigma)$ for each $X \in \mathcal{X}$. Intuitively, given $X \in \mathcal{X}$ and $\mu \in \Delta$, the value $\gamma \circ \mathfrak{R}_X(\mu)$ can be seen as an evaluation, determined by $\succsim^{\mathrm{d}}$, of the distribution $X_{\#}\mu$.

To couple the risk component $\succsim^{\mathrm{d}}$ and the ambiguity component $\succsim^{\mathrm{m}}$ into a preference relation $\succsim$, we assume that the DM's preference relation satisfies

$$X \succsim Y \iff \gamma \circ \mathfrak{R}_X \succsim^{\mathrm{m}} \gamma \circ \mathfrak{R}_Y, \quad X, Y \in \mathcal{X}.$$

Therefore, the representation functional ρ in (4.6) aggregates the evaluation maps as

$$X \succsim Y \iff \rho(\gamma \circ \mathfrak{R}_X) \geq \rho(\gamma \circ \mathfrak{R}_Y), \quad X, Y \in \mathcal{X}. \tag{4.7}$$

We refer to any preference relation $\succsim$ satisfying (4.7) as an *ETA preference relation*. It is clear that any ETA preference relation satisfies Axiom (M). Additionally, if we take

$$\rho(\eta) = \eta(\mu), \quad \eta \in \mathcal{X}(\Delta, \Sigma),$$

for some $\mu \in \Delta$, then the ETA preference relation in (4.7) is probabilistically sophisticated.

Many standard decision models under ambiguity can be formulated as ETA preference relations, and we list four below. Most of these models were first developed and axiomatized in the setting of Anscombe and Aumann [1963]; here we translate them to the Savage setting. In each example, we use the preference relation on distributions derived from the statistic $\mathbb{E}_u$ for some vNM utility u, which corresponds to the expected utility model of von Neumann and Morgenstern [1947]. Remark that the certainty-equivalent statistic associated with $\mathbb{E}_u$ is the functional $u^{-1} \circ \mathbb{E}_u$, where $u^{-1} : \mathrm{im}(u) \to \mathbb{R}$ is the inverse of u. (a) The maxmin expected utility model of Gilboa and Schmeidler [1989] corresponds to

$$\rho(\eta) = \inf_{\mu \in \mathcal{Q}} u(\eta(\mu)), \quad \eta \in \mathcal{X}(\Delta, \Sigma),$$

where $\mathcal{Q} \subseteq \Delta$. (b) The smooth ambiguity model of Klibanoff et al. [2005] corresponds to

$$\rho(\eta) = \int_{\Delta} v(\eta) \mathrm{d}\pi, \quad \eta \in \mathcal{X}(\Delta, \Sigma),$$

where v is a vNM utility, and $\pi \in \mathcal{M}_1(\Delta, \Sigma)$. (c) The variational preference model of Maccheroni et al. [2006] corresponds to

$$\rho(\eta) = \inf_{\mu \in \Delta} \left(u(\eta(\mu)) + c(\mu) \right), \quad \eta \in \mathcal{X}(\Delta, \Sigma),$$

where $c : \Delta \to [0, \infty]$ satisfies $c(\mu) < \infty$ for some $\mu \in \Delta$. (d) The α-maxmin model of Ghirardato et al. [2004] corresponds to

$$\rho(\eta) = \alpha \inf_{\mu \in \mathcal{Q}} u(\eta(\mu)) + (1 - \alpha) \sup_{\mu \in \mathcal{Q}} u(\eta(\mu)), \quad \eta \in \mathcal{X}(\Delta, \Sigma),$$

where $\mathcal{Q} \subseteq \Delta$ and $\alpha \in [0, 1]$.

In the following section, we propose an alternative to the ETA framework.

4.3.2 The aggregate-then-evaluate framework

In the ATE framework, to describe the ambiguity component of the preferences, the DM fixes a mapping $\mathfrak{D} : \mathcal{X} \to \mathcal{D}$, which we call an *act-to-distribution mapping*. Intuitively, the DM uses an act-to-distribution mapping to aggregate their ambiguity regarding an act into a single distribution. Given $\mu \in \Delta$, the *probabilistic act-to-distribution mapping*, denoted by $\mathfrak{D}_\mu$, is defined by $\mathfrak{D}_\mu(X) = X_{\#}\mu$ for all $X \in \mathcal{X}$. Probabilistic act-to-distribution mappings are used when the DM does not perceive ambiguity, or is indifferent to it. When ambiguity is perceived by, and matters to, the DM, the chosen act-to-distribution mapping should differ from all probabilistic ones. The following example clarifies this idea.

Example 4.1. Assume the DM believes the models $\mathcal{Q} \subseteq \Delta$ are relevant to the decision-making process. We can define the act-to-distribution mapping $\mathfrak{D}_\mathcal{Q}$ by

$$\mathfrak{D}_\mathcal{Q}(X) = \bigwedge_{\text{fsd}} X_{\#}\mathcal{Q}, \quad X \in \mathcal{X},$$

where $X_{\#}\mathcal{Q} = \{X_{\#}\mu : \mu \in \mathcal{Q}\}$. To see this mapping is well-defined, given $X \in \mathcal{X}$, find $a, b \in \mathbb{R}$ such that $a \geq X \geq b$. For all $\mu \in \Delta$, since $\mu(a \geq X \geq b) = 1$, we have $\text{supp}(X_{\#}\mu) \subset [b, a]$. Therefore, the set $\mathcal{X}_{\#}\mathcal{Q}$ has uniformly bounded support and $\bigwedge_{\text{fsd}} X_{\#}\mathcal{Q}$ exists. As long as $\mathcal{Q}$ is not a singleton, the act-to-distribution mapping $\mathfrak{D}_\mathcal{Q}$ does not coincide with any probabilistic one. As we will see later, $\mathfrak{D}_\mathcal{Q}$ represents a form of ambiguity aversion.

The following are natural properties for act-to-distribution mappings, and they are

all satisfied by probabilistic act-to-distribution mappings. The first two properties are straightforward and need no further motivation.

(R1) For all $X, Y \in \mathcal{X}$, if $X \geq Y$ then $\mathfrak{D}(X) \geq_{\text{fsd}} \mathfrak{D}(Y)$.

(R2) For all bounded $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$, if $X_n \to X$ pointwise then $\mathfrak{D}(X_n) \to \mathfrak{D}(X)$ in distribution.

To motivate the final property, suppose a DM has an act-to-distribution mapping $\mathfrak{D}$, and let $f : \mathbb{R} \to \mathbb{R}$ be strictly increasing and continuous, interpreted as the map that transforms monetary units to f-units. For any act $X \in \mathcal{X}$, the DM can induce a distribution over f-units in two ways. First, they obtain the monetary distribution $\mathfrak{D}(X)$ and then transform this distribution via f, yielding $f_{\#}(\mathfrak{D}(X))$. Alternatively, they may first transform the act by applying f to obtain the f-unit-valued act $f(X)$ and then apply $\mathfrak{D}$ to obtain the distribution $\mathfrak{D}(f(X))$. The following property requires that these two procedures yield the same distribution.

(R3) For all $X \in \mathcal{X}$, if $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing and continuous, then $\mathfrak{D}(f(X)) = f_{\#}(\mathfrak{D}(X))$.

To couple the risk component $\succsim^{\text{d}}$ and the ambiguity component $\mathfrak{D}$ into a preference relation $\succsim$, the DM starts by fixing a statistic γ that represents $\succsim^{\text{d}}$. The preference relation $\succsim$ of the DM can be represented by

$$X \succsim Y \iff \gamma \circ \mathfrak{D}(X) \geq \gamma \circ \mathfrak{D}(Y), \quad X, Y \in \mathcal{X}. \tag{4.8}$$

We refer to any preference relation $\succsim$ given by (4.8) for any statistic $\gamma : \mathcal{D} \to \mathbb{R}$ and any act-to-distribution mapping $\mathfrak{D} : \mathcal{X} \to \mathcal{D}$ as an *ATE preference relation*. It is clear that any preference relation $\succsim$ satisfying (4.8), with $\mathfrak{D}$ satisfying Property (R1), must also satisfy Axiom (M). Additionally, if we take $\mathfrak{D}$ to be a probabilistic act-to-distribution mapping, then the preference relation $\succsim$ is probabilistically sophisticated.

In what follows, we present a method for constructing act-to-distribution mappings inspired by Sarin and Wakker [1994]. The key distinction is that our definition applies to all acts, rather than being restricted to those with a finite range. Given a continuous capacity ν, the *Choquet act-to-distribution mapping*, denoted by $\mathfrak{D}_\nu$, is defined by

$$S_{\mathfrak{D}_\nu(X)}(x) = \nu(X > x) \text{ for all } x \in \mathbb{R}.$$

The mapping $\mathfrak{D}_\nu$ is well-defined because ν is continuous and survival functions uniquely characterize elements of $\mathcal{D}$. It is straightforward to show that if ν is a probability measure, then $\mathfrak{D}_\nu$ is a probabilistic act-to-distribution mapping. Also, by the well-known layer-cake representation, we have

$$\begin{aligned}\mathbb{E} \circ \mathfrak{D}_\nu(X) &= \int_0^\infty S_{\mathfrak{D}_\nu(X)}(x)\mathrm{d}x + \int_{-\infty}^0 (S_{\mathfrak{D}_\nu(X)}(x) - 1)\mathrm{d}x \\ &= \int_0^\infty \nu(X > x)\mathrm{d}x + \int_{-\infty}^0 (\nu(X > x) - 1) = \int_\Omega X \mathrm{d}\nu, \quad X \in \mathcal{X}.\end{aligned}$$

Preference relations $\succsim$ satisfying (4.8) with $\mathfrak{D}$ given by a Choquet act-to-distribution mapping are said to be *Choquet ATE* preference relations. We have the following theorem characterizing all Choquet act-to-distribution mappings.

Theorem 4.1. *Let $\mathfrak{D}$ be an act-to-distribution mapping. Then $\mathfrak{D}$ satisfies Properties (R1)–(R3) if and only if $\mathfrak{D}$ is a Choquet act-to-distribution mapping.*

Theorem 4.1 is related to Liu et al. [2021, Theorem 1], which characterizes transformations $T : \mathcal{D} \to \mathcal{D}$ satisfying

$$S_{T(Q)} = g \circ S_Q, \quad Q \in \mathcal{D}$$

for a distortion function g, with a property similar to (R3). The main difference between the two results is that we consider mappings defined on acts, whereas Liu et al. [2021] studied transformations defined on distributions.

Given a continuous capacity ν and a vNM utility u, by Theorem 4.1, we have

$$\mathbb{E}_u \circ \mathfrak{D}_\nu(X) = \mathbb{E}(u_\# \mathfrak{D}_\nu(X)) = \mathbb{E} \circ \mathfrak{D}_\nu(u(X)) = \int_\Omega u(X)\mathrm{d}\nu, \quad X \in \mathcal{X}.$$

Therefore, preference relations given by the Choquet expected utility model of Schmeidler [1989] with a continuous capacity are examples of Choquet ATE preference relations.

4.4 Axioms and representation for the Choquet ATE model

We next state and discuss several additional axioms for preference relations $\succsim$, which lead to a main representation result in Theorem 4.2.

For the rest of this chapter, we will work in a Savage setting with an exogenously given family $\mathcal{G} \subseteq \mathcal{F}$ of unambiguous events. Intuitively, this assumption helps us identify the influence of pure risk on preference relations and disentangle its effects from those of ambiguity. The set of unambiguous events will form a λ-system, as stated in the following assumption.

Assumption 4.1. The set of unambiguous events $\mathcal{G}$ is a *λ-system*. That is:

(i) $\Omega \in \mathcal{G}$;

(ii) if $A \in \mathcal{G}$ then $A^c \in \mathcal{G}$;

(iii) if the sequence $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{G}$ is pairwise disjoint then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$.

The use of λ-systems for unambiguous events is common in the literature; see Zhang [2002], Marinacci [2002], and Epstein [2004]. Note that a σ-algebra is always a λ-system, but the converse is not true. Compared with σ-algebras, a λ-system lacks the closure property under intersection. That is, the intersection of two unambiguous events may be ambiguous. For instance, in a bivariate model, the marginal distributions may be unambiguous but the joint distribution can still be ambiguous. Therefore, we deliberately avoid the stronger assumption that $\mathcal{G}$ is a sub-σ-algebra.

Intuitively, an event is unambiguous if the DM can assign it a unique likelihood; here we use likelihood rather than probability because we do not have a σ-algebra. Property (i) reflects that the DM can always assess the likelihood of the whole space. Property (ii) says that if $A \in \mathcal{F}$ is unambiguous with the likelihood p_A, then its complement A^c is also unambiguous with the assigned likelihood $1 - p_A$. Finally, Property (iii) is motivated by the fact that the likelihood of the countable union of pairwise disjoint events should equal the sum of their individual likelihoods.

Given the interpretations of $\mathcal{G}$ in Assumption 4.1, we can associate a likelihood mapping to $\mathcal{G}$, which acts like a probability measure.

Assumption 4.2. There exists a function $\mathbb{P} : \mathcal{G} \to [0, 1]$ such that $\mathbb{P}(\Omega) = 1$ and for every pairwise disjoint sequence $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{G}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

We call $\mathbb{P}$ the *exogenous likelihood mapping*.

When $\mathcal{G}$ is a σ-algebra, the exogenous likelihood mapping $\mathbb{P}$ is a probability measure on $\mathcal{G}$. We will refer to this setting, consisting of Assumptions 4.1 and 4.2, as the *Savage setting with pure risk*.

We write $\mathcal{X}(\mathcal{G})$ for the subset of $\mathcal{X}$ consisting of all acts X such that $X^{-1}(B) \in \mathcal{G}$ for every $B \in \mathcal{B}(\mathbb{R})$. By the definition of the exogenous likelihood mapping, each $X \in \mathcal{X}(\mathcal{G})$ induces a distribution $X_{\#}\mathbb{P} \in \mathcal{D}$ via

$$X_{\#}\mathbb{P}(B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

Although the exogenous likelihood mapping is not a probability measure in the usual sense, as $\mathcal{G}$ is not a σ-algebra, the induced object $X_{\#}\mathbb{P}$ is a compactly supported Borel probability measure on $\mathbb{R}$ for all $X \in \mathcal{X}(\mathcal{G})$. Therefore, any partial order $\geq$ on $\mathcal{D}$ induces a preorder $\geq^{\mathbb{P}}$ on $\mathcal{X}(\mathcal{G})$ by

$$X \geq^{\mathbb{P}} Y \iff X_{\#}\mathbb{P} \geq Y_{\#}\mathbb{P}, \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

Under the Savage setting with pure risk, since acts $X \in \mathcal{X}(\mathcal{G})$ exhibit no distributional ambiguity, the preference relation $\succsim$ restricted to $\mathcal{X}(\mathcal{G})$ should be determined solely by the distributions of these acts under the exogenous likelihood mapping. This brings us to our second axiom of risk conformity.

(RC) Risk conformity: For all $X, Y \in \mathcal{X}(\mathcal{G})$, if $X =^{\mathbb{P}}_{\mathrm{fsd}} Y$ then $X \simeq Y$.

Baillon et al. [2025] and Shen et al. [2026] studied Axiom (RC) for general decision models and risk measures when $\mathcal{G}$ is further assumed to be a σ-algebra; we only need $\mathcal{G}$ to be a λ-system. If the preference relation $\succsim$ satisfies Axiom (RC), then the restriction of $\succsim$ to $\mathcal{X}(\mathcal{G})$, called the *pure-risk preferences* of $\succsim$, corresponds to a preference relation over distributions.

We say that an act-to-distribution mapping $\mathfrak{D}$ is *risk conforming* if for all $X \in \mathcal{X}(\mathcal{G})$, it holds that $\mathfrak{D}(X) = X_{\#}\mathbb{P}$. Let $\succsim$ be an ATE preference relation. In the Savage setting with pure risk, to remain consistent with the interpretation of an act-to-distribution mapping, $\mathfrak{D}$ in (4.8) should be risk conforming. This follows as acts in $\mathcal{X}(\mathcal{G})$ have no distributional ambiguity. A capacity ν is *risk conforming* if $\nu|_{\mathcal{G}} = \mathbb{P}$. In this case, risk conformity means that, for unambiguous events, the capacity coincides with the exogenous likelihood mapping. When ν describes the DM's subjective non-additive evaluation regarding the likelihoods of events, risk conformity is essential, as the DM believes that $\mathbb{P}$ specifies the likelihood of unambiguous events. Let ν be a continuous capacity. It is straightforward to show that ν is risk conforming if and only if the Choquet act-to-distribution mapping $\mathfrak{D}_\nu$ is risk conforming. Finally, to allow for probabilistically sophisticated preference relations

that are consistent with the exogenous likelihood mapping, we will make the following minor structural assumption.

Assumption 4.3. There exists a risk-conforming $\mu \in \Delta$.

Assumption 4.3 means that one could find a probability model μ that rationalizes the behavior of the DM on $\mathcal{G}$. This allows for the existence of probabilistically sophisticated preferences with a belief agreeing with $\mathbb{P}$ on $\mathcal{X}(\mathcal{G})$.

The next two axioms are regularity requirements.

(SRM) Strict risk monotonicity: For all $X, Y \in \mathcal{X}(\mathcal{G})$, if $X >^{\mathbb{P}}_{\text{fsd}} Y$ then $X \succ Y$.

(C) Continuity: For all $X \in \mathcal{X}$, $\{Y \in \mathcal{X} : X \succsim Y\}$ and $\{Y \in \mathcal{X} : Y \succsim X\}$ are B-closed.

Axiom (SRM) reflects the fact that the DM uses the exogenous likelihood mapping to associate distributions with acts in $\mathcal{X}(\mathcal{G})$. Thus, a stronger form of monotonicity holds for these acts using these distributions. Axiom (C) is a reformulation of the compact continuity in Chew and Mao [1995] for acts and is commonly used in the literature on risk measures; see Föllmer and Schied [2016].

The axioms discussed thus far concern only certain notions of regularity and essentially impose no specific functional forms on the preference relations.

The final conceptual tool we will need is matching probabilities, as defined below. We will write $A \succsim B$ for $A, B \in \mathcal{F}$ to mean $\mathbb{1}_A \succsim \mathbb{1}_B$, where $\mathbb{1}_A$ denotes the binary act which yields 1 if $\omega \in A$ and 0 otherwise.

Definition 4.1. Given a preference relation $\succsim$, a function $\nu : \mathcal{F} \to [0, 1]$ is a $\succsim$-*matching probability* if for all $A \in \mathcal{F}$, there exists $R_A \in \mathcal{G}$ such that $\nu(A) = \mathbb{P}(R_A)$ and $A \simeq R_A$.

Intuitively, a $\succsim$-matching probability uses the restriction of the preference relation $\succsim$ on the set $\{\mathbb{1}_A : A \in \mathcal{F}\}$ to describe ambiguity on the level of events. Specifically, it assigns to each event a "pseudo-likelihood" by equating it, under the preference $\succsim$, to an unambiguous event and adopting the likelihood of that event. This idea requires the Savage setting with pure risk, since the definition is only conceptually meaningful when $\mathcal{G}$ contains unambiguous events. Finally, to allow for arbitrary precision in the identification of matching probabilities, we will, henceforth, assume the following.

Assumption 4.4. There exists a collection of events $\{S_r\}_{r \in [0,1]} \subseteq \mathcal{G}$ such that $S_{r'} \supseteq S_r$ whenever $r' \geq r$, and $\mathbb{P}(S_r) = r$ for all $r \in [0, 1]$.

Remark 4.2. If $\mathcal{G}$ is a σ-algebra, the exogenous likelihood mapping $\mathbb{P}$ is a probability measure, and Assumption 4.3 is equivalent to the non-atomicity of $\mathbb{P}$. The assumption of non-atomicity is common in risk management.

As the following proposition shows, our regularity axioms imply the existence and uniqueness of matching probabilities.

Proposition 4.3. *For every preference relation $\succsim$ satisfying Axioms (M), (RC), (SRM), and (C), there exists a unique $\succsim$-matching probability ν. Furthermore, ν is a continuous risk-conforming capacity, and*

$$A \succsim B \iff \nu(A) \geq \nu(B), \quad A, B \in \mathcal{F}.$$

As we will see in Theorem 4.2, matching probabilities will fully determine the ambiguity component of the preferences. To discuss the risk component of the preferences, we have that these regularity axioms are sufficient to represent the pure-risk preferences via a statistic.

Proposition 4.4. *If the preference relation $\succsim$ satisfies Axioms (M), (RC), (SRM), and (C), then there exists a statistic γ such that*

$$X \succsim Y \iff \gamma(X_{\#}\mathbb{P}) \geq \gamma(Y_{\#}\mathbb{P}), \quad X, Y \in \mathcal{X}(\mathcal{G}). \tag{4.9}$$

Moreover, the statistic γ is ordinally unique, and there exists a unique certainty-equivalent statistic γ satisfying (4.9).

For any preference relation $\succsim$ satisfying Axioms (M), (RC), (SRM), and (C), we refer to this unique certainty-equivalent statistic in Proposition 4.4 as the $\succsim$-*CES*.

The final axiom, adapted from Sarin and Wakker [1994], is used to couple the risk component of preferences, represented by the $\succsim$-CES, with the ambiguity component, represented by the $\succsim$-matching probability. Before introducing it, we require some preliminary definitions. Given a preference relation $\succsim$, we can define the relation $\succsim^*$ given by

$$X \succsim^* Y \iff \{X > x\} \succsim \{Y > x\} \text{ for all } x \in \mathbb{R}, \quad X, Y \in \mathcal{X}.$$

This allows us to define our final axiom.

(CD) Cumulative dominance: For all $X, Y \in \mathcal{X}$, if $X \simeq^* Y$ then $X \simeq Y$.

The following result shows that the above axioms fully characterize Choquet ATE preference relations in the Savage setting with pure risk.

Theorem 4.2. *The preference relation $\succsim$ satisfies Axioms (M), (RC), (SRM), (C), and (CD) if and only if $\succsim$ is a Choquet ATE preference relation, that is, there exists a statistic γ and a risk-conforming capacity ν such that*

$$X \succsim Y \iff \gamma \circ \mathfrak{D}_\nu(X) \geq \gamma \circ \mathfrak{D}_\nu(Y). \tag{4.10}$$

Moreover, in (4.10), the statistic γ is ordinally unique, and the capacity ν is unique and coincides with the $\succsim$-matching probability.

Example 4.2. Let $\succsim$ be a Choquet ATE preference relation. Assume that γ is equal to the statistic $\mathbb{E}_u^g$ from (4.5), corresponding to the rank-dependent utility model. We have

$$\mathbb{E}_u^g \circ \mathfrak{D}_\nu(X) = \mathbb{E}^g(u_\# \mathfrak{D}_\nu(X)) = \mathbb{E}^g(\mathfrak{D}_\nu(u(X))) = \int_\Omega u(X) \mathrm{d}(g \circ \nu), \quad X \in \mathcal{X}.$$

In this ambiguity model, the DM begins by assigning non-additive likelihoods to events, captured by ν, which incorporates their beliefs and attitudes toward ambiguity. Afterwards, the pure-risk preferences, described by an RDU model, are incorporated by distorting this capacity via g, as in the case of pure risk. Finally, the DM evaluates acts $X \in \mathcal{X}$ via the Choquet integral of $u(X)$ with respect to the capacity $g \circ \nu$. This ambiguity model was originally introduced in Tversky and Fox [1995]. Additional axiomatizations for this specific model are given by Buchak [2025] and Van Oosten and Wang [2025a].

Example 4.3. Allowing for general statistics in Choquet ATE preference relations leads to new decision models under ambiguity. For example, given $a \in \mathbb{R}\backslash\{0\}$, let u_a be the vNM utility function defined by $u_a(x) = e^{ax}$. For $a \in \mathbb{R}$, define the certainty-equivalent statistic γ_a by

$$\gamma_a(Q) = \frac{1}{a} \log\left(\mathbb{E}_{u_a}(Q)\right) \text{ if } a \in \mathbb{R}\backslash\{0\}, \text{ and } \gamma_0(Q) = \mathbb{E}(Q), \quad Q \in \mathcal{D}.$$

The statistic γ_a is the entropic risk measure (see Föllmer and Schied, 2002). Given a Borel probability measure κ on $\mathbb{R}$, Mu et al. [2024] studied a statistic given by

$$\gamma_\kappa(Q) = \int_{\mathbb{R}} \gamma_a(Q) \kappa(\mathrm{d}a), \quad Q \in \mathcal{D}.$$

If $\succsim$ is a Choquet ATE preference relation with the $\succsim$-CES given by γ_κ, then

$$X \succsim Y \iff \int_{\mathbb{R}} \frac{1}{a} \log \left(\int_\Omega e^{aX} \mathrm{d}\nu \right) \kappa(\mathrm{d}a) \geq \int_{\mathbb{R}} \frac{1}{a} \log \left(\int_\Omega e^{aY} \mathrm{d}\nu \right) \kappa(\mathrm{d}a), \quad X, Y \in \mathcal{X}, \tag{4.11}$$

where ν is the $\succsim$-matching probability. The model (4.11) has a natural interpretation. To account for ambiguity, the DM first evaluates each act using the certainty equivalent under the Choquet expected utility model with utility function u_a, for every $a \in \mathbb{R}$. These values are then averaged with respect to the measure κ. An axiomatization of this model can be obtained by taking the axioms of Theorem 4.2 and adding the following axiom: for all $X, Y, Z \in \mathcal{X}(\mathcal{G})$ with Z independent of both X and Y under the exogenous likelihood mapping, we have

$$X \succsim Y \iff X + Z \succsim Y + Z.$$

This follows directly from Mu et al. [2024, Theorem 1], which characterizes the class of statistics γ_κ.

4.5 Risk and ambiguity attitudes

In this section, we study how attitudes toward pure risk and ambiguity can be separated within the ATE framework. We begin by isolating the risk component of the preferences and recalling several classical risk attitudes. We then introduce comparative and absolute ambiguity attitudes and conclude by analyzing how risk and ambiguity attitudes can lead to preferences for diversification.

4.5.1 Risk attitudes

In this section, we will call the Axioms (M), (RC), (SRM), and (C) the regularity axioms. Let $\succsim$ be a preference relation satisfying the regularity axioms. Risk attitudes for $\succsim$ correspond to properties of the pure-risk preferences associated with $\succsim$. Intuitively, risk attitudes should be represented by properties of the $\succsim$-CES, which isolates the risk component of $\succsim$. An independent study of risk attitudes is necessary because their interaction with ambiguity attitudes (properties of the act-to-distribution mapping) plays a crucial role in determining the properties of the overall preference relation, as we will see in Section 4.5.3. This section also connects to and extends previous studies on the connection between the propensity for diversification and strong risk aversion in the pure-risk setting;

see Dekel [1989] and Chateauneuf and Lakhnati [2007]. We now recall several standard risk attitudes commonly discussed in the literature.

The first risk attitude we will discuss is the strong risk aversion as studied by Rothschild and Stiglitz [1970].

(SRA) Strong risk aversion: For all $X, Y \in \mathcal{X}(\mathcal{G})$, if $X \geq^{\mathbb{P}}_{\text{ssd}} Y$ then $X \succsim Y$.

Given a preference relation $\succsim$ satisfying the regularity axioms, it is clear that $\succsim$ satisfies Property (SRA) if and only if the $\succsim$-CES γ is *monotonic with respect to* $\geq_{\text{ssd}}$, i.e., for all $Q, P \in \mathcal{D}$ satisfying $Q \geq_{\text{ssd}} P$, $\gamma(Q) \geq \gamma(P)$.

The next risk attitude we discuss is risk diversification, reformulated from the main property in Dekel [1989].

(RD) Risk diversification: For all $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$ such that $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$,

$$X \simeq Y \implies \lambda X + (1-\lambda)Y \succsim X.$$

If the preference relation of the DM satisfies Property (RD), then the DM prefers mixing the outcomes of equally preferred acts with unambiguous distributions. This is consistent with the intuitive notion of diversification. We will discuss the overall preference for diversification in Section 4.5.3. Note that, given $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$, it may not hold that $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$, because the dependence structure between X and Y may be ambiguous. To reflect on the risk attitude, we need to assume that the "diversified act" $\lambda X + (1-\lambda)Y$ is also unambiguous for a proper comparison. In the special case that X and Y are comonotonic and unambiguous, their diversified act is always unambiguous, as shown in the next result.

Proposition 4.5. *Given comonotonic $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$, it holds that $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$.*

Proposition 4.5 is conceptually appealing, as it is consistent with the interpretation of $\mathcal{G}$: If the comonotonic acts X and Y have unambiguous distributions, that is, $X, Y \in \mathcal{X}(\mathcal{G})$, then their joint distribution is also unambiguous. This follows because comonotonicity fixes the dependency structure between X and Y. Consequently, the distribution of the random variable $\lambda X + (1-\lambda)Y$ should also be unambiguous for all $\lambda \in [0,1]$. This result is useful in formulating Property (RDC) below.

(RDC) Risk diversification for comonotonic acts: For all comonotonic $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$,

$$X \simeq Y \implies \lambda X + (1-\lambda)Y \succsim X.$$

Property (RDC) was studied by Chateauneuf and Tallon [2002]. It is clear that Property (RDC) is weaker than Property (RD).

Given a preference relation $\succsim$ satisfying the regularity axioms, one goal is to characterize Property (RD) through properties of the associated $\succsim$-CES. A compelling candidate is a quasiconcavity property. Therefore, we must first introduce an appropriate mixture operation on the set of distributions $\mathcal{D}$.

Given $Q, P \in \mathcal{D}$, denote by $Q \oplus P$ the element of $\mathcal{D}$ with a quantile function equal to the sum of those of P and Q, that is,

$$\mathrm{VaR}_\alpha(Q \oplus P) = \mathrm{VaR}_\alpha(Q) + \mathrm{VaR}_\alpha(P) \text{ for all } \alpha \in (0,1).$$

Given $Q \in \mathcal{D}$ and non-negative $a \in \mathbb{R}$ denote by $a \otimes Q$ the element of $\mathcal{D}$ satisfying

$$\mathrm{VaR}_\alpha(a \otimes Q) = a\mathrm{VaR}_\alpha(Q) \text{ for all } \alpha \in (0,1).$$

As discussed in Liu et al. [2021, Proposition 3], the triplet $(\mathcal{D}, \oplus, \otimes)$ is a convex cone. We say that a statistic γ is *comonotonic quasiconcave* if for all $Q, P \in \mathcal{D}$ and $\lambda \in [0,1]$,

$$\gamma\left(\lambda \otimes Q \oplus (1-\lambda) \otimes P\right) \geq \min\{\gamma(Q), \gamma(P)\}.^{2}$$

The following proposition demonstrates that the above operations naturally appear when considering Choquet act-to-distribution mappings.

Proposition 4.6. *Let ν be a continuous capacity. For comonotonic $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$, we have*

$$\mathfrak{D}_\nu(X+Y) = \mathfrak{D}_\nu(X) \oplus \mathfrak{D}_\nu(Y) \quad \textit{and} \quad \mathfrak{D}_\nu(\lambda X) = \lambda \otimes \mathfrak{D}_\nu(X).$$

We are now ready to present the main theorem of the section.

Theorem 4.3. *Let $\succsim$ be a preference relation satisfying the regularity axioms. The following are equivalent:*

(i) The preference relation $\succsim$ satisfies Property (RD);

[2] Here $\otimes$ takes precedence over $\oplus$ in the order of operations.

(ii) the preference relation $\succsim$ satisfies Properties (SRA) and (RDC);

(iii) the $\succsim$-CES is monotonic with respect to $\geq_{\text{ssd}}$ and comonotonic quasiconcave.

For a preference relation $\succsim$ satisfying the regularity axioms, Theorem 4.3 implies that $\succsim$ satisfies Property (RD) if and only if the $\succsim$-CES satisfies monotonicity with respect to $\geq_{\text{ssd}}$ and comonotonic quasiconcavity. In Dekel [1989], probabilistic quasiconcavity together with $\geq_{\text{ssd}}$-monotonicity was shown to be sufficient for Property (RD). However, probabilistic quasiconcavity is not necessary for Property (RD); instead, we show that comonotonic quasiconcavity is the relevant condition.

Example 4.4. Let $\mathbb{E}_u^g$ be the statistic in (4.5) from the RDU model. A direct consequence of Schmidt and Zank [2008, Corollary 3] is that $\mathbb{E}_u^g$ is monotonic with respect to $\geq_{\text{ssd}}$ if and only if u is concave and g is convex. Additionally, a consequence of Wakker and Yang [2019, Theorem 3] is that $\mathbb{E}_u^g$ in (4.5) is comonotonic quasiconcave if and only if u is concave. Therefore, if $\succsim$ is a preference relation satisfying the regularity axioms and the $\succsim$-CES is given by $\mathbb{E}_u^g$, $\succsim$ satisfies Property (RD) if and only if $\succsim$ satisfies Property (SRA). This helps explain why Property (RDC) has received limited attention in the literature, since Property (SRA) often implies Property (RDC) for many statistics commonly used in decision-making under pure risk.

We conclude this section by discussing when risk attitudes extend to properties of the overall preference relation in the case of Choquet ATE preference relations. The first example of this phenomenon is regarding Property (RDC).

Proposition 4.7. *Let $\succsim$ be a Choquet ATE preference relation. If $\succsim$ satisfies Property (RDC) then for all comonotonic $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$,*

$$X \simeq Y \implies \lambda X + (1-\lambda)Y \succsim X.$$

The remaining properties pertain to statistics γ. Nonetheless, it is easy to see that these properties could just as well be formulated in terms of pure-risk preferences. We say that a statistic γ is *constant additive* if for all $Q \in \mathcal{D}$ and $a \in \mathbb{R}$, $\gamma(Q \oplus \delta_a) = \gamma(Q) + a$; *positively homogeneous* if for all $Q \in \mathcal{D}$ and non-negative $a \in \mathbb{R}$, $\gamma(a \otimes Q) = a\gamma(Q)$; biseparable if there exists a vNM utility u and a statistic $\tilde{\gamma}$ satisfying constant additivity and positive homogeneity such that

$$\gamma(Q) = \tilde{\gamma}(u_{\#}Q), \quad Q \in \mathcal{D}. \tag{4.12}$$

Constant additivity and positive homogeneity are common properties discussed in the literature of risk measures; see Föllmer and Schied [2016]. Biseparability is a property of

biseparable preference relations. Traditionally, they are considered one of the weakest classes of preference relations, encompassing many famous decision models under ambiguity. For more on biseparable preference relations, see Ghirardato and Marinacci [2001].

Proposition 4.8. *Let $\succsim$ be a Choquet ATE preference relation.*

(i) If the $\succsim$-CES satisfies constant additivity, then for all $a \in \mathbb{R}$ we have

$$X \succsim Y \iff X + a \succsim Y + a, \quad X, Y \in \mathcal{X}.$$

(ii) If the $\succsim$-CES satisfies positive homogeneity, then for all positive $a \in \mathbb{R}$ we have

$$X \succsim Y \iff aX \succsim aY, \quad X, Y \in \mathcal{X}.$$

(iii) If the $\succsim$-CES satisfies biseparability, then there exists a vNM function u and $I : \mathcal{X} \to \mathbb{R}$ satisfying $I(X) \geq I(Y)$ for all $X, Y \in \mathcal{X}$ with $X \geq Y$ and $I(aX + b) = aI(X) + b$ for all $X \in \mathcal{X}$, non-negative $a \in \mathbb{R}$, and $b \in \mathbb{R}$, such that

$$X \succsim Y \iff I(u(X)) \geq I(u(Y)), \quad X, Y \in \mathcal{X}.$$

A key consequence of Proposition 4.8 is that, for a Choquet ATE preference relation $\succsim$, the $\succsim$-CES being biseparable implies that the preference relation $\succsim$ is biseparable in the sense of Ghirardato and Marinacci [2001].

4.5.2 Comparative and absolute ambiguity attitudes

We next define comparative ambiguity attitudes for Choquet ATE preference relations. This will then allow us to define absolute ambiguity attitudes once a normalization for ambiguity neutrality is provided.

Given the Choquet ATE preference relations $\succsim_1$ and $\succsim_2$, we say that $\succsim_2$ is *more ambiguity averse than* $\succsim_1$ if

$$R \succsim_1 A \implies R \succsim_2 A, \qquad R \in \mathcal{G} \text{ and } A \in \mathcal{F}.$$

The interpretation of this definition is quite natural: As events $R \in \mathcal{G}$ have no ambiguity, they can serve as the benchmark for comparing ambiguous events. This is similar to the classic notion of comparative risk aversion (Pratt, 1964, Yaari, 1969), where constant acts serve as the benchmark.

We have the following theorem characterizing comparative ambiguity attitudes.

Theorem 4.4. *Let $\succsim_1$ and $\succsim_2$ be Choquet ATE preference relations with matching probabilities ν_1 and ν_2 respectively. Then, $\succsim_2$ is more ambiguity averse than $\succsim_1$ if and only if $\nu_1(A) \geq \nu_2(A)$ for all $A \in \mathcal{F}$.*

In Theorem 4.4, the $\succsim_1$-CES and the $\succsim_2$-CES may not be the same. This is consistent with the fact that the ambiguity component in the Choquet ATE model is separated from the risk component. To get a full comparative implication on all acts, instead of only binary ones, we need to force the risk components of both preference relations to be identical, as shown in the next result.

Proposition 4.9. *Let $\succsim_1$ and $\succsim_2$ be Choquet ATE preference relations. Then, the comparative implication*

$$X \succsim_1 Y \ (X \succ_1 Y) \implies X \succsim_2 Y \ (X \succ_2 Y), \quad \text{for all } X \in \mathcal{X}(\mathcal{G}) \text{ and } Y \in \mathcal{X} \tag{4.13}$$

holds if and only if the $\succsim_1$-CES is equal to the $\succsim_2$-CES and $\succsim_2$ is more ambiguity averse than $\succsim_1$.

