Shadow Probability Theory for Asset Pricing under Ambiguity

Yehuda Izhakian\footnote{Department of Economics, Stern School of Business, New York University, yud@stern.nyu.edu} $^\dagger$\footnote{We benefited from discussions with Menachem Brenner, Itzhak Gilboa, Yakov Amihud, Avi Wohl, Saggi Katz, Azi Ben-Rephael Menachem Abudy and Shimon Benninga.}

October 22, 2011

Abstract

Assuming that probabilities (capacities) of events are random, this paper introduces a novel model of decision making under ambiguity, called Shadow probability theory, a generalization of the Choquet expected utility. In this model, probabilities of observable events in a subordinated outcome-space are dominated by second-order unobservable events in a directing probability-space. The level of ambiguity, and the decision maker’s attitude toward it, are measured with respect to the directing space. Risk and risk attitude, on the other hand, apply to the subordinated space, as in classical expected utility theory. The desired distinction between preferences and beliefs and between risk and ambiguity is then obtained. A measure of ambiguity is naturally carried over on these settings. The present paper proves that in most cases ambiguity cannot be diversified. Counterintuitively, adding an ambiguous lottery to a portfolio of lotteries usually increases its ambiguity. This result, which implies that full diversification is not always optimal, challenges the common notion in the financial literature that investors should hold a fully diversified portfolio. Using this model of choice, the current paper generalizes the classical asset pricing theory by incorporating ambiguous probabilities. It proposes a well defined ambiguity premium that can be measured empirically.

Keywords: Ambiguity, Ambiguity Measure, Ambiguity Aversion, Uncertainty, Knightian Uncertainty, Prospect Theory, Cumulative Prospect Theory, Ellsber Paradox, Ambiguity Premium.

JEL Classification Numbers: C44, C65, D81, D83, G11, G12.
1 Introduction

Risk is defined as a situation in which the event to be realized is *a-priori* unknown, but the odds of all possible events are perfectly known. In reality, of course, these odds can rarely be known precisely. Ambiguity, or *Knightian uncertainty*, refers to conditions in which not only the event to be realized is *a-priori* unknown, but where the odds of events are also either not uniquely assigned or are unknown. The neoclassical finance literature, dealing with risk tolerance, commonly ignores the presence of ambiguity, and assumes that decision makers (DMs) are able to precisely estimate the probability distribution of returns on assets. The main reason for this oversight is simply a lack of theoretical methods which are empirically testable.

The objective of this paper is to provide additional methods that enable studying the implications of ambiguity in regards to finance. The main theoretical contribution of this paper to the existing literature is that it introduces a novel, empirically applicable, ambiguity measure. The second contribution is that it addresses the question of whether, and in what circumstances, ambiguity can be diversified in a manner similar to risk. The third contribution is that this paper generalizes the classical asset pricing theory to incorporate ambiguity, providing a well-defined ambiguity premium which is completely distinguished from risk, and can be tested empirically.

Assuming that probabilities of events are random, this paper presents a novel model of decision making, called *Shadow probability theory* (Shadow theory or SPT, for short), which aims to capture the multi-dimensional nature of uncertainty. In this model, there are two different tiers of uncertainty, one with respect to consequences (outcomes) and the other with respect to the probabilities of these consequences. Each tier is modeled by a separate space. Observable outcomes are dictated by events in a *subordinated outcome-space* (possibly) having a random probability measure, which in turn is dominated by a second-order unobservable uncertain lottery (prospect) in a *directing probability-space*, equipped with second-order probabilities.\(^1\)

This structure attains a complete separation of risk from ambiguity as regards both aspects: beliefs and preferences. The level of ambiguity, and the DM’s attitude toward it, are measured with respect to the directing space. Risk and risk attitudes, on the other hand, apply to the subordinated space, as in classical expected utility theory. Ambiguous probabilities in this model can be considered as a new dimension of risk attitude: probabilistic sensitivity, i.e., the nonlinear ways in which individuals may process probabilities. Perceived probabilities, in this structure, are nonadditive. Ambiguity aversion results in a sub-additive subjective probability measure, while ambiguity seeking results in a super-additive measure. A measure of ambiguity, distinguished from preferences, is then naturally carried over on these settings, and the desired distinction between preferences and beliefs, and between risk and ambiguity is obtained.