As Proposition 4.9 demonstrates, our notion of comparative ambiguity attitudes for Choquet ATE preference relations generalizes the comparative ambiguity attitudes of Epstein [2004] given by (4.13).

Theorem 4.4 suggests a natural way to define comparative ambiguity attitudes for general act-to-distribution mappings. To see this, observe that for two continuous capacities ν_1 and ν_2, the condition $\nu_1(A) \geq \nu_2(A)$ for all $A \in \mathcal{F}$ is equivalent to the requirement that $\mathfrak{D}_{\nu_1}(X) \geq_{\text{fsd}} \mathfrak{D}_{\nu_2}(X)$ for all $X \in \mathcal{X}$. Therefore, given two act-to-distribution mappings $\mathfrak{D}_1$ and $\mathfrak{D}_2$, we say that $\mathfrak{D}_2$ *is more ambiguity averse than* $\mathfrak{D}_1$ if

$$\mathfrak{D}_1(X) \geq_{\text{fsd}} \mathfrak{D}_2(X), \quad X \in \mathcal{X}.$$

The next step is to define a normalization for ambiguity neutrality.

(AN) Ambiguity neutrality: For all $A, B, C \in \mathcal{F}$ satisfying $(A \cup B) \cap C = \varnothing$,

$$A \succsim B \iff A \cup C \succsim B \cup C.$$

The following proposition characterizes Property (AN). The proof of this proposition follows a similar argument as the proof of Van Oosten and Wang [2025a, Theorem 2].

Proposition 4.10. *Let $\succsim$ satisfy Axioms (RC), (M), (SRM), and (C). Then $\succsim$ satisfies Property (AN) if and only if the $\succsim$-matching probability ν is a probability measure.*

Finally, we say that a Choquet ATE preference relation $\succsim$ is *ambiguity averse* if it is more ambiguity averse than a Choquet ATE preference relation satisfying Property (AN). A straightforward consequence of Theorem 4.4 and Proposition 4.10 is that given a Choquet ATE preference relation $\succsim$, ambiguity aversion is equivalent to the $\succsim$-matching probability being balanced.

Once again, Proposition 4.10 suggests a natural way to define ambiguity aversion for general act-to-distribution mappings. Given the act-to-distribution mapping $\mathfrak{D}$, we say that $\mathfrak{D}$ is *ambiguity neutral* if there exists $\mu \in \Delta$ such that $\mathfrak{D}(X) = X_{\#}\mu$ for all $X \in \mathcal{X}$. That is, an act-to-distribution mapping is ambiguity neutral if and only if it is a probabilistic act-to-distribution mapping. Additionally, we say that an act-to-distribution mapping is ambiguity averse if it is more ambiguity averse than an ambiguity-neutral act-to-distribution mapping.

4.5.3 Diversification seeking and aggregation concavity of act-to-distribution mappings

We now extend the definition of Property (RD) to all acts. This notion was formalized in the seminal contributions of Gilboa and Schmeidler [1989] and Schmeidler [1989],[3] and is sometimes also described as the convexity of preferences.

(D) Diversification: For all $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$,

$$X \simeq Y \implies \lambda X + (1-\lambda)Y \succsim X.$$

Clearly, Property (D) is stronger than Property (RD). However, there does not exist an equivalent proposition to Proposition 4.7 in the case of Property (RD) extending to Property (D). To substantiate this claim, we provide a counterexample in Section 4.11. The following example motivates a property for act-to-distribution mappings that allows Property (RD) to imply Property (D).

Given $\mu \in \Delta$ and a partial order $\geq$ on $\mathcal{D}$, we can induce a preorder $\geq^{\mu}$ on $\mathcal{X}$ by

$$X \geq^{\mu} Y \iff X_{\#}\mu \geq Y_{\#}\mu, \quad X, Y \in \mathcal{X}.$$

[3] In Gilboa and Schmeidler [1989] and Schmeidler [1989], this property is referred to as uncertainty aversion.

Example 4.5. Consider a DM that perceives no ambiguity and describes their uncertainty with the unique $\mu \in \Delta$. Additionally, assume that, to aid in the decision-making, the DM uses a statistic γ that is monotonic with respect to $\geq_{\text{ssd}}$ and comonotonic quasiconcave. Therefore, the DM's preference relation $\succsim$ is given by

$$X \succsim Y \iff \gamma \circ \mathfrak{D}_\mu(X) \geq \gamma \circ \mathfrak{D}_\mu(Y), \quad X, Y \in \mathcal{X}.$$

Establishing that $\succsim$ satisfies Property (D) follows by invoking the well-known comonotonic sum inequality; see Dhaene et al. [2002, Theorem 7]. This inequality states that, given $X, Y \in \mathcal{X}$ and comonotonic $X_c, Y_c \in \mathcal{X}$ such that $X =^{\mu}_{\text{fsd}} X_c$, $Y =^{\mu}_{\text{fsd}} Y_c$, and $X + Y \geq^{\mu}_{\text{cv}} X_c + Y_c$. Using Proposition 4.6, we can rewrite this inequality in terms of the probabilistic act-to-distribution mapping as follows:

$$\mathfrak{D}_\mu(X + Y) \geq_{\text{cv}} \mathfrak{D}_\mu(X_c + Y_c) = \mathfrak{D}_\mu(X_c) \oplus \mathfrak{D}_\mu(Y_c) = \mathfrak{D}_\mu(X) \oplus \mathfrak{D}_\mu(Y). \tag{4.14}$$

Given $\lambda \in [0, 1]$, we can use (4.14) and Proposition 4.6 to obtain the inequality

$$\mathfrak{D}_\mu(\lambda X + (1-\lambda)Y) \geq_{\text{cv}} \mathfrak{D}_\mu(\lambda X) \oplus \mathfrak{D}_\mu((1-\lambda)Y) = \lambda \otimes \mathfrak{D}_\mu(X) \oplus (1-\lambda) \otimes \mathfrak{D}_\mu(Y).$$

Since $Q \geq_{\text{cv}} P$ implies $Q \geq_{\text{ssd}} P$ for all $Q, P \in \mathcal{D}$, we get

$$\mathfrak{D}_\mu\left(\lambda X + (1-\lambda)Y\right) \geq_{\text{ssd}} \lambda \otimes \mathfrak{D}_\mu(X) \oplus (1-\lambda) \otimes \mathfrak{D}_\mu(Y). \tag{4.15}$$

Therefore, using (4.15) and the fact that γ is monotonic with respect to $\geq_{\text{ssd}}$ and comonotonic quasiconcave, for all $X, Y \in \mathcal{X}$ satisfying $X \simeq Y$ and $\lambda \in [0, 1]$,

$$\begin{aligned} \gamma\left(\mathfrak{D}_\mu\left(\lambda X + (1-\lambda)Y\right)\right) &\geq \gamma\left(\lambda \otimes \mathfrak{D}_\mu(X) \oplus (1-\lambda) \otimes \mathfrak{D}_\mu(Y)\right) \\ &\geq \min\left\{\gamma(\mathfrak{D}_\mu(X)), \gamma(\mathfrak{D}_\mu(Y))\right\} = \gamma(\mathfrak{D}_\mu(X)). \end{aligned} \tag{4.16}$$

Thus, $\lambda X + (1-\lambda)Y \succsim X$, which shows that $\succsim$ satisfies Property (D).

Therefore, motivated by (4.15) from Example 4.5, we introduce the following property for act-to-distribution mappings $\mathfrak{D}$, which we call *aggregation concavity*:

$$\mathfrak{D}\left(\lambda X + (1-\lambda)Y\right) \geq_{\text{ssd}} \lambda \otimes \mathfrak{D}(X) \oplus (1-\lambda) \otimes \mathfrak{D}(Y). \tag{4.17}$$

Let $\succsim$ be an ATE preference relation represented by (4.8). If γ is monotone with respect to $\geq_{\text{ssd}}$ and comonotonic quasiconcave, and if $\mathfrak{D}$ is aggregation concave, then $\succsim$ satisfies Property (D). This follows by the same reasoning as in (4.16). A natural question, which the following theorem answers, is the characterization of when Choquet act-to-distribution

mappings are aggregation concave.

Theorem 4.5. *Let ν be a continuous capacity. Then $\mathfrak{D}_\nu$ is aggregation concave if and only if ν is supermodular.*

Aggregation concavity of an act-to-distribution mapping reflects a form of ambiguity aversion, since it implies that a strongly risk-averse DM views the distribution associated with the mixture of ambiguous acts as more attractive than the comonotonic mixture of the distributions associated with each act separately. This idea is highlighted by the fact that a Choquet act-to-distribution mapping is ambiguity averse if it is aggregation concave. This follows from Theorem 4.5 and the fact that supermodular capacities are balanced; see Shapley [1971]. However, as the following example shows, this implication does not extend to all act-to-distribution mappings.

Example 4.6. Given distinct $\mu_1, \mu_2 \in \Delta$, let $\mathfrak{D}$ be the act-to-distribution mapping given by

$$\mathrm{VaR}_\alpha(\mathfrak{D}(X)) = \frac{1}{2}\mathrm{VaR}_\alpha(\mathfrak{D}_{\mu_1}(X)) + \frac{1}{2}\mathrm{VaR}_\alpha(\mathfrak{D}_{\mu_2}(X)) \text{ for all } \alpha \in (0,1).$$

We show in Section 4.11 that $\mathfrak{D}$ is aggregation concave and not ambiguity averse.

Given a continuous capacity ν, it is clear that the Choquet act-to-distribution mapping $\mathfrak{D}_\nu$ is aggregation concave if and only if for all $X_1, \ldots, X_n \in \mathcal{X}$ and $\lambda_1, \ldots, \lambda_n \in [0,1]$ satisfying $\sum_{k=1}^n \lambda_k = 1$, we have

$$\mathfrak{D}_\nu\left(\sum_{k=1}^n \lambda_k X_k\right) \geq_{\mathrm{ssd}} \bigoplus_{k=1}^n \lambda_k \otimes \mathfrak{D}_\nu(X_k). \tag{4.18}$$

By restricting the form of the acts $\sum_{k=1}^n \lambda_k X_k$ in (4.18), we can define weaker notions of concavity for Choquet act-to-distribution mappings. Motivated by the work of Hartmann and Kauffeldt [2026], we can use this idea to characterize the family of Choquet act-to-distribution mappings presented in Example 4.1. Given $A \in \mathcal{F}$ and $x, y \in \mathbb{R}$, xAy denotes the binary act which yields x if $\omega \in A$ and y otherwise.

Proposition 4.11. *It holds that $\mathfrak{D}_\nu = \mathfrak{D}_{\mathfrak{C}(\nu)}$ if and only if (4.18) holds for all $\sum_{k=1}^n \lambda_k X_k = xAy$ where $A \in \mathcal{F}$ and $x, y \in \mathbb{R}$.*

Finally, using similar arguments as the proof for Proposition 4.11, one can show that $\mathfrak{D}_\nu$ is ambiguity averse if and only if (4.18) holds for all $\sum_{k=1}^n \lambda_k X_k = x$ where $x \in \mathbb{R}$. This is consistent with the idea of sure diversification from Chateauneuf and Tallon [2002].

4.5.4 The distributionally robust interpretation

The framework of distributionally robust optimization extends the maxmin expected utility model of Gilboa and Schmeidler [1989] by allowing for general statistics; see Delage et al. [2019] and Fadina et al. [2024] for approaches along this line in the context of risk measures. A preference relation $\succsim$ is called a *distributionally robust* preference relation if there exists a collection of risk-conforming probability measures $\mathcal{Q}$[4] and a statistic γ such that

$$X \succsim Y \iff \min_{\mu\in\mathcal{Q}} \gamma(X_{\#}\mu) \geq \min_{\mu\in\mathcal{Q}} \gamma(Y_{\#}\mu), \quad X, Y \in \mathcal{X}.$$

Intuitively, a DM with a distributionally robust preference relation views the set $\mathcal{Q}$ as a collection of equally plausible models describing their uncertainty, while the statistic γ represents their pure-risk preferences. In this decision model, $\mathcal{Q}$ represents the ambiguity component of $\succsim$, and γ represents the risk component of $\succsim$. Distributionally robust preference relations represent a strong form of ambiguity aversion, as discussed by Ghirardato et al. [2004].

Theorem 4.6. *Let $\succsim$ be a Choquet ATE preference relation. Then $\succsim$ is a distributionally robust preference relation if and only if the $\succsim$-matching probability is supermodular.*

Since any distributionally robust preference relation is an example of an ETA preference relation, Theorem 4.6 establishes that the ETA and ATE frameworks are capable of generating a broad class of common preference relations.

4.6 Conclusion

The first conceptual contribution of this chapter is to show that the role of ambiguity in a decision model can be determined by where ambiguity enters relative to the evaluation of pure risk. By formalizing the ETA and ATE frameworks in the Savage setting, we show that the classical ambiguity models can be seen as instantiations of two general thought processes. This perspective clarifies that differences between existing models arise from the structural choice of whether ambiguity is incorporated by model-specific evaluations or in the distributional representation of acts themselves. In turn, this structural distinction identifies the precise primitives that represent ambiguity (aggregators in the ETA case

[4]Here, we suppose each $\mu \in \mathcal{Q}$ is risk conforming to ensure we remain in the Savage setting with pure-risk.

and act-to-distribution mappings in the ATE case) and shows how they interact with the statistic associated with pure risk.

After introducing the two frameworks, we focused exclusively on the ATE framework. A key benefit of the ATE perspective is that act-to-distribution mappings offer a richer, more flexible way to incorporate ambiguity than the ETA perspective, thus expanding the modeling domain of ambiguity attitudes by permitting them to reshape many important distributional characteristics, e.g., tail behavior, variability, and asymmetry. Such flexibility is not immediately available in the ETA framework, where it is unclear—or considerably more complicated to determine—how model-based aggregation can address these important concerns. The new ATE framework allows us to construct new decision models and opens new directions for future research. As the main ATE model analyzed in this chapter, the Choquet ATE model offers great flexibility in modeling both the risk attitudes and the ambiguity attitudes. This model is only a natural first step in modeling transforms from an ambiguous act to a distribution, and there are many more possibilities for general act-to-distribution mappings to explore, drawing on tools from statistical theory.

4.7 Distorted capacities

This section introduces distorted capacities, which are a generalization of distorted probabilities, and presents several results that will be used in the proofs provided in subsequent sections.

Given a distortion function g and a capacity ν, we call $g \circ \nu$ a *distorted capacity*. If the distortion function g is convex (resp. concave) and the capacity ν is supermodular (resp. submodular), then the distorted capacity $g \circ \nu$ is supermodular (resp. submodular); see Van Oosten and Wang [2025a, Proposition 13].

In the proofs in the following sections, we will make use of the two following families of distortion functions. Given $\alpha \in (0,1)$, define the distortion functions $g_\alpha(\beta) = \mathbb{1}_{[1-\alpha,1]}(\beta)$ and $h_\alpha(\beta) = (\beta/\alpha - (1-\alpha)/\alpha)^+$ for all $\beta \in (0,1)$. Given a capacity ν and $\alpha \in (0,1)$, we can define the following functionals given by the Choquet integral with respect to the distorted capacity:

$$\mathrm{G}_\alpha^\nu(X) = \int X \mathrm{d}(g_\alpha \circ \nu) \quad \text{and} \quad \mathrm{H}_\alpha^\nu(X) = \int X \mathrm{d}(h_\alpha \circ \nu), \quad X \in \mathcal{X}.$$

Since h_α is concave for every $\alpha \in (0,1)$, $h_\alpha \circ \nu$ is supermodular if ν is supermodular. Therefore, if ν is supermodular, then H_α^ν is concave for every $\alpha \in (0,1)$ by Marinacci and Montrucchio [2004, Corollary 4.2].

Proposition 4.12. *Let ν be a continuous capacity. For all $\alpha \in (0,1)$, we have*

$$\mathrm{G}^{\nu}_{\alpha}(X) = \mathrm{VaR}_{\alpha}(\mathfrak{D}_{\nu}(X)) \quad \textit{and} \quad \mathrm{H}^{\nu}_{\alpha}(X) = \frac{1}{\alpha}\int_0^{\alpha} \mathrm{G}^{\nu}_{\beta}(X)\mathrm{d}\beta, \quad X \in \mathcal{X}.$$

Proof. Fix $X \in \mathcal{X}$, let $x_0 = \inf\{x \in \mathbb{R} : \nu(X > x) < 1-\alpha\}$. For $\epsilon > 0$, we know that $\nu(X > x_0 + \epsilon) < 1-\alpha$ and $\nu(X > x_0 - \epsilon) \geq 1-\alpha$. Therefore, $g_{\alpha}(\nu(X > x_0+\epsilon)) = 0$ and $g_{\alpha}(\nu(X > x_0 - \epsilon)) = 1$. Thus,

$$g_{\alpha}(\nu(X > x)) = \begin{cases} \mathbb{1}_{(-\infty, x_0)}(x) & \text{if } \nu(X > x_0) \geq 1-\alpha \\ \mathbb{1}_{(-\infty, x_0]}(x) & \text{if } \nu(X > x_0) < 1-\alpha \end{cases}, \quad x \in \mathbb{R}.$$

By the definition of the Choquet integral, $\mathrm{G}^{\nu}_{\alpha}(X) = x_0$. Since, by definition, $S_{\mathfrak{D}_{\nu}(X)}(x) = \nu(X > x)$ for all $x \in \mathbb{R}$, we have $\mathrm{G}^{\nu}_{\alpha}(X) = \mathrm{VaR}_{\alpha}(\mathfrak{D}_{\nu}(X))$.

Let $\delta \in [0,1]$, we have

$$\frac{1}{\alpha}\int_0^{\alpha} g_{\beta}(\delta)\mathrm{d}\beta = \frac{1}{\alpha}\int_0^{\alpha} \mathbb{1}_{[1-\delta,1]}(\beta)\mathrm{d}\beta = (\delta/\alpha - (1-\alpha)/\alpha)^{+} = h_{\alpha}(\delta).$$

Since $\varphi : \mathcal{X} \to \mathbb{R}$ defined by $\varphi(X) = \frac{1}{\alpha}\int_0^{\alpha} \mathrm{G}^{\nu}_{\beta}(X)\mathrm{d}\beta$ is monotone, comonotonic additive, and normalized, $\varphi(X) = \int_{\Omega} X \mathrm{d}\tilde{\nu}$ for some capacity $\tilde{\nu}$. For all $A \in \mathcal{F}$,

$$\tilde{\nu}(A) = \varphi(\mathbb{1}_A) = \frac{1}{\alpha}\int_0^{\alpha} \mathrm{G}^{\nu}_{\beta}(\mathbb{1}_A)\mathrm{d}\beta = \frac{1}{\alpha}\int_0^{\alpha} g_{\beta}(\nu(A))\mathrm{d}\beta = h_{\alpha}(\nu(A)).$$

Therefore, $\varphi(X) = \mathrm{H}^{\nu}_{\alpha}(X)$ for all $X \in \mathcal{X}$. □

4.8 Proof accompanying Section 4.2

Proof of Proposition 4.1. Let γ and $\tilde{\gamma}$ define the same preference relation on distributions $\succsim^{\mathrm{d}}$ in (4.4). Define $f : \mathbb{R} \to \mathbb{R}$ and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \gamma(\delta_x) \text{ and } \tilde{f}(x) = \tilde{\gamma}(\delta_x), \quad x \in \mathbb{R}.$$

Since γ and $\tilde{\gamma}$ are statistics, both f and $\tilde{f}$ are strictly increasing and continuous. We claim that $\mathrm{im}(\gamma) = \mathrm{im}(f)$. Clearly, $\mathrm{im}(f) \subseteq \mathrm{im}(\gamma)$. Let $Q \in \mathcal{D}$. Since Q has compact support, there exists $x, y \in \mathbb{R}$ such that $\delta_x \geq_{\mathrm{fsd}} Q \geq_{\mathrm{fsd}} \delta_y$. Therefore, $f(x) \geq \gamma(Q) \geq f(y)$. As f is continuous, by the intermediate value theorem, there exists $z \in \mathbb{R}$ such that

$f(z) = \gamma(Q)$. Thus, $\text{im}(\gamma) \subseteq \text{im}(f)$ and $\text{im}(\gamma) = \text{im}(f)$. A similar argument will show that $\text{im}(\tilde{\gamma}) = \text{im}(\tilde{f})$. Define the statistics $\psi : \mathcal{D} \to \mathbb{R}$ and $\tilde{\psi} : \mathcal{D} \to \mathbb{R}$ by

$$\psi(Q) = f^{-1} \circ \gamma(Q) \text{ and } \tilde{\psi}(Q) = \tilde{f}^{-1} \circ \tilde{\gamma}(Q), \quad Q \in \mathcal{D},$$

where $f^{-1} : \text{im}(f) \to \mathbb{R}$ and $\tilde{f}^{-1} : \text{im}(\tilde{f}) \to \mathbb{R}$ denote the inverses of f and $\tilde{f}$, respectively. It is straightforward to show that ψ and $\tilde{\psi}$ are certainty-equivalent statistics that satisfy (4.4). Note that this additionally shows the existence of a certainty-equivalent statistic representing $\succsim^{\text{d}}$. Given $Q \in \mathcal{D}$, $Q \simeq^{\text{d}} \delta_{\psi(Q)}$, which implies $\psi(Q) = \tilde{\psi}(Q)$. Thus, we have $\tilde{\gamma} = \tilde{f} \circ f^{-1} \circ \gamma$. Additionally, if γ and $\tilde{\gamma}$ were certainty-equivalent statistics, then $\gamma = \tilde{\gamma}$ since $f(x) = \tilde{f}(x) = x$ for all $x \in \mathbb{R}$. □

4.9 Proofs, discussions, and additional results accompanying Section 4.3

Proof of Proposition 4.2. Fix $\eta \in \mathcal{X}(\Delta, \Sigma)$. We claim the set $\mathcal{I}_1(\eta) = \{c \in \mathbb{R} : \eta \succsim^{\text{m}} c\}$ must be of the form $(-\infty, a_0(\eta)]$ for some $a_0(\eta) \in \mathbb{R}$. Let $a \in \mathcal{I}_1(\eta)$ and $b \in \mathbb{R}$ satisfy $a > b$, then, by Property (M3), $\eta \succsim^{\text{m}} a \succ^{\text{m}} b$ and $b \in \mathcal{I}_1(\eta)$. Find $a \in \mathbb{R}$ such that $a \geq \eta$. Therefore, by Property (M1), $a \succsim^{\text{m}} \eta$. If $b \in \mathbb{R}$ satisfies $b > a$, then, by Property (M3), $b \succ^{\text{m}} a \succsim^{\text{m}} \eta$ and $b \notin \mathcal{I}_1(\eta)$. Therefore, $\sup(\mathcal{I}_1(\eta))$ exists. Denote by $a_0(\eta) = \sup(\mathcal{I}_1(\eta))$. By Property (M2), $a_0(\eta) \in \mathcal{I}_1(\eta)$. Therefore, $\mathcal{I}_1(\eta) = (-\infty, a_0(\eta)]$. Similarly, we can show that $\mathcal{I}_2(\eta) = \{c \in \mathbb{R} : c \succsim^{\text{m}} \eta\}$ is of the form $[b_0(\eta), \infty)$ for some $b_0(\eta) \in \mathbb{R}$. It must hold that $a_0(\eta) \geq b_0(\eta)$ as $\succsim$ is a preference relation on $\mathcal{X}(\Delta, \Sigma)$. Therefore, $\eta \simeq^{\text{m}} a_0(\eta)$.

Define the functional $\rho : \mathcal{X}(\Delta, \Sigma) \to \mathbb{R}$ by $\rho(\eta) = a_0(\eta)$. By Property (M3), it is clear that

$$\eta \succsim^{\text{m}} \zeta \iff \rho(\eta) \geq \rho(\zeta), \quad \eta, \zeta \in \mathcal{X}(\Delta, \Sigma).$$

By Property (M1), $\rho(\eta) \geq \rho(\zeta)$ for all $\eta, \zeta \in \mathcal{X}(\Delta, \Sigma)$ satisfying $\eta \geq \zeta$. By property (M3), $\rho(a) > \rho(b)$ for all $a, b \in \mathbb{R}$ satisfying $a > b$. □

Proposition 4.13. *Let γ be a statistic. Then for all $X \in \mathcal{X}$, $\gamma \circ \mathfrak{R}_X \in \mathcal{X}(\Delta, \Sigma)$.*

Proof. Fix $X \in \mathcal{X}$. Find $a, b \in \mathbb{R}$ such that $a \geq X \geq b$. Given $\mu \in \Delta$, we have

$$\delta_a \geq_{\text{fsd}} \mathfrak{R}_X(\mu) \geq_{\text{fsd}} \delta_b.$$

Therefore, since γ is a statistic, we have $\gamma(\delta_a) \geq \gamma \circ \mathfrak{R}_X(\mu) \geq \gamma(\delta_b)$. Since μ was arbitrary, $\gamma(\delta_a) \geq \gamma \circ \mathfrak{R}_X \geq \gamma(\delta_b)$. Therefore, $\gamma \circ \mathfrak{R}_X$ is bounded. To show that $\gamma \circ \mathfrak{R}_X$ is Σ-measurable,

we will show that it is continuous. Let $(\mu_\lambda)_{\lambda\in\Lambda} \subseteq \Delta$ be a net weak* converging to $\mu \in \Delta$. Given the bounded and continuous $f : \mathbb{R} \to \mathbb{R}$, we have

$$\lim_\lambda \int_{\mathbb{R}} f \mathrm{d}(\mathfrak{R}_X(\mu_\lambda)) = \lim_\lambda \int_\Omega f(X) \mathrm{d}\mu_\lambda = \int_\Omega f(X) \mathrm{d}\mu = \int_{\mathbb{R}} f \mathrm{d}(\mathfrak{R}(\mu)).$$

Therefore, $\mathfrak{R}_X(\mu_\lambda) \to \mathfrak{R}_X(\mu)$ in distribution. Define $\mathcal{D}_0 \subseteq \mathcal{D}$ by

$$\mathcal{D}_0 = \{Q \in \mathcal{D} : \mathrm{supp}(Q) \subseteq [\gamma(\delta_b), \gamma(\delta_a)]\}.$$

Equipping $\mathcal{D}_0$ with the topology consistent with convergence in distribution, we have that $\gamma|_{\mathcal{D}_0}$ is continuous as γ is a statistic. Since $\mathrm{im}(\mathfrak{R}_X) \subseteq \mathcal{D}_0$, we have $\lim_\lambda \gamma \circ \mathfrak{R}_X(\mu_\lambda) = \gamma \circ \mathfrak{R}_X(\mu)$. Therefore, $\gamma \circ \mathfrak{R}_X$ is Σ-measurable. □

Lemma 4.1. *Let ν be a continuous capacity. If $X, Y \in \mathcal{X}$ satisfy $X \geq Y$, then $\mathfrak{D}_\nu(X) \geq_{\mathrm{fsd}} \mathfrak{D}_\nu(Y)$.*

Proof. For all $x \in \mathbb{R}$, $\{X > x\} \supseteq \{Y > x\}$. Thus, for all $x \in \mathbb{R}$,

$$S_{\mathfrak{D}_\nu(X)}(x) = \nu(X > x) \geq \nu(Y > x) = S_{\mathfrak{D}_\nu(Y)}(x).$$

Therefore, $\mathfrak{D}_\nu(X) \geq_{\mathrm{fsd}} \mathfrak{D}_\nu(Y)$. □

Lemma 4.2. *Let ν be a continuous capacity. If $(Z_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ is a bounded sequence pointwise converging to $Z \in \mathcal{X}$, then $\mathfrak{D}_\nu(Z_n) \to \mathfrak{D}_\nu(Z)$ in distribution.*

Proof. For $n \in \mathbb{N}$, define $X_n = \sup_{m\geq n} Z_m$. It holds that $(X_n)_{n\in\mathbb{N}}$ is a bounded sequence such that $X_n \downarrow Z$ pointwise. We claim that $\mathfrak{D}_\nu(X_n) \to \mathfrak{D}_\nu(Z)$ in distribution. Denote by $\mathfrak{X} \subseteq \mathbb{R}$ the set of continuity points for $S_{\mathfrak{D}_\nu(Z)}$. Given $x \in \mathfrak{X}$, since

$$\{Z > x\} \subseteq \bigcap_{n=1}^\infty \{X_n > x\} \subseteq \{Z \geq x\},$$

it holds that $\nu(Z > x) = \nu(Z \geq x) \geq \nu(\bigcap_{n=1}^\infty \{X_n > x\}) \geq \nu(Z > x)$. Additionally, as ν is continuous, we have

$$S_{\mathfrak{D}_\nu(Z)}(x) = \nu(Z > x) = \nu\left(\bigcap_{n=1}^\infty \{X_n > x\}\right) = \lim_{n\to\infty} \nu(X_n > x) = \lim_{n\to\infty} S_{\mathfrak{D}_\nu(X_n)}(x)$$

Therefore $\mathfrak{D}_\nu(X_n) \to \mathfrak{D}_\nu(Z)$ in distribution.

For $n \in \mathbb{N}$, define $Y_n = \inf_{m \geq n} Z_m$. It holds that $(Y_n)_{n\in\mathbb{N}}$ is a bounded sequence such that $Y_n \uparrow Z$ pointwise. Since ν is continuous,

$$\lim_{n\to\infty} S_{\mathfrak{D}_\nu(Y_n)}(x) = \lim_{n\to\infty} \nu(Y_n > x) = \nu\left(\bigcup_{n=1}^{\infty}\{Y_n > x\}\right) = \nu(Z > x) = S_{\mathfrak{D}_\nu(Z)}(x), \quad x \in \mathbb{R}.$$

Thus, $\mathfrak{D}_\nu(Y_n) \to \mathfrak{D}_\nu(Z)$ in distribution.

Finally, since $X_n \geq Z_n \geq Y_n$ for all $n \in \mathbb{N}$, by Lemma 4.1, we have that

$$\mathfrak{D}_\nu(X_n) \geq_{\text{fsd}} \mathfrak{D}_\nu(Z_n) \geq_{\text{fsd}} \mathfrak{D}_\nu(Y_n).$$

Therefore, if $x \in \mathbb{R}$ is a continuity point of $S_{\mathfrak{D}_\nu(Z)}$, by the above results, we have

$$\begin{aligned} S_{\mathfrak{D}_\nu(Z)}(x) = \lim_{n\to\infty} S_{\mathfrak{D}_\nu(X_n)}(x) &\geq \limsup_{n\to\infty} S_{\mathfrak{D}_\nu(Z_n)}(x) \\ &\geq \liminf_{n\to\infty} S_{\mathfrak{D}_\nu(Z_n)}(x) \geq \lim_{n\to\infty} S_{\mathfrak{D}_\nu(Y_n)}(x) = S_{\mathfrak{D}_\nu(Z)}(x). \end{aligned}$$

Therefore, $\mathfrak{D}_\nu(Z_n) \to \mathfrak{D}_\nu(Z)$ in distribution. □

Proof of Theorem 4.1. Assume that $\mathfrak{D}$ satisfies Properties (R1)–(R3). Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and increasing. We will first show that

$$\mathfrak{D}(f(X)) = f_\#(\mathfrak{D}(X))^{[5]}, \quad X \in \mathcal{X}. \tag{4.19}$$

For each $n \in \mathbb{N}$, define the function $f_n : \mathbb{R} \to \mathbb{R}$ by $f_n(x) = f(x)+\exp(x-n)$. The following are immediately clear. (1) For all $x \in \mathbb{R}$, $\lim_{n\to\infty} f_n(x) = f(x)$. (2) For all $n \in \mathbb{N}$, f_n is a strictly increasing and continuous function. (3) For all $X \in \mathcal{X}$, $(f_n(X))_{n\in\mathbb{N}}$ is bounded. Let $Q \in \mathcal{D}$, we claim that $(f_n)_\# Q \to f_\# Q$ in distribution. To see this, let $g : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Then, by the dominated convergence theorem, we have

$$\int_{\mathbb{R}} g\mathrm{d}((f_n)_\# Q) = \int_{\mathbb{R}} g \circ f_n \mathrm{d}Q \to \int_{\mathbb{R}} g \circ f \mathrm{d}Q = \int_{\mathbb{R}} g\mathrm{d}(f_\# Q).$$

Therefore, $(f_n)_\# Q \to f_\# Q$ in distribution. By Property (R2) and (R3), for all $X \in \mathcal{X}$, we have

$$\mathfrak{D}(f(X)) = \lim_{n\to\infty} \mathfrak{D}(f_n(X)) = \lim_{n\to\infty} (f_n)_\#(\mathfrak{D}(X)) = f_\#(\mathfrak{D}(X)).$$

[5]Note that Property (R3) only guarantees this equality for continuous and strictly increasing functions.

Next, we claim that

$$\operatorname{supp}(\mathfrak{D}(X)) \subseteq \overline{\operatorname{im}(X)}, \quad X \in \mathcal{X}, \tag{4.20}$$

where $\overline{\operatorname{im}(X)}$ denotes the closure of $\operatorname{im}(X)$. To see this claim, for the sake of contradiction, assume that (4.20) is false. That is, assume there exists $X \in \mathcal{X}$ and $x_0 \in \operatorname{supp}(\mathfrak{D}(X))$ such that $x_0 \notin \overline{\operatorname{im}(X)}$. To ease notation, let $Q_X = \mathfrak{D}(X)$. Therefore, we can find $\epsilon > 0$ such that for $I = (x_0 - \epsilon, x_0 + \epsilon)$, we have $Q_X(I) > 0$ and $X^{-1}(I) = \varnothing$. Fix $\delta > 0$ satisfying $\delta < \epsilon$. Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x_0 - \epsilon & \text{if } x_0 - \epsilon < x < x_0 + \delta \\ \frac{2\epsilon}{\epsilon - \delta}(x - x_0 - \delta) + x_0 - \epsilon & \text{if } x_0 + \delta \leq x < x_0 + \epsilon \\ x & \text{if } x \leq x_0 - \epsilon \text{ or } x \geq x_0 + \epsilon \end{cases}.$$

Let g denote the identity function. We have that $f(X) = g(X)$. Therefore, by (4.19), $f_{\#}Q_X = g_{\#}Q_X$. However, we have

$$\begin{aligned} [f_{\#}Q_X](\{x_0 - \epsilon\}) &= Q_X([x_0 - \epsilon, x_0 + \delta)) \\ &= Q_X(\{x_0 - \epsilon\}) + Q_X((x_0 - \epsilon, x_0 + \delta)) \\ &> Q_X(\{x_0 - \epsilon\}) = [g_{\#}Q_X](\{x_0 - \epsilon\}), \end{aligned}$$

which is a contradiction. In the previous inequality, we use the fact that $x_0 \in \operatorname{supp}(Q_X)$, implying that $Q_X((x_0 - \epsilon, x_0 + \delta)) > 0$.

Let $X, Y \in \mathcal{X}$ be comonotonic, by Denneberg [1994, Proposition 4.5], there exist increasing and continuous $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ such that $f(X+Y) = X$, $g(X+Y) = Y$, and $f(x) + g(x) = x$ for all $x \in \mathbb{R}$. Therefore, by (4.19) and Liu et al. [2024, Theorem 1], for all $\alpha \in (0, 1)$,

$$\begin{aligned} \operatorname{VaR}_\alpha(\mathfrak{D}(X) \oplus \mathfrak{D}(Y)) &= \operatorname{VaR}_\alpha(\mathfrak{D}(X)) + \operatorname{VaR}_\alpha(\mathfrak{D}(Y)) \\ &= \operatorname{VaR}_\alpha(f_{\#}(\mathfrak{D}(X+Y))) + \operatorname{VaR}_\alpha(g_{\#}(\mathfrak{D}(X+Y))) \\ &= f(\operatorname{VaR}_\alpha(\mathfrak{D}(X+Y))) + g(\operatorname{VaR}_\alpha(\mathfrak{D}(X+Y))) \\ &= \operatorname{VaR}_\alpha(\mathfrak{D}(X+Y)). \end{aligned}$$

Thus $\mathfrak{D}(X+Y) = \mathfrak{D}(X) \oplus \mathfrak{D}(Y)$. For $X \in \mathcal{X}$ and positive $a \in \mathbb{R}$, by Property (R3), we have $\mathfrak{D}(aX) = a \otimes \mathfrak{D}(X)$. If $a = 0$, we have $\mathfrak{D}(aX) = a \otimes \mathfrak{D}(X)$ by (4.20).

Given $A \in \mathcal{F}$, by (4.20), we know that there exists a unique $a_A \in [0, 1]$ such that

$$\mathfrak{D}(\mathbb{1}_A) = a_A \delta_1 + (1 - a_A)\delta_0.$$

Define the function $\nu : \mathcal{F} \to [0,1]$ by $\nu(A) = a_A$. By (4.20), $\nu(\varnothing) = 0$ and $\nu(\Omega) = 1$. By Property (R1), for all $A, B \in \mathcal{F}$ satisfying $A \subseteq B$, $\nu(A) \leq \nu(B)$. Therefore, ν is a capacity. Let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ be an increasing sequence. Define $A = \bigcup_{n=1}^{\infty} A_n$. By Property (R2), we have $\mathfrak{D}(\mathbb{1}_{A_n}) \to \mathfrak{D}(\mathbb{1}_A)$ in distribution. Therefore, since $\mathbb{E}$ is a statistic,

$$\lim_{n\to\infty} \nu(A_n) = \lim_{n\to\infty} \mathbb{E}(\mathfrak{D}(\mathbb{1}_{A_n})) = \mathbb{E}(\mathfrak{D}(\mathbb{1}_A)) = \nu(A).$$

Thus, ν is upwards continuous. A similar proof will show that ν is downwards continuous. Thus, ν is continuous. Remark that for all $A \in \mathcal{F}$, $\mathfrak{D}(\mathbb{1}_A) = \mathfrak{D}_\nu(\mathbb{1}_A)$.

We claim that for all non-negative simple $X \in \mathcal{X}$, $\mathfrak{D}(X) = \mathfrak{D}_\nu(X)$. Fix a non-negative simple $X \in \mathcal{X}$. We can find $N \in \mathbb{N}$, $(a_n)_{n=1}^N \subseteq \mathbb{R}$ satisfying $0 < a_1 < \cdots < a_N$, and $(A_n)_{n=1}^N \subseteq \mathcal{F}$ satisfying $\Omega = A_1 \supset A_2 \supset \cdots \supset A_N \neq \varnothing$ such that $X = \sum_{n=1}^N (a_n - a_{n-1})\mathbb{1}_{A_n}$, where $a_0 = 0$. For all $\alpha \in (0,1)$,

$$\begin{aligned}
\mathrm{VaR}_\alpha(\mathfrak{D}(X)) &= \mathrm{VaR}_\alpha\left(\bigoplus_{n=1}^N (a_n - a_{n-1}) \otimes \mathfrak{D}(\mathbb{1}_{A_n})\right) \\
&= \sum_{n=1}^N (a_n - a_{n-1})\mathrm{VaR}_\alpha(\mathfrak{D}(\mathbb{1}_{A_n})) \\
&= \sum_{n=1}^N (a_n - a_{n-1})\mathrm{VaR}_\alpha(\mathfrak{D}_\nu(\mathbb{1}_{A_n})) = \mathrm{VaR}_\alpha(\mathfrak{D}_\nu(X)).
\end{aligned}$$

Therefore, $\mathfrak{D}(X) = \mathfrak{D}_\nu(X)$.

We claim that for all non-negative $X \in \mathcal{X}$, $\mathfrak{D}(X) = \mathfrak{D}_\nu(X)$. Fix a non-negative $X \in \mathcal{X}$. Find a sequence $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ of non-negative simple acts such that $X_{n+1} \geq X_n \geq 0$ for all $n \in \mathbb{N}$ and $X_n \to X$ pointwise. By Property (R2) and Lemma 4.2, we have

$$\mathfrak{D}(X) = \lim_{n\to\infty} \mathfrak{D}(X_n) = \lim_{n\to\infty} \mathfrak{D}_\nu(X_n) = \mathfrak{D}_\nu(X).$$

Therefore, $\mathfrak{D}(X) = \mathfrak{D}_\nu(X)$.

Finally, we claim that for general $X \in \mathcal{X}$, $\mathfrak{D}(X) = \mathfrak{D}_\nu(X)$. Fix $X \in \mathcal{X}$. We can find $a \in \mathbb{R}$ such that $X + a \geq 0$. Since $X + a$ and $-a$ are comonotonic, we have

$$\mathfrak{D}(X) = \mathfrak{D}(X + a - a) = \mathfrak{D}(X + a) \oplus (-a) = \mathfrak{D}_\nu(X + a) \oplus (-a) = \mathfrak{D}_\nu(X).$$

Therefore, $\mathfrak{D}(X) = \mathfrak{D}_\nu(X)$.

Conversely, let ν be a continuous capacity. The fact that $\mathfrak{D}_\nu$ satisfies Properties (R1)

and (R2) follows from Lemma 4.1 and Lemma 4.2, respectively. Let u be a vNM utility function. By Liu et al. [2024, Theorem 1] and Proposition 4.12, for all $\alpha \in (0,1)$,

$$\mathrm{VaR}_\alpha(\mathfrak{D}_\nu(u(X))) = G^\nu_\alpha(u(X)) = u(G^\nu_\alpha(X)) = u(\mathrm{VaR}_\alpha(\mathfrak{D}_\nu(X)) = \mathrm{VaR}_\alpha(u_\#(\mathfrak{D}_\nu(X))).$$

Therefore, $\mathfrak{D}_\nu(u(X)) = u_\#(\mathfrak{D}_\nu(X))$. Thus $\mathfrak{D}_\nu$ satisfies Property (R3). □

4.10 Proofs accompanying Section 4.4

The following Lemma is a variant of Van Oosten and Wang [2025a, Proposition 4]. The same arguments in that proof can be used to prove the following lemma.

Lemma 4.3. *Given a preference relation $\succsim$ satisfying Axioms (M), (SRM), and (C), for each $X \in \mathcal{X}$, there exists a unique $c_X \in \mathbb{R}$ such that $X \simeq c_X$. Moreover, if $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ is a bounded sequence converging pointwise to $X \in \mathcal{X}$, $\lim_{n\to\infty} c_{X_n} = c_X$.*

Let $\mathcal{J}$ denote the sub-σ-algebra generated by the sets $\{S_r\}_{r\in[0,1]}$ from Assumption 4.4, and define

$$\mathcal{X}(\mathcal{J}) = \{X \in \mathcal{X} : X^{-1}(B) \in \mathcal{J} \text{ for all } B \in \mathcal{B}(\mathbb{R})\}.$$

Lemma 4.4. *It holds that $\mathcal{J} \subseteq \mathcal{G}$ and $(\Omega, \mathcal{J}, \mathbb{P}|_{\mathcal{J}})$ is an atomless probability space.*

Proof. Since the sets $\{S_r\}_{r\in[0,1]}$ forms a π-system, we have, by the π-λ theorem, $\mathcal{J} \subseteq \mathcal{G}$. By the definition of the exogenous likelihood mapping $\mathbb{P}$, $(\Omega, \mathcal{J}, \mathbb{P}|_{\mathcal{J}})$ is a probability space.

For the sake of contradiction, assume there exists an atom $A \in \mathcal{J}$. That is, $a := \mathbb{P}(A) > 0$ and for all $B \in \mathcal{J}$ satisfying $B \subseteq A$, it holds that either $\mathbb{P}(B) = a$ or $\mathbb{P}(A\backslash B) = a$. Define the function $f : [0,1] \to \{0, a\}$ by $f(r) = \mathbb{P}(A \cap S_r)$. The function f is well-defined as A is an atom. Let $b = \inf\{r \in [0,1] : f(r) = a\}$, which exists as $f(1) = a$.

We claim that $b > 0$. To see this, for the sake of contradiction, assume that $b = 0$. Therefore, there exists a sequence $(r_n)_{n\in\mathbb{N}} \subseteq (0,1]$ such that $r_n \downarrow 0$ and $f(r_n) = a$. Find $m \in \mathbb{N}$ such that $r_m < a$. We get the contradiction

$$a = f(r_m) = \mathbb{P}(A \cap S_{r_m}) \leq \mathbb{P}(S_{r_m}) = r_m < a.$$

Therefore, $b > 0$.

We claim that $b < 1$. To see this, for the sake of contradiction, assume that $f(r) = 0$ for all $r \in [0,1)$. Therefore, for all $r \in [0,1)$, $\mathbb{P}(S^c_r \cap A) = a$. However, taking $r_0 = 1 - a/2$,

we get the contradiction

$$a = \mathbb{P}(S_{r_0}^c \cap A) \leq \mathbb{P}(S_{r_0}^c) = 1 - r_0 = a/2.$$

Therefore, $b < 1$.

Find $\epsilon_0 > 0$ such that $[b - \epsilon_0, b + \epsilon_0] \subseteq [0, 1]$. Fix $\epsilon \in (0, \epsilon_0)$, since $f(b + \epsilon) = a$ and $f(b - \epsilon) = 0$,

$$a = f(b + \epsilon) - f(b - \epsilon) = \mathbb{P}(A \cap (S_{b+\epsilon} \backslash S_{b-\epsilon})) \leq \mathbb{P}(S_{b+\epsilon} \backslash S_{b-\epsilon}) = 2\epsilon. \tag{4.21}$$

As (4.21) holds for arbitrary $\epsilon \in (0, \epsilon_0)$, we have $a \leq 0$, which is a contradiction. Therefore $(\Omega, \mathcal{J}, \mathbb{P}|_{\mathcal{J}})$ is atomless. □

The following corollary follows from Föllmer and Schied [2016, Proposition A.31].

Corollary 4.1. *There exists $U : \Omega \to (0, 1)$ such that $U \in \mathcal{X}(\mathcal{J}) \subseteq \mathcal{X}(\mathcal{G})$ and $U_{\#}\mathbb{P}$ is a standard uniform distribution.*

Proof of Proposition 4.3. Given $A \in \mathcal{F}$, define

$$L_A = \{r \in [0, 1] : A \succsim S_r\} \quad \text{and} \quad U_A = \{r \in [0, 1] : S_r \succsim A\}.$$

By Axioms (RC) and (M), we have that $A \succsim \varnothing \simeq S_0$ and $S_1 \simeq \Omega \succsim A$. Therefore, both L_A and U_A are non-empty. Define $l_A = \sup(L_A)$ and $u_A = \inf(U_A)$. We claim that $l_A = u_A$. If $l_A = 1$, then for all $r \in [0, 1)$, there exists $r' \in (r, 1)$ such that $A \succsim S_{r'}$. Therefore, by Axiom (SRM), $A \succsim S_{r'} \succ S_r$, and $r \notin U_A$. Thus, $u_A = 1$. If $l_A = 0$, then for all $r \in (0, 1]$, $S_r \succ A$, and $r \in U_A$. Thus, $u_A = 0$. Finally, if $l_A \in (0, 1)$, for all $r \in [0, l_A)$, there exists $r' \in (r, l_A)$ such that $A \succsim S_{r'}$. By Axiom (SRM), $A \succsim S_{r'} \succ S_r$, and $r \notin U_A$. Therefore, $u_A \geq l_A$. Additionally, as $S_r \succ A$ for all $r \in (l_A, 1]$, we have that $r \in U_A$. Therefore, $u_A \leq l_A$. Define $r_A = l_A = u_A$.

Take $(r_n)_{n\in\mathbb{N}} \subseteq [0, r_A]$ such that $A \succsim S_{r_n}$ for all $n \in \mathbb{N}$ and $r_n \uparrow r_A$. Define $S \in \mathcal{J}$ by $S = \bigcup_{n=1}^{\infty} S_{r_n}$. By Axiom (C), we have that $A \succsim S$. Additionally,

$$\mathbb{P}(S) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} S_{r_n}\right) = \lim_{n\to\infty} \mathbb{P}(S_{r_n}) = \lim_{n\to\infty} r_n = r_A.$$

Therefore, by axiom (RC), $A \succsim S \simeq S_{r_A}$. A similar argument will show that $S_{r_A} \succsim A$.

Define $\nu : \mathcal{F} \to [0, 1]$ by $\nu(A) = r_A$. Since, for all $A \in \mathcal{F}$, $A \simeq S_{r_A}$, ν is a $\succsim$-matching probability. Let $\tilde{\nu} : \mathcal{F} \to [0, 1]$ be a $\succsim$-matching probability. Given $A \in \mathcal{F}$, by

definition, there exists $R_A \in \mathcal{G}$ such that $A \simeq R_A$ and $\tilde{\nu}(A) = \mathbb{P}(R_A)$. By Axiom (RC), $A \simeq R_A \simeq S_{\tilde{\nu}(A)}$. By the definition of L_A and U_A, $\nu(A) = u_A \leq \tilde{\nu}(A) \leq l_A = \nu(A)$. Since $A \in \mathcal{F}$ was arbitrary, $\nu = \tilde{\nu}$. Therefore, ν is the unique $\succsim$-matching probability.

By Axioms (M) and (SRM), we have

$$A \succsim B \iff S_{\nu(A)} \succsim S_{\nu(B)} \iff \nu(A) \geq \nu(B), \quad A, B \in \mathcal{F}. \tag{4.22}$$

Additionally, by Axiom (M), (4.22) implies that ν is a capacity. Let $A \in \mathcal{G}$, by Axiom (RC), $A \simeq S_{\mathbb{P}(A)}$. By the definition of L_A and U_A, $\nu(A) = u_A \leq \mathbb{P}(A) \leq l_A = \nu(A)$. Since $A \in \mathcal{G}$ was arbitrary, ν is risk conforming.

Finally, we need to show that ν is continuous. Let $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ be an increasing sequence and define $A = \bigcup_{n=1}^{\infty} A_n$. By (4.22), $(\nu(A_n))_{n\in\mathbb{N}}$ is an increasing sequence. Let $r = \lim_{n\to\infty} \nu(A_n)$. By Axiom (M), $A \succsim A_n \simeq S_{\nu(A_n)}$ for all $n \in \mathbb{N}$. By Axiom (C), $A \succsim \bigcup_{n=1}^{\infty} S_{\nu(A_n)}$. Since $\bigcup_{n=1}^{\infty} S_{\nu(A_n)} \in \mathcal{J}$ and

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} S_{\nu(A_n)}\right) = \lim_{n\to\infty} \mathbb{P}\left(S_{\nu(A_n)}\right) = \lim_{n\to\infty} \nu(A_n) = r,$$

by Axiom (RC), $A \succsim S_r$. Additionally, by Axiom (M), $S_r \succsim S_{\nu(A_n)} \simeq A_n$ for all $n \in \mathbb{N}$. Therefore, by Axiom (C), $S_r \succsim A$. By the definition of L_A and U_A, $\nu(A) = u_A \leq r \leq l_A = \nu(A)$. Thus, $\nu(A) = r = \lim_{n\to\infty} \nu(A_n)$ and ν is upward continuous. A similar proof will show that ν is downward continuous. Therefore, ν is continuous. □

Proof of Proposition 4.4. Define the function $I : \mathcal{X}(\mathcal{G}) \to \mathbb{R}$ by $I(X) = c_X$, where c_X is the unique certainty equivalent from Lemma 4.3. Let $U \in \mathcal{X}(\mathcal{J})$ be the act from Corollary 4.1. Given $Q \in \mathcal{D}$, it is routine to check that $q_Q(U) \in \mathcal{X}(\mathcal{G})$ and $(q_Q(U))_{\#}\mathbb{P} = Q$. Define the functional $\gamma : \mathcal{D} \to \mathbb{R}$ by $\gamma(Q) = I(q_Q(U))$. Since $\succsim$ satisfies Axiom (RC), we have that $I(X) = \gamma(X_{\#}\mathbb{P})$ for all $X \in \mathcal{X}(\mathcal{G})$. Therefore,

$$X \succsim Y \iff I(X) \geq I(Y) \iff \gamma(X_{\#}\mathbb{P}) \geq \gamma(Y_{\#}\mathbb{P}), \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

We claim that γ is a statistic.

Let $Q, P \in \mathcal{D}$ satisfy $Q >_{\text{fsd}} P$. By Axiom (SRM), we have $q_Q(U) \succ q_P(U)$. Therefore,

$$\gamma(Q) = I(q_Q(U)) > I(q_P(U)) = \gamma(P).$$

Let $(Q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ be a sequence with uniformly bounded support converging in distribution to $Q \in \mathcal{D}$. Let $C \subseteq (0, 1)$ denote the points of discontinuity for q_Q. Since q_Q

is increasing, C is at most countable. Therefore, C is Borel measurable. Define the set $A = U^{-1}(C) \in \mathcal{G}$, and the acts $X_n = \mathbb{1}_{A^c} q_{Q_n}(U)$ and $X = \mathbb{1}_{A^c} q_Q(U)$. For all $B \in \mathcal{B}(\mathbb{R})$, we have

$$X^{-1}(B) = \begin{cases} U^{-1}(q_Q^{-1}(B) \cap C^c) \cup U^{-1}(C) & \text{if } 0 \in B \\ U^{-1}(q_Q^{-1}(B) \cap C^c) & \text{if } 0 \notin B \end{cases}.$$

Since $U^{-1}(q_Q^{-1}(B) \cap C^c) \cap U^{-1}(C) = \varnothing$, we have that $X^{-1}(B) \in \mathcal{G}$. Therefore, $X \in \mathcal{X}(\mathcal{G})$. A similar argument will show that $X_n \in \mathcal{X}(\mathcal{G})$ for all $n \in \mathbb{N}$. As $U_{\#}\mathbb{P}$ is a standard uniform distribution, $\mathbb{P}(A) = \mathbb{P}(U \in C) = 0$. Therefore, $X_n =^{\mathbb{P}}_{\text{fsd}} q_{Q_n}(U)$ for all $n \in \mathbb{N}$ and $X =^{\mathbb{P}}_{\text{fsd}} q_Q(U)$. By Axiom (RC), we have that $X_n \simeq q_{Q_n}(U)$ for all $n \in \mathbb{N}$ and $X \simeq q_Q(U)$. We claim that $X_n \to X$ pointwise. Let $\omega \in A^c$, then, by the definition of A, $U(\omega)$ is a point of continuity of q_Q. Therefore,

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} q_{Q_n}(U(\omega)) = q_Q(U(\omega)) = X(\omega),$$

as $Q_n \to Q$ in distribution. If $\omega \in A$, for all $n \in \mathbb{N}$, $X_n(\omega) = X(\omega) = 0$. Therefore, $X_n \to X$ pointwise. Since $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}(\mathcal{G})$ is a bounded sequence, by Lemma 4.3, we have

$$\lim_{n\to\infty} \gamma(Q_n) = \lim_{n\to\infty} I(q_{Q_n}(U)) = \lim_{n\to\infty} I(X_n) = I(X) = I(q_Q(U)) = \gamma(Q).$$

Thus γ is a statistic.

The proof of the uniqueness claim for γ in (4.10) follows from Proposition 4.1. □

Proof of Theorem 4.2. Define the function $I : \mathcal{X} \to \mathbb{R}$ by $I(X) = c_X$, where c_X is the unique certainty equivalent from Lemma 4.3. Let ν denote the $\gtrsim$-matching probability. By Proposition 4.3, ν is continuous risk-conforming capacity. We claim that for all $X, Y \in \mathcal{X}$,

$$\mathfrak{D}_\nu(X) = \mathfrak{D}_\nu(Y) \implies I(X) = I(Y).$$

Fix $X, Y \in \mathcal{X}$ satisfying $\mathfrak{D}_\nu(X) = \mathfrak{D}_\nu(Y)$. Therefore, for all $x \in \mathbb{R}$, $\nu(X > x) = \nu(Y > x)$. By Proposition 4.3, $\{X > x\} \simeq \{Y > x\}$. As $x \in \mathbb{R}$ was general, by Axiom (CD), $X \simeq Y$ and $I(X) = I(Y)$. Therefore, there exists $\gamma : \mathcal{D} \to \mathbb{R}$ such that $I(X) = \gamma \circ \mathfrak{D}_\nu(X)$ for all $X \in \mathcal{X}$. Let $U \in \mathcal{X}(\mathcal{J})$ be the act from Corollary 4.1. As ν is risk conforming, we have that

$$\gamma(Q) = \gamma((q_Q(U))_{\#}\mathbb{P}) = \gamma \circ \mathfrak{D}_\nu(q_Q(U)) = I(q_Q(U)), \quad Q \in \mathcal{D}.$$

As this is the same functional from the proof of Proposition 4.4, γ is statistic. Thus,

$$X \gtrsim Y \iff I(X) \geq I(Y) \iff \gamma \circ \mathfrak{D}_\nu(X) \geq \gamma \circ \mathfrak{D}_\nu(Y), \quad X, Y \in \mathcal{X}.$$

For the converse, assume that there exists a statistic γ and a risk-conforming continuous capacity ν such that (4.10) holds. As γ is a statistic, by Lemma 4.1, $\succsim$ satisfies Axiom (M). Since ν is risk conforming, it is clear that $\succsim$ satisfies Axiom (RC). Since ν is risk conforming and γ is a statistic, it is clear that $\succsim$ satisfies Axiom (SRM). Fix $X \in \mathcal{X}$, and let $(Z_n)_{n\in\mathbb{N}} \subseteq \{Y \in \mathcal{X} : X \succsim Y\}$ be a bounded sequence converging pointwise to $Z \in \mathcal{X}$. Since γ is a statistic, by Lemma 4.2,

$$\gamma \circ \mathfrak{D}_\nu(Z) = \lim_{n\to\infty} \gamma \circ \mathfrak{D}_\nu(Z_n) \leq \gamma \circ \mathfrak{D}_\nu(X).$$

Therefore, $X \succsim Z$ and $\{Y \in \mathcal{X} : X \succsim Y\}$ is B-closed. A similar proof will show that $\{Y \in \mathcal{X} : Y \succsim X\}$ is B-closed. Therefore, $\succsim$ satisfies Axiom (C).

Define the function $\varphi : [0,1] \to \mathbb{R}$ by $\varphi(x) = \gamma(x\delta_1 + (1-x)\delta_0)$. As γ is a statistic, φ is strictly increasing and continuous. Let $X, Y \in \mathcal{X}$ satisfy $X \simeq^* Y$. Therefore, given $x \in \mathbb{R}$,

$$\varphi(\nu(X > x)) = \gamma \circ \mathfrak{D}_\nu\left(\mathbb{1}_{\{X>x\}}\right) = \gamma \circ \mathfrak{D}_\nu\left(\mathbb{1}_{\{Y>x\}}\right) = \varphi(\nu(Y > x)).$$

Therefore, $\nu(X > x) = \nu(Y > x)$ for all $x \in \mathbb{R}$, and $\mathfrak{D}_\nu(X) = \mathfrak{D}_\nu(Y)$. Thus, $\gamma \circ \mathfrak{D}_\nu(X) = \gamma \circ \mathfrak{D}_\nu(Y)$, and $X \simeq Y$. Therefore $\succsim$ satisfies Axiom (CD).

The proof of the uniqueness claim for γ in (4.10) follows from Proposition 4.1. Finally, we will show that ν in (4.10) is the $\succsim$-matching probability. Let $A \in \mathcal{F}$ and $R_A \in \mathcal{G}$ with $\mathbb{P}(R_A) = \nu(A)$. Then, using the function φ defined above,

$$\gamma \circ \mathfrak{D}_\nu(\mathbb{1}_A) = \varphi(\nu(A)) = \varphi(\mathbb{P}(R_A)) = \varphi(\nu(R_A)) = \gamma \circ \mathfrak{D}_\nu(\mathbb{1}_{R_A}).$$

Therefore $A \simeq R_A$ and ν is the $\succsim$-matching probability. □

4.11 Proofs and additional results accompanying Section 4.5

Given a family of subsets $\mathfrak{S}$ of some set, let $\sigma(\mathfrak{S})$ denote the smallest σ-algebra containing the family $\mathfrak{S}$.

Proof of Proposition 4.5. Let $X, Y \in \mathcal{X}$ be comonotonic. Define $\mathcal{B}_0(\mathbb{R}^2) = \{(x,\infty) \times (y,\infty) : x, y \in \mathbb{R}\}$. By Folland [1999, Proposition 1.4], $\sigma(\mathcal{B}_0(\mathbb{R}^2)) = \mathcal{B}(\mathbb{R}^2)$, where $\mathcal{B}(\mathbb{R}^2)$ denotes the Borel σ-algebra on $\mathbb{R}^2$. Define the mapping $\tau : \Omega \to \mathbb{R}^2$ by $\tau(\omega) = (X(\omega), Y(\omega))$.

Define $\mathcal{H} = \{\tau^{-1}(B) : B \in \mathcal{B}_0(\mathbb{R}^2)\}$. Fix $x, y \in \mathbb{R}$, we have that

$$\tau^{-1}((x,\infty) \times (y,\infty)) = \{X > x\} \cap \{Y > y\}.$$

Since X and Y are comonotonic, by Denneberg [1994, Proposition 4.5], either $\{X > x\} \subseteq \{Y > y\}$ or $\{Y > y\} \subseteq \{X > x\}$. Therefore, $\tau^{-1}((x,\infty) \times (y,\infty)) \in \mathcal{G}$, and $\mathcal{H} \subseteq \mathcal{G}$. Since $\mathcal{B}_0(\mathbb{R}^2)$ is a π-system, $\mathcal{H}$ is a π-system. Therefore, by the π-λ theorem, $\sigma(\mathcal{H}) \subseteq \mathcal{G}$. However, we have

$$\begin{aligned}\left\{\tau^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^2)\right\} &= \left\{\tau^{-1}(B) : B \in \sigma(\mathcal{B}_0(\mathbb{R}^2))\right\} \\ &= \sigma\left(\left\{\tau^{-1}(B) : B \in \mathcal{B}_0(\mathbb{R}^2)\right\}\right) = \sigma(\mathcal{H}) \subseteq \mathcal{G}.\end{aligned}$$

Define $\varphi : \mathbb{R}^2 \to \mathbb{R}$ by $\varphi(x,y) = x + y$. Since $\varphi^{-1}(B) \in \mathcal{B}(\mathbb{R}^2)$ for all $B \in \mathcal{B}(\mathbb{R})$, we have that $(X+Y)^{-1}(B) = \tau^{-1}(\varphi^{-1}(B)) \in \mathcal{G}$ for all $B \in \mathcal{B}(\mathbb{R})$. Thus, $X + Y \in \mathcal{X}(\mathcal{G})$.

Let $X, Y \in \mathcal{X}(\mathcal{G})$ be comonotonic and $\lambda \in [0,1]$. Since λX and $(1-\lambda)Y$ are comonotonic, $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$ by the previous result. □

Proof of Proposition 4.6. Let $X, Y \in \mathcal{X}$ be comonotonic. Then, by Proposition 4.12, for all $\alpha \in (0,1)$ it holds that

$$\begin{aligned}\mathrm{VaR}_\alpha(\mathfrak{D}_\nu(X+Y)) = \mathrm{G}^\nu_\alpha(X+Y) = \mathrm{G}^\nu_\alpha(X) + \mathrm{G}^\nu_\alpha(Y) &= \mathrm{VaR}_\alpha(\mathfrak{D}_\nu(X)) + \mathrm{VaR}_\alpha(\mathfrak{D}_\nu(Y)) \\ &= \mathrm{VaR}_\alpha(\mathfrak{D}_\nu(X) \oplus \mathfrak{D}_\nu(Y)),\end{aligned}$$

as X and Y are comonotonic and G^ν_α is comonotonic additive. Given $\lambda \in [0,1]$, an identical proof will show that $\mathfrak{D}_\nu(\lambda X) = \lambda \otimes \mathfrak{D}_\nu(X)$ since $\mathrm{G}^\nu_\alpha(aX) = a\mathrm{G}^\nu_\alpha(X)$ for all $X \in \mathcal{X}$ and non-negative $a \in \mathbb{R}$. □

We now introduce a property that strengthens Property (RD), originally proposed in Chateauneuf and Lakhnati [2007], which will be utilized in the proof of Theorem 4.3.

(RDI) Risk diversification for identically distributed acts: For all $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$ such that $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$,

$$X =^{\mathbb{P}}_{\mathrm{fsd}} Y \implies \lambda X + (1-\lambda)Y \succsim X.$$

Clearly, for a preference relation satisfying Axiom (RC), Property (RDI) is implied by Property (RD). The following proposition generalizes Chateauneuf and Lakhnati [2007, Theorem 4.2].

Proposition 4.14. *Let $\succsim$ be a preference relation satisfying the regularity axioms. Then $\succsim$ satisfies Property (SRA) if and only if $\succsim$ satisfies Property (RDI).*

Before providing the proof of Proposition 4.14, we need the following lemma.

Lemma 4.5. *Suppose that $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$ satisfy $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$. Then there exist $X', Y' \in \mathcal{X}(\mathcal{J})$ such that $X =^{\mathbb{P}}_{\text{fsd}} X'$, $Y =^{\mathbb{P}}_{\text{fsd}} Y'$, and $\lambda X + (1-\lambda)Y =^{\mathbb{P}}_{\text{fsd}} \lambda X + (1-\lambda)Y$.*

Proof. By Assumption 4.3, there exists a risk-conforming $\mu \in \Delta$. Since $(\Omega, \mathcal{J}, \mathbb{P}|_{\mathcal{J}})$ is atomless (Lemma 4.4), by Delbaen [2002, Theorem 11.7.5], there exists $X', Y' \in \mathcal{X}(\mathcal{J})$ such that $X_{\#}\mu = X'_{\#}\mathbb{P}$, $Y_{\#}\mu = Y'_{\#}\mathbb{P}$, and $(\lambda X + (1-\lambda)Y)_{\#}\mu = (\lambda X' + (1-\lambda)Y')_{\#}\mathbb{P}$. The result follows since μ is risk conforming. □

Proof of Proposition 4.14. Let γ denote the $\succsim$-CES. By Proposition 4.4,

$$X \succsim Y \iff \gamma(X_{\#}\mathbb{P}) \geq \gamma(Y_{\#}\mathbb{P}), \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

Assume that $\succsim$ satisfies Property (RDI). Since the restriction of $\succsim$ to $\mathcal{X}(\mathcal{J})$ satisfies the conditions of the preference relations in Chateauneuf and Lakhnati [2007], it follows from Chateauneuf and Lakhnati [2007, Theorem 4.2] that, for all $X, Y \in \mathcal{X}(\mathcal{J})$, if $X \geq^{\mathbb{P}}_{\text{ssd}} Y$ then $X \succsim Y$. Let $U \in \mathcal{X}(\mathcal{J})$ be the act from Corollary 4.1. Given $Q, P \in \mathcal{D}$ satisfying $Q \geq_{\text{ssd}} P$, we have $q_Q(U) \geq^{\mathbb{P}}_{\text{ssd}} q_P(U)$. Therefore, $q_Q(U) \succsim q_P(U)$, and

$$\gamma(Q) = \gamma((q_Q(U))_{\#}\mathbb{P}) \geq \gamma((q_P(U))_{\#}\mathbb{P}) = \gamma(P).$$

Thus, γ is monotonic with respect to $\geq_{\text{ssd}}$, and $\succsim$ satisfies Property (SRA).

Conversely, assume that $\succsim$ satisfies Property (SRA). Since the restriction of $\succsim$ to $\mathcal{X}(\mathcal{J})$ satisfies the conditions of the preference relations in Chateauneuf and Lakhnati [2007], it follows from Chateauneuf and Lakhnati [2007, Theorem 4.2] that, for all $X, Y \in \mathcal{X}(\mathcal{J})$ and $\lambda \in [0,1]$,

$$X =^{\mathbb{P}}_{\text{fsd}} Y \implies \lambda X + (1-\lambda)Y \succsim X.$$

Let $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$ satisfy $\lambda X + (1-\lambda)Y \in \mathcal{X}(\mathcal{G})$. Let $X', Y' \in \mathcal{X}(\mathcal{J})$ be the acts from Lemma 4.5. Since $X' =^{\mathbb{P}}_{\text{fsd}} Y'$, we have that $\lambda X' + (1-\lambda)Y' \succsim X'$. Therefore,

$$\gamma((\lambda X + (1-\lambda)Y)_{\#}\mathbb{P}) = \gamma((\lambda X' + (1-\lambda)Y')_{\#}\mathbb{P}) \geq \gamma(X'_{\#}\mathbb{P}) = \gamma(X_{\#}\mathbb{P}).$$

Therefore $\lambda X + (1-\lambda)Y \succsim X$, and $\succsim$ satisfies Property (RDI). □

For a preference relation satisfying the regularity axioms, Property (RD) implies Property (SRA). As shown by Dekel [1989], the two properties are not equivalent in general. Therefore, we need the following results to help bridge this gap.

Lemma 4.6. *Let $X, Y \in \mathcal{X}(\mathcal{G})$ be comonotonic and $\lambda \in [0,1]$. Then*

$$(\lambda X + (1-\lambda)Y)_{\#}\mathbb{P} = \lambda \otimes X_{\#}\mathbb{P} \oplus (1-\lambda) \otimes Y_{\#}\mathbb{P}.$$

Proof. By Assumption 4.3, there exists a risk-conforming $\mu \in \Delta$. By Proposition 4.6, for all comonotonic $X, Y \in \mathcal{X}(\mathcal{G})$ and $\lambda \in [0,1]$, we have

$$\begin{aligned}(\lambda X + (1-\lambda)Y)_{\#}\mathbb{P} = \mathfrak{D}_{\mu}(\lambda X + (1-\lambda)Y) &= \lambda \otimes \mathfrak{D}_{\mu}(X) \oplus (1-\lambda) \otimes \mathfrak{D}_{\mu}(Y) \\ &= \lambda \otimes X_{\#}\mathbb{P} \oplus (1-\lambda) \otimes Y_{\#}\mathbb{P}.\end{aligned}$$

Therefore, $(\lambda X + (1-\lambda)Y)_{\#}\mathbb{P} = \lambda \otimes X_{\#}\mathbb{P} \oplus (1-\lambda) \otimes Y_{\#}\mathbb{P}$. □

Comonotonic quasiconcavity of the $\succsim$-CES characterizes Property (RDC).

Proposition 4.15. *Let $\succsim$ be a preference relation satisfying the regularity axioms. Then $\succsim$ satisfies Property (RDC) if and only if the $\succsim$-CES is comonotonic quasiconcave.*

Proof. Let γ denote the $\succsim$-CES. By Proposition 4.4,

$$X \succsim Y \iff \gamma(X_{\#}\mathbb{P}) \geq \gamma(Y_{\#}\mathbb{P}), \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

Assume that $\succsim$ satisfies Property (RDC). Fix $Q, P \in \mathcal{D}$, without loss of generality, we may assume that $\gamma(Q) \geq \gamma(P)$. Given $\lambda \in [0,1]$, we need to show that $\gamma(\lambda \otimes Q \oplus (1-\lambda) \otimes P) \geq \gamma(P)$. Let $U \in \mathcal{X}(\mathcal{J})$ be the act from Corollary 4.1. Define $X_Q = q_Q(U)$ and $X_P = q_P(U)$, which are both elements of $\mathcal{X}(\mathcal{G})$. We have that $(X_Q)_{\#}\mathbb{P} = Q$ and $(X_P)_{\#}\mathbb{P} = P$. Since $\gamma(Q) \geq \gamma(P)$, this implies that $X_Q \succsim X_P$.

We claim that there exists $c_0 \in [0, \infty)$ such that $X_Q - c_0 \simeq X_P$. Define

$$c_0 = \sup\{c \in [0,\infty) : X_Q - c \succsim X_P\}.$$

Let $c_1 = \|X_Q\| + \|X_P\|$. We have that $X_Q - c_1 \leq -\|X_P\| \leq X_P$. By Axioms (M) and (SRM), for all $c > c_1$, $X_P \succ X_Q - c$. Thus $\{c \in [0,\infty) : X_Q - c \succsim X_P\} \subseteq [0, c_1]$ and $c_0 \in [0,\infty)$. A simple consequence of Axiom (C) is that $X_Q - c_0 \succsim X_P$. Additionally, by the definition of c_0, $X_P \succ X_Q - c$ for all $c > c_0$. Thus, by Axiom (C), $X_P \succsim X_Q - c_0$. Therefore, $X_P \simeq X_Q - c_0$.

It is clear that X_Q and X_P are comonotonic. Therefore, $X_Q - c_0$ and X_P are comonotonic and, by Axiom (M) and Property (RDC),

$$\lambda X_Q + (1-\lambda)X_P \succsim \lambda(X_Q - c_0) + (1-\lambda)X_P \succsim X_P.$$

Thus, by Lemma 4.6, we have

$$\gamma(\lambda \otimes Q \oplus (1-\lambda) \otimes P) = \gamma((\lambda X_Q + (1-\lambda)X_P)_{\#}\mathbb{P}) \geq \gamma((X_P)_{\#}\mathbb{P}) = \gamma(P).$$

Conversely, assume that γ is comonotonic quasiconcave. Let $X, Y \in \mathcal{X}(\mathcal{G})$ be comonotonic such that $X \simeq Y$ and $\lambda \in [0,1]$. As λX and $(1-\lambda)Y$ are comonotonic, by Lemma 4.6 and the fact that γ is comonotonic quasiconcave,

$$\begin{aligned}\gamma\left((\lambda X + (1-\lambda)Y)_{\#}\mathbb{P}\right) &= \gamma\left(\lambda \otimes X_{\#}\mathbb{P} \oplus (1-\lambda) \otimes Y_{\#}\mathbb{P}\right) \\ &\geq \min\left\{\gamma(X_{\#}\mathbb{P}), \gamma(Y_{\#}\mathbb{P})\right\} = \gamma(X_{\#}\mathbb{P}).\end{aligned}$$

Therefore, $\lambda X + (1-\lambda)Y \succsim X$, and $\succsim$ satisfies Property (RDC). □

We are now ready to present the proof of Theorem 4.3.

Proof of Theorem 4.3. (i) $\Longrightarrow$ (ii): It is clear that $\succsim$ satisfies Property (RDC). Moreover, it is clear that $\succsim$ satisfies Property (RDI). Therefore, by Proposition 4.14, $\succsim$ satisfies Property (SRA).

(ii) $\Longrightarrow$ (iii): Follows from Proposition 4.15.