Given an objective, or a subjective, classification of consequences as loss or gain, we prove that the level of ambiguity can be measured by the variance of the probability of loss, which is equal to

\(^1\)Uncertain lottery is defined as a lottery whose probabilities are not uniquely assigned.
the variance of the probability of gain. Formally, our measure of ambiguity is given by

\[ \aleph^2 = 4 \text{Var}[P_L] = 4 \text{Var}[P_G], \]

where \( P_L \) and \( P_G \) are the random probabilities of loss and gain, respectively, and the variance is taken with respect to the second-order probabilities. The intuition behind this new measure is that ambiguity is caused by a perturbation of probabilities with respect to a (objective or subjective) reference point, which distinguishes losses from gains. The main advantage of this measure is that it can be computed easily from data and can be used as an explanatory variable in empirical tests.\(^3\)

To illustrate the new insight gained by Shadow theory about the nature of ambiguity and how individuals perceive it, let us consider the Ellsberg’s (1961)[26] three-color urn experiment as an example.\(^4\) Ellsberg suggested the following two-part experiment. Consider an urn with 90 colored balls; 30 of them are red and the other 60 are either black or yellow. In each part of the experiment, before one ball is drawn at random from the urn, a DM is offered two alternative bets; winning the bet entitles her to a sum of $9. In the first part, the DM has to choose between two alternative bets: the drawn ball is red (\( R \)) or the drawn ball is black (\( B \)). Then, in the second part, the DM has to choose between betting on: the drawn ball is red or yellow (\( RY \)) or alternatively the drawn ball is black or yellow (\( BY \)). Behavioral experiments have demonstrated that individuals usually prefer (\( R \)) over (\( B \)), but (\( BY \)) over (\( RY \)). Such preferences are usually attributed to lack of information about likelihoods (ambiguity) and to the DM’s aversion toward it.\(^5\)

Let us now try and explain the choices that emerge from the Ellsberg experiment, in view of Shadow theory. The subordinated space is defined by the events of drawing different balls, i.e., \( \{R, Y, B\} \), with random probabilities. In the first part of the experiment, from the DM’s point of view, the probability of (\( B \)) can obtain one of the possible values \( \frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \ldots, \frac{60}{90} \). The precise probability of (\( B \)) is dominated by a second-order unobservable event in a directing-space. Such an event can be, for example: “The experimenter put \( x \) black balls in the urn.” The DM, however, does not have any additional information indicating which of the possible probabilities is more likely and thus she assigns an equal weight to each possibility. The DM considers strictly positive outcomes ($9) as a gain; and otherwise, ($0) as a loss.

In the first part of the experiment, computing the variance of the probability of gain from (\( B \)) to obtain the measure of ambiguity indicates that the level of ambiguity is \( \aleph^2 [B] = 0.1530 \). The alternative in this part of the experiment is (\( R \)), for which the ambiguity level is \( \aleph^2 [R] = 0 \), since its probability, \( \frac{1}{3} \), is precisely known to the DM. Clearly, since the expected outcome of (\( R \)) and (\( B \)) is

\(^2\)Aleph, \( \aleph^2 \), is the first letter of the Hebrew alphabet.

\(^3\)Brenner and Izhakian (2011)[11] for example.

\(^4\)Following intuitive arguments by Knight (1921)[63], the expected utility paradigm was challenged by Ellsberg’s (1961)[26] experiment.

\(^5\)In expected utility theory, the DM’s assessments of the likelihoods of \( R, B \) and \( Y \) can be described by some probability measure \( P \). The DM is assumed to prefer a greater chance of winning $9 to a smaller chance of winning $9, such that the choices above imply that \( P (R) > P (B) \) and \( P (B \cup Y) > P (R \cup Y) \). However, since \( R, B \) and \( Y \) are mutually exclusive events, no such probability measure exists; hence, it is considered a paradox.
identical, an ambiguity-averse DM prefers \((R)\) over \((B)\); that is, she chooses the bet with the lower level of ambiguity. In the second part of the experiment, the probability of gain from \((RY)\) can take one of the possible values \(\frac{30}{99}, \frac{31}{99}, \ldots, \frac{90}{99}\), which in turn also implies an ambiguity level of \(\aleph^2 \{RY\} = 0.1530\). The probability of gain from the alternative, \((BY)\), is exactly \(\frac{2}{3}\), which implies \(\aleph^2 \{BY\} = 0\). Obviously, an ambiguity-averse DM prefers \((BY)\) over \((RY)\).\(^6\) Table 1 is a stylized description of this example.