(iii) $\Longrightarrow$ (i): Fix $X, Y \in \mathcal{X}(\mathcal{G})$ satisfying $X \succsim Y$ and $\lambda \in [0,1]$. Let $X', Y' \in \mathcal{X}(\mathcal{J})$ be the acts given by Lemma 4.5. By the comonotonic sum inequality (see Dhaene et al. [2002, Theorem 7]), there exists $X'_c, Y'_c \in \mathcal{X}(\mathcal{J})$ such that $X' =^{\mathbb{P}}_{\text{fsd}} X'_c$, $Y' =^{\mathbb{P}}_{\text{fsd}} Y'_c$, and

$$\begin{aligned}(\lambda X' + (1-\lambda)Y')_{\#}\mathbb{P} \geq_{\text{ssd}} (\lambda X'_c + (1-\lambda)Y'_c)_{\#}\mathbb{P} &= \lambda \otimes (X'_c)_{\#}\mathbb{P} \oplus (1-\lambda) \otimes (Y'_c)_{\#}\mathbb{P} \\ &= \lambda \otimes X_{\#}\mathbb{P} \oplus (1-\lambda) \otimes Y_{\#}\mathbb{P}.\end{aligned}$$

Therefore, since the $\succsim$-CES γ is monotonic with respect to $\geq_{\text{ssd}}$ and comonotonic quasiconcave, we have

$$\begin{aligned}\gamma((\lambda X + (1-\lambda)Y)_{\#}\mathbb{P}) = \gamma((\lambda X' + (1-\lambda)Y')_{\#}\mathbb{P}) &\geq \gamma(\lambda \otimes X_{\#}\mathbb{P} \oplus (1-\lambda) \otimes Y_{\#}\mathbb{P}) \\ &\geq \min\{\gamma(X_{\#}\mathbb{P}), \gamma(Y_{\#}\mathbb{P})\} = \gamma(X_{\#}\mathbb{P}).\end{aligned}$$

Therefore, by Proposition 4.4, $\lambda X + (1-\lambda)Y \succsim X$. Thus, $\succsim$ satisfies Property (RD). □

Proof of Proposition 4.7. This is a direct consequence of Proposition 4.6 and Proposition 4.15. □

Proof of Proposition 4.8. Results (i) and (ii) follow directly from Proposition 4.6. To prove (iii), let γ denote the $\succsim$-CES and ν denote the $\succsim$-matching probability. Since γ is biseparable, there exists a vNM utility u and a statistic $\tilde{\gamma}$ satisfying constant additivity and positive homogeneity such that (4.12) holds. Define $I : \mathcal{X} \to \mathbb{R}$ by $I(X) = \tilde{\gamma} \circ \mathfrak{D}_\nu(X)$. It is clear that $I(X) \geq I(Y)$ for all $X, Y \in \mathcal{X}$ with $X \geq Y$. By Proposition 4.6, $I(aX+b) = aI(X)+b$ for all $X \in \mathcal{X}$, non-negative $a \in \mathbb{R}$, and $b \in \mathbb{R}$. By Theorem 4.1, we have

$$I(u(X)) = \tilde{\gamma} \circ \mathfrak{D}_\nu(u(X)) = \tilde{\gamma}(u_\#(\mathfrak{D}_\nu(X))) = \gamma \circ \mathfrak{D}_\nu(X), \quad X \in \mathcal{X}.$$

Therefore,

$$X \succsim Y \iff \gamma \circ \mathfrak{D}_\nu(X) \geq \gamma \circ \mathfrak{D}_\nu(Y) \iff I(u(X)) \geq I(u(Y)), \quad X, Y \in \mathcal{X},$$

and the proof is complete. □

Proof of Theorem 4.4. Assume that $\succsim_2$ is more ambiguity averse than $\succsim_1$. Let $A \in \mathcal{F}$ and $R_1 \in \mathcal{G}$ such that $A \simeq_1 R_1$. By the definition of comparative ambiguity attitudes, $R_1 \succsim_2 A$. Find $R_2 \in \mathcal{G}$ such that $A \simeq_2 R_2$. As $R_1 \succsim_2 R_2$, it holds that $\mathbb{P}(R_1) \geq \mathbb{P}(R_2)$ and $\nu_1(A) \geq \nu_2(A)$. To show the converse, let $R \in \mathcal{G}$ and $A \in \mathcal{F}$ satisfy $R \succsim_1 A$. Furthermore, let $R_1, R_2 \in \mathcal{G}$ satisfy $A \simeq_1 R_1$ and $A \simeq_2 R_2$. Since $\nu_1(A) \geq \nu_2(A)$ and $R \succsim_1 A$, $\mathbb{P}(R) \geq \mathbb{P}(R_1) \geq \mathbb{P}(R_2)$. Therefore, $R \succsim_2 R_2 \simeq_2 A$. □

Proof of Proposition 4.9. Assume that the $\succsim_1$-CES is equal to the $\succsim_2$-CES and $\succsim_2$ is more ambiguity averse than $\succsim_1$. Denote the shared certainty-equivalent statistic by γ. Denote the $\succsim_1$-matching probability by ν_1 and the $\succsim_2$-matching probability by ν_2. By Theorem 4.4, for all $A \in \mathcal{F}$, $\nu_1(A) \geq \nu_2(A)$. Therefore, for all $X \in \mathcal{X}$, $\nu_1(X > x) \geq \nu_2(X > x)$ for all $x \in \mathbb{R}$. Thus, for all $X \in \mathcal{X}$, $\mathfrak{D}_{\nu_1}(X) \geq_{\text{fsd}} \mathfrak{D}_{\nu_2}(X)$. Since γ is a statistic, for all $X \in \mathcal{X}$, we have $\gamma \circ \mathfrak{D}_{\nu_1}(X) \geq \gamma \circ \mathfrak{D}_{\nu_2}(X)$.

Let $X \in \mathcal{X}(\mathcal{G})$ and $Y \in \mathcal{X}$ satisfy $X \succsim_1 Y$ ($X \succ_1 Y$). We have

$$\gamma \circ \mathfrak{D}_{\nu_2}(X) = \gamma(X_\#\mathbb{P}) = \gamma \circ \mathfrak{D}_{\nu_1}(X) \overset{(>)}{\geq} \gamma \circ \mathfrak{D}_{\nu_1}(Y) \geq \gamma \circ \mathfrak{D}_{\nu_2}(Y).$$

Therefore, $X \succsim_2 Y$ ($X \succ_2 Y$).

For the converse, assume that (4.13) holds true. Since (4.13) is stronger than the definition of comparative ambiguity aversion, $\succsim_2$ is more ambiguity averse than $\succsim_1$. Let

$X, Y \in \mathcal{X}(\mathcal{G})$. If $X \succsim_1 Y$, then $X \succsim_2 Y$ by (4.13). If $X \succsim_2 Y$, then $X \succsim_1 Y$ or else there would be a contradiction with (4.13). Thus,

$$X \succsim_1 Y \iff X \succsim_2 Y, \quad X, Y \in \mathcal{X}(\mathcal{G}).$$

Therefore, the $\succsim_1$-CES and the $\succsim_2$-CES both represent the Choquet ATE preference relation $\succsim_1$ on $\mathcal{X}(\mathcal{G})$. Therefore, by Proposition 4.1, the $\succsim_1$-CES equals the $\succsim_2$-CES. □

The following example shows that Property (RD) does not necessarily imply Property (D). In the following example, given $A \in \mathcal{F}$ and $\mu \in \Delta$, we use $\mu(A|\mathcal{H})$ to denote the conditional probability of A under μ given the sub-σ-algebra $\mathcal{H}$. Additionally, in the following example, we will assume that $\mathcal{G}$ is a sub-σ-algebra.

Example 4.7. Assume there exists a risk-conforming $\mu \in \Delta$ and a sub-σ-algebra $\mathcal{H}$, independent of $\mathcal{G}$ under μ, such that there exists $A_0 \in \mathcal{H}$ satisfying $\mu(A_0) = 1/2$. Let $\tilde{g}$ be a concave distortion function satisfying $\tilde{g}(1/2) > 1/2$. Define the distortion function $\tilde{f} = \tilde{g} \circ \tilde{g}$. Let $g = \tilde{g}^{-1}$ and $f = \tilde{f}^{-1}$, it is clear that g is convex. Also, $g(\tilde{f}(1/2)) > 1/2$. Define the capacity

$$\tilde{\nu}(A) = \int_\Omega f(\mu(A|\mathcal{H}))\mathrm{d}\mu, \quad A \in \mathcal{F}.$$

An application of both the monotone convergence theorem and the conditional monotone convergence theorem will verify that $\tilde{\nu}$ is continuous. As $\mathcal{G}$ and $\mathcal{H}$ are independent, it is straightforward to show that for all $A \in \mathcal{G}$, $\tilde{\nu}(A) = f(\mu(A)) = f(\mathbb{P}(A))$. Also, for all $A \in \mathcal{H}$, $\tilde{\nu}(A) = \mu(A)$. Define the capacity ν by $\nu(A) = \tilde{f} \circ \tilde{\nu}(A)$ for all $A \in \mathcal{F}$. It is clear that ν is risk conforming and $\nu(A) = \tilde{f} \circ \mu(A)$ for all $A \in \mathcal{H}$. As f is continuous, ν is continuous. Define the Choquet ATE preference relation $\succsim$ by

$$X \succsim Y \iff \mathbb{E}^g \circ \mathfrak{D}_\nu(X) \iff \mathbb{E}^g \circ \mathfrak{D}_\nu(Y).$$

Since g is convex, by Marinacci and Montrucchio [2004, Corollary 4.2], $\succsim$ satisfies Property (RD). It is clear that $\mathbb{1}_{A_0} \simeq \mathbb{1}_{A_0^c}$. We have

$$\begin{aligned}
&\mathbb{E}^g \circ \mathfrak{D}_\nu \left((1/2)\mathbb{1}_{A_0} + (1/2)\mathbb{1}_{A_0^c} \right) \\
&\quad = \mathbb{E}^g \left(\delta_{(1/2)} \right) \\
&\quad = 1/2 < g\left(\tilde{f}(1/2) \right) = \mathbb{E}^g \left(\tilde{f}(\mu(A_0))\delta_1 + \left(1 - \tilde{f}(\mu(A_0)) \right) \delta_0 \right) = \mathbb{E}^g \circ \mathfrak{D}_\nu \left(\mathbb{1}_{A_0} \right).
\end{aligned}$$

Therefore, $\mathbb{1}_{A_0} \succ (1/2)\mathbb{1}_{A_0} + (1/2)\mathbb{1}_{A_0^c}$ and $\succsim$ does not satisfy Property (D).

Proof of Therorem 4.5. Assume that $\mathfrak{D}_\nu$ is aggregation concave. Fix $A, B \in \mathcal{F}$. By Propo-

sition 4.6, and the fact that $\mathbb{1}_{A\cup B}$ and $\mathbb{1}_{A\cap B}$ are comonotonic, we have

$$\begin{aligned}\mathfrak{D}_\nu((1/2)\mathbb{1}_A + (1/2)\mathbb{1}_B) &= \mathfrak{D}_\nu((1/2)\mathbb{1}_{A\cap B} + (1/2)\mathbb{1}_{A\cup B}) \\ &= (1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_{A\cup B}) \oplus (1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_{A\cap B}).\end{aligned}$$

Furthermore, since $\mathfrak{D}_\nu$ is concave, it holds that

$$\mathfrak{D}_\nu((1/2)\mathbb{1}_A + (1/2)\mathbb{1}_B) \geq_{\text{ssd}} (1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_A) \oplus (1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_B).$$

The following two properties of $\mathbb{E}$ are immediately clear. (a) $\mathbb{E}$ is monotonic with respect to $\geq_{\text{ssd}}$. (b) For all $Q, P \in \mathcal{D}$ and non-negative $a, b \in \mathbb{R}$, it holds $\mathbb{E}(a\otimes Q \oplus b\otimes P) = a\mathbb{E}(Q) + b\mathbb{E}(P)$. Let $A \in \mathcal{F}$, since $\mathfrak{D}_\nu(\mathbb{1}_A) = \nu(A)\delta_1 + (1-\nu(A))\delta_0$, we have that $\mathbb{E}(\mathfrak{D}_\nu(\mathbb{1}_A)) = \nu(A)$. Therefore, we have

$$\begin{aligned}\frac{1}{2}\nu(A\cap B) + \frac{1}{2}\nu(A\cup B) &= \mathbb{E}\left[(1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_{A\cup B}) \oplus (1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_{A\cap B})\right] \\ &= \mathbb{E}\left[\mathfrak{D}_\nu((1/2)\mathbb{1}_A + (1/2)\mathbb{1}_B)\right] \\ &\geq \mathbb{E}\left[(1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_A) \oplus (1/2)\otimes\mathfrak{D}_\nu(\mathbb{1}_B)\right] \\ &= \frac{1}{2}\nu(A) + \frac{1}{2}\nu(B).\end{aligned}$$

Thus, $\nu(A\cap B) + \nu(A\cup B) \geq \nu(A) + \nu(B)$ and ν is supermodular.

For the converse, let $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$. Since ν is supermodular, H^ν_α is concave for all $\alpha \in (0,1)$; see Section 4.7. Therefore, by Proposition 4.12, we have

$$\begin{aligned}\int_0^\alpha \mathrm{VaR}_\beta\left(\mathfrak{D}_\nu\left(\lambda X + (1-\lambda)Y\right)\right)\mathrm{d}\beta &= \int_0^\alpha \mathrm{G}^\nu_\beta\left(\lambda X + (1-\lambda)Y\right)\mathrm{d}\beta \\ &= \alpha\mathrm{H}^\nu_\alpha\left(\lambda X + (1-\lambda)Y\right) \\ &\geq \alpha\lambda\mathrm{H}^\nu_\alpha(X) + \alpha(1-\lambda)\mathrm{H}^\nu_\alpha(Y) \\ &= \lambda\int_0^\alpha \mathrm{G}^\nu_\beta\left(X\right)\mathrm{d}\beta + (1-\lambda)\int_0^\alpha \mathrm{G}^\nu_\beta\left(Y\right)\mathrm{d}\beta \\ &= \int_0^\alpha \lambda\mathrm{VaR}_\beta\left(\mathfrak{D}_\nu(X)\right) + (1-\lambda)\mathrm{VaR}_\beta\left(\mathfrak{D}_\nu(Y)\right)\mathrm{d}\beta \\ &= \int_0^\alpha \mathrm{VaR}_\beta\left(\lambda\otimes\mathfrak{D}_\nu(X) \oplus (1-\lambda)\otimes\mathfrak{D}_\nu(Y)\right)\mathrm{d}\beta.\end{aligned}$$

Thus, by Shaked and Shanthikumar [2007, Theorem 4.A.3], $\mathfrak{D}_\nu$ is aggregation concave. □

Proof of the claims in Example 4.6. Since μ_1 and μ_2 are supermodular, $\mathrm{H}_\alpha^{\mu_1}$ and $\mathrm{H}_\alpha^{\mu_2}$ are concave for all $\alpha \in (0,1)$; see Section 4.7. Let $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$. we have

$$
\begin{aligned}
&\int_0^\alpha \mathrm{VaR}_\beta\left(\mathfrak{D}\left(\lambda X + (1-\lambda)Y\right)\right) \mathrm{d}\beta \\
&= \frac{1}{2}\int_0^\alpha \mathrm{VaR}_\beta\left(\mathfrak{D}_{\mu_1}\left(\lambda X + (1-\lambda)Y\right)\right) \mathrm{d}\beta + \frac{1}{2}\int_0^\alpha \mathrm{VaR}_\beta\left(\mathfrak{D}_{\mu_2}\left(\lambda X + (1-\lambda)Y\right)\right) \mathrm{d}\beta \\
&= \frac{1}{2}\int_0^\alpha \mathrm{G}_\beta^{\mu_1}\left(\lambda X + (1-\lambda)Y\right) \mathrm{d}\beta + \frac{1}{2}\int_0^\alpha \mathrm{G}_\beta^{\mu_2}\left(\lambda X + (1-\lambda)Y\right) \mathrm{d}\beta \\
&= \frac{\alpha}{2}\mathrm{H}_\alpha^{\mu_1}\left(\lambda X + (1-\lambda)Y\right) + \frac{\alpha}{2}\mathrm{H}_\alpha^{\mu_2}\left(\lambda X + (1-\lambda)Y\right) \\
&\geq \frac{\alpha\lambda}{2}\mathrm{H}_\alpha^{\mu_1}(X) + \frac{\alpha(1-\lambda)}{2}\mathrm{H}_\alpha^{\mu_1}(Y) + \frac{\alpha\lambda}{2}\mathrm{H}_\alpha^{\mu_2}(X) + \frac{\alpha(1-\lambda)}{2}\mathrm{H}_\alpha^{\mu_2}(Y) \\
&= \frac{\lambda}{2}\int_0^\alpha \mathrm{G}_\beta^{\mu_1}(X) \mathrm{d}\beta + \frac{1-\lambda}{2}\int_0^\alpha \mathrm{G}_\beta^{\mu_1}(Y) \mathrm{d}\beta + \frac{\lambda}{2}\int_0^\alpha \mathrm{G}_\beta^{\mu_2}(X) \mathrm{d}\beta + \frac{1-\lambda}{2}\int_0^\alpha \mathrm{G}_\beta^{\mu_2}(Y) \mathrm{d}\beta \\
&= \frac{\lambda}{2}\int_0^\alpha \mathrm{VaR}_\beta\left(\mathfrak{D}_{\mu_1}(X)\right) + \mathrm{VaR}_\beta\left(\mathfrak{D}_{\mu_2}(X)\right) \mathrm{d}\beta \\
&\quad + \frac{1-\lambda}{2}\int_0^\alpha \mathrm{VaR}_\beta\left(\mathfrak{D}_{\mu_1}(Y)\right) + \mathrm{VaR}_\beta\left(\mathfrak{D}_{\mu_2}(Y)\right) \mathrm{d}\beta \\
&= \int_0^\alpha \lambda \mathrm{VaR}_\beta\left(\mathfrak{D}(X)\right) + (1-\lambda)\mathrm{VaR}_\beta\left(\mathfrak{D}(Y)\right) \mathrm{d}\beta \\
&= \int_0^\alpha \mathrm{VaR}_\beta\left(\lambda \otimes \mathfrak{D}(X) \oplus (1-\lambda) \otimes \mathfrak{D}(Y)\right) \mathrm{d}\beta.
\end{aligned}
$$

Thus, by Shaked and Shanthikumar [2007, Theorem 4.A.3], $\mathfrak{D}$ is aggregation concave.

For the sake of contradiction, assume that $\mathfrak{D}$ is ambiguity averse, that is, assume there exists a $\mu \in \Delta$ such that $\mathfrak{D}_\mu(X) \geq_{\mathrm{fsd}} \mathfrak{D}(X)$ for all $X \in \mathcal{X}$. Find $A \in \mathcal{F}$ satisfying $\mu_1(A) > \mu_2(A)$. We claim that $\mu(A) \geq \mu_1(A)$. If $\mu(A) = 1$, this is clear. If $\mu(A) < 1$, define $\alpha' = 1 - \mu(A)$. Given $\epsilon \in (0, \alpha')$, it is straightforward to show that $\mathrm{VaR}_{\alpha'-\epsilon}(\mathfrak{D}_\mu(\mathbb{1}_A)) = 0$. Therefore, $0 = \mathrm{VaR}_{\alpha'-\epsilon}(\mathfrak{D}(\mathbb{1}_A))$. Since for all α,

$$
\mathrm{VaR}_\alpha(\mathfrak{D}(\mathbb{1}_A)) = \begin{cases} 1 & \text{if } \alpha \in [1-\mu_2(A), 1) \\ 1/2 & \text{if } \alpha \in [1-\mu_1(A), 1-\mu_2(A)) \\ 0 & \text{if } \alpha \in (0, 1-\mu_1(A)), \end{cases}
$$

we have $1-\mu_1(A) > \alpha'-\epsilon$. Since ϵ was arbitrary, $1-\mu_1(A) \geq \alpha'$. Therefore, $\mu(A) \geq \mu_1(A)$.

A similar argument will show that $\mu(A^c) \geq \mu_2(A^c)$. Therefore,

$$1 = \mu(A) + \mu(A^c) \geq \mu_1(A) + \mu_2(A^c) > \mu_2(A) + \mu_2(A^c) = 1,$$

a contradiction. Thus, by contradiction, $\mathfrak{D}$ is not ambiguity averse. □

Lemma 4.7. *Let ν be a continuous capacity. Then ν is exact if and only if $\mathfrak{D}_\nu = \mathfrak{D}_{\mathfrak{C}(\nu)}$.*

Proof. Assume that ν is exact. Given $X \in \mathcal{X}$, by Mao et al. [2025, Proposition 1], we have

$$S_{\mathfrak{D}_{\mathfrak{C}(\nu)}(X)}(x) = \min_{\mu \in \mathfrak{C}(\nu)} \mu(X > x) = \nu(X > x) = S_{\mathfrak{D}_\nu(X)}(x), \quad x \in \mathbb{R}.$$

Therefore, $\mathfrak{D}_\nu = \mathfrak{D}_{\mathfrak{C}(\nu)}$.

For the converse, let $A \in \mathcal{F}$. By Mao et al. [2025, Proposition 1], we have

$$\nu(A) = \nu(\mathbb{1}_A > 0) = S_{\mathfrak{D}_\nu(\mathbb{1}_A)}(0) = S_{\mathfrak{D}_{\mathfrak{C}(\nu)}(\mathbb{1}_A)}(0) = \min_{\mu \in \mathfrak{C}(\nu)} \mu(\mathbb{1}_A > 0) = \min_{\mu \in \mathfrak{C}(\nu)} \mu(A).$$

Thus, ν is exact. □

Proof of Proposition 4.11. Assume that $\mathfrak{D}_\nu = \mathfrak{D}_{\mathfrak{C}(\nu)}$. Let $X_1, \dots, X_n \in \mathcal{X}$ and $\lambda_1, \dots, \lambda_n \in [0,1]$ with $\sum_{k=1}^n \lambda_k = 1$ satisfy $\sum_{k=1}^n \lambda_k X_k = xAy$ for some $A \in \mathcal{F}$ and $x, y \in \mathbb{R}$. By Lemma 4.7, ν is exact. Therefore, there exists $\mu' \in \mathfrak{C}(\nu)$ satisfying $\mu'(A) = \nu(A)$. Without loss of generality, we may assume $x \geq y$. It is clear that $\mathfrak{D}_\nu(xAy) = \mathfrak{D}_{\mu'}(xAy)$. Also, by the comonotonic sum inequality, we have

$$\mathfrak{D}_{\mu'}(xAy) = \mathfrak{D}_{\mu'}\left(\sum_{k=1}^n \lambda_k X_k\right) \geq_{\mathrm{ssd}} \bigoplus_{k=1}^n \lambda_k \otimes \mathfrak{D}_{\mu'}(X_k).$$

Finally, since $\mu' \in \mathfrak{C}(\nu)$, we have that $\mathfrak{D}_{\mu'}(X_k) \geq_{\mathrm{fsd}} \mathfrak{D}_\nu(X_k)$ for all k. Therefore,

$$\bigoplus_{k=1}^n \lambda_k \otimes \mathfrak{D}_{\mu'}(X_k) \geq_{\mathrm{fsd}} \bigoplus_{k=1}^n \lambda_k \otimes \mathfrak{D}_\nu(X_k).$$

As $\geq_{\mathrm{fsd}}$ is stronger than $\geq_{\mathrm{ssd}}$, we have that (4.18) is true.

For the converse, by Lemma 4.7, we only need to show that ν is exact. Fix

$$A_1, \dots, A_n, A \in \mathcal{F}$$

and non-negative $c_1, \dots, c_n, c \in \mathbb{R}$ satisfying $\mathbb{1}_A \geq \sum_{k=1}^n c_k \mathbb{1}_{A_k} - c$. Find $A_{n+1}, \dots A_m \in \mathcal{F}$ and non-negative $c_{n+1}, \dots, c_m \in \mathbb{R}$ such that $\mathbb{1}_A = \sum_{k=1}^m c_k \mathbb{1}_{A_k} - c$, where $m \geq n$. Define $a = \sum_{k=1}^m c_k$. Therefore, by (4.18),

$$\begin{aligned}(1/a) \otimes [(\nu(A))\delta_{c+1} + (1 - \nu(A))\delta_c] &= (1/a) \otimes \mathfrak{D}_\nu \left((c+1)Ac\right) \\ &\geq_{\text{ssd}} \bigoplus_{k=1}^m (c_k/a) \otimes \mathfrak{D}_\nu(\mathbb{1}_{A_k}) \\ &= \bigoplus_{k=1}^m (c_k/a) \otimes [(\nu(A_k))\delta_1 + (1 - \nu(A_k))\delta_0] \,.\end{aligned}$$

Since

$$\begin{aligned}\mathbb{E}\left((1/a) \otimes [(\nu(A))\delta_{c+1} + (1 - \nu(A))\delta_c]\right) &= (1/a)(\nu(A)(c+1) + (1 - \nu(A))c) \\ &= (1/a)(\nu(A) + c), \\ \mathbb{E}\left(\bigoplus_{k=1}^m (c_k/a) \otimes [(\nu(A_k))\delta_1 + (1 - \nu(A_k))\delta_0]\right) &= \sum_{k=1}^m (c_k/a)\nu(A_k),\end{aligned}$$

and $\mathbb{E}$ is monotone with respect to $\geq_{\text{ssd}}$, we have

$$\nu(A) \geq \sum_{k=1}^m c_k \nu(A_k) - c \geq \sum_{k=1}^n c_k \nu(A_k) - c.$$

Therefore, by Schmeidler [1972], ν is exact. □

Proof of Theorem 4.6. Since $\succsim$ is a Choquet ATE preference relation, we have that

$$X \succsim Y \iff \gamma \circ \mathfrak{D}_\nu(X) \geq \gamma \circ \mathfrak{D}_\nu(Y), \quad X, Y \in \mathcal{X}, \tag{4.23}$$

where γ is the $\succsim$-CES.

Assume that $\succsim$ is a distributionally robust preference relation. Therefore, by definition, there exists a collection of risk-conforming probability measures $\mathcal{Q}$ and a statistic $\tilde{\gamma}$ such that

$$X \succsim Y \iff \min_{\mu \in \mathcal{Q}} \tilde{\gamma}(X_{\#}\mu) \geq \min_{\mu \in \mathcal{Q}} \tilde{\gamma}(Y_{\#}\mu), \quad X, Y \in \mathcal{X}. \tag{4.24}$$

Without loss of generality, we may assume that $\tilde{\gamma}$ is a certainty-equivalent statistic. Fix $Q \in \mathcal{D}$. Since $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, we can find $X \in \mathcal{X}(\mathcal{G})$ such that $Q = X_{\#}\mathbb{P} = \mathfrak{D}_\nu(X)$. By (4.23), we have that $X \simeq \gamma(Q)$. Thus, by (4.24), we have $\gamma(Q) = \tilde{\gamma}(Q)$. Since Q was

arbitrary, we have that $\gamma = \tilde{\gamma}$. Given $X \in \mathcal{X}$, by (4.23), $X \simeq \gamma \circ \mathfrak{D}_\nu(X)$. Thus, by (4.24), we have

$$\min_{\mu \in \mathcal{Q}} \gamma(X_\# \mu) = \gamma \circ \mathfrak{D}_\nu(X). \tag{4.25}$$

Define $\varphi : [0,1] \to [0,1]$ by $\varphi(x) = \gamma(x\delta_1 + (1-x)\delta_0)$. Since γ is a certainty-equivalent statistic, it is easy to show that φ is a distortion function. Let $A \in \mathcal{F}$, by (4.25) we have

$$\begin{aligned}
\varphi(\nu(A)) = \gamma(\nu(A)\delta_1 + (1-\nu(A))\delta_0) &= \gamma \circ \mathfrak{D}_\nu(\mathbb{1}_A) \\
&= \min_{\mu \in \mathcal{Q}} \gamma((\mathbb{1}_A)_\# \mu) \\
&= \min_{\mu \in \mathcal{Q}} \gamma(\mu(A)\delta_1 + (1-\mu(A))\delta_0) \\
&= \min_{\mu \in \mathcal{Q}} \varphi(\mu(A)) = \varphi\left(\min_{\mu \in \mathcal{Q}} \mu(A)\right).
\end{aligned}$$

Since φ is invertible, we have $\nu(A) = \min_{\mu \in \mathcal{Q}} \mu(A)$. Additionally, we have $\mu(A) \geq \nu(A)$ for all $\mu \in \mathcal{Q}$ and $A \in \mathcal{F}$. Thus $\mathcal{Q} \subseteq \mathfrak{C}(\nu)$. We claim that for each $X \in \mathcal{X}$, there exists $\mu_0 \in \mathcal{Q}$ such that $\mathfrak{D}_\nu(X) = X_\# \mu_0$. To see this note that for all $\mu \in \mathcal{Q}$, we have $\mu(X > x) \geq \nu(X > x)$ for all $x \in \mathbb{R}$. Therefore,

$$X_\# \mu \geq_{\text{fsd}} \mathfrak{D}_\nu(X), \quad \mu \in \mathcal{Q}.$$

Additionally, by (4.25), there exists $\mu_0 \in \mathcal{Q}$ such that $\gamma(X_\# \mu_0) = \gamma \circ \mathfrak{D}_\nu(X)$. If $X_\# \mu_0 >_{\text{fsd}} \mathfrak{D}_\nu(X)$, it would hold that $\gamma(X_\# \mu_0) > \gamma \circ \mathfrak{D}_\nu(X)$, which is a contradiction. Therefore, $\mathfrak{D}_\nu(X) = X_\# \mu_0$. Given $A, B \in \mathcal{F}$ satisfying $A \subseteq B$, we have that there exists a $\mu_0 \in \mathcal{Q}$ such that

$$\begin{aligned}
&\mu(A)\delta_2 + (\mu(B) - \mu(A))\delta_1 + (1-\mu(B))\delta_0 \\
&\qquad = (\mathbb{1}_A + \mathbb{1}_B)_\# \mu_0 = \mathfrak{D}_\nu(\mathbb{1}_A + \mathbb{1}_B) = \nu(A)\delta_2 + (\nu(B) - \nu(A))\delta_1 + (1-\nu(B))\delta_0.
\end{aligned}$$

Therefore, $\nu(A) = \mu(A)$ and $\nu(B) = \mu(B)$. Since $\mu \in \mathfrak{C}(\nu)$, by Marinacci and Montrucchio [2004, Theorem 4.7], we have that ν is supermodular.

Conversely, assume that ν is supermodular. We claim that

$$\gamma \circ \mathfrak{D}_\nu(X) = \min_{\mu \in \mathfrak{C}(\nu)} \gamma(X_\# \mu), \quad X \in \mathcal{X}. \tag{4.26}$$

Since for all $\mu \in \mathfrak{C}(\nu)$, μ is risk conforming, (4.26) shows that $\succsim$ is a distributionally robust preference relation. To show (4.26), start by fixing $X \in \mathcal{X}$. As $\{X > x\}_{x \in \mathbb{R}}$ is a chain, by Marinacci and Montrucchio [2004, Theorem 4.7], there exists $\mu_0 \in \mathfrak{C}(\nu)$ such that for all

$\mu \in \mathfrak{C}(\nu)$, we have

$$\mu(X > x) \geq \nu(X > x) = \mu_0(X > x), \quad x \in \mathbb{R}.$$

Therefore, for all $\mu \in \mathfrak{C}(\nu)$, $X_{\#}\mu \geq_{\text{fsd}} \mathfrak{D}_\nu(X) = X_{\#}\mu_0$. Thus,

$$\gamma(X_{\#}\mu) \geq \gamma \circ \mathfrak{D}_\nu(X) = \gamma(X_{\#}\mu_0), \quad \mu \in \mathfrak{C}(\nu).$$

Finally, as X was arbitrary, (4.26) holds. □

Chapter 5

Distributionally robust optimization under Bayesian ambiguity sets

5.1 Introduction

Decision functionals are instrumental tools for decision-making under uncertainty problems found in economics, finance, insurance, and operations research. Put simply, decision functionals are mappings from choice objects or *acts* (given by real-valued random variables in this chapter) to the real numbers. In practice, such functionals are constructed statistically: a decision-maker (DM) fixes a *statistic* (a mapping defined on distributions) and a *first-order model* (a probability measure), then evaluates an act by first computing its distribution under the first-order model and subsequently applying the statistic to that distribution. For example, statistical decision functionals can be found in Markowitz [1952] for portfolio selection, Savage [1954] for decision theory, and Artzner et al. [1999] for financial risk management.

A critical shortcoming of the aforementioned statistical process is that it presupposes a DM can fix a unique first-order model to describe their uncertainty. Following Knight [1921], one distinguishes between *pure risk*, governed by a known first-order model, and *ambiguity* (also called Knightian uncertainty or model uncertainty), which arises when the first-order model cannot be uniquely identified. This distinction is not merely philosophical: experimental evidence, beginning with the well-known Ellsberg [1961] paradox, demonstrates that economic participants distinguish between pure risk and ambiguity. In practical applications, ambiguity arises because the available data and background knowledge are insufficient to uniquely determine the first-order model. Furthermore, a DM who

ignores ambiguity may make decisions that are insufficiently robust. The 2008 financial crisis provided a stark reminder that failing to account for the ambiguity in the underlying first-order model can lead to catastrophic failures (e.g., McNeil et al. [2015]).

Two broad paradigms have emerged for constructing decision functionals that incorporate ambiguity. In both of these paradigms, we assume the DM fixes a statistic to describe how they would make decisions in the absence of ambiguity.

The first one, known as *distributionally robust optimization (DRO)*, addresses ambiguity by constructing an *ambiguity set*, i.e., a collection of relevant first-order models. To define the decision functional, the DM first applies the statistic to the distributions of the act under each first-order model in the ambiguity set, and then selects the worst-case value. This paradigm is motivated by the maxmin expected utility model of Gilboa and Schmeidler [1989] and has subsequently found broad application in financial risk management [El Ghaoui et al., 2003, Natarajan et al., 2008, Zhu and Fukushima, 2009], operations research [Delage and Ye, 2010, Ben-Tal et al., 2013], and actuarial science [Blanchet et al., 2019, Birghila and Pflug, 2019]. For axiomatizations of these decision functionals, see Delage et al. [2019] and Fadina et al. [2024]. A general limitation of DRO is that it may not accommodate learning, that is, updating the perception of ambiguity (in this case, the ambiguity set) as new information becomes available.

The second paradigm, rooted in the Bayesian statistical literature, quantifies ambiguity with a second-order model, i.e., a probability measure over first-order models. To construct decision functionals in this paradigm, the second-order model is averaged to yield the predictive (first-order) model, and the DM applies the statistic to the distribution of the act under the predictive model. This idea has been used in financial risk management [Ardia, 2008, Ardia et al., 2017], actuarial science [Klugman, 1992, Cairns, 2000], and portfolio allocation [Kandel and Stambaugh, 1996, Feng and He, 2022]. Unlike DRO, the Bayesian paradigm has a natural way to incorporate learning. Such learning is accomplished through Bayes' theorem, which underpins the widespread adoption of Bayesian statistics. However, a key limitation of this approach is that the predictive model reduces all ambiguity to a single averaged first-order model. As a result, it cannot capture general attitudes toward ambiguity, since decision-making based on a single first-order model is, by definition, ambiguity-neutral.

A major contribution of the decision theory literature is the smooth ambiguity model of Klibanoff et al. [2005]. In this model, as in the Bayesian paradigm, the DM represents ambiguity through a second-order model. Unlike the general setting considered previously, this model uses the expected utility statistic of von Neumann and Morgenstern [1947]. The decision functional is then obtained by transforming the first-order model-specific expected utilities through a function that captures the DM's attitude toward ambiguity, and inte-

grating the resulting values with respect to the second-order model. When the function encoding ambiguity attitudes coincides with the utility function used in the expected utility statistic, the resulting decision functional coincides with the expected utility statistic evaluated at the predictive distribution of the acts, thereby recovering the Bayesian paradigm. These decision functionals have been used extensively in the literature (Gollier, 2011, Ju and Miao, 2012, Maccheroni et al., 2013) because they provide a clear separation between pure risk, ambiguity perception, and ambiguity attitude. However, as we will discuss in detail later, the smooth model's conceptual justification depends crucially on properties that rely on the statistic being the expected utility statistic and do not generalize to arbitrary statistics.

The goal of this chapter is to develop a framework for constructing ambiguity-sensitive decision functionals that preserve the main conceptual strengths of the smooth ambiguity model while allowing for general statistics. Our approach operates within the DRO paradigm, in which we construct ambiguity sets based on second-order models. This process is formalized with the notion of an ambiguity-reduction mapping, that is, a rule that transforms a second-order model into an ambiguity set in a way that reflects both ambiguity perception and ambiguity attitude. The resulting worst-case decision functionals have an interpretable separation between pure-risk preferences, ambiguity perception, and ambiguity attitude, return to the Bayesian paradigm under ambiguity neutrality, and can be applied to statistics important in risk management, finance, and actuarial science. In addition to the theoretical framework, we provide computational methods for evaluating the resulting decision functionals and illustrate their implementation through numerical examples in portfolio optimization. These examples show how the proposed framework can be used in practice and how ambiguity can influence optimal decisions.

The remainder of the chapter is organized as follows. Section 5.2 introduces the setting, notation, and basic objects used throughout the chapter. Section 5.3 reviews the smooth ambiguity model, explains why its key properties fail to extend to general statistics, and introduces ambiguity-reduction mappings together with our canonical construction. Section 5.4 develops tractable formulas and presents numerical examples in the context of portfolio optimization. Section 5.5 connects these decision functionals to the existing literature.

5.2 Preliminaries

5.2.1 General notation

Given measurable spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$, a measurable mapping $X : \Omega \to E$, and a probability measure μ on $(\Omega, \mathcal{F})$, the *pushforward probability measure* $X_{\#}\mu$ on $(E, \mathcal{E})$ is defined by

$$X_{\#}\mu(A) = \mu(X^{-1}(A)), \quad A \in \mathcal{E}.$$

A *Polish measurable space* is a measurable space $(E, \mathcal{E})$ where E is a Polish topological space and $\mathcal{E}$ is its Borel σ-algebra. We will sometimes denote the Borel-σ-algebra by $\mathcal{B}(E)$. Given a Polish measurable space $(E, \mathcal{E})$ and a probability measure μ on $(E, \mathcal{E})$, the *support* of μ is the smallest closed set $C \subseteq E$ such that $\mu(E) = 1$. Given a Polish measurable space $(E, \mathcal{E})$ and a sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ on $(E, \mathcal{E})$ we say that $(\mu_n)_{n\in\mathbb{N}}$ *converges in distribution* to some probability measure μ on $(E, \mathcal{E})$, denoted by $\mu_n \Rightarrow \mu$, if for all continuous and bounded $f : E \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \int_E f \mathrm{d}\mu_n = \int_E f \mathrm{d}\mu.$$

5.2.2 Our setting

Throughout the chapter, we fix a Polish measurable space $(\Omega, \mathcal{F})$, which is interpreted as the set of future states of the world. An *act* is a real-valued measurable function on $(\Omega, \mathcal{F})$. In this chapter, an act is to be viewed in the loss-profit sense (positive values correspond to losses or bad outcomes). We denote by $\mathcal{X}$ the set of all acts, by $\mathcal{X}_b$ the set of all bounded acts, by $\mathcal{C}$ the set of all continuous acts, and by $\mathcal{C}_b$ the set of all continuous bounded acts. Let Δ denote the set of probability measures on $(\Omega, \mathcal{F})$, which we call the *first-order models*. Given $X \in \mathcal{X}$ and $\mu \in \Delta$, we use $\mathbb{E}^\mu[X]$, to denote the Lebesgue integral of X under μ. Given $X \in \mathcal{X}$, define the function $\psi_X : \Delta \to \mathbb{R} \cup \{-\infty, \infty\}$ by $\psi_X(\mu) = \mathbb{E}^\mu[X]$.[1] We equip Δ with the coarsest topology τ where the maps $\{\psi_X\}_{X\in\mathcal{C}_b}$ are continuous, i.e., the weak topology. Additionally, we equip Δ with the σ-algebra generated by τ, denoted by Σ. Since $(\Omega, \mathcal{F})$ is a Polish measurable space, we know that (Δ, Σ) is a Polish measurable space; see Billingsley [2013, Theorem 6.8]. We will use Π to denote the set of probability measures on (Δ, Σ), which we call the *second-order models*.

[1] When the $\psi_X(\mu)$ is undefined, set it equal to zero.

For $\pi \in \Pi$, let $L^1(\pi)$ denote the set of acts $X \in \mathcal{X}$ satisfying $\int_\Delta \mathbb{E}^\nu[|X|]\pi(\mathrm{d}\nu) < \infty$. We verify that ψ_X is Σ-measurable for all $X \in \mathcal{X}$ in Section 5.12. For $\pi \in \Pi$, we define the *predictive model* $\mu_\pi \in \Delta$ by

$$\mu_\pi(A) = \int_\Delta \nu(A)\pi(\mathrm{d}\nu), \quad A \in \mathcal{F}.$$

An application of the dominated-convergence theorem will verify that $\mu_\pi \in \Delta$. We have the following proposition regarding predictive models.

Proposition 5.1. *For $\pi \in \Pi$ and $X \in L^1(\pi)$, we have*

$$\int_\Delta \mathbb{E}^\nu[X]\pi(\mathrm{d}\nu) = \mathbb{E}^{\mu_\pi}[X].$$

We additionally equip Π with the binary relation $\geq_{\mathrm{cx}}$ defined by

$$\tilde{\pi} \geq_{\mathrm{cx}} \pi \iff \int_\Delta f\mathrm{d}\tilde{\pi} \geq \int_\Delta f\mathrm{d}\pi, \quad \tilde{\pi}, \pi \in \Pi,$$

whenever these integrals are well-defined, for all $f : \Delta \to \mathbb{R} \cup \{\infty\}$ that are proper, convex, and τ-lower semicontinuous.

Theorem 5.1. *The binary relation $\geq_{\mathrm{cx}}$ is a partial order.*

The partial order $\geq_{\mathrm{cx}}$ can be interpreted as a way to quantify when one second-order model exhibits greater perceived ambiguity than another, analogous to the classical convex order used to compare pure risk; see Rothschild and Stiglitz [1970]. The following proposition helps reinforce this interpretation.

Proposition 5.2. *For all $\pi \in \Pi$, $\pi \geq_{\mathrm{cx}} \delta_{\mu_\pi}$.*

Given $X, Y \in \mathcal{X}$, we write $X \geq Y$ if $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$. Given $\mathcal{Y} \subseteq \mathcal{X}$ containing the constants, a *decision functional* is a mapping $\rho : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ such that for all $X, Y \in \mathcal{Y}$,

$$X \geq Y \implies \rho(X) \geq \rho(Y)$$

and $\rho(c) = c$ for all $c \in \mathbb{R}$.

Denote by $\mathcal{D}$ the set of all Borel probability measures on $\mathbb{R}$, and $\mathcal{D}_1$ the subset of $\mathcal{D}$ with finite first moment. We will refer to $\mathcal{D}$ as the set of *distributions*. For $X \in \mathcal{X}$, define the mapping $\iota_X : \Delta \to \mathcal{D}$ by $\iota_X(\mu) = X_\# \mu$. Given $Q \in \mathcal{D}$, we define the *survival function of*

Q by $S_Q(x) = Q((x,\infty))$ for all $x \in \mathbb{R}$. For $Q, P \in \mathcal{D}$, we write $Q \geq_{\text{fsd}} P$ if $S_Q(x) \geq S_P(x)$ for all $x \in \mathbb{R}$. We have that $\geq_{\text{fsd}}$ is a partial order on $\mathcal{D}$.

In this chapter, a *statistic* is a mapping $\gamma : \mathcal{D}_1 \to \mathbb{R} \cup \{\infty\}$ such that for all $Q, P \in \mathcal{D}$,

$$Q \geq_{\text{fsd}} P \implies \gamma(Q) \geq \gamma(P)$$

and $\gamma(\delta_c) = c$ for all $c \in \mathbb{R}$, where δ_c denotes the Dirac measure at c. Given a statistic γ, $X \in \mathcal{X}$, and $\mu \in \Delta$, we will use $\gamma(X|\mu)$ in place of $\gamma(X_{\#}\mu)$. The following are examples of statistics commonly found in the literature.

(a) The expected value statistic

$$\mathbb{E}(Q) = \int_{\mathbb{R}} xQ(\mathrm{d}x), \quad Q \in \mathcal{D}_1.$$

(b) The expected loss statistic of von Neumann and Morgenstern [1947]

$$\mathbb{E}_\ell(Q) = \ell^{-1}\left(\int_{\mathbb{R}} \ell \mathrm{d}Q\right) = \ell^{-1}\left(\mathbb{E}(\ell_{\#}Q)\right), \quad Q \in \mathcal{D}_1,$$

where $\ell : \mathbb{R} \to \mathbb{R}$ is continuous, convex, and increasing, called a *loss function*. We use the expected loss convention, as opposed to the usual expected utility, because we are working with loss-profit acts.

(c) The dual statistic of Yaari [1987]

$$\mathrm{D}_g(Q) = \int_0^\infty g(S_Q(x))\mathrm{d}x + \int_{-\infty}^0 (g(S_Q(x)) - 1)\mathrm{d}x, \quad Q \in \mathcal{D}_1,$$

where $g : [0,1] \to [0,1]$ is continuous, increasing and $g(x) \geq x$ for all $x \in [0,1]$, called a *distortion function*.

Under the distortion $g_\alpha(x) = \frac{x}{1-\alpha} \wedge 1$, where $\alpha \in [0,1)$, the dual statistic D_{g_α} coincides with *Expected Shortfall* of Acerbi and Tasche [2002]; hereby denoted by ES_α. Expected Shortfall is widely used in financial practice and serves as the standard risk measure for market risk within the banking regulatory framework of BCBS [2019]. Under the distortion $h_\alpha = \mathbb{1}_{(1-\alpha,1]}$, where $\alpha \in (0,1)$, the dual statistic D_{h_α} coincides with the *Value-at-Risk*; hereby denoted by VaR_α.

5.3 Decision functionals incorporating second-order models

In this section, we review the smooth ambiguity model of Klibanoff et al. [2005] and discuss its failure to generalize to arbitrary statistics. Afterward, we define ambiguity-reduction mappings that yield Bayesian ambiguity sets. We show that the worst-case decision functionals induced by these Bayesian ambiguity sets satisfy the conceptually appealing properties of the smooth model while allowing for general statistics. Finally, we provide a class of ambiguity-reduction mappings that will be used in the numerical implementation later in the chapter.

5.3.1 The smooth model of ambiguity

In this section, we review the decision functionals from the smooth model of ambiguity introduced by Klibanoff et al. [2005]. Afterward, we will discuss the inability of these functionals to accommodate general statistics while preserving the nice consequences of the original model.

Consider a DM who fixes the expected loss statistic $\mathbb{E}_\ell$, for some loss function ℓ. We will interpret the statistic $\mathbb{E}_\ell$ as the functional the DM would use to make decisions in the absence of ambiguity, that is, if they had a unique first-order model. Therefore, we will say that $\mathbb{E}_\ell$ determines the *pure-risk preferences* of the DM. From here, we need notions of ambiguity perceptions and ambiguity attitudes. Ambiguity perceptions will be modeled by the set of second-order models Π. Ambiguity attitudes will be modeled by the set

$$\Lambda = \{\phi \circ \ell \mid \phi : \mathbb{R} \to \mathbb{R} \text{ is convex and strictly increasing}\}.$$

To allow for the comparison of these ambiguity attitudes, we equip Λ with the partial order

$$\varphi \succeq \tilde{\varphi} \iff \varphi = \phi \circ \tilde{\varphi}$$

for some convex and strictly increasing $\phi : \mathbb{R} \to \mathbb{R}$. Clearly $\phi \succeq \ell$ for all $\varphi \in \Lambda$. We will see that the minimal element ℓ represents the attitude that is neutral to ambiguity. This setup only allows for a discussion regarding ambiguity-averse attitudes, but the following ideas can easily be generalized to include ambiguity-seeking attitudes.

Given $\pi \in \Pi$ and $\varphi \in \Lambda$, define the decision functional $\rho(\cdot; \pi, \varphi) : \mathcal{X}_b \to \mathbb{R}$ by

$$\rho(X, \pi, \varphi) = \varphi^{-1}\left(\int_\Delta \varphi\left(\mathbb{E}_\ell\left(X|\nu\right)\right)\pi(\mathrm{d}\nu)\right), \quad X \in \mathcal{X}_b. \tag{5.1}$$

The following observations of these decision functionals provide the theoretical justification for their widespread use.

(1) **Clear separation between pure risk and ambiguity:** Separation between pure-risk preferences ($\mathbb{E}_\ell$), ambiguity attitude (φ), and ambiguity perception (π).

(2) **Law-invariance under epistemic certainty:** for all $\mu \in \Delta$ and $\varphi \in \Lambda$,

$$\rho(X; \delta_\mu, \varphi) = \mathbb{E}_\ell(X|\mu), \quad X \in \mathcal{X}_b.$$

(3) **Law-invariance at the minimal element:** for all $\pi \in \Pi$,

$$\rho(X; \pi, \ell) = \mathbb{E}_\ell(X|\mu_\pi), \quad X \in \mathcal{X}_b.$$

(4) **Monotonicity in ambiguity attitudes:** if $\varphi, \tilde{\varphi} \in \Lambda$ satisfy $\varphi \succeq \tilde{\varphi}$, then for all $\pi \in \Pi$,

$$\rho(X; \pi, \varphi) \geq \rho(X; \pi, \tilde{\varphi}), \quad X \in \mathcal{X}_b.$$

(5) **Monotonicity in ambiguity perceptions:** if $\pi, \tilde{\pi} \in \Pi$ satisfy $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, then for every $\varphi \in \Lambda$,

$$\rho(X; \tilde{\pi}, \varphi) \geq \rho(X; \pi, \varphi), \quad X \in \mathcal{X}_b.$$

The fact that Observations (1)–(4) are true is clear; however, Observation (5) needs verification, which we do in Section 5.12.

The family of decision functionals given in (5.1) is conceptually appealing because of Observations (1)–(5). However, a key limitation is that some of these observations rely on the assumption that the pure-risk preferences are given by $\mathbb{E}_\ell$. In many practical applications, this necessity can be restrictive. For example, under no ambiguity, a bank must calculate capital reserves using Expected Shortfall ES_α due to regulatory requirements. This naturally raises the following question: can the functional form of the smooth ambiguity model be adapted to construct an 'ambiguity aware' Expected Shortfall that incorporates a Bayesian prior, preserves Observations (1)–(5) above, and reduces to Expected Shortfall in the no-ambiguity or ambiguity neutral case?

A natural first approach is to replace the expected loss statistic $\mathbb{E}_\ell$ in (5.1) with Expected Shortfall ES_α. Such a modification is appealing, as ambiguity models originating in decision theory are often adapted to other areas through precisely this type of substitution; see, for instance, how the maxmin expected utility model of Gilboa and Schmeidler [1989] is used to motivate the DRO paradigm. However, this replacement causes Observations (3) and (5) to no longer hold. This follows from the following proposition. We verify in Section 5.12 that, given $X \in \mathcal{X}_b$, $\mathrm{ES}_\alpha \circ \iota_X$ is Σ-measurable.

Proposition 5.3. *If $\tilde{\pi}, \pi \in \Pi$ satisfy $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, then*

$$\int_\Delta \mathrm{ES}_\alpha(X|\nu)\tilde{\pi}(\mathrm{d}\nu) \leq \int_\Delta \mathrm{ES}_\alpha(X|\nu)\pi(\mathrm{d}\nu), \quad X \in \mathcal{X}_b$$

Specifically, by Proposition 5.2 and Proposition 5.3, for all $\pi \in \Pi$,

$$\int_\Delta \mathrm{ES}_\alpha(X|\nu)\pi(\mathrm{d}\nu) \leq \mathrm{ES}_\alpha(X|\mu_\pi), \quad X \in \mathcal{X}_b.$$

Consequently, if banks were permitted to incorporate a second-order model in the calculation of capital reserves, they would have an incentive to adopt a highly dispersed second-order model and compute the average Expected Shortfall under that second-order model, rather than determine capital reserves based on the predictive model. Therefore, the resulting decision functional may inadvertently incentivize ambiguity-seeking behavior. From a regulatory perspective, this is highly problematic, as banks should exhibit ambiguity aversion to enhance the resilience of the financial system during periods of heightened ambiguity in the economy.

5.3.2 Ambiguity-reduction mappings

The goal of this section is to define and motivate ambiguity-reduction mappings. By reduction, we mean transforming second-order models into a set of competing first-order models or an ambiguity set. Afterward, we will construct decision functionals within the DRO paradigm that return to the Bayesian paradigm under ambiguity neutrality. The motivation for this work is two-fold. (I) These ambiguity-reduction mappings allow us to construct decision functionals that have Observations (1)–(5) while also allowing for general statistics. (II) Traditionally, DRO models have problems regarding updating, i.e., how to change your ambiguity set in the face of new information. However, constructing ambiguity sets using second-order models enables updating via Bayesian updating.

Let $(\Lambda, \succeq)$ be a partially ordered set with a minimal element λ_0. Similar to the previous section, the set Λ encodes different ambiguity-averse attitudes. The following properties clarify this interpretation.

Definition 5.1. A map $\mathcal{U} : \Pi \times \Lambda \longrightarrow 2^{\Delta}$ is called an *ambiguity-reduction mapping* if it satisfies:

(1) **Degeneracy consistency:** for every $\mu \in \Delta$ and $\lambda \in \Lambda$, $\mathcal{U}(\delta_\mu, \lambda) = \{\mu\}$.

(2) **Ambiguity neutrality at the least element:** for every $\pi \in \Pi$, $\mathcal{U}(\pi, \lambda_0) = \{\mu_\pi\}$.

(3) **Monotonicity in ambiguity aversion:** if $\lambda, \tilde{\lambda} \in \Lambda$ satisfy $\tilde{\lambda} \succeq \lambda$, then for every $\pi \in \Pi$,
$$\mathcal{U}(\pi, \lambda) \subseteq \mathcal{U}(\pi, \tilde{\lambda}).$$

(4) **Monotonicity in ambiguity perception:** if $\pi, \tilde{\pi} \in \Pi$ satisfy $\tilde{\pi} \geq_{\text{cx}} \pi$, then for every $\lambda \in \Lambda$,
$$\mathcal{U}(\pi, \lambda) \subseteq \mathcal{U}(\tilde{\pi}, \lambda).$$

Property (1) says that if there is no epistemic uncertainty, i.e., the perception of no ambiguity, then no robustification is necessary. Property (2) says that the minimal element λ_0 corresponds to the reduction of the second-order model (π) to a single model (μ_π), i.e., an ambiguity-neutral reduction. Property (3) is to remain consistent with the intuitive idea that Λ parametrizes ambiguity-aversion attitudes. Property (4) says that the reduced ambiguity sets should be larger if the perception of ambiguity increases, regardless of ambiguity attitudes.

An important conceptual feature is that attitudes toward ambiguity are fully determined by the ambiguity-reduction mapping. As a result, the ambiguity set produced by this mapping captures two distinct dimensions: ambiguity perception, reflected by the fact that the second-order model is an input, and attitudes toward ambiguity. A related concept appears in Gajdos et al. [2008], where the authors introduce *selected probability-possibility sets*. Like our ambiguity-reduction mappings, their selected probability-possibility sets take an object representing ambiguity perception as an input and output an ambiguity set. The key distinction lies in how ambiguity perception is modeled: we use a second-order model as an input, whereas they use the initial structured ambiguity set, i.e., the ambiguity set consistent with the constraints of the decision problem. Additionally, they take the same interpretation that this reduction encodes attitudes toward ambiguity.

Ambiguity sets constructed from a second-order probability have recently been discussed by Dellaporta et al. [2024] and Dellaporta et al. [2025]. In this work, they define

the mappings, $\mathcal{U}_{\text{PP}}$ and $\mathcal{U}_{\text{PE}}$, both with domain $\Pi \times [0, \infty)$, by

$$\mathcal{U}_{\text{PP}}(\pi, \epsilon) = \{\mu \in \Delta \mid \text{KL}(\mu|\mu_\pi) \leq \epsilon\}$$
$$\mathcal{U}_{\text{PE}}(\pi, \epsilon) = \left\{\mu \in \Delta \;\middle|\; \int_\Delta \text{KL}(\mu|\nu)\pi(\mathrm{d}\nu) \leq \epsilon\right\},$$

where KL denotes the Kullback–Leibler divergence. We verify in Section 5.12 that the mapping $\nu \mapsto \text{KL}(\mu|\nu)$ is Σ-measurable for all $\mu \in \Delta$. The mappings $\mathcal{U}_{\text{PP}}$ and $\mathcal{U}_{\text{PE}}$ have a similar structure to our ambiguity-reduction mappings, but it is easy to see that neither satisfies all the properties of an ambiguity-reduction mapping. For example, neither satisfies Property (1). This failure follows from the fact that $\mathcal{U}_{\text{PP}}$ and $\mathcal{U}_{\text{PE}}$ were introduced to address problems related to model misspecification; see Shapiro et al. [2023] and Cerreia-Vioglio et al. [2025]. We will discuss model misspecification later in Section 5.5.3.

Given a statistic γ, an ambiguity-reduction mapping $\mathcal{U}$, $\pi \in \Pi$, and $\lambda \in \Lambda$, define the decision functional $\rho(\cdot; \pi, \lambda) : \mathcal{X}_b \to \mathbb{R}$ by

$$\rho(X; \pi, \lambda) = \sup_{\nu \in \mathcal{U}(\pi,\lambda)} \gamma(X|\nu), \quad X \in \mathcal{X}_b. \tag{5.2}$$

This model clearly separates pure-risk preferences (γ), ambiguity attitude (λ), and ambiguity perception (π). Additionally, it is not hard to see that Observations (2)–(5) are also satisfied by the decision functional in (5.2).

5.3.3 The canonical example

Let Υ denote the set of all concave distortion functions g such that there exists $a \in \mathbb{N}$ with

$$g(x) \leq ax, \quad x \in [0, 1]. \tag{5.3}$$

Define the partial order $\succeq$ on Υ by pointwise dominance:

$$g \succeq \tilde{g} \iff g(x) \geq \tilde{g}(x), \quad x \in [0, 1].$$

The least element is $g_0(x) = x$ for all $x \in [0, 1]$.

Given $\pi \in \Pi$ and $g \in \Upsilon$, define the collection of second-order models

$$\mathcal{R}(\pi, g) = \{\xi \in \Pi \mid \xi(B) \leq g(\pi(B)) \text{ for all } B \in \Sigma\}.$$

Since g is concave and continuous, the capacity $g \circ \pi$ is submodular and continuous; see

Marinacci and Montrucchio [2004]. Furthermore, this implies $\mathcal{R}(\pi, g) \neq \varnothing$ and, whenever the right-hand side is well-defined (i.e., at least one of the two integrals is finite),

$$\max_{\xi \in \mathcal{R}(\pi,g)} \int_\Delta f(\nu)\xi(\mathrm{d}\nu) = \int_0^\infty g(\pi(f > x))\mathrm{d}x + \int_{-\infty}^0 (g(\pi(f > x)) - 1)\mathrm{d}x, \tag{5.4}$$

where $f : \Delta \to \mathbb{R}$ is Σ-measurable. The value on the right-hand side of (5.4) is known as the Choquet integral (see Denneberg [1994]). Henceforth, we denote this value by $\int_\Delta f(\nu) g \circ \pi(\mathrm{d}\nu)$.

Define the mapping $\mathcal{S} : \Pi \times \Upsilon \to 2^\Delta$ by

$$\mathcal{S}(\pi, g) = \{\mu_\xi \mid \xi \in \mathcal{R}(\pi, g)\}. \tag{5.5}$$

Theorem 5.2. *The mapping $\mathcal{S}$ from (5.5) is an ambiguity-reduction mapping.*

We require the restriction given in (5.3) to have the following proposition.

Proposition 5.4. *Given $\pi \in \Pi$, $g \in \Upsilon$, and $X \in L^1(\pi)$, we have that $X_\# \mu \in \mathcal{D}_1$ for all $\mu \in \mathcal{S}(\pi, g)$.*

5.4 Numerical examples

In this section, we turn from the theoretical development of Bayesian ambiguity sets to their practical computation. Our goal is to show that the proposed decision functionals are computationally tractable in applied settings. To do this, we first develop alternative formulas to evaluate these decision functionals under different statistics within the canonical class of ambiguity-reduction mappings introduced earlier. We then numerically illustrate the resulting formulas through a portfolio optimization example.

5.4.1 Parametrization

To make these ideas more tractable, we will work with the Polish measurable space $(\Theta, \mathcal{A})$, which is interpreted as the *parameter space*. Additionally, we will have a mapping $\mathfrak{P} : \Theta \to \Delta$, called a *parametrization*. We will assume that

(1) $\mathfrak{P}$ is $\mathcal{A}/\Sigma$-measurable;

(2) $\theta = \tilde{\theta} \iff \mathfrak{P}(\theta) = \mathfrak{P}(\tilde{\theta})$;

(3) $\mathcal{A} = \mathfrak{P}^{-1}(\Sigma)$.

It is well-known (see Bertsekas and Shreve [1996, Proposition 7.25]) that Condition (1) is equivalent to requiring that, for all $A \in \mathcal{F}$, the mapping $\theta \mapsto [\mathfrak{P}(\theta)](A)$ is $\mathcal{A}$-measurable. In other words, $\mathfrak{P}$ can be viewed as a probability kernel. Condition (3) may be hard to check in practice; however, by the Lusin-Souslin theorem (Kechris 2012, Theorem 15.1), if $\mathfrak{P}$ is continuous, Condition (3) holds.

We will use $\Delta(\Theta)$ to denote the set of probability measures on $(\Theta, \mathcal{A})$. Through the parametrization, an element $\pi \in \Delta(\Theta)$ induces a second-order model $\mathfrak{P}_{\#}\pi$. Therefore, we may view $\Delta(\Theta)$ as a subset of second-order models. Conceptually, there are two different ways to define ambiguity sets given $\pi \in \Delta(\Theta)$ and $g \in \Upsilon$. The first is to take $\mathcal{S}(\mathfrak{P}_{\#}\pi, g)$. The second begins by defining

$$\mathcal{R}_0(\pi, g) = \{\xi \in \Delta(\Theta) \mid \xi(A) \leq g(\pi(A)) \text{ for all } A \in \mathcal{A}\},$$

and then defining the ambiguity set $\mathcal{S}_0(\pi, g)$ as the set of all first-order models given by

$$\mu(A) = \int_{\Theta} \mu_\theta(A)\xi(\mathrm{d}\theta), \quad A \in \mathcal{F},$$

for some $\xi \in \mathcal{R}_0(\pi, g)$, where $\mu_\theta = \mathfrak{P}(\theta)$ for all $\theta \in \Theta$. The following proposition confirms that these methods produce the same ambiguity set.

Proposition 5.5. *Let $\pi \in \Delta(\Theta)$ and $g \in \Upsilon$. We have that*

$$\mathcal{R}(\mathfrak{P}_{\#}\pi, g) = \mathfrak{P}_{\#}(\mathcal{R}_0(\pi, g)) := \left\{\mathfrak{P}_{\#}\xi \;\middle|\; \xi \in \mathcal{R}_0(\pi, g)\right\} \quad \textit{and} \quad \mathcal{S}(\mathfrak{P}_{\#}\pi, g) = \mathcal{S}_0(\pi, g).$$

Therefore, with a slight abuse of notation, given $\pi \in \Delta(\Theta)$ and $g \in \Upsilon$, we denote the set $\mathcal{R}_0(\pi, g)$ by $\mathcal{R}(\pi, g)$ and the set $\mathcal{S}_0(\pi, g)$ by $\mathcal{S}(\pi, g)$, in order to maintain consistency with the notation used thus far.

5.4.2 Convergence result

Given $g \in \Upsilon$ and $\pi \in \Delta(\Theta)$, in the following algorithms, we will need to numerically approximate the decision functional

$$\int_{\Theta} \mathbb{E}^{\mu_\theta}[X] g \circ \pi(\mathrm{d}\theta) = \sup_{\mu \in S(\pi, g)} \mathbb{E}^{\mu}[X], \quad X \in L^1(\pi).$$

Many of the numerical schemes found in the literature can produce a stochastic process $(\theta_n)_{n\in\mathbb{N}}$ with values in the parameter space Θ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f(\theta_k) = \int_\Theta f \mathrm{d}\pi \quad \kappa\text{-a.s.}, \tag{5.6}$$

where $f : \Theta \to \mathbb{R}$ is measurable and satisfies $\int_\Theta |f|\mathrm{d}\pi < \infty$, and $\kappa \in \Delta(\Theta^{\mathbb{N}})$ is the probability measure induced by the stochastic process on path space. For example, many of the schemes used to sample from the posterior probability measure in Bayesian statistics satisfy this property; see Meyn and Tweedie [2012]. Additionally, if we are able to sample π directly, we may take $\kappa = \pi^{\mathbb{N}}$, in which case, (5.6) follows directly from the law of large numbers. For the numerical examples in the next sections, we assume that π can be sampled directly. Nevertheless, we formulate the next theorem under the more general condition (5.6) in order to motivate the use of MCMC-type algorithms in settings where direct sampling from π is not available.

Theorem 5.3. *Given a stochastic process $(\theta_n)_{n\in\mathbb{N}}$ satisfying (5.6) and $X \in L_1(\pi)$, define $Y_n = \mathbb{E}^{\mu_{\theta_n}}[X]$. We have that*

$$\lim_{n\to\infty}\sum_{k=1}^{n}\left(g\left(\frac{n-k+1}{n}\right) - g\left(\frac{n-k}{n}\right)\right) Y_{k:n} = \int_\Theta \mathbb{E}^{\mu_\theta}[X] g\circ\pi(\mathrm{d}\theta), \quad \kappa\text{-a.s.},$$

where $Y_{k:n}$ denotes the k^{th} order statistic of the random vector $(Y_1, \ldots, Y_n)$.

5.4.3 Alternative formulas and numerical results

In this section, we derive numerically tractable formulas for (5.2) under the canonical ambiguity-reduction mapping. We do so for three specific classes of statistics. After presenting each formula, we implement it numerically using Theorem 5.3.

In all numerical implementations, we assume that $(\Omega, \mathcal{F}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. The parameter space is taken to be

$$\Theta = \mathbb{R}^2 \times (0,\infty)^2 \times [0,1],$$

with the parametrization

$$\Theta \ni (m_1, m_2, \sigma_1, \sigma_2, c) \mapsto \mathrm{MVN}_2\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & c\sigma_1\sigma_2 \\ c\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

Let X_1 and X_2 denote the acts given by the marginal projections, representing the single-period percentage losses of two financial positions. Our objective is to evaluate

$$\sup_{\nu \in \mathcal{S}(\pi, g_\beta)} \gamma(wX_1 + (1-w)X_2|\nu), \tag{5.7}$$

where $w \in [0,1]$, $\beta \in [0,1)$, and $g_\beta(x) = \frac{x}{1-\beta} \wedge 1$ for all $x \in [0,1]$. Since $g_\beta \succeq g_{\beta'} \iff \beta \geq \beta'$, the parameter β corresponds to ambiguity aversion attitudes, with larger values of β corresponding to greater ambiguity aversion. Finally, in each numerical implementation, we consider the three second-order models listed in Table 5.1.

Note that the distributions of the acts X_1 and X_2 under the predictive model are the same for both π_1 and π_2 in Table 5.1. Additionally, under π_1, we perceive no ambiguity for the act X_1. Therefore, in the numerical implementations later, the decision functional in (5.7) should assign the same value to X_1 ($w = 1$) and X_2 ($w = 0$) under ambiguity neutrality ($\beta = 0$). However, as we become more ambiguity averse (β increases), the decision functional should assign a higher value to X_2 ($w = 0$) than to X_1 ($w = 1$).

Table 5.1: Second-order models used in the numerical implementations

	π_1	π_2	π_3
m_1	$\delta_{-0.05}$	$\mathrm{N}(-0.05, 0.01)$	$\mathrm{N}(-0.05, 0.01)$
m_2	$\mathrm{N}(-0.05, 0.02)$	$\mathrm{N}(-0.05, 0.02)$	$\mathrm{N}(-0.1, 0.03)$
σ_1	$\delta_{0.01\sqrt{5}}$	$\delta_{0.02}$	$\delta_{0.02}$
σ_2	$\delta_{0.01}$	$\delta_{0.01}$	$\mathrm{EXP}(1/50)$
c	$\mathrm{U}(-0.5, 0.5)$	$\mathrm{U}(-0.5, 0.5)$	$\mathrm{U}(-0.5, 0.5)$

The first class of statistics we will provide an algorithm for is the optimized certainty equivalents of Ben-Tal and Teboulle [2007]. Let $h : \mathbb{R} \to \mathbb{R}$ be convex, continuous, and nondecreasing. Define the statistic

$$\gamma_h^{\mathrm{OCE}}(Q) = \inf_{m \in \mathbb{R}} \left\{ m + \int_{\mathbb{R}} h(x-m) Q(\mathrm{d}x) \right\}, \quad Q \in \mathcal{D}_1.$$

Expected Shortfall ES_α provides a classic example of an optimized certainty equivalent. It arises when $h(x) = x_+$ for all $x \in \mathbb{R}$. In this case, the optimized certainty equivalent

representation coincides with the well-known Rockafellar and Uryasev [2002] formulation of Expected Shortfall.

To provide an alternative formula for the worst-case optimized certainty equivalent, we need the following definition.

Definition 5.2. Given $\pi \in \Delta(\Theta)$, we say that $X \in L^1(\pi)$ is *π-almost uniformly integrable* if for all $\epsilon > 0$, there exists $U \in \mathcal{A}$ with $\pi(U) > 1 - \epsilon$ such that

$$\lim_{K\to\infty} \sup_{\theta\in U} \mathbb{E}^{\mu_\theta}\left[|X|\mathbb{1}_{\{|X|>K\}}\right] = 0.$$

The following example provides sufficient conditions for π-almost uniform integrability. Using this example, it is straightforward to check that $wX_1 + (1-w)X_2$ is π-almost uniformly integrable for each $w \in [0,1]$ and $\pi \in \{\pi_1, \pi_2, \pi_3\}$.

Example 5.1. Given $\pi \in \Delta(\Theta)$, assume that $X \in \mathcal{X}$ satisfies

$$\int_\Theta \mathbb{E}^{\mu_\theta}[X^2]\pi(\mathrm{d}\theta) < \infty. \tag{5.8}$$

It is clear that $X \in L^1(\pi)$. For each $n \in \mathbb{N}$, define $U_n = \{\theta \in \Theta \mid \mathbb{E}^{\mu_\theta}[X^2] \le n\}$. Given $\epsilon > 0$, find $n_0 \in \mathbb{N}$ such that $\pi(U_{n_0}) > 1 - \epsilon$; guaranteed to exist by (5.8). Therefore,

$$\begin{aligned}\lim_{K\to\infty} \sup_{\theta\in U_{n_0}} \left(\mathbb{E}^{\mu_\theta}[X\mathbb{1}_{\{|X|>K\}}]\right)^2 &\le \lim_{K\to\infty} \sup_{\theta\in U_{n_0}} \mathbb{E}^{\mu_\theta}[X^2]\mu_\theta(|X| > K) \\ &\le \lim_{K\to\infty} \sup_{\theta\in U_{n_0}} \left(\mathbb{E}^{\mu_\theta}[X^2]\right)^2/K^2 \le \lim_{K\to\infty} n_0^2/K^2 = 0.\end{aligned}$$

Therefore, X is π-almost uniformly integrable.

We have the following alternative formula for (5.2) under the canonical ambiguity-reduction mapping when the statistic is given by an optimized certainty equivalent.

Theorem 5.4. *Let $\pi \in \Delta(\Theta)$ and $g \in \Upsilon$, $X \in \mathcal{C}$. If, for all $m \in \mathbb{R}$, $h(X-m)$ is π-almost uniformly integrable, then*

$$\sup_{\nu\in\mathcal{S}(\pi,g)} \gamma_h^{\mathrm{OCE}}(X|\nu) = \inf_{m\in\mathbb{R}} \left\{ m + \int_\Theta \mathbb{E}^{\mu_\theta}[h(X-m)]\, g\circ\pi(\mathrm{d}\theta) \right\}.$$

Remark 5.1. When $h(x) = x_+$ for all $x \in \mathbb{R}$, the condition that $h(X-m)$ is π-almost uniformly integrable for each $m \in \mathbb{R}$ is implied by X being π-almost uniformly integrable.

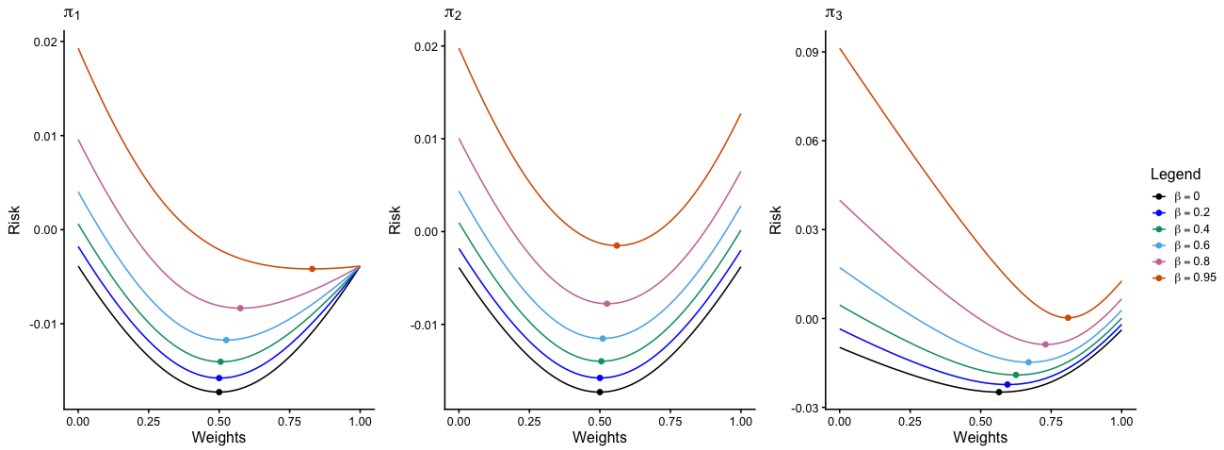


Figure 5.1: Worst-case Expected Shortfall

Figure 5.1 presents the values of the decision functional (5.7) when $\gamma = \mathrm{ES}_{0.95}$. Under the second-order model π_1 (left panel), the ambiguity-neutral case ($\beta = 0$) is minimized by an equal allocation between X_1 and X_2. This is consistent with the preference for equal diversification, since X_1 and X_2 have the same distribution under the predictive model. As β increases, however, the minimizing portfolio shifts monotonically toward a full allocation in X_1. This is consistent with interpreting β as a parameter of ambiguity aversion: under π_1, ambiguity is perceived in the distribution of X_2, but ambiguity is not perceived in the distribution of X_1.

This story changes in the center and right panels, corresponding to the second-order models π_2 and π_3 respectively. In both cases, ambiguity is perceived in the distributions of X_1 and X_2. For π_2, increasing β produces little deviation from the fact that an equal allocation between X_1 and X_2 minimizes the decision functional. This is because the ambiguity associated with the two assets is roughly comparable. The case of π_3 is markedly different. There, the predictive distribution of X_2 is roughly as favorable as that of X_1, but π_3 perceives greater ambiguity in X_2 than in X_1 (e.g., the standard deviation of X_2 is ambiguous). As a result, when ambiguity aversion increases, the minimizer shifts toward a larger allocation in X_1.