<table>
<thead>
<tr>
<th></th>
<th>(R)</th>
<th>(Y)</th>
<th>(B)</th>
<th>Prob</th>
<th>E</th>
<th>(\aleph^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((R))</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>(\frac{1}{3})</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>((B))</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>(\frac{0}{99}, \frac{1}{99}, \frac{2}{99}, \ldots, \frac{60}{99})</td>
<td>3</td>
<td>0.1530</td>
</tr>
<tr>
<td>((RY))</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>(\frac{30}{99}, \frac{31}{99}, \frac{32}{99}, \ldots, \frac{90}{99})</td>
<td>6</td>
<td>0.1530</td>
</tr>
<tr>
<td>((BY))</td>
<td>0</td>
<td>9</td>
<td>9</td>
<td>(\frac{2}{3})</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: The Ellsberg example

The differentiation of ambiguity from risk and the measure of ambiguity, suggested by our model, allows prospects to be ranked by the criteria of ambiguity. This provides a way to address important questions that arise regarding the nature of ambiguity, in general, and the nature of the aggregate ambiguity of prospect portfolios, in particular. We prove that in most cases ambiguity cannot be diversified. On the contrary, adding a prospect to a portfolio of prospects usually increases its ambiguity. The reason for this is that, depending on the classification of consequences as a loss or a gain, the variance of outcomes usually has a negative effect on the level of ambiguity, such that decreasing variance by adding a prospect to a portfolio increases its ambiguity.

The intuition for this result is that ambiguity in our model is measured by the volatility of probability of loss; thus, it is positively affected by the amplitude of the probability density functions. Adding a prospect to a portfolio of prospects decreases its variances, the probability density functions becomes steeper, the amplitude of the cumulative probability increases and, thus, the ambiguity level also increases. To provide more insight, consider the case where the possible variances increase. In such a case, the probability density functions becomes flatter, the amplitude of the cumulative probability decreases and, thus, the ambiguity level also decreases. In the extreme case, when the variances tend to infinity, the probability density functions tend to the uniform distribution; thus, all the probabilities are known and the level of ambiguity is, therefore, zero.

The negative relation between the variance of return and ambiguity has considerable financial implications. The existing financial literature basically asserts that investors should hold a fully risk-diversified portfolio. Our results, however, imply that this is usually not true for ambiguity. Increasing the number of assets in a portfolio decreases its variance, but at the same time increases its level of ambiguity, such that for an ambiguity-averse DM it’s not necessarily optimal to hold a fully diversified portfolio. This insight may shed some light on various puzzling financial phenomena.

For example, it has been documented that individuals tend to hold very small portfolios (3-4 stocks)\(^6\) This paper proves that such a DM is willing to pay an ambiguity premium for replacing the ambiguous bet \((RY)\) with the risky bet \((BY)\).
Goetzmann and Kumar (2008)[42]). This phenomena can potentially be explained by the existence of ambiguity. Other empirical phenomena that the financial literature refers to as puzzles - under the rational expectations hypothesis, i.e., the assumption that there exists an objective probability law governing the state process and that DMs know this law, which coincides with their subjective beliefs, include: the equity premium puzzle (Mehra and Prescott (1985)[76]), the risk-free rate puzzle (Weil (1989)[106]), the phenomenon wherein the observed equity volatility is too high to be justified by changes in the fundamental (Shiller (1981)[92]), and the home bias puzzle (Coval and Moskowitz (1999)[22]).

Having established a well defined ambiguity measure, this paper uses it to generalize the asset pricing theory to incorporate ambiguity. The point of departure of this generalization is acknowledging the fact that, in reality, asset pricing is conducted under conditions of ambiguity, so that the price of an asset is determined not only by its level of risk and by the DMs’ attitude toward that risk, but also by its level of ambiguity and by the DMs’ attitude toward this ambiguity, in general. This demonstrates that the price of an asset has three components: the price of time, or the pure interest rate on a risk-free asset, the price of risk, an additional expected return per unit of risk borne, and the price of ambiguity, a second additional expected return per unit of ambiguity borne.

In an economy that faces ambiguity, three types of premiums are observed. The risk premium is the premium that a DM is willing to pay for replacing a risky bet with its expected outcome. The ambiguity premium is the premium that a DM is willing to pay for replacing an ambiguous bet with a risky bet with an identical expected outcome. The uncertainty premium is the total premium that a DM is willing to pay for replacing an ambiguous bet with its expected outcome, i.e., the accumulation of risk premium and ambiguity premium.\footnote{This paper uses the term uncertainty to describe the aggregation of risk and ambiguity.} We present a generalization of Pratt’s risk premium to uncertainty and prove that it can be separated into risk premium and ambiguity premium. We propose a well defined ambiguity premium, completely distinguished from risk and attitude toward risk, which can be computed from the data.

We are not aware of any prior study that conducts direct empirical tests on models of decision making under ambiguity with data other than through parametric fitting and calibrations, except for the study of Brenner and Izhakian (2011)[11].\footnote{Uppal and Wang (2003)[100], Epstein and Schneider (2008)[27], and Ju and Miao (2011)[59], for example calibrate their model to the data.}\footnote{A few papers suggest attributing different explanatory variables to ambiguity. For example, Anderson et al. (2009)[3] attributed the degree of disagreement of professional forecasters to ambiguity, while Erbas and Mirakhor (2007)[31] suggest using the World Bank institutional quality indexes as a proxy for ambiguity.} Using Shadow theory, as proposed in this paper, Brenner and Izhakian (2011)[11] empirically show that ambiguity, measured by $\mathbb{A}^2$, has a significant negative impact on the market-portfolio’s return. They use the return on the S&P 500 stock index as a proxy for the market-portfolio. Applying a time series analysis, they prove that the representative investor in the stock market is, in fact, an ambiguity-lover.