Let $\ell : \mathbb{R} \to \mathbb{R}$ be a loss function. Let $\pi \in \Delta(\Theta)$ and $g \in \Upsilon$, one readily obtains the following equivalent formula.

$$\sup_{\nu \in \mathcal{S}(\pi, g)} \mathbb{E}_\ell(X|\nu) = \ell^{-1}\left(\int_\Theta \mathbb{E}^{\mu_\theta}[\ell(X)] g \circ \pi(\mathrm{d}\theta)\right).$$

This specific decision functional is an example of the decision functionals from the second-order dual expected utility model of Giraud [2014]. In that work, the author also introduces their model as an alternative to the smooth model, in which the second-order aggregation of the expected loss is given by the dual model of Yaari [1987]. When the loss function ℓ is given by $\ell(x) = e^{\theta x}$ for all $x \in \mathbb{R}$, where $\theta > 0$, the statistic $\mathbb{E}_\ell$ coincides with the entropic risk measure ER_θ; see Föllmer and Schied [2002].

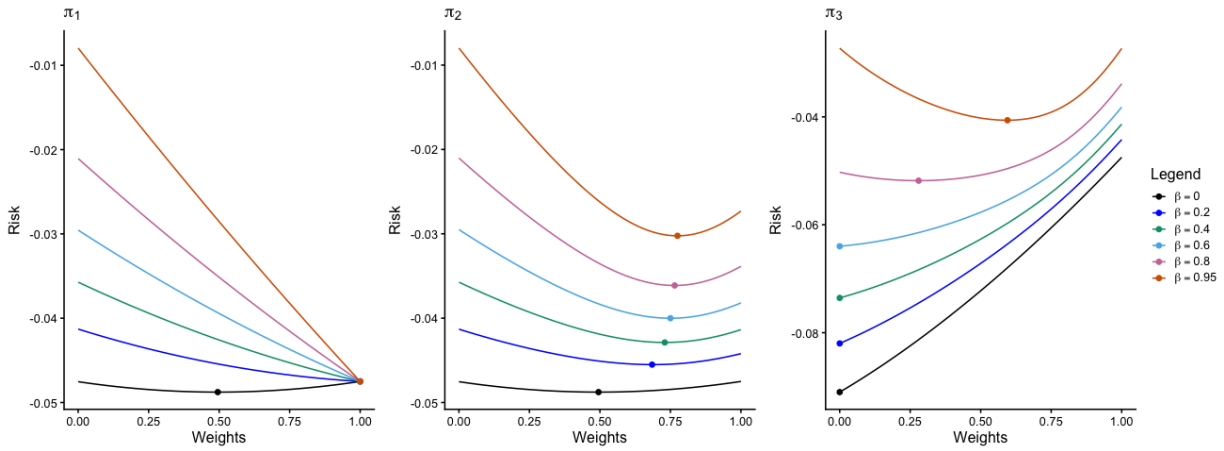


Figure 5.2: Worst-case entropic risk measure

Figure 5.2 presents the values of the decision functional (5.7) when $\gamma = \mathrm{ER}_{10}$. An interesting contrast between the numerical results in Figure 5.1 and those in Figure 5.2 happens under the second-order model π_3 (right panel). For low values of β, the minimizing portfolio consists of a full allocation in X_2. As β increases, however, the minimizing portfolio shifts toward X_1. Intuitively, this happens since the standard deviation of X_2 is ambiguous. This differs from the previous results, since the statistic ER_{10} evaluated on the predictive distribution of X_2 is much smaller than when evaluated on the predictive distribution of X_1.

The last statistic we will provide an algorithm for is the value at risk. The worst-case value at risk was recently rigorously studied in Liu et al. [2024].

Proposition 5.6. *Let $\pi \in \Delta(\Theta)$, $g \in \Upsilon$, and $\alpha \in (0,1)$. We have $\sup_{\nu \in \mathcal{S}(\pi,g)} \mathrm{VaR}_\alpha(X|\nu)$ is finite, and*

$$\sup_{\nu \in \mathcal{S}(\pi,g)} \mathrm{VaR}_\alpha(X|\nu) = \inf\left\{ x \in \mathbb{R} \;\middle|\; \int_\Theta \mu_\theta(X > x) g \circ \pi(\mathrm{d}\theta) \leq 1 - \alpha \right\}, \quad X \in \mathcal{X}.$$

Figure 5.3 presents the values of the decision functional (5.7) when $\gamma = \text{VaR}_{0.95}$. As we can see, the results for the Value-at-Risk are qualitatively similar to the Expected Shortfall results presented in Figure 5.1.

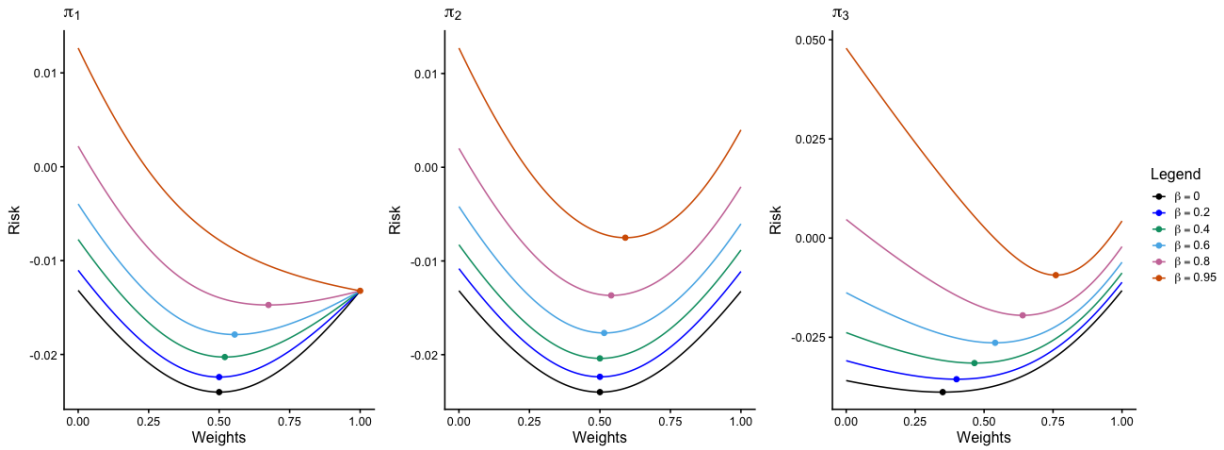


Figure 5.3: Worst-case Value-at-Risk

5.5 Connections to existing literature

In this section, we present several technical results that complement the chapter's main analysis. These results highlight the relevance of the decision functionals in (5.2) under the canonical ambiguity-reduction mapping to broader questions and concerns in the literature.

5.5.1 Unambiguous events

Let $\tau_{\mathcal{D}}$ denote the topology on $\mathcal{D}$ associated with convergence in distribution and let $\Sigma_{\mathcal{D}}$ be the σ-algebra generated by $\tau_{\mathcal{D}}$. It follows that $(\mathcal{D}, \Sigma_{\mathcal{D}})$ is a Polish measurable space.

Proposition 5.7. *Given $X \in \mathcal{X}$, the mapping $\iota_X : \Delta \to \mathcal{D}$ is $\Sigma/\Sigma_{\mathcal{D}}$-measurable.*

Given $\pi \in \Pi$ and $X \in \mathcal{X}$, the probability measure $(\iota_X)_{\#}\pi$ on $(\mathcal{D}, \Sigma_{\mathcal{D}})$ represents the epistemic belief regarding the distribution of X. Given $\pi \in \Pi$, define

$$\mathfrak{C}(\pi) = \{X \in L^1(\pi) \mid (\iota_X)_{\#}\pi = \delta_Q \text{ for some } Q \in \mathcal{D}_1\}.$$

Intuitively, $\mathfrak{C}(\pi)$ represents the set of acts with an unambiguous distribution under the second-order model $\pi \in \Pi$. Additionally, for $\pi \in \Pi$, define the set

$$\mathfrak{U}(\pi) = \{A \in \mathcal{F} : \mathbb{1}_A \in \mathfrak{C}(\pi)\}.$$

We can view $\mathfrak{U}(\pi)$ as the set of unambiguous events.

Proposition 5.8. *Given $\pi \in \Pi$, the set $\mathfrak{U}(\pi)$ is a λ-system, that is,*

(i) $\Omega \in \mathfrak{U}(\pi)$;

(ii) if $A \in \mathfrak{U}(\pi)$ *then* $A^c \in \mathfrak{U}(\pi)$;

(iii) if the sequence $(A_n)_{n\in\mathbb{N}} \subseteq \mathfrak{U}(\pi)$ *is pairwise disjoint then* $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{U}(\pi)$.

Moreover, $\mathfrak{C}(\pi) = \{X \in L^1(\pi) \mid X^{-1}(B) \in \mathfrak{U}(\pi) \text{ for all } B \in \mathcal{B}(\mathbb{R})\}$.

The use of a λ-system for unambiguous events is common in the literature; see Zhang [2002], Marinacci [2002], Epstein [2004], and Van Oosten and Wang [2025b]. Given $\pi \in \Pi$ and $X \in \mathfrak{C}(\pi)$, let $Q_X^\pi \in \mathcal{D}_1$ denote the distribution satisfying $(\iota_X)_\# \pi = \delta_{Q_X^\pi}$. The following proposition confirms that, when an act has an unambiguous distribution, the decision functional given in (5.2) under the canonical ambiguity-reduction mapping reduces to calculating the statistic on this unambiguous distribution.

Proposition 5.9. *Let γ be a statistic, $\pi \in \Pi$, and $g \in \Upsilon$. We have that*

$$\sup_{\nu \in \mathcal{S}(\pi, g)} \gamma(X|\nu) = \gamma(Q_X^\pi), \quad X \in \mathfrak{C}(\pi).$$

Decision functionals in (5.2) under the canonical ambiguity-reduction mapping are invariant to epistemic distributional beliefs, as stated in the following proposition.

Proposition 5.10. *Let $\pi \in \Pi$ be atomless, $g \in \Upsilon$, and γ be a statistic. We have,*

$$(\iota_X)_\# \pi = (\iota_Y)_\# \pi \implies \sup_{\mu \in \mathcal{S}(\pi, g)} \gamma(X|\mu) = \sup_{\mu \in \mathcal{S}(\pi, g)} \gamma(Y|\mu), \quad X, Y \in L^1(\pi).$$

5.5.2 Bayesian consistency

In this section and the following one, we will assume that the parametrization $\mathfrak{P}$ is continuous. Denote by $\mathcal{D}_c$ the set of all Borel probability measures on $\mathbb{R}$ with compact

support. Additionally, in this section and the next, statistics will have domain $\mathcal{D}_c$ rather than $\mathcal{D}_1$. A collection $\mathcal{Q} \subseteq \mathcal{D}_c$ is said to have *uniformly bounded support* if there exists a compact set that contains the support of each $Q \in \mathcal{Q}$. We say that a statistic γ is *Lebesgue continuous* if for all sequences $(Q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}_c$ with uniformly bounded support satisfying $Q_n \Rightarrow Q$ for some $Q \in \mathcal{D}_c$, we have that $\lim_{n\to\infty} \gamma(Q_n) = \gamma(Q)$.

Proposition 5.11. *Let γ be a Lebesgue continuous statistic, $\pi \in \Delta(\Theta)$, and $g \in \Upsilon$. Then, for all $X \in \mathcal{X}_b$, there exists $\mu_X \in \mathcal{S}(\pi, g)$ such that*

$$\gamma(X|\mu_X) = \sup_{\nu\in\mathcal{S}(\pi,g)} \gamma(X|\nu).$$

In Bayesian statistics, consistency concerns the asymptotic behavior of the posterior distribution (second-order model). For example, suppose we observe an IID sequence of random variables $(\zeta_k)_{k\in\mathbb{N}}$ taking values in Ω. Assume that $\zeta_k \sim \mu_{\theta_0}$ for some $\theta_0 \in \Theta$. Under certain regularity conditions, see Ghosal and van der Vaart [2017], we have that the sequence of second-order models $(\pi_n)_{n\in\mathbb{N}}$ given by the posterior distribution (π_n is obtained after realizing $(\zeta_k)_{k=1}^n$) satisfies

$$\pi_n \Rightarrow \delta_{\theta_0} \quad \mu_{\theta_0}^\infty\text{-a.s.}$$

The following proposition verifies that, for acts in $\mathcal{C}_b$, the decision functional (5.2), under the canonical ambiguity-reduction mapping, asymptotically agrees with the correct value.

Proposition 5.12. *Let $(\pi_n)_{n\in\mathbb{N}} \subseteq \Delta(\Theta)$ be a sequence such that $\pi_n \Rightarrow \delta_{\theta_0}$ for some $\theta_0 \in \Theta$, $g \in \Upsilon$, and γ be a Lebesgue continuous statistic. Then, we have*

$$\lim_{n\to\infty} \max_{\nu\in\mathcal{S}(\pi_n,g)} \gamma(X|\nu) = \gamma(X|\mu_{\theta_0}), \quad X \in \mathcal{C}_b.$$

5.5.3 Model misspecification

Model misspecification refers to the case in which the parametric model $(\Theta, \mathfrak{P})$ fails to contain the true first-order model that generates the random phenomenon. In the Bayesian literature, there are results on the asymptotic behavior of the posterior distribution (second-order model). For example, as in the previous section, suppose we observe an IID sequence of random variables $(\zeta_k)_{k\in\mathbb{N}}$ taking values in Ω. We now consider the misspecified case, meaning that $\zeta_n \sim \mu_0$, but, different from the previous section, there does not exist a parameter $\theta_0 \in \Theta$ such that $\mu_0 = \mu_{\theta_0}$. However, assume that the mapping

$$\Theta \to \mathbb{R} \cup \{\infty\} : \theta \mapsto \mathrm{KL}(\mu_\theta|\mu_0),$$

admits a unique minimizer. We denote this minimizer by θ_0. Under certain regularity conditions, see Berk [1966], we have that the sequence of second-order models $(\pi_n)_{n\in\mathbb{N}}$ given by the posterior distribution (π_n is obtained after realizing $(\zeta_k)_{k=1}^n$) satisfies

$$\pi_n \Rightarrow \delta_{\theta_0} \quad \mu_0^\infty\text{-a.s.}$$

Therefore, a natural way for a DM to address model misspecification is to first robustify their statistic. The following proposition discusses how this can be done.

Proposition 5.13. *Given a statistic γ, atomless $\mu \in \Delta$, and $\epsilon > 0$, there exists a statistic $\gamma_\epsilon^{\mathrm{KL}}$ satisfying*

$$\gamma_\epsilon^{\mathrm{KL}}(X|\mu) = \sup_{\nu\in \mathrm{KL}(\mu;\epsilon)} \gamma(X|\nu), \quad X \in \mathcal{X}_b,$$

where $\mathrm{KL}(\mu;\epsilon) = \{\nu \in \Delta \mid \mathrm{KL}(\nu|\mu) \leq \epsilon\}$. *Moreover, if γ is Lebesgue continuous, then $\gamma_\epsilon^{\mathrm{KL}}$ is Lebesgue continuous.*

Example 5.2. Given $\epsilon > 0$, the statistic $\mathbb{E}_\epsilon^{\mathrm{KL}}$ is known as the entropic Value-at-Risk, see Föllmer and Knispel [2011] and Ahmadi-Javid [2012].

Assuming the DM uses γ to represent their pure-risk preferences, if they were worried about model misspecification, given $\pi \in \Delta(\Theta)$, $g \in \Upsilon$, and $\epsilon > 0$, they could define the decision functional

$$\rho(X;\pi,g,\epsilon) = \sup_{\nu\in\mathcal{S}(\pi,g)} \gamma_\epsilon^{\mathrm{KL}}(X|\nu), \quad X \in \mathcal{X}_b. \tag{5.9}$$

The parameter ϵ in (5.9) represents the DM's fear of model misspecification. For example, if it holds that $\epsilon \geq \mathrm{KL}(\mu_{\theta_0}|\mu_0)$ and γ is Lebesgue continuous, then, by Proposition 5.12, we have

$$\lim_{n\to\infty} \rho(X;\pi_n,g,\epsilon) = \gamma_\epsilon^{\mathrm{KL}}(X|\mu_{\theta_0}) \geq \gamma(X|\mu_0), \quad X \in \mathcal{C}_b. \tag{5.10}$$

Therefore, the DM does not asymptotically underestimate the true value of their statistic under model misspecification.

Independent of the asymptotic result, the decision functional in (5.9) is interesting from a conceptual point of view. This follows because the decision functional separates the perception of ambiguity (π), the attitude toward ambiguity (g), and the attitude toward model misspecification (ϵ). The joint incorporation of attitudes toward ambiguity and model misspecification into decision functionals was discussed in Cerreia-Vioglio et al. [2025].

In Shapiro et al. [2023], the authors also study decision functionals in a Bayesian context that address model misspecification. In that work, given $\pi \in \Delta(\Theta)$ and $\epsilon > 0$, the authors

define the decision functional

$$\rho(X;\pi,\epsilon) = \int_{\Theta} \mathbb{E}_{\epsilon}^{\mathrm{KL}}(X|\mu_\theta)\pi(\mathrm{d}\theta), \quad X \in \mathcal{X}_b. \tag{5.11}$$

It is not hard to show that for all $X \in \mathcal{C}_b$, a similar consistency result to (5.10) can be shown to hold for the decision functional given in (5.11). However, there is no mechanism in the decision functional (5.11) to account for attitudes toward ambiguity. Moreover, we have the following proposition. We verify in Section 5.12 that, given $X \in \mathcal{X}_b$, $\mathbb{E}_{\epsilon}^{\mathrm{KL}} \circ \iota_X$ is Σ-measurable.

Proposition 5.14. *If* $\tilde{\pi}, \pi \in \Pi$ *satisfy* $\tilde{\pi} \geq_{\mathrm{cx}} \pi$*, then*

$$\int_{\Delta} \mathbb{E}_{\epsilon}^{\mathrm{KL}}(X|\nu)\tilde{\pi}(\mathrm{d}\nu) \leq \int_{\Delta} \mathbb{E}_{\epsilon}^{\mathrm{KL}}(X|\nu)\pi(\mathrm{d}\nu), \quad X \in \mathcal{X}_b$$

Therefore, the decision functional in (5.11) reflects an ambiguity-seeking attitude, as discussed in Section 5.3. Although the decision functionals in (5.9) and (5.11) exhibit a similar asymptotic property, they can lead to substantially different decisions in finite samples.

5.6 Conclusion

This chapter suggests that Bayesian and robust approaches to ambiguity may not be seen as alternatives, but as complementary ingredients of a richer theory of decision-making under ambiguity. The central message is that the role of second-order beliefs in ambiguity-sensitive decision-making extends beyond the smooth model and can be generalized to arbitrary statistics within a DRO framework. Seen this way, the key modeling question is not only what a decision-maker believes, but also how those beliefs are converted into ambiguity sets. This shift in perspective opens the door to more flexible constructions of decision functionals under ambiguity, where ambiguity attitudes, statistical learning, and model misspecification can be treated simultaneously. We hope it also helps motivate a broader study of how second-order beliefs can be transformed into ambiguity sets.

5.7 Additional results needed for the proofs

Given $n \in \mathbb{N}$, denote by $\mathcal{X}_n$ the set of all $X \in \mathcal{X}_b$ such that $n \geq X \geq -n$. Additionally, denote by $\mathcal{C}_n$ the set of all $X \in \mathcal{C}_b$ such that $n \geq X \geq -n$.

Lemma 5.1. *Given $n \in \mathbb{N}$ and $\mathcal{Y} \subseteq \mathcal{X}_n$, if $\mathcal{Y}$ is closed under pointwise convergence and $\mathcal{C}_n \subseteq \mathcal{Y}$, then $\mathcal{Y} = \mathcal{X}_n$.*

Proof. Define $\mathcal{Y}_0 = \{X \in \mathcal{X} : (X \wedge n) \vee (-n) \in \mathcal{Y}\}$. Since $\mathcal{Y}$ is closed under pointwise convergence, $\mathcal{Y}_0$ is closed under pointwise convergence. Since $\mathcal{C}_n \subseteq \mathcal{Y}$, $\mathcal{C} \subseteq \mathcal{Y}_0$. Therefore, by Kechris [2012, Theorem 11.6], $\mathcal{Y}_0 = \mathcal{X}$. Therefore, $\mathcal{Y} = \mathcal{X}_n$. □

Denote by τ_b the coarsest topology on Δ where the maps $\{\psi_X\}_{X \in \mathcal{X}_b}$ are continuous.

Proposition 5.15. *Let γ be a Lebesgue continuous statistic. Given $X \in \mathcal{C}_b$, $\gamma \circ \iota_X$ is τ-continuous. Given $X \in \mathcal{X}_b$, $\gamma \circ \iota_X$ is τ_b-continuous.*

Proof. Given $a, b \in \mathbb{R}$ satisfying $a \leq b$, denote the set of Borel measures on $\mathbb{R}$ with support contained in the interval $[a, b]$ by $\mathcal{D}_{[a,b]}$. Denote by $\tau_{\mathcal{D}_{[a,b]}}$ the subspace topology on $\mathcal{D}_{[a,b]}$ from the topology $\tau_{\mathcal{D}}$. As γ is Lebesgue continuous, $\gamma|_{\mathcal{D}_{[a,b]}}$ is $\tau_{\mathcal{D}_{[a,b]}}$-continuous.

Let $X \in \mathcal{C}_b$, we claim that ι_X is $\tau/\tau_{\mathcal{D}}$-continuous. To see this, let $(\mu_k)_{k \in \mathbb{N}} \subseteq \Delta$ be a sequence such that $\mu_k \Rightarrow \mu$ for some $\mu \in \Delta$. Let $f : \mathbb{R} \to \mathbb{R}$ be bounded and continuous, we have

$$\lim_{k \to \infty} \int_{\mathbb{R}} f \mathrm{d}\iota_X(\mu_k) = \lim_{k \to \infty} \mathbb{E}^{\mu_k}[f(X)] = \mathbb{E}^{\mu}[f(X)] = \int_{\mathbb{R}} f \mathrm{d}\iota_X(\mu).$$

Therefore, $\iota_X(\mu_k) \Rightarrow \iota_X(\mu)$, and ι_X is $\tau/\tau_{\mathcal{D}}$-continuous. Since $\iota_X(\Delta) \subseteq \mathcal{D}_{[a,b]}$, where $a \leq X \leq b$, we have that $\gamma \circ \iota_X$ is τ-continuous. A similar argument will show that, for all $X \in \mathcal{X}_b$, $\gamma \circ \iota_X$ is τ_b-continuous. □

Proposition 5.16. *Let γ be a Lebesgue continuous statistic and $X \in \mathcal{X}_b$. The mapping $\gamma \circ \iota_X$ is Σ-measurable.*

Proof. Fix $n \in \mathbb{N}$ and define

$$\mathcal{Y} = \{X \in \mathcal{X}_n \mid \gamma \circ \iota_X \text{ is } \Sigma\text{-measurable}\}.$$

By Proposition 5.15, we have that $\mathcal{C}_n \subseteq \mathcal{Y}$. Let $(X_k)_{k \in \mathbb{N}} \subseteq \mathcal{Y}$ converge pointwise to some $X \in \mathcal{X}_n$. Since γ is Lebesgue continuous, $n \geq X_k \geq -n$ for all $k \in \mathbb{N}$, and $(X_k)_{\#}\nu \Rightarrow X_{\#}\nu$

for all $\nu \in \Delta$, we have that

$$\lim_{k\to\infty} \gamma \circ \iota_{X_k}(\nu) = \lim_{k\to\infty} \gamma(X_k|\nu) = \gamma(X|\nu) = \gamma \circ \iota_X(\nu), \quad \nu \in \Delta.$$

Since $\gamma \circ \iota_{X_k}$ is Σ-measurable for all $k \in \mathbb{N}$, $\gamma \circ \iota_X$ is Σ-measurable. By Lemma 5.1, we have that $\mathcal{Y} = \mathcal{X}_n$. Since $n \in \mathbb{N}$ was arbitrary, the proof is complete. □

Proposition 5.16 guarantees that the integrals in the following proposition are well-defined.

Proposition 5.17. *Let γ be a Lebesgue continuous statistic and $\varphi : \mathbb{R} \to \mathbb{R}$ be continuous. Assume that for all $\tilde{\pi}, \pi \in \Pi$ satisfying $\tilde{\pi} \geq_{\mathrm{cx}} \pi$ we have*

$$\int_\Delta \varphi(\gamma(X|\nu))\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta \varphi(\gamma(X|\nu))\pi(\mathrm{d}\nu), \quad X \in \mathcal{C}_b.$$

Then, for all $\tilde{\pi}, \pi \in \Pi$ satisfying $\tilde{\pi} \geq_{\mathrm{cx}} \pi$ we have

$$\int_\Delta \varphi(\gamma(X|\nu))\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta \varphi(\gamma(X|\nu))\pi(\mathrm{d}\nu), \quad X \in \mathcal{X}_b.$$

Proof. Fix $n \in \mathbb{N}$ and define

$$\mathcal{Y} = \left\{ X \in \mathcal{X}_n \;\middle|\; \int_\Delta \varphi(\gamma(X|\nu))\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta \varphi(\gamma(X|\nu))\pi(\mathrm{d}\nu) \text{ for all } \tilde{\pi}, \pi \in \Pi \text{ s.t. } \tilde{\pi} \geq_{\mathrm{cx}} \pi \right\}.$$

By assumption we have that $\mathcal{C}_n \subseteq \mathcal{Y}$. Let $(X_k)_{k\in\mathbb{N}} \subseteq \mathcal{Y}$ converge pointwise to some $X \in \mathcal{X}_n$. Since γ is Lebesgue continuous, $n \geq X_k \geq -n$ for all $k \in \mathbb{N}$, and $(X_k)_\# \nu \Rightarrow X_\# \nu$ for all $\nu \in \Delta$, we have that

$$\lim_{k\to\infty} \gamma(X_k|\nu) = \gamma(X|\nu), \quad \nu \in \Delta.$$

Given $\tilde{\pi}, \pi \in \Pi$ satisfying $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, by the dominated convergence theorem, we have

$$\begin{aligned}\int_\Delta \varphi(\gamma(X|\nu))\tilde{\pi}(\mathrm{d}\nu) = \lim_{k\to\infty} \int_\Delta \varphi(\gamma(X_k|\nu))\tilde{\pi}(\mathrm{d}\nu) &\geq \lim_{k\to\infty} \int_\Delta \varphi(\gamma(X_k|\nu))\pi(\mathrm{d}\nu) \\ &= \int_\Delta \varphi(\gamma(X|\nu))\pi(\mathrm{d}\nu).\end{aligned}$$

Since $\pi, \tilde{\pi}$ were arbitrary, we have that $X \in \mathcal{Y}$. By Lemma 5.1, we have that $\mathcal{Y} = \mathcal{X}_n$. Since $n \in \mathbb{N}$ was arbitrary, the proof is complete. □

Corollary 5.1. *Given convex $\varphi : \mathbb{R} \to \mathbb{R}$ and $\tilde{\pi}, \pi \in \Pi$ satisfying $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, we have*

$$\int_\Delta \varphi(\mathbb{E}^\nu[X])\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta \varphi(\mathbb{E}^\nu[X])\pi(\mathrm{d}\nu), \quad X \in \mathcal{X}_b.$$

Proof. Let $X \in \mathcal{C}_b$. Since, by Proposition 5.15, $\mathbb{E} \circ \iota_X$ is τ-continuous, we have that $\nu \mapsto \varphi(\mathbb{E}^\nu[X])$ is proper, convex, and τ-continuous. Therefore, by the definition $\geq_{\mathrm{cx}}$, for all $\tilde{\pi}, \pi \in \Pi$ satisfying $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, we have

$$\int_\Delta \varphi(\mathbb{E}^\nu[X])\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta \varphi(\mathbb{E}^\nu[X])\pi(\mathrm{d}\nu).$$

Since $X \in \mathcal{C}_b$ was arbitrary, by Proposition 5.17, the proof is complete. □

5.8 Proofs accompanying Section 5.2

Proof of Proposition 5.1. If $X = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$, where $\{A_k\}_{k=1}^n \subseteq \mathcal{F}$ partitions Ω (i.e., X is simple), then we have

$$\int_\Delta \mathbb{E}^\nu[X]\pi(\mathrm{d}\nu) = \int_\Delta \sum_{k=1}^n a_k \nu(A_k)\pi(\mathrm{d}\nu) = \sum_{k=1}^n a_k \mu_\pi(A_k) = \mathbb{E}^{\mu_\pi}[X].$$

If $X \in \mathcal{X}$ is non-negative, find an increasing sequence of simple acts $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}$ that converges pointwise to X. By three applications of the monotone convergence theorem, we have

$$\int_\Delta \mathbb{E}^\nu[X]\pi(\mathrm{d}\nu) = \int_\Delta \lim_{n\to\infty} \mathbb{E}^\nu[X_n]\pi(\mathrm{d}\nu) = \lim_{n\to\infty} \int_\Delta \mathbb{E}^\nu[X_n]\pi(\mathrm{d}\nu) = \lim_{n\to\infty} \mathbb{E}^{\mu_\pi}[X_n] = \mathbb{E}^{\mu_\pi}[X].$$

If $X \in L^1(\pi)$ is general, first note that

$$\begin{aligned}\max\left\{\int_\Delta \mathbb{E}^\nu[X_+]\pi(\mathrm{d}\nu), \int_\Delta \mathbb{E}^\nu[X_-]\pi(\mathrm{d}\nu)\right\} &\leq \int_\Delta \left(\mathbb{E}^\nu[X_+] + \mathbb{E}^\nu[X_-]\right)\pi(\mathrm{d}\nu) \\ &= \int_\Delta \mathbb{E}^\nu[|X|]\pi(\mathrm{d}\nu) < \infty.\end{aligned}$$

Therefore, $\mathbb{E}^\nu[X_+] < \infty$ π-a.s., $\mathbb{E}^\nu[X_-] < \infty$ π-a.s., and $\mathbb{E}^\nu[X] = \mathbb{E}^\nu[X_+] - \mathbb{E}^\nu[X_-]$ π-a.s.

Thus,

$$\int_\Delta \mathbb{E}^\nu[X]\pi(\mathrm{d}\nu) = \int_\Delta \mathbb{E}^\nu[X_+]\pi(\mathrm{d}\nu) - \int_\Delta \mathbb{E}^\nu[X_-]\pi(\mathrm{d}\nu) = \mathbb{E}^{\mu_\pi}[X_+] - \mathbb{E}^{\mu_\pi}[X_-] = \mathbb{E}^{\mu_\pi}[X],$$

and the proof is complete. □

Given a collection $\mathcal{Z} \subseteq \mathcal{F}$, we denote by $\sigma(\mathcal{Z})$ the σ-algebra generated by $\mathcal{Z}$. Given a collection $\mathcal{Y} \subseteq \mathcal{X}$, we denote by $\sigma(\mathcal{Y})$ the coarsest σ-algebra making each $X \in \mathcal{Y}$ measurable.

Proof of Theorem 5.1. By Bertsekas and Shreve [1996, Proposition 7.25], we know that

$$\Sigma = \sigma(\{\psi_X\}_{X \in \mathcal{X}_b}).$$

Let $\pi, \tilde{\pi} \in \Pi$ satisfy $\pi \geq_{\mathrm{cx}} \tilde{\pi}$ and $\tilde{\pi} \geq_{\mathrm{cx}} \pi$. Let $\Gamma \subseteq \Sigma$ denote the π-system consisting of all finite intersections of sets of the form $\psi_X^{-1}(B)$, where $X \in \mathcal{X}_b$ and $B \in \mathcal{B}(\mathbb{R})$. Clearly,

$$\Sigma = \sigma(\{\psi_X\}_{X \in \mathcal{X}_b}) = \sigma(\Gamma).$$

If we can show that π and $\tilde{\pi}$ agree on Γ, then, by the π-λ theorem, $\pi = \tilde{\pi}$. To this end, let $\{X_k\}_{k=1}^n \subseteq \mathcal{X}_b$ and $\{B_k\}_{k=1}^n \subseteq \mathcal{B}(\mathbb{R})$. Let $t \in \mathbb{R}^n$ and define $Z(t) = \sum_{k=1}^n t_k \psi_{X_k}$. By Corollary 5.1, for all convex $\varphi : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned}\int_\Delta \varphi\left(Z(t)\right) \mathrm{d}\pi = \int_\Delta \varphi\left(\mathbb{E}^\nu\left[\sum_{k=1}^n t_k X_k\right]\right) \pi(\mathrm{d}\nu) &= \int_\Delta \varphi\left(\mathbb{E}^\nu\left[\sum_{k=1}^n t_k X_k\right]\right) \tilde{\pi}(\mathrm{d}\nu) \\ &= \int_\Delta \varphi\left(Z(t)\right) \mathrm{d}\tilde{\pi}.\end{aligned}$$

Since φ was an arbitrary convex function, we have $\left(Z(t)\right)_\# \pi = \left(Z(t)\right)_\# \tilde{\pi}$. Since $t \in \mathbb{R}^n$ was arbitrary, $(\psi_{X_1}, \ldots, \psi_{X_n})_\# \pi = (\psi_{X_1}, \ldots, \psi_{X_n})_\# \tilde{\pi}$. Therefore,

$$\pi\left(\bigcap_{k=1}^n \psi_{X_k}^{-1}(B_k)\right) = \tilde{\pi}\left(\bigcap_{k=1}^n \psi_{X_k}^{-1}(B_k)\right).$$

Thus, π and $\tilde{\pi}$ agree on Γ, and the proof is complete. □

Proof of Proposition 5.2. We will use Vesely [2017, Theorem 2] to prove the claim. First note that Δ can be embedded into the vector space of signed measures with finite total

variation $\mathcal{M}$. Additionally, we can equip $\mathcal{M}$ with the coarsest topology $\tilde{\tau}$ making the mappings

$$\mathcal{M} \to \mathbb{R} : \nu \mapsto \mathbb{E}^{\nu}[X], \quad X \in \mathcal{C}_b,$$

continuous. Clearly, $\tilde{\tau}|_{\Delta} = \tau$.

We claim that $\Delta \subseteq \mathcal{M}$ is $\tilde{\tau}$-closed. To prove the claim, let $(\mu_\eta)_{\eta \in \mathcal{N}} \subseteq \Delta$ be a net converging to some $\mu \in \mathcal{M}$ in the topology $\tilde{\tau}$. Clearly, $\mu(\Omega) = 1$. Let d be a metric for Ω, $C \in \mathcal{F}$ be closed, and $\epsilon > 0$. Define, $X_\epsilon^C \in \mathcal{C}_b$ by

$$X_\epsilon^C(\omega) = \left(1 - \frac{d(\omega, A)}{\epsilon}\right)_+, \quad \omega \in \Omega,$$

where $d(\omega, A) = \inf\{d(\omega, \tilde{\omega}) \mid \tilde{\omega} \in A\}$. Since $X_\epsilon^C \geq 0$, we have that $\mathbb{E}^{\mu}[X_\epsilon^A] \geq 0$. Therefore,

$$\mu(C) = \lim_{\epsilon \to 0^+} \mathbb{E}^{\mu}[X_\epsilon^C] \geq 0.$$

For the sake of contradiction, assume there exists $A \in \mathcal{F}$ such that $\mu(A) < 0$. By inner regularity of μ, there exists a closed $C \in \mathcal{F}$ such that $\mu(C) < 0$, which is a contradiction. Therefore, $\mu(A) \geq 0$ for all $A \in \mathcal{F}$. Therefore, $\mu \in \Delta$ and we have proven the earlier claim.

Since Δ is $\tilde{\tau}$-closed, Δ is evenly convex, as defined in Vesely [2017]. A consequence of Proposition 5.1 is that, for all $\pi \in \Pi$, μ_π is the barycenter of π; see Vesely [2017] for the definition of the barycenter. Let $\pi \in \Pi$ and $f : \Delta \to \mathbb{R} \cup \{\infty\}$ be proper, convex, and τ-lower semicontinuous. Then, by Vesely [2017, Theorem 2], we have

$$\int_\Delta f(\nu)\pi(\mathrm{d}\nu) \geq f(\mu_\pi) = \int_\Delta f(\nu)\delta_{\mu_\pi}(\mathrm{d}\nu).$$

Therefore, since f was arbitrary, $\pi \geq_{\mathrm{cx}} \delta_{\mu_\pi}$. □

5.9 Proofs accompanying Section 5.3

Proof of Proposition 5.3. Let $X \in \mathcal{C}_b$. As ES_α is Lebesgue continuous, by Proposition 5.15, $\mathrm{ES}_\alpha \circ \iota_X$ is τ-continuous. Therefore, $-\mathrm{ES}_\alpha \circ \iota_X$ is proper, convex, and τ-continuous. Therefore, by the definition $\geq_{\mathrm{cx}}$, for all $\tilde{\pi}, \pi \in \Pi$ satisfying $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, we have

$$\int_\Delta -\mathrm{ES}_\alpha(X|\nu)\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta -\mathrm{ES}_\alpha(X|\nu)\pi(\mathrm{d}\nu).$$

Since $X \in \mathcal{C}_b$ was arbitrary, by Proposition 5.17, we have

$$\int_\Delta -\mathrm{ES}_\alpha(X|\nu)\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta -\mathrm{ES}_\alpha(X|\nu)\pi(\mathrm{d}\nu), \quad X \in \mathcal{X}_b,$$

and the proof is complete. □

Lemma 5.2. *Let $\pi \in \Pi$ and $g \in \Upsilon$, then $\mathcal{S}(\pi, g)$ is convex, τ_b-compact, and τ-compact.*

Proof. Convexity is clear, as $\mathcal{R}(\pi, g)$ is convex. Let $(\mu_\eta)_{\eta\in\mathcal{N}} \subseteq \mathcal{S}(\pi, g)$ be a net. We claim that there exists a subnet that converges in τ_b to some $\mu \in \mathcal{S}(\pi, g)$. Let $(\xi_\eta)_{\eta\in\mathcal{N}} \subseteq \mathcal{R}(\pi, g)$ be a net such that $\mu_{\xi_\eta} = \mu_\eta$ for all $\eta \in \mathcal{N}$. Since $g \circ \pi$ is submodular and continuous, by [Marinacci and Montrucchio, 2004, Theorem 4.2] and Bredon [1993, Theorem 7.14], there exists a subnet $(\xi_{\varphi(\beta)})_{\beta\in\mathcal{B}}$ and $\xi \in \mathcal{R}(\pi, g)$ such that

$$\lim_\beta \int_\Delta f \mathrm{d}\xi_{\varphi(\beta)} = \int_\Delta f \mathrm{d}\xi$$

for all bounded Σ-measurable $f : \Delta \to \mathbb{R}$. Define $\mu = \mu_\xi \in \mathcal{S}(\pi, g)$. We have that

$$\lim_\beta \mathbb{E}^{\mu_{\varphi(\beta)}}[X] = \lim_\beta \int_\Delta \mathbb{E}^\nu[X]\xi_{\varphi(\beta)}(\mathrm{d}\nu) = \int_\Delta \mathbb{E}^\nu[X]\xi(\mathrm{d}\nu) = \mathbb{E}^\mu[X], \quad X \in \mathcal{X}_b.$$

Therefore, $\mu_{\varphi(\beta)} \to \mu$ in τ_b. Since the original net $(\mu_\eta)_{\eta\in\mathcal{N}}$ was arbitrary, by Bredon [1993, Theorem 7.14], $\mathcal{S}(\pi, g)$ is τ_b-compact. Since $\tau \subseteq \tau_b$, $\mathcal{S}(\pi, g)$ is τ-compact. □

Proof of Theorem 5.2. The fact that $\mathcal{S}$ satisfies Properties (1)–(3) is clear. To show Property (4), let $\pi, \tilde{\pi} \in \Pi$ with $\tilde{\pi} \geq_{\mathrm{cx}} \pi$.