Our Shadow theory relies on the Choquet expected utility (CEU) of Schmeidler (1989)[86] assuming that probabilities (capacities) are random and governed by second-order events, where capacities mean
probabilities, possibly nonadditive, i.e., the sum of probabilities can be either smaller or greater than 1. In his seminal work, Schmeidler (1989) introduced an axiomatic derivation which paved the way for modeling decision making under uncertainty. In their pioneering studies, introduce the idea that, in the presence of ambiguity, the probabilities that reflect the DM’s willingness to bet cannot be additive. To enable the integration of nonadditive probabilities they apply the Choquet (1955) integral for nonnegative functions.

Shadow theory combines the concept of nonadditivity with the insight that the attitude towards gains and loss is not symmetric, introduced in prospect theory of (1979). However, while prospect theory applies reference-dependent to preference toward risk and toward utility, Shadow theory applies it to probabilities and to preferences toward ambiguity. Prospect theory assumes that the DM’s preferences toward risk are reference-dependent, with a different risk attitude toward losses and toward gains, such that it gives rise to different evaluations, in terms of utility, of losses and gains. Our model applies reference-dependence to differentiate between the probabilities of a gain and a loss, and applies it to characterize the DM’s preferences concerning these obscure probabilities.

Shadow theory is not the first to integrate the concepts of nonadditive probabilities and asymmetry of gains and losses. Cumulative prospect theory (CPT) of Tversky and Kahneman (1992) also applies a two sides CEU to gains and to losses. As in prospect theory, CPT assumes a DM having a different risk attitude toward losses and toward gains. Capacities, in CPT, are assumed to have a different weighting schemes for losses and for gains. In CPT, nonadditivity is reflected in different decision weights (capacities) for losses and for gains, i.e., a different probabilistic distortion function for losses and for gains.

Probabilities in our model are subjective reductions of objective random probabilities, guided by the DM’s available information and subjective beliefs. Nonadditivity is obtained from the DM’s (reference-dependent) preferences toward ambiguity, applied to the directing space. In our analysis, we assume homogenous risk preferences for losses and for gains, no loss aversion, and a homogenous probability weighting function for losses and for gains. Under these settings, subadditivity is a result of ambiguity aversion, and superadditivity is a result of ambiguity seeking, as in Gilboa (1987) and Schmeidler (1989). Nonadditivity is sustained either for homogenous ambiguity preferences for losses and for gains, or for reference-dependent preferences to ambiguity, unless the DM is ambiguity-neutral. In the case of an ambiguity-neutral DM, pure-loss prospects or pure-gain prospects, our model reverts back to the classical expected utility model.

Capacities are also referred to as decision weights or probability weights. This paper usually uses the term probability in a broad sense, i.e., can possibly be nonadditive and can be either subjective or objective.

Inspired by Knight (1921), his main philosophical assertion is that a DM can be rational; yet, not being Bayesian.

See also, Schmeidler (1982), Gilboa (1985) and Schmeidler (1986).

A typical DM, according to both prospect theory, is assumed to exhibit risk aversion for gains, risk seeking for losses, and loss aversion modeled by a steeper value function for losses than for gains.

Using the perception of rank-dependent and cumulative functionals proposed by Weymark (1981), Quiggin (1982), Yaari (1987) and Schmeidler (1989), cumulative prospect theory generalizes the original prospect theory of Kahneman and Tversky (1979) from risk to uncertainty. Relying on cumulative functionals, the newer version modifies the probability weighting functionals to relax the theoretical limitations of the original prospect theory, such that it always satisfies stochastic dominance and supports prospects taking a large (or an infinite) number of outcomes.
Shadow theory relies on the axiomatic foundation proposed by Wakker (2010)[103] for both preferences - toward risk, over the subordinated space, and toward ambiguity, over the directing space. Applying this axiomatization to preferences over the subordinated space proves the existence of a functional representation of the DM’s preferences toward risk. Applying it to preferences over the directing space proves the existence of a functional representation of the DM’s preferences toward ambiguity. Shadow theory shows that the weighting probability functions (capacities) in CPT are not arbitrary and can be explained by the presence of ambiguity and the DM’s preferences toward it.

The rest of the paper is organized as follows: Section 2 reviews the basic principles of CPT and extends it to introduce our new Shadow theory for decision making under uncertainty. Using this model, Section 3 suggests a new measure of ambiguity and characterizes DMs’ attitudes toward