Fix $X \in \mathcal{X}_b$. If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, by Corollary 5.1, we have that

$$\int_{\mathbb{R}} \varphi(x)(\psi_X)_\#\tilde{\pi}(\mathrm{d}x) = \int_\Delta \varphi(\psi_X(\nu))\tilde{\pi}(\mathrm{d}\nu) \geq \int_\Delta \varphi(\psi_X(\nu))\pi(\mathrm{d}\nu) = \int_{\mathbb{R}} \varphi(x)(\psi_X)_\#\pi(\mathrm{d}x).$$

Therefore, $(\psi_X)_\#\tilde{\pi} \geq_{\mathrm{cx}} (\psi_X)_\#\pi$, where $\geq_{\mathrm{cx}}$ denotes the traditional convex order; see Shaked and Shanthikumar [2007, Section 3.A]. Therefore, we have

$$\begin{aligned}\sup_{\nu\in\mathcal{S}(\tilde{\pi},g)} \mathbb{E}^\nu[X] = \sup_{\xi\in\mathcal{R}(\tilde{\pi},g)} \int_\Delta \mathbb{E}^\nu[X]\xi(\mathrm{d}\nu) &= \int_\Delta \mathbb{E}^\nu[X] g \circ \tilde{\pi}(\mathrm{d}\nu) \\ &= \int_\Delta \psi_X(\nu) g \circ \tilde{\pi}(\mathrm{d}\nu)\end{aligned}$$

$$= \mathrm{D}_g((\psi_X)_{\#}\tilde{\pi})$$
$$\geq \mathrm{D}_g((\psi_X)_{\#}\pi) = \sup_{\nu \in \mathcal{S}(\pi,g)} \mathbb{E}^{\nu}[X],$$

where the inequality follows from the fact that $\mathrm{D}_g(Q) \geq \mathrm{D}_g(P)$ whenever $Q \geq_{\mathrm{cx}} P$, since g is concave; see Shaked and Shanthikumar [2007, Theorem 3.A.7].

Let $\mu \in \mathcal{S}(\pi, g)$, we have that

$$\mathbb{E}^{\mu}[X] \leq \sup_{\nu \in \mathcal{S}(\pi,g)} \mathbb{E}^{\nu}[X] \leq \sup_{\nu \in \mathcal{S}(\tilde{\pi},g)} \mathbb{E}^{\nu}[X], \quad X \in \mathcal{X}_b. \tag{5.12}$$

For the sake of contradiction, assume that $\mu \notin \mathcal{S}(\tilde{\pi}, g)$. As, by Lemma 5.2, $\mathcal{S}(\tilde{\pi}, g)$ is convex and τ-compact, by the Hahn-Banach separation theorem, there exists $Y \in \mathcal{C}_b$ such that

$$\sup_{\nu \in \mathcal{S}(\tilde{\pi},g)} \mathbb{E}^{\nu}[Y] < \mathbb{E}^{\mu}[Y],$$

which is a contradiction of (5.12). Therefore, $\mu \in \mathcal{S}(\tilde{\pi}, g)$, and $\mathcal{S}(\pi, g) \subseteq \mathcal{S}(\tilde{\pi}, g)$. □

Proof of Proposition 5.4. Let $\mu \in \mathcal{S}(\pi, g)$. Therefore, we can find $\xi \in \mathcal{R}(\pi, g)$ such that $\mu = \mu_{\xi}$. Let $a \in \mathbb{N}$ be given such that $g(x) \leq ax$ for all $x \in [0, 1]$. Fix $X \in L^1(\pi)$, by the definition of $\mathcal{R}(\pi, g)$, it holds that

$$\int_{\Delta} \mathbb{E}^{\nu}[|X|]\xi(\mathrm{d}\nu) \leq \int_{\Delta} \mathbb{E}^{\nu}[|X|]g \circ \pi(\mathrm{d}\nu) \leq a \int_{\Delta} \mathbb{E}^{\nu}[|X|]\pi(\mathrm{d}\nu) < \infty.$$

Therefore $X \in L^1(\xi)$ and, by Proposition 5.1, $\mathbb{E}^{\mu}[|X|] = \int_{\Delta} \mathbb{E}^{\nu}[|X|]\xi(\mathrm{d}\nu) < \infty$. Thus, $X_{\#}\mu \in \mathcal{D}_1$. □

5.10 Proofs and additional results accompanying Section 5.4

Proof of Proposition 5.5. It is clear that

$$\mathfrak{P}_{\#}(\mathcal{R}_0(\pi, g)) \subseteq \mathcal{R}(\mathfrak{P}_{\#}\pi, g).$$

Let $\xi \in \mathcal{R}(\mathfrak{P}_{\#}\pi, g)$. It is clear that $\xi \ll \mathfrak{P}_{\#}\pi$. Let $\zeta = \frac{\mathrm{d}\xi}{\mathrm{d}\mathfrak{P}_{\#}\pi}$. Define

$$\tilde{\zeta} : \Theta \to [0, \infty) : \theta \mapsto \zeta(\mathfrak{P}(\theta)),$$

and $\tilde{\xi} \in \Delta(\Theta)$ by $\tilde{\xi}(A) = \int_A \tilde{\zeta} \mathrm{d}\pi$ for all $A \in \mathcal{A}$. Let $A \in \mathcal{A}$, by Condition (3), there exists $B \in \Sigma$ such that $\mathfrak{P}^{-1}(B) = A$. Therefore,

$$\tilde{\xi}(A) = \int_A \tilde{\zeta} \mathrm{d}\pi = \int_B \zeta \mathrm{d}(\mathfrak{P}_{\#}\pi) = \xi(B) \leq g(\mathfrak{P}_{\#}\pi(B)) = g(\pi(A)),$$

and $\tilde{\xi} \in \tilde{\mathcal{R}}(\pi, g)$. Since $\mathfrak{P}_{\#}\tilde{\xi} = \xi$, we have that $\xi \in \mathfrak{P}_{\#}(\mathcal{R}_0(\pi, g))$. Therefore,

$$\mathfrak{P}_{\#}(\mathcal{R}_0(\pi, g)) = \mathcal{R}(\mathfrak{P}_{\#}\pi, g).$$

Given $\xi \in \mathcal{R}_0(\pi, g)$, since

$$\int_\Theta \mu_\theta(A)\xi(\mathrm{d}\theta) = \int_\Delta \nu(A)\mathfrak{P}_{\#}\xi(\mathrm{d}\nu), \quad A \in \mathcal{F},$$

we have that $\mathcal{S}(\pi, g) = \mathcal{S}_0(\pi, g)$. □

Proof of Theorem 5.3. Define the function $s_n : \mathbb{R} \times \Theta^{\mathbb{N}} \to [0, 1]$ by

$$s_n(x, \zeta) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{(x,\infty)}(Y_k(\zeta)),$$

and the function $s : \mathbb{R} \to [0, 1]$ by $s(x) = \pi(\{\theta \in \Theta \mid \mathbb{E}^{\mu_\theta}[X] > x\})$. We have that

$$\begin{aligned}
&\sum_{k=1}^{n} \left(g\left(\frac{n-k+1}{n}\right) - g\left(\frac{n-k}{n}\right) \right) Y_{k:n}(\zeta) \\
&= \int_0^\infty g(s_n(x, \zeta))\mathrm{d}x + \int_{-\infty}^0 (g(s_n(x; \zeta)) - 1)\mathrm{d}x, \quad \zeta \in \Theta^{\mathbb{N}}.
\end{aligned}$$

For each $q \in \mathbb{Q}$, let $A_q \in \mathcal{A}^{\mathbb{N}}$ be a set with $\kappa(A_q) = 1$ and $\lim_{n\to\infty} s_n(q, \zeta) = s(q)$ for all $\zeta \in A_q$. Define $A = \bigcap_{q\in\mathbb{Q}} A_q \in \mathcal{A}^{\mathbb{N}}$. Clearly, $\kappa(A) = 1$. We claim that for all $\zeta \in A$, $s_n(\cdot, \zeta) \to s$ Lebesgue-a.e. To see this, let $x \in \mathbb{R}$ be a continuity point of s. Find sequences $(q_m^-)_{m\in\mathbb{N}} \subseteq \mathbb{Q}$ and $(q_m^+)_{m\in\mathbb{N}} \subseteq \mathbb{Q}$ such that $q_m^- \uparrow x$ and $q_m^+ \downarrow x$. We have that

$$s(q_m^-) = \lim_{n\to\infty} s_n(q_m^-, \zeta) \leq \liminf_{n\to\infty} s_n(x, \zeta) \leq \limsup_{n\to\infty} s_n(x, \zeta) \leq \lim_{n\to\infty} s_n(q_m^+, \zeta) = s(q_m^+).$$

Therefore, by virtue of x be a continuity point of s, we have that

$$s(x) = \liminf_{n\to\infty} s_n(x,\zeta) = \limsup_{n\to\infty} s_n(x,\zeta).$$

Thus, $\lim_{n\to\infty} s_n(x,\zeta) = s(x)$. Since the set of discontinuity points is countable, the original claim follows.

For $q \in \mathbb{N}$, let $B_q \in \mathcal{A}^{\mathbb{N}}$ be a set with $\kappa(B_q) = 1$ and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} (Y_k(\zeta) - q)_+ = \int_\Theta (\mathbb{E}^{\mu_\theta}[X] - q)_+ \pi(\mathrm{d}\theta), \quad \zeta \in B_q.$$

For $q \in \mathbb{N}$, let $C_q \in \mathcal{A}^{\mathbb{N}}$ be a set with $\kappa(C_q) = 1$ and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} (-Y_k(\zeta) - q)_+ = \int_\Theta (-\mathbb{E}^{\mu_\theta - q}[X])_+ \pi(\mathrm{d}\theta), \quad \zeta \in C_q.$$

Define $D = (\bigcap_{q\in\mathbb{N}} B_q) \cap (\bigcap_{q\in\mathbb{N}} C_q) \cap A \in \mathcal{A}^{\mathbb{N}}$. Clearly, $\kappa(D) = 1$.

Fix $\zeta \in D$, we claim that

$$\begin{aligned}
&\lim_{n\to\infty} \left(\int_0^\infty g(s_n(x,\zeta))\mathrm{d}x + \int_{-\infty}^0 (g(s_n(x;\zeta)) - 1)\mathrm{d}x \right) \\
&= \int_0^\infty g(s(x))\mathrm{d}x + \int_{-\infty}^0 (g(s_n(x;\zeta)) - 1)\mathrm{d}x.
\end{aligned}$$

Given $\epsilon > 0$, find $M \in \mathbb{N}$ such that

$$\int_\Theta (\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) < \epsilon \quad \text{and} \quad \int_\Theta (-\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) < \epsilon.$$

Find $N \in \mathbb{N}$ such that for all $n \geq N$, we have

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{k=1}^{n} (Y_k(\zeta) - M)_+ - \int_\Theta (\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) \right| < \epsilon, \\
&\left| \frac{1}{n}\sum_{k=1}^{n} (-Y_k(\zeta) - M)_+ - \int_\Theta (-\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) \right| < \epsilon, \\
&\left| \int_0^M g(s_n(x,\zeta))\mathrm{d}x - \int_0^M g(s(x))\mathrm{d}x \right| < \epsilon,
\end{aligned}$$

$$\left|\int_{-M}^{0} g(s_n(x,\zeta))\mathrm{d}x - \int_{-M}^{0} g(s(x))\mathrm{d}x\right| < \epsilon,$$

where the dominated convergence theorem is invoked to get the last two inequalities. Let $a \in \mathbb{N}$ such that $g(x) \leq ax$ for all $x \in [0,1]$. For all $n \geq N$, we have that

$$\begin{aligned}
&\left|\sum_{k=1}^{n}\left(g\left(\frac{n-k+1}{n}\right) - g\left(\frac{n-k}{n}\right)\right) Y_{k:n}(\zeta) - \int_{\Theta} \mathbb{E}^{\mu_\theta}[X] g \circ \pi(\mathrm{d}\theta)\right| \\
&= \left|\int_{0}^{\infty} g(s_n(x,\zeta))\mathrm{d}x + \int_{-\infty}^{0} (g(s_n(x,\zeta)) - 1)\mathrm{d}x - \int_{0}^{\infty} g(s(x))\mathrm{d}x - \int_{-\infty}^{0} (g(s(x)) - 1)\mathrm{d}x\right| \\
&\leq \left|\int_{0}^{M} g(s_n(x,\zeta))\mathrm{d}x - \int_{0}^{M} g(s(x))\mathrm{d}x\right| \\
&\quad + \left|\int_{-M}^{0} (g(s_n(x,\zeta)) - 1)\mathrm{d}x - \int_{-M}^{0} (g(s(x)) - 1)\mathrm{d}x\right| \\
&\quad + \left|\int_{M}^{\infty} g(s_n(x,\zeta))\mathrm{d}x - \int_{M}^{\infty} g(s(x))\mathrm{d}x\right| \\
&\quad + \left|\int_{-\infty}^{-M} (g(s_n(x,\zeta)) - 1)\mathrm{d}x - \int_{-\infty}^{-M} (g(s(x)) - 1)\mathrm{d}x\right| \\
&< \left|\int_{-M}^{0} g(s_n(x,\zeta))\mathrm{d}x - \int_{-M}^{0} g(s(x))\mathrm{d}x\right| + \int_{M}^{\infty} g(s_n(x,\zeta))\mathrm{d}x + \int_{M}^{\infty} g(s(x))\mathrm{d}x \\
&\quad + \int_{-\infty}^{-M} (1 - g(s_n(x,\zeta)))\mathrm{d}x + \int_{-\infty}^{-M} (1 - g(s(x)))\mathrm{d}x + \epsilon \\
&\leq a\int_{M}^{\infty} s_n(x,\zeta)\mathrm{d}x + a\int_{M}^{\infty} s(x)\mathrm{d}x + \int_{-\infty}^{-M} (1 - s_n(x,\zeta))\mathrm{d}x + \int_{-\infty}^{-M} (1 - (s(x))\mathrm{d}x + 2\epsilon \\
&= \frac{a}{n}\sum_{k=1}^{n} (Y_k(\zeta) - M)_+ + a\int_{\Theta} (\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) + \frac{1}{n}\sum_{k=1}^{n} (-Y_k(\zeta) - M)_+ \\
&\quad + \int_{\Theta} (-\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) + 2\epsilon \\
&\leq \frac{a}{n}\sum_{k=1}^{n} (Y_k(\zeta) - M)_+ + \frac{1}{n}\sum_{k=1}^{n} (-Y_k(\zeta) - M)_+ + 3\epsilon + a\epsilon \\
&\leq a\int_{\Theta} (\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) + \int_{\Theta} (-\mathbb{E}^{\mu_\theta}[X] - M)_+ \pi(\mathrm{d}\theta) + 4\epsilon + 2a\epsilon \\
&\leq 5\epsilon + 3a\epsilon = (5 + 3a)\epsilon.
\end{aligned}$$

Since $\zeta \in D$ was arbitrary, the proof is complete. □

Lemma 5.3. *For $\pi \in \Delta(\Theta)$, $g \in \Upsilon$, and π-almost uniformly integrable $X \in L^1(\pi)$, we have that*

$$\lim_{K\to\infty} \sup_{\nu\in\mathcal{S}(\pi,g)} \mathbb{E}^{\nu}[|X|\mathbb{1}_{|X|>K}] = 0.$$

That is, the set of distributions $\{X_{\#}\nu : \nu \in \mathcal{S}(\pi, g)\} \subseteq \mathcal{D}_1$ is uniformly integrable.

Proof. Let the sequence $(U_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ satisfy $\pi(U_n) > 1 - 1/n$ and

$$\lim_{K\to\infty} \sup_{\theta\in U_n} \mathbb{E}^{\mu_\theta}\left[|X|\mathbb{1}_{\{|X|>K\}}\right] = 0$$

for all $n \in \mathbb{N}$. Without loss of generality, we may assume that $U_n \subseteq U_{n+1}$ for all $n \in \mathbb{N}$. Since, for all $n, \tilde{n} \in N$, the mappings

$$\theta \mapsto \mathbb{E}^{\mu_\theta}[|X|]\mathbb{1}_{U_n^c}(\theta) \quad \text{and} \quad \nu \mapsto \mathbb{E}^{\mu_\theta}[|X|]\mathbb{1}_{U_{\tilde{n}}^c}(\theta)$$

are comonotonic (see Marinacci and Montrucchio [2004]), by Marinacci and Montrucchio [2004, Theorem 4.7], there exists $\xi \in \mathcal{R}(\pi, g)$ such that

$$\int_{U_n^c} \mathbb{E}^{\mu_\theta}[|X|]g \circ \pi(\mathrm{d}\theta) = \int_{U_n^c} \mathbb{E}^{\mu_\theta}[|X|]\xi(\mathrm{d}\theta), \quad n \in \mathbb{N}.$$

Therefore, by the dominated convergence theorem, we have

$$\lim_{n\to\infty} \int_{U_n^c} \mathbb{E}^{\mu_\theta}[|X|]g \circ \pi(\mathrm{d}\theta) = \lim_{n\to\infty} \int_{U_n^c} \mathbb{E}^{\mu_\theta}[|X|]\xi(\mathrm{d}\theta) = \int_{U^c} \mathbb{E}^{\mu_\theta}[|X|]\xi(\mathrm{d}\theta) = 0,$$

where $U = \bigcup_{n=1}^{\infty} U_n$, and the last equality follows from the fact that $\pi(U) = 1$.

Fix $\epsilon > 0$, and find $N \in \mathbb{N}$ such that $\int_{U_N^c} \mathbb{E}^{\mu_\theta}[|X|]g \circ \pi(\mathrm{d}\theta) < \epsilon$. We have that

$$\begin{aligned}
&\sup_{\nu\in\mathcal{S}(\pi,g)} \mathbb{E}^{\nu}[|X|\mathbb{1}_{|X|>K}] \\
&= \sup_{\xi\in\mathcal{R}(\pi,g)} \int_{\Theta} \mathbb{E}^{\mu_\theta}[|X|\mathbb{1}_{|X|>K}]\xi(\mathrm{d}\theta) \\
&\leq \sup_{\xi\in\mathcal{R}(\pi,g)} \int_{U_N} \mathbb{E}^{\mu_\theta}[|X|\mathbb{1}_{|X|>K}]\xi(\mathrm{d}\theta) + \sup_{\xi\in\mathcal{R}(\pi,g)} \int_{U_N^c} \mathbb{E}^{\mu_\theta}[|X|\mathbb{1}_{|X|>K}]\xi(\mathrm{d}\theta) \\
&\leq \sup_{\theta\in U_N} \mathbb{E}^{\mu_\theta}[|X|\mathbb{1}_{|X|>K}] + \int_{U_N^c} \mathbb{E}^{\mu_\theta}[|X|]g \circ \pi(\mathrm{d}\theta)
\end{aligned}$$

$$\leq \sup_{\theta \in U_N} \mathbb{E}^{\mu_\theta}[|X| \mathbb{1}_{|X|>K}] + \epsilon.$$

Therefore,

$$\limsup_{K \to \infty} \sup_{\nu \in \mathcal{S}(\pi,g)} \mathbb{E}^{\nu}[|X| \mathbb{1}_{|X|>K}] \leq \epsilon.$$

Since $\epsilon > 0$ was arbitrary, the proof is complete. □

Proof of Theorem 5.4. Define the function $R : \mathcal{S}(\pi, g) \times \mathbb{R} \to \mathbb{R}$ by

$$R(\nu, m) = m + \mathbb{E}^{\nu}[h(X - m)].$$

The fact that R is well-defined follows from Proposition 5.4.

Given $m \in \mathbb{R}$, let $R_m : \mathcal{S}(\pi, g) \to \mathbb{R}$ be given by $R_m(\nu) = R(\nu, m)$. Clearly, R_m is affine for all $m \in \mathbb{R}$. We claim that R_m is continuous for all $m \in \mathbb{R}$ when we equip $\mathcal{S}(\pi, g)$ with the subspace topology $\tau|_{\mathcal{S}(\pi,g)}$. To see this, let $(\nu_n)_{n\in\mathbb{N}} \subseteq \mathcal{S}(\pi, g)$ be a sequence converging to some $\nu \in \mathcal{S}(\pi, g)$. Note that we only need to check sequences since τ is metrizable. Fix $m \in \mathbb{R}$, and define $Y^{(m)} = h(X - m) \in L^1(\pi)$, the sequence $(Q_n^{(m)})_{n\in\mathbb{N}} \subseteq \mathcal{D}_1$ by $Q_n^{(m)} = Y_{\#}^{(m)}\nu_n$ and $Q^{(m)} = Y_{\#}^{(m)}\nu$. Since $X \in \mathcal{C}$ and h is continuous, it is clear that $Q_n^{(m)} \Rightarrow Q^{(m)}$. Since $Y^{(m)}$ is π-almost uniformly integrable, by Lemma 5.3, $\{Q_n^{(m)}\}_{n\in\mathbb{N}}$ is uniformly integrable. Therefore,

$$\begin{aligned}\lim_{n\to\infty} R_m(\nu_n) = m + \lim_{n\to\infty} \mathbb{E}^{\nu_n}[Y^{(m)}] &= m + \lim_{n\to\infty} \mathbb{E}(Q_n^{(m)}) \\ &= m + \mathbb{E}(Q^{(m)}) = m + \mathbb{E}^{\nu}[Y^{(m)}] = R_m(\nu).\end{aligned}$$

Given $\nu \in \mathcal{S}(\pi, g)$, let $R_\nu : \mathbb{R} \to \mathbb{R}$ be given by $R_\nu(m) = R(\nu, m)$. Clearly, R_ν is convex for all $\nu \in \mathcal{S}(\pi, g)$. Since $Y_{\#}^{(m)}\nu \in \mathcal{D}_1$ for all $m \in \mathbb{R}$ and $\nu \in \mathcal{S}(\pi, g)$, an application of the dominated convergence theorem will confirm that R_ν is continuous.

As $\mathcal{S}(\pi, g)$ is convex and τ-compact (Lemma 5.2), by Sion's minmax theorem (see Mertens et al. [2015, Theorem I.1.1]), it holds that

$$\begin{aligned}\sup_{\nu\in\mathcal{S}(\pi,g)} \gamma_h^{\mathrm{OCE}}(X|\nu) = \sup_{\nu\in\mathcal{S}(\pi,g)} \inf_{m\in\mathbb{R}} R(\nu, m) &= \inf_{m\in\mathbb{R}} \sup_{\nu\in\mathcal{S}(\pi,g)} R(\nu, m) \\ &= \inf_{m\in\mathbb{R}} \left\{ m + \sup_{\nu\in\mathcal{S}(\pi,g)} \mathbb{E}^{\nu}[h(X - m)] \right\} \\ &= \inf_{m\in\mathbb{R}} \left\{ m + \int_\Theta \mathbb{E}^{\mu_\theta}[h(X - m)] g \circ \pi(\mathrm{d}\theta) \right\}.\end{aligned}$$

Therefore, the proof is complete. □

Proof of Proposition 5.6. Let $X \in \mathcal{X}$, we have

$$\begin{aligned}\lim_{x\to\infty} \sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > x) &= \lim_{x\to\infty} \sup_{\xi\in\mathcal{R}(\pi,g)} \int_\Theta \mu_\theta(X > x)\xi(\mathrm{d}\theta) \\ &= \lim_{x\to\infty} \int_\Theta \mu_\theta(X > x) g \circ \pi(\mathrm{d}\theta) \\ &= \lim_{x\to\infty} \int_0^1 g(\pi(\{\theta \in \Theta \mid \mu_\theta(X > x) > y\}))\mathrm{d}y.\end{aligned}$$

Since $\lim_{x\to\infty} g(\pi(\{\theta \in \Theta \mid \mu_\theta(X > x) > y\})) = 0$ for all $y \in [0,1]$, by the dominated convergence theorem, we have that

$$\lim_{x\to\infty} \sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > x) = 0.$$

Find $z_0 \in \mathbb{R}$ such that $\sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > z_0) \leq 1-\alpha$. Therefore, $\nu(X > z_0) \leq 1-\alpha$ for all $\nu \in \mathcal{S}(\pi,g)$, and $\mathrm{VaR}_\alpha(X|\nu) \leq z_0$ for all $\nu \in \mathcal{S}(\pi,g)$. Thus, $\sup_{\nu\in\mathcal{S}(\pi,g)} \mathrm{VaR}_\alpha(X|\nu)$ is finite.

Define

$$x_0 = \sup_{\nu\in\mathcal{S}(\pi,g)} \mathrm{VaR}_\alpha(X|\nu) \quad \text{and} \quad y_0 = \inf\left\{x \in \mathbb{R} \;\middle|\; \sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > x) \leq 1-\alpha\right\}.$$

Let $\epsilon > 0$. By definition, we have that

$$\sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > y_0 + \epsilon) \leq 1-\alpha.$$

Therefore, $\mathrm{VaR}_\alpha(X|\nu) \leq y_0 + \epsilon$ for all $\nu \in \mathcal{S}(\pi,g)$, $x_0 \leq y_0 + \epsilon$. Since $\epsilon > 0$ was arbitrary, $x_0 \leq y_0$.

Let $\epsilon > 0$. By definition, we have that

$$\nu(X > x_0 + \epsilon) \leq 1-\alpha, \quad \nu \in \mathcal{S}(\pi,g).$$

Therefore, $\sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > x_0 + \epsilon) \leq 1-\alpha$, and $y_0 \leq x_0 + \epsilon$. Since $\epsilon > 0$ was arbitrary, $y_0 \leq x_0$.

Observing that

$$\sup_{\nu\in\mathcal{S}(\pi,g)} \nu(X > x) = \sup_{\xi\in\mathcal{R}(\pi,g)} \int_\Theta \mu_\theta(X > x)\xi(\mathrm{d}\theta) = \int_\Theta \mu_\theta(X > x) g \circ \pi(\mathrm{d}\theta), \quad x \in \mathbb{R}.$$

completes the proof. □

5.11 Proofs and additional results accompanying Section 5.5

Proof of Proposition 5.7. Fix $X \in \mathcal{C}$. Let $(\mu_n)_{n\in\mathbb{N}} \subseteq \Delta$ such that $\mu_n \Rightarrow \mu$, and $f : \mathbb{R} \to \mathbb{R}$ be continuous and bounded. We have that

$$\lim_{n\to\infty} \int_{\mathbb{R}} f d\iota_X(\mu_n) = \lim_{n\to\infty} \mathbb{E}^{\mu_n}[f(X)] = \mathbb{E}^{\mu}[f(X)] = \int_{\mathbb{R}} f \mathrm{d}\iota_X(\mu).$$

Therefore, $\iota_X(\mu_n) \Rightarrow \iota_X(\mu)$, and ι_X is continuous.

Define the set $\mathcal{W} = \{X \in \mathcal{X} \mid \iota_X \text{ is } \Sigma/\Sigma_{\mathcal{D}}\text{-measurable}\}$. By the previous observation, $\mathcal{C} \subseteq \mathcal{W}$. It is clear that $\mathcal{W}$ is closed under sequences that converge pointwise. Therefore, by Kechris [2012, Theorem 11.6], $\mathcal{X} \subseteq \mathcal{W}$, and $\mathcal{X} = \mathcal{W}$. □

Proof of Proposition 5.8. It is clear that $A \in \mathfrak{U}(\pi)$ if and only if there exists $c_A \in [0, 1]$ such that $\nu(A) = c\ \pi(\nu)$-a.s. Therefore, it is straightforward to see that $\mathfrak{U}(\pi)$ is a λ-system.

Let $X \in \mathfrak{C}(\pi)$. Then, $(\iota_X)_\# \pi = \delta_{Q_X^\pi}$ and

$$\nu(X^{-1}(B)) = X_\#\nu(B) = [\iota_X(\nu)](B) = Q_X^\pi(B)\ \pi(\nu)\text{-a.s.}, \quad B \in \mathcal{B}(\mathbb{R}).$$

Therefore, $X^{-1}(B) \in \mathfrak{U}(\pi)$ for all $B \in \mathcal{B}(\mathbb{R})$.

Conversely, assume that $X \in L^1(\pi)$ satisfies $X^{-1}(B) \in \mathfrak{U}(\pi)$ for all $B \in \mathcal{B}(\mathbb{R})$. Given $q \in \mathbb{Q}$, let $A_q \in \Sigma$ satisfy $\pi(A_q) = 1$ and, for all $\nu \in A_q$, $\nu(X \le q) = c_q$. Let $A = \bigcap_{q\in\mathbb{Q}} A_q$. We have $\pi(A) = 1$ and, for all $\nu \in A$ and $q \in \mathbb{Q}$, $\nu(X \le q) = c_q$. Find the unique $Q \in \mathcal{D}_1$ such that $Q((-\infty, q]) = c_q$ for all $q \in \mathbb{Q}$. Therefore, for all $\nu \in A$, $X_\#\nu = Q$. Thus, $(\iota_X)_\#\pi = \delta_Q$ and $X \in \mathfrak{C}(\pi)$. □

Proof of Proposition 5.9. Let $X \in \mathfrak{C}(\pi)$, we claim that $X_\#\mu = Q_X^\pi$ for all $\mu \in \mathcal{S}(\pi, g)$. Let $\mu \in \mathcal{S}(\pi, g)$ and $\xi \in \mathcal{R}(\pi, g)$ such that $\mu_\xi = \mu$. We know that $\pi(\iota_X^{-1}(\{Q_X^\pi\}) = 1$ and

$$\xi\left(\left(\iota_X^{-1}\left(\{Q_X^\pi\}\right)\right)^c\right) \le g\left(\pi\left(\left(\iota_X^{-1}\left(\{Q_X^\pi\}\right)\right)^c\right)\right) = 0.$$

Therefore, $(\iota_X)_{\#}\xi = \delta_{Q_X^\pi}$ and

$$X_{\#}\mu(B) = \int_\Delta [\iota_X(\nu)](B)\xi(\mathrm{d}\nu) = \int_{\mathcal{D}} Q(B)\delta_{Q_X^\pi}(\mathrm{d}Q) = Q_X^\pi(B), \quad B \in \mathcal{B}(\mathbb{R}).$$

Therefore, $X_{\#}\mu = Q_X^\pi$ and, since $\mu \in \mathcal{S}(\pi, g)$ was arbitrary, the proof is complete. □

Proof of Proposition 5.10. We claim that $\{X_{\#}\nu : \nu \in \mathcal{S}(\pi, g)\} \subseteq \{Y_{\#}\nu : \nu \in \mathcal{S}(\pi, g)\}$. Let $\mu \in \mathcal{S}(\pi, g)$, and $\xi \in \mathcal{R}(\pi, g)$ such that $\mu = \mu_\xi$. Since, π is atomless, and $(\mathcal{D}, \Sigma_{\mathcal{D}})$ is a Polish measurable space, by Kallenberg [1997, Corollary 5.11], there exists $T : \Delta \to \Delta$ that is Σ/Σ-measurable such that $T_{\#}\pi = \pi$ and $\iota_X = \iota_Y \circ T$ π-a.s. Define $\xi' = T_{\#}\xi$, we have

$$\xi'(A) = \xi(T^{-1}(A)) \leq g(\pi(T^{-1}(A))) = g(\pi(A)), \quad A \in \Sigma.$$

Therefore, $\xi' \in \mathcal{R}(\pi, g)$. Given $B \in \mathcal{B}(\mathbb{R})$, we have

$$Y_{\#}\mu_{\xi'}(B) = \int_\Delta [\iota_Y(\nu)](B)\xi'(\mathrm{d}\nu) = \int_\Delta [\iota_X(\nu)](B)\xi(\mathrm{d}\nu) = X_{\#}\mu_\xi(B) = X_{\#}\mu(B).$$

Thus, $Y_{\#}\mu_{\xi'} = X_{\#}\mu$. Since $\mu_{\xi'} \in \mathcal{S}(\pi, g)$, we have shown the claim. The other set inclusion follows from the same argument. □

Proposition 5.11 is a corollary of Lemma 5.2 and Proposition 5.18 (presented below).

Proposition 5.18. *Let $\mathcal{P} \subseteq \Delta$ be τ_b-compact and γ be a Lebesgue continuous statistic. Then, for all $X \in \mathcal{X}_b$, there exists $\mu_X \in \mathcal{P}$ such that*

$$\gamma(X|\mu_X) = \sup_{\nu \in \mathcal{P}} \gamma(X|\nu).$$

Proof. Let $X \in \mathcal{X}_b$. By Proposition 5.15, $\gamma \circ \iota_X$ is τ_b-continuous. Since $\mathcal{P}$ is τ_b-compact, by the extreme-value theorem, there exists $\mu_X \in \mathcal{P}$ such that

$$\gamma(X|\mu_X) = \gamma \circ \iota_X(\mu_X) = \sup_{\nu \in \mathcal{P}} \gamma \circ \iota_X(\nu) = \sup_{\nu \in \mathcal{P}} \gamma(X|\nu).$$

Therefore, the proof is complete. □

Proof of Proposition 5.12. Fix $X \in \mathcal{C}_b$. For each $n \in \mathbb{N}$, let $\mu_n \in \mathcal{S}(\pi_n, g)$ be the first-order model satisfying

$$\max_{\nu \in \mathcal{S}(\pi_n, g)} \gamma(X|\nu) = \gamma(X|\mu_n).$$

For each $n \in \mathbb{N}$, find $\xi_n \in \mathcal{R}(\pi_n, g)$ satisfying $\mu_{\xi_n} = \mu_n$.

We claim that $\xi_n \Rightarrow \delta_{\theta_0}$. For $n \in \mathbb{N}$, since $\xi_n \in \mathcal{R}(\pi_n, g)$, we have

$$g^*(\pi_n(A)) \leq \xi_n(A) \leq g(\pi_n(A)), \quad A \in \mathcal{A},$$

where $g^*(x) = 1 - g(1-x)$ for all $x \in [0,1]$. Let $h : \Theta \to \mathbb{R}$ be bounded and continuous. It is clear that $h_{\#}\pi_n \Rightarrow h_{\#}\delta_{\mu_0}$. Thus, for Lebesgue-a.e. $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \pi_n(h > x) = \lim_{n\to\infty} h_{\#}\pi_n((x,\infty)) = h_{\#}\delta_{\mu_0}((x,\infty)) = \mathbb{1}_{(-\infty,h(\mu_0))}(x).$$

By the dominated convergence theorem, for all distortion functions $\tilde{g}$, we have

$$\begin{aligned}
\lim_{n\to\infty} \int_\Theta h \mathrm{d}(\tilde{g} \circ \pi_n) &= \lim_{n\to\infty} \left(\int_0^\infty \tilde{g}(\pi_n(h > x))\mathrm{d}x + \int_{-\infty}^0 (\tilde{g}(\pi_n(h > x)) - 1)\mathrm{d}x \right) \\
&= \int_0^\infty \tilde{g}\left(\mathbb{1}_{(-\infty,h(\mu_0))}(x)\right) \mathrm{d}x + \int_{-\infty}^0 \left(\tilde{g}\left(\mathbb{1}_{(-\infty,h(\mu_0))}(x)\right) - 1\right) \mathrm{d}x \\
&= \int_0^\infty \mathbb{1}_{(-\infty,h(\mu_0))}(x)\mathrm{d}x + \int_{-\infty}^0 \left(\mathbb{1}_{(-\infty,h(\mu_0))}(x) - 1\right) \mathrm{d}x = h(\theta_0).
\end{aligned}$$

Since

$$\int_\Theta h \mathrm{d}(g^* \circ \pi_n) \leq \int_\Theta h \mathrm{d}\xi_n \leq \int_\Theta h \mathrm{d}(g \circ \pi_n),$$

we have that $\lim_{n\to\infty} \int_\Theta h \mathrm{d}\xi_n = h(\theta_0)$. Thus, $\xi_n \Rightarrow \delta_{\theta_0}$.

We claim that $\mu_n \Rightarrow \mu_{\theta_0}$. Let $Y \in \mathcal{X}_c$, we have

$$\lim_{n\to\infty} \mathbb{E}^{\mu_n}[Y] = \lim_{n\to\infty} \int_\Theta \mathbb{E}^{\mu_\theta}[Y]\xi_n(d\theta) = \int_\Theta \mathbb{E}^{\mu_\theta}[Y]\delta_{\theta_0}(\mathrm{d}\theta) = \mathbb{E}^{\mu_{\theta_0}}[Y],$$

where we are using the fact that the parametrization is continuous. Therefore, $\mu_n \Rightarrow \mu_{\theta_0}$.

Since $X \in \mathcal{C}_b$, $X_{\#}\mu_n \Rightarrow X_{\#}\mu_{\theta_0}$. Since $(X_{\#}\mu_n)_{n\in\mathbb{N}}$ has uniformly bounded support and γ is Lebesgue continuous,

$$\lim_{n\to\infty} \max_{\nu \in \mathcal{S}(\pi_n, g)} \gamma(X|\nu) = \lim_{n\to\infty} \gamma(X|\mu_n) = \gamma(X|\mu_{\theta_0}).$$

Therefore, the proof is complete. □

Proposition 5.13 is a corollary of Proposition 5.19 and Proposition 5.20; presented and proved below. Given $\mu \in \Delta$, we define $\Delta(\mu) = \{\nu \in \Delta \mid \nu \ll \mu\}$. Given $\mu \in \Delta$ and

$X, Y \in \mathcal{X}$, we say $X \overset{\mathrm{d}}{=}_\mu Y$ if $X_\# \mu =_{\mathrm{fsd}} Y_\# \mu$. Given $\mu \in \Delta$, we say that the decision functional $\rho : \mathcal{X}_b \to \mathbb{R}$ is *μ-law invariant* if $\rho(X) = \rho(Y)$ for all $X, Y \in \mathcal{X}_b$ satisfying $X \overset{\mathrm{d}}{=}_\mu Y$. Clearly, if $\mu \in \Delta$ is atomless, a decision functional ρ is μ-law invariant if and only if there exists a statistic γ such that $\rho(X) = \gamma(X|\mu)$ for all $X \in \mathcal{X}_b$. Given $\mu \in \Delta$ and $\nu \in \Delta(\mu)$, define

$$\mathcal{T}(\nu|\mu) = \left\{ \tilde{\nu} \in \Delta(\mu) \;\middle|\; \frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\mu} \overset{\mathrm{d}}{=}_\mu \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\}.$$

Given $\mu \in \Delta$ and $\mathcal{P} \subseteq \Delta(\mu)$, we say that $\mathcal{P}$ is *μ-law invariant* if for all $\nu \in \mathcal{P}$, $\mathcal{T}(\nu|\mu) \subseteq \mathcal{P}$. Given $\mu \in \Delta$, we say that $X, Y \in \mathcal{X}_b$ are *μ-comonotonic* if there exists $A \in \mathcal{F}$ with $\mu(A) = 1$ such that

$$(X(\omega) - X(\tilde{\omega}))(Y(\omega) - Y(\tilde{\omega})), \quad \omega, \tilde{\omega} \in A.$$

Proposition 5.19. *Let $\mu \in \Delta$ be atomless and $\mathcal{P} \subseteq \Delta(\mu)$ be μ-law invariant. Then, for all statistics γ, the decision functional*

$$\rho(X) = \sup_{\nu \in \mathcal{P}} \gamma(X|\nu), \quad X \in \mathcal{X}_b,$$

is μ-law invariant.

Proof. Given $X \in \mathcal{X}_b$, define

$$\mathcal{P}(X) = \left\{ \nu \in \mathcal{P} \;\middle|\; \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \text{ and } X \text{ are } \mu\text{-comonotonic} \right\}.$$

Since $\mathcal{P}$ is μ-law invariant, $\mathcal{T}(\nu|\mu) \subseteq \mathcal{P}$ for all $\nu \in \mathcal{P}$. Additionally, since μ is atomless, given $\nu \in \mathcal{P}$ and $X \in \mathcal{X}_b$, we have that $\mathcal{T}(\nu|\mu) \cap \mathcal{P}(X) \neq \varnothing$; see Föllmer and Schied [2016, Lemma A.25]. Let $\nu \in \mathcal{P}$, $X \in \mathcal{X}_b$, and $\tilde{\nu} \in \mathcal{T}(\nu|\mu) \cap \mathcal{P}(X)$. By the Hardy–Littlewood inequality, we have that

$$\nu(X > x) = \mathbb{E}^\mu \left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \mathbb{1}_{(x,\infty)}(X) \right] \leq \mathbb{E}^\mu \left[\frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\mu} \mathbb{1}_{(x,\infty)}(X) \right] = \tilde{\nu}(X > x), \quad x \in \mathbb{R}.$$

Thus, $X_\# \nu \leq_{\mathrm{fsd}} X_\# \tilde{\nu}$ and $\gamma(X|\nu) \leq \gamma(X|\tilde{\nu})$. Therefore, we have

$$\begin{aligned} \rho(X) = \sup_{\nu \in \mathcal{P}} \gamma(X|\nu) &= \sup_{\nu \in \mathcal{P}} \sup_{\tilde{\nu} \in \mathcal{T}(\nu|\mu)} \gamma(X|\tilde{\nu}) \\ &= \sup_{\nu \in \mathcal{P}} \sup_{\tilde{\nu} \in \mathcal{T}(\nu|\mu) \cap \mathcal{P}(X)} \gamma(X|\tilde{\nu}) = \sup_{\nu \in \mathcal{P}(X)} \gamma(X|\nu), \quad X \in \mathcal{X}_b. \end{aligned}$$

Let $X, Y \in \mathcal{X}_b$ satisfy $X \overset{\mathrm{d}}{=}_\mu Y$. Let $\nu \in \mathcal{P}(X)$ and $\tilde{\nu} \in \mathcal{T}(\nu|\mu) \cap \mathcal{P}(Y)$. Since $\left(X, \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \overset{\mathrm{d}}{=}_\mu \left(Y, \frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\mu}\right)$, we have that

$$\nu(X > x) = \mathbb{E}^\mu\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathbb{1}_{(x,\infty)}(X)\right] = \mathbb{E}^\mu\left[\frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\mu}\mathbb{1}_{(x,\infty)}(Y)\right] = \tilde{\nu}(Y > x), \quad x \in \mathbb{R}.$$

Thus, $X_\# \nu =_{\mathrm{fsd}} Y_\# \tilde{\nu}$ and $\gamma(X|\nu) = \gamma(Y|\tilde{\nu})$. Since $\nu \in \mathcal{P}(X)$ was arbitrary,

$$\rho(X) = \sup_{\nu \in \mathcal{P}(X)} \gamma(X|\nu) \leq \sup_{\nu \in \mathcal{P}(Y)} \gamma(Y|\nu) = \rho(Y).$$

A similar argument will show that $\rho(Y) \leq \rho(X)$. Therefore, $\rho(X) = \rho(Y)$ and ρ is μ-law invaraint. □

We say that the decision functional $\rho : \mathcal{X}_b \to \mathbb{R}$ is Lebesgue continuous if for bounded sequences $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}_b$ that converge pointwise to some $X \in \mathcal{X}_b$, we have that

$$\lim_{n\to\infty} \rho(X_n) = \rho(X).$$

Clearly, if $\mu \in \Delta$ is atomless, then the decision functional $\rho : \mathcal{X}_b \to \mathbb{R}$ defined by $\rho(X) = \gamma(X|\mu)$, for some statistic γ, is Lebesgue continuous if and only if γ is Lebesgue continuous.

Proposition 5.20. *Let $\mathcal{P} \subseteq \Delta$ be τ_b-compact and γ be Lebesgue continuous statistic, then the decision functional*

$$\rho(X) = \sup_{\nu \in \mathcal{P}} \gamma(X|\nu), \quad X \in \mathcal{X}_b,$$

is Lebesgue continuous.

Proof. It is well-known that ρ is Lebesgue continuous if and only if for all pointwise-increasing or pointwise-decreasing sequences $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{X}_b$, we have that $\lim_{n\to\infty} \rho(X_n) = \rho(\lim_{n\to\infty} X_n)$. Clearly, since γ is Lebesgue continuous, the above holds for all pointwise-increasing sequences. The hard part is showing that it is true for all pointwise-decreasing sequences. Therefore, let $(X_n)_{n\in\mathbb{N}}$ be a pointwise-decreasing sequence and define $X = \lim_{n\to\infty} X_n$.

Let Φ_{BL} denote the set of all Lipschitz continuous functions $f : \mathbb{R} \to \mathbb{R}$ that are bounded by, and have Lipschitz constant less than, 1. It is well-known, see Dudley [2002, Theorem 11.3.3], that

$$d_{\mathrm{BL}} : \mathcal{D}^2 \to [0, \infty) : (Q, P) \mapsto \sup_{f \in \Phi_{\mathrm{BL}}} \left| \int_{\mathbb{R}} f(x) Q(\mathrm{d}x) - \int_{\mathbb{R}} f(x) P(\mathrm{d}x) \right|,$$

is a metric and metrizes $\tau_{\mathcal{D}}$. We claim that

$$\lim_{n\to\infty} \sup_{\nu\in\mathcal{P}} d_{\mathrm{BL}}((X_n)_{\#}\nu, X_{\#}\nu) = 0.$$

Since $\mathcal{P}$ is τ_b-compact, by Marinacci and Montrucchio [2004, Lemma 4.1] and Dunford and Schwartz [1988, Theorem IV.9.2], there exists $\mu \in \Delta$ such that for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$\mu(A) < \delta \implies \sup_{\nu\in\mathcal{P}} \nu(A) \leq \epsilon, \quad A \in \mathcal{F}.$$

Let $\epsilon > 0$, by Egorov's theorem, we can find $A \in \mathcal{F}$ and $N \in \mathbb{N}$ such that $\mu(A) > 1 - \delta$ and for all $n \geq \mathbb{N}$, we have $|X_n(\omega) - X(\omega)| \leq \epsilon$ for all $\omega \in A$. Therefore, since $\mu(A^c) < \delta$, for all $n \geq \mathbb{N}$,

$$\begin{aligned}
&\sup_{\nu\in\mathcal{P}} d_{\mathrm{BL}}((X_n)_{\#}\nu, X_{\#}\nu) \\
&\leq \sup_{\nu\in\mathcal{P}} \sup_{f\in\Phi_{\mathrm{BL}}} \mathbb{E}^{\nu}[|f(X_n) - f(X)|] \\
&\leq \sup_{\nu\in\mathcal{P}} \sup_{f\in\Phi_{\mathrm{BL}}} \mathbb{E}^{\nu}[|f(X_n) - f(X)|\,\mathbb{1}_A] + \sup_{\nu\in\mathcal{P}} \sup_{f\in\Phi_{\mathrm{BL}}} \mathbb{E}^{\nu}[|f(X_n) - f(X)|\,\mathbb{1}_{A^c}] \\
&\leq \sup_{\nu\in\mathcal{P}} \mathbb{E}^{\nu}[|X_n - X|\,\mathbb{1}_A] + 2\sup_{\nu\in\mathcal{P}} \nu(A^c) \leq 3\epsilon.
\end{aligned}$$

Find $a, b \in \mathbb{R}$ such that $a \leq X_n \leq b$ for all $n \in \mathbb{N}$. Let $\mathcal{D}_{[a,b]}$ denote the set of Borel measures on $\mathbb{R}$ with support contained in the interval $[a, b]$. Clearly, for all $n \in \mathbb{N}$, $(X_n)_{\#}\mu \in \mathcal{D}_{[a,b]}$ for all $\mu \in \Delta$. Additionally, $X_{\#}\mu \in \mathcal{D}_{[a,b]}$ for all $\mu \in \Delta$. By Prokhorov's theorem, the set $\mathcal{D}_{[a,b]}$ is compact under $\tau_{\mathcal{D}}$. Since, $\gamma|_{\mathcal{D}_{[a,b]}}$ is $\tau_{\mathcal{D}}$-continuous (as γ is Lebesgue continuous), for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$\mathrm{d}_{\mathrm{BL}}(Q, P) < \delta \implies |\gamma(Q) - \gamma(P)| < \epsilon, \quad Q, P \in \mathcal{D}_{[a,b]}. \tag{5.13}$$

Given $\epsilon > 0$, find $\delta > 0$ such that (5.13) holds. Now find $N \in \mathbb{N}$ such that

$$\sup_{\nu\in\mathcal{P}} d_{\mathrm{BL}}((X_n)_{\#}\nu, X_{\#}\nu) < \delta$$

for all $n \geq N$. For each $n \geq N$, let $\mu_n \in \mathcal{P}$ satisfy $\sup_{\nu\in\mathcal{P}}(X_n|\nu) = \gamma(X|\mu_n)$.

Therefore,

$$\rho(X_n) - \rho(X) \leq \gamma(X_n|\mu_n) - \gamma(X|\mu_n) < \epsilon.$$

Thus $\lim_{n\to\infty}\rho(X_n)=\rho(X)$, and ρ is Lebesgue continuous. □

Proof of Proposition 5.14. Given $\mu\in\Delta$, by Ahmadi-Javid [2012, Theorem 3.3],

$$\mathbb{E}_\epsilon^{\mathrm{KL}}(X|\mu)=\inf_{z>0}\left(z\log\left(\mathbb{E}^\mu\left[e^{X/z-\epsilon}\right]\right)\right),\quad X\in\mathcal{X}_b.$$

Therefore, it is clear that, for all $X\in\mathcal{X}_b$, $\mathbb{E}_\epsilon^{\mathrm{KL}}\circ\iota_X$ is concave.

Let $X\in\mathcal{C}_b$. As $\mathbb{E}_\epsilon^{\mathrm{KL}}$ is Lebesgue continuous, by Proposition 5.15, $\mathbb{E}_\epsilon^{\mathrm{KL}}\circ\iota_X$ is τ-continuous. Therefore, $-\mathbb{E}_\epsilon^{\mathrm{KL}}\circ\iota_X$ is proper, convex, and τ-continuous. Therefore, by the definition $\geq_{\mathrm{cx}}$, for all $\tilde\pi,\pi\in\Pi$ satisfying $\tilde\pi\geq_{\mathrm{cx}}\pi$, we have

$$\int_\Delta -\mathbb{E}_\epsilon^{\mathrm{KL}}(X|\nu)\tilde\pi(\mathrm{d}\nu)\geq\int_\Delta -\mathbb{E}_\epsilon^{\mathrm{KL}}(X|\nu)\pi(\mathrm{d}\nu).$$

Since $X\in\mathcal{C}_b$ was arbitrary, by Proposition 5.17, we have

$$\int_\Delta -\mathbb{E}_\epsilon^{\mathrm{KL}}(X|\nu)\tilde\pi(\mathrm{d}\nu)\geq\int_\Delta -\mathbb{E}_\epsilon^{\mathrm{KL}}(X|\nu)\pi(\mathrm{d}\nu),\quad X\in\mathcal{X}_b,$$

and the proof is complete. □

5.12 Omitted verifications

Verification 5.1. *The mapping ψ_X is Σ-measurable for all $X\in\mathcal{X}$.*

Proof. If $X\in\mathcal{X}$ has finite range, by Bertsekas and Shreve [1996, Proposition 7.25], the mapping ψ_X is Σ-measurable.

If $X\in\mathcal{X}$ is non-negative, find an increasing sequence of simple acts $(X_n)_{n\in\mathbb{N}}\subseteq\mathcal{X}$ that converges pointwise to X. By the monotone convergence theorem, we have

$$\lim_{n\to\infty}\psi_{X_n}(\mu)=\lim_{n\to\infty}\mathbb{E}^\mu[X_n]=\mathbb{E}^\mu[X]=\psi_X(\mu),\quad\mu\in\Delta.$$

Since ψ_{X_n} is Σ-measurable for all $n\in\mathbb{N}$, ψ_X is Σ-measurable.

If $X\in\mathcal{X}$ is arbitrary, define

$$A_1=\{\mu\in\Delta:\mathbb{E}^\mu[X_+]=\infty\}\quad\text{and}\quad A_2=\{\mu\in\Delta:\mathbb{E}^\mu[X_-]=\infty\}.$$

Taking $A=A_1\cap A_2$, we have $\psi_X=(\psi_{X_+}-\psi_{X_-})\mathbb{1}_{A^c}$, which is Σ-measurable. □

Verification 5.2. *If $\pi, \tilde{\pi} \in \Pi$ satisfy $\tilde{\pi} \geq_{\mathrm{cx}} \pi$, then for every $\varphi \in \Lambda$,*

$$\rho(X; \tilde{\pi}, \varphi) \geq \rho(X; \pi, \varphi), \quad X \in \mathcal{X}_b,$$

where $\rho(\cdot; \pi, \varphi) : \mathcal{X}_b \to \mathbb{R}$ is the decision functional defined in (5.1).

Proof. This follows directly from Corollary 5.1. □

Verification 5.3. *For all $X \in \mathcal{X}_b$, the mapping $\mathrm{ES}_\alpha \circ \iota_X$ is Σ-measurable.*

Proof. Since ES_α is Lebesgue continuous, by Proposition 5.16, $\mathrm{ES}_\alpha \circ \iota_X$ is Σ-measurable for $X \in \mathcal{X}_b$. □

Verification 5.4. *The mapping $\nu \mapsto \mathrm{KL}(\mu|\nu)$ is Σ-measurable for all $\mu \in \Delta$.*

Proof. Since $(\Omega, \mathcal{F})$ is a Polish measurable space, by Donsker and Varadhan [1983, Equation (2.1)], we know that for all $\mu, \nu \in \Delta$,

$$\mathrm{KL}(\mu|\nu) = \sup_{X \in \mathcal{C}_b} \left(\mathbb{E}^\mu[X] - \log\left(\mathbb{E}^\nu\left[e^X\right]\right)\right).$$

Given $\mu \in \Delta$ and $X \in \mathcal{C}_b$, since the mapping

$$\Delta \to \mathbb{R} : \nu \mapsto \mathbb{E}^\mu[X] - \log\left(\mathbb{E}^\nu\left[e^X\right]\right)$$

is τ-continuous, the mapping $\nu \mapsto \mathrm{KL}(\mu|\nu)$ is τ-lower semicontinuous. Since τ-lower semicontinuity implies Σ-measurability, we have shown the original claim. □

Verification 5.5. *For all $X \in \mathcal{X}_b$, the mapping $\mathbb{E}_\epsilon^{\mathrm{KL}} \circ \iota_X$ is Σ-measurable.*

Proof. Since $\mathbb{E}_\epsilon^{\mathrm{KL}}$ is Lebesgue continuous, by Proposition 5.16, $\mathbb{E}_\epsilon^{\mathrm{KL}} \circ \iota_X$ is Σ-measurable for $X \in \mathcal{X}_b$. □

Chapter 6

Conclusion and future work

This thesis develops and analyzes new decision models under ambiguity that accommodate arbitrary pure-risk projections. Non-EU decision models (e.g., rank-dependent utility, coherent and convex risk measures, and non-linear premium principles) are useful in the pure-risk setting and important models in many fields (e.g., mathematical finance, quantitative risk management, actuarial science, and operations research). However, the dominant ambiguity models in the literature reduce to EU in the absence of ambiguity, creating a disconnect between these models and the pure-risk models that practitioners and regulators already use. This thesis presents new models that close that gap. By allowing the pure-risk projection to be specified freely, the models developed here can be used without compromising pure-risk evaluation.

Because the decision models developed in this thesis are general, the most natural direction for future work is to study specific instances. In particular, the statistical literature offers a wealth of tools for specifying the components of the representation functionals in a way that integrates ambiguity with statistical learning. Such specifications would bring these models into the applied settings where decision models are ultimately meant to be used. For example, in the spirit of Chapter 5, one could study act-to-distribution mappings (as developed in Chapter 4) that take a second-order model as input. This new process would connect the ATE framework to a Bayesian setting and provide an alternative model to the one presented in Chapter 5.

Finally, an important avenue for future work is applying these models to concrete decision problems, such as portfolio allocation, insurance pricing, and risk sharing. As these models allow for arbitrary pure-risk projections, one can take the pure-risk models used in different applications and, in some sense, add ambiguity for free. This makes the models attractive to practitioners seeking to incorporate ambiguity into their work.

References

M. Abdellaoui and P. P. Wakker. Savage for dummies and experts. *Journal of Economic Theory*, 186:104991, 2020.

M. Abdellaoui, F. Vossmann, and M. Weber. Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science*, 51(9): 1384–1399, 2005.

M. Abdellaoui, H. Bleichrodt, E. Kemel, and O. L'Haridon. Measuring beliefs under ambiguity. *Operations Research*, 69(2):599–612, 2021.

C. Acerbi and D. Tasche. Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388, 2002.

A. Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123, 2012.

M. Allais. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4):503–546, 1953.

M. Amarante and E. Filiz. Ambiguous events and maxmin expected utility. *Journal of Economic Theory*, 134(1):1–33, 2007.

F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963.

D. Ardia. *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

D. Ardia, J. Kolly, and D. A. Trottier. The impact of parameter and ambiguity on market risk predictions from GARCH-type models. *Journal of Forecasting*, 36(7):808–823, 2017.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

A. Baillon, Z. Huang, A. Selim, and P. P. Wakker. Measuring ambiguity attitudes for all (natural) events. *Econometrica*, 86(5):1839–1858, 2018.

A. Baillon, H. Bleichrodt, C. Li, and P. P. Wakker. Source theory: A tractable and positive ambiguity theory. *Management Science*, 71(10):8767–8782, 2025.

N. Bäuerle and A. Müller. Stochastic orders and risk measures: Consistency and bounds. *Insurance: Mathematics and Economics*, 38(1):132–148, 2006.

BCBS. Minimum capital requirements for market risk. Technical report, Bank for International Settlements, Basel, February 2019. URL https://www.bis.org/bcbs/publ/d457.htm.

F. Bellini, P. Koch-Medina, C. Munari, and G. Svindland. Law-invariant functionals that collapse to the mean. *Insurance: Mathematics and Economics*, 98:83–91, 2021.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

R. H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58, 1966.

C. Bernard, X. Jiang, and R. Wang. Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, 54:93–108, 2014.

D. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*, volume 5. Athena Scientific, 1996.

P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.

C. Birghila and G. C. Pflug. Optimal XL-insurance under wasserstein-type ambiguity. *Insurance: Mathematics and Economics*, 88:30–43, 2019.

J. Blanchet, H. Lam, Q. Tang, and Z. Yuan. Robust actuarial risk analysis. *North American Actuarial Journal*, 23(1):33–63, 2019.

G. Bredon. *Topology and Geometry*. Springer, Berlin, 1993.

L. Buchak. A unified treatment of risk and ambiguity within a rank-dependent framework. *Theory and Decision*, 99:529–556, 2025.

A. J. Cairns. A discussion of parameter and ambiguity in insurance. *Insurance: Mathematics and Economics*, 27(3):313–330, 2000.

E. Castagnoli, G. Cattelan, F. Maccheroni, C. Tebaldi, and R. Wang. Star-shaped risk measures. *Operations Research*, 70(5):2637–2654, 2022.

S. Cerreia-Vioglio, P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi. Rational preferences under ambiguity. *Economic Theory*, 48(2):341–375, 2011.

S. Cerreia-Vioglio, L. P. Hansen, F. Maccheroni, and M. Marinacci. Making decisions under model misspecification. *Review of Economic Studies*, page rdaf046, 2025.

A. Chateauneuf and G. Lakhnati. From sure to strong diversification. *Economic Theory*, 32(3):511–522, 2007.

A. Chateauneuf and J.-M. Tallon. Diversification, convex preferences and non-empty core in the Choquet expected utility model. *Economic Theory*, 19(3):509–523, 2002.

S. Chew and M. Mao. A Schur concave characterization of risk aversion for non-expected utility preferences. *Journal of Economic Theory*, 67(2):402–435, 1995.

B. de Finetti. Sul significato soggettivo della probabilità. *Fundamenta Mathematicae*, 17: 298–329, 1931. Translated into English by Khale, M. as *On the Subjective Meaning of Probability*. In Monari, P. and Cocchi, D. (eds.) (1993), *Probabilità e Induzione*, 291–321, Clueb, Bologna.

E. Dekel. Asset demands without the independence axiom. *Econometrica*, 57(6):163–169, 1989.

E. Delage and Y. Li. Minimizing risk exposure when the choice of a risk measure is ambiguous. *Management Science*, 64(1):327–344, 2018.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

E. Delage, D. Kuhn, and W. Wiesemann. "dice"-sion-making under uncertainty: When can a random decision reduce risk? *Management Science*, 65(7):3282–3301, 2019.

F. Delbaen. Coherent risk measures, 2000. Lecture Notes, Pisa.

F. Delbaen. Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics*, pages 1–37. Springer, Berlin, Heidelberg, 2002.

F. Delbaen. Commonotonicity and time-consistency for Lebesgue-continuous monetary utility functions. *Finance and Stochastics*, 25:597–614, 2021.

C. Dellaporta, P. O'Hara, and T. Damoulas. Decision making under the exponential family: Distributionally robust optimisation with Bayesian ambiguity sets. *arXiv*, 2024. arXiv:2411.16829.

C. Dellaporta, P. O'Hara, and T. Damoulas. Decision making under model misspecification: DRO with robust Bayesian ambiguity sets. *arXiv*, 2025. arXiv:2505.03585.

D. Denneberg. *Non-additive Measure and Integral*. Springer Science & Business Media, 1994.

K. Detlefsen and G. Scandolo. Conditional and dynamic convex risk measures. *Finance and Stochastics*, 9:539–561, 2005.

J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

J. Dhaene, B. Stassen, K. Barigou, D. Linders, and Z. Chen. Fair valuation of insurance liabilities: Merging actuarial judgement and market-consistency. *Insurance: Mathematics and Economics*, 76:14–27, 2017.

S. Dimmock, R. Kouwenberg, and P. P. Wakker. Ambiguity attitudes in a large representative sample. *Management Science*, 62(5):1363–1380, 2016.

M. D. Donsker and S. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Communications on Pure and Applied Mathematics*, 36 (2):183–212, 1983.

S. Drapeau and M. Kupper. Risk preferences and their robust representation. *Mathematics of Operations Research*, 38(1):28–62, 2013.

R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

N. Dunford and J. T. Schwartz. *Linear Operators, Part 1: General Theory*. John Wiley & Sons, 1988.

L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization. *Operations Research*, 51:543–556, 2003.

D. Ellsberg. Risk, ambiguity and Savage axioms. *Quarterly Journal of Economics*, 75: 643–669, 1961.

P. Embrechts, G. Puccetti, and L. Rüschendorf. Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, 37(8):2750–2764, 2013.

P. Embrechts, B. Wang, and R. Wang. Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4):763–790, 2015.

P. Embrechts, T. Mao, Q. Wang, and R. Wang. Bayes risk, elicitability, and the Expected Shortfall. *Mathematical Finance*, 31:1190–1217, 2021.

L. G. Epstein. A definition of uncertainty aversion. In I. Gilboa, editor, *Uncertainty in Economic Theory*, pages 171–208. Routledge, New York, NY, 2004.

L. G. Epstein and M. Schneider. Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31, 2003.

L. G. Epstein and J. Zhang. Subjective probabilities on subjectively unambiguous events. *Econometrica*, 69(2):265–306, 2001.

T. Fadina, Y. Liu, and R. Wang. A framework for measures of risk under uncertainty. *Finance and Stochastics*, 28:363–390, 2024.

G. Feng and J. He. Factor investing: A Bayesian hierarchical approach. *Journal of Econometrics*, 230(1):183–200, 2022.

D. Filipović, M. Kupper, and N. Vogelpoth. Approaches to conditional risk. *SIAM Journal on Financial Mathematics*, 3(1):402–432, 2012.

G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, Hoboken, second edition, 1999.

H. Föllmer and T. Knispel. Entropic risk measures: Coherence vs. convexity, model ambiguity, and robust large deviations. *Stochastics and Dynamics*, 11(2-3):333–351, 2011.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002.

H. Föllmer and A. Schied. *Stochastic Finance. An Introduction in Discrete Time*. Walter de Gruyter, Berlin, fourth edition, 2016.

M. Frittelli and E. Rosazza Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26:1473–1486, 2002.

M. Frittelli and E. Rosazza Gianin. Law invariant convex risk measures. *Advances in Mathematical Economics*, 7:33–46, 2005.

T. Gajdos, T. Hayashi, J. M. Tallon, and J. C. Vergnaud. Attitude toward imprecise information. *Journal of Economic Theory*, 140(1):27–65, 2008.

P. Ghirardato and M. Marinacci. Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research*, 26(4):864–890, 2001.

P. Ghirardato and M. Marinacci. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, 102(2):251–289, 2002.

P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi. A subjective spin on roulette wheels (full version). Technical Report 1127, Caltech Social Science Working Paper, August 2001. SSRN: 278235.

P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi. A subjective spin on roulette wheels. *Econometrica*, 71(6):1897–1908, 2003.

P. Ghirardato, F. Maccheroni, and M. Marinacci. Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, 118(2):133–173, 2004.

S. Ghosal and A. W. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*, volume 44. Cambridge University Press, 2017.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

R. Giraud. Second order beliefs models of choice under imprecise risk: Nonadditive second order beliefs versus nonlinear second order utility. *Theoretical Economics*, 9(3):779–816, 2014.

C. Gollier. Portfolio choices and asset prices: The comparative statics of ambiguity aversion. *The Review of Economic Studies*, 78(4):1329–1344, 2011.

F. Gul and W. Pesendorfer. Expected uncertain utility theory. *Econometrica*, 82(1):1–39, 2014.

X. Han, Q. Wang, R. Wang, and J. Xia. Cash-subadditive risk measures without quasi-convexity. *Mathematics of Operations Research*, 2025.

L. P. Hansen and T. J. Sargent. Robust control and model uncertainty. *American Economic Review*, 91(2):60–66, 2001.

L. P. Hansen and T. J. Sargent. Structured ambiguity and model misspecification. *Journal of Economic Theory*, 199:105–165, 2022.

L. Hartmann and T. F. Kauffeldt. Binary diversification characterizes exact capacities. *Operations Research*, 74(1):377–389, 2026.

E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, 9:49–71, 2006.

N. Ju and J. Miao. Ambiguity, learning, and asset returns. *Econometrica*, 80(2):559–591, 2012.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, NY, 1997.

S. Kandel and R. F. Stambaugh. On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance*, 51(2):385–424, 1996.

A. Kechris. *Classical Descriptive Set Theory*. Springer Science & Business Media, 2012.

P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under uncertainty. *Econometrica*, 73(6):1849–1892, 2005.

P. Klibanoff, M. Marinacci, and S. Mukerji. Definitions of ambiguous events and the smooth ambiguity model. *Economic Theory*, 48(2):399–424, 2011.

S. A. Klugman. *Bayesian Statistics in Actuarial Science: With Emphasis on Credibility*. Kluwer Academic Publishers, Boston, 1992.

F. H. Knight. *Risk, Uncertainty and Profit*, volume 31. Houghton Mifflin, 1921.

M. Kupper and W. Schachermayer. Representation results for law invariant time consistent functions. *Mathematics and Financial Economics*, 2(3):189–210, 2009.

S. Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001.

R. J. Laeven, E. Rosazza Gianin, and M. Zullino. Law-invariant return and star-shaped risk measures. *Insurance: Mathematics and Economics*, 117:140–153, 2024.

F.-B. Liebrich. Are reference measures of law-invariant functionals unique? *Insurance: Mathematics and Economics*, 129:129–141, 2024.

F. B. Liebrich and C. Munari. Law-invariant functionals that collapse to the mean: Beyond convexity. *Mathematics and Financial Economics*, 16(3):447–480, 2022.

F. Liu and R. Wang. A theory for measures of tail risk. *Mathematics of Operations Research*, 46(3):1109–1128, 2021.

F. Liu, T. Mao, R. Wang, and L. Wei. Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, 47(3):2494–2519, 2022.

P. Liu, A. Schied, and R. Wang. Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research*, 46(4):1490–1512, 2021.

P. Liu, T. Mao, and R. Wang. Quantiles under ambiguity and risk sharing. *arXiv*, 2024. arXiv:2412.19546.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

F. Maccheroni, M. Marinacci, and D. Ruffino. Alpha as ambiguity: Robust mean-variance portfolio analysis. *Econometrica*, 81(3):1075–1113, 2013.

M. J. Machina and D. Schmeidler. A more robust definition of subjective probability. *Econometrica*, 60(4):745–780, 1992.

T. Mao and R. Wang. Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, 11(1):169–200, 2020.

T. Mao, R. Wang, and Q. Wu. Model aggregation for risk evaluation and robust optimization. *Management Science*, 2025. forthcoming.

M. Marinacci. Probabilistic sophistication and multiple priors. *Econometrica*, 70(2):755–764, 2002.

M. Marinacci and L. Montrucchio. Introduction to the mathematics of ambiguity. In I. Gilboa, editor, *Uncertainty in Economic Theory*, pages 46–107. Routledge, New York, NY, 2004.

H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ, revised edition, 2015.

J.-F. Mertens, S. Sorin, and S. Zamir. *Repeated Games*. Cambridge University Press, 2015.

S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

X. Mu, L. Pomatto, P. Strack, and O. Tamuz. Monotone additive statistics. *Econometrica*, 92(4):995–1031, 2024.

K. Natarajan, D. Pachamanova, and M. Sim. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54(3):573–585, 2008.

K. Nehring. Capacities and probabilistic beliefs: A precarious coexistence. *Mathematical Social Sciences*, 38(2):197–213, 1999.

M. Nendel, F. Riedel, and M. D. Schmeck. A decomposition of general premium principles into risk and deviation. *Insurance: Mathematics and Economics*, 100:193–209, 2021.

A. Pelsser and M. Stadje. Time-consistent and market-consistent evaluations. *Mathematical Finance*, 24:25–65, 2014.

J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32:122–136, 1964.

J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343, 1982.

F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112(2):185–200, 2004.

R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviation in risk analysis. *Finance and Stochastics*, 10:51–74, 2006.

M. Rothschild and J. E. Stiglitz. Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3):225–243, 1970.

R. Sarin and P. Wakker. A simple axiomatization of nonadditive expected utility. *Econometrica*, 60:1255–1272, 1992.

R. Sarin and P. Wakker. A general result for quantifying beliefs. *Econometrica*, 62(3): 683–685, 1994.

R. Sarin and P. Wakker. A single-stage approach to anscombe and aumann's expected utility. *The Review of Economic Studies*, 64:399–409, 1997.

L. J. Savage. *The Foundations of Statistics*. John Wiley & Sons, 1954.

D. Schmeidler. Cores of exact games, I. *International Journal of Game Theory*, 1(4): 275–288, 1972.

D. Schmeidler. Integral representation without additivity. *Proceedings of the American Mathematical Society*, 97(2):255–261, 1986.

D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.

U. Schmidt and H. Zank. Risk aversion in cumulative prospect theory. *Management Science*, 54(1):208–216, 2008.

M. Shaked and J. G. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.

A. Shapiro. On kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, 38:142–152, 2013.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics, 2021.

A. Shapiro, E. Zhou, and Y. Lin. Bayesian distributionally robust optimization. *SIAM Journal on Optimization*, 33(2):1279–1304, 2023.

L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1(1):11–26, 1971.

J. Shen, Y. Shen, B. Wang, and R. Wang. Distributional compatibility for change of measures. *Finance and Stochastics*, 23(3):761–794, 2019.

Y. Shen, Z. Van Oosten, and R. Wang. Partial law invariance and risk measures. *Management Science*, 2026. forthcoming.

T. Strzalecki. Axiomatic foundations of multiplier preferences. *Econometrica*, 79(1):47–73, 2011.

G. Svindland. Continuity properties of law-invariant (quasi-)convex risk functions on L^∞. *Mathematics and Financial Economics*, 3:39–43, 2010.

A. Tversky and C. R. Fox. Weighing risk and uncertainty. *Psychological Review*, 102(2): 269–283, 1995.

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

Z. Van Oosten and R. Wang. Choquet rank-dependent utility with an exogenous unambiguous source. *arXiv*, 2025a. 2509.10788.

Z. Van Oosten and R. Wang. Aggregate then evaluate. *arXiv*, 2025b. arXiv:2512.03396.

L. Vesely. Jensen's integral inequality in locally convex spaces. *Annals. Series on Science of Mathematics*, 9(2):136–144, 2017.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, second edition, 1947.

P. P. Wakker. *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press, 2010.

P. P. Wakker and J. Yang. A powerful tool for analyzing concave/convex utility and weighting functions. *Journal of Economic Theory*, 181:143–159, 2019.

A. Wald. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, 46(2):265–280, 1945.

A. Wald. Statistical decision functions. *The Annals of Mathematical Statistics*, 20(2): 165–205, 1949.

R. Wang and J. F. Ziegel. Scenario-based risk evaluation. *Finance and Stochastics*, 25: 725–756, 2021.

R. Wang and R. Zitikis. An axiomatic foundation for the Expected Shortfall. *Management Science*, 67:1413–1429, 2021.

R. Wang, Y. Wei, and G. E. Willmot. Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, 45(3):993–1015, 2020.

Z. Wang and G. J. Klir. *Generalized Measure Theory*, volume 25. Springer Science & Business Media, 2010.

M. E. Yaari. Some remarks on measures of risk aversion and on their uses. *Journal of Economic Theory*, 1(3):315–329, 1969.

M. E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.

J. Zhang. Subjective ambiguity, expected utility and Choquet expected utility. *Economic Theory*, 20(1):159–181, 2002.

S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.

C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific, Singapore, 2002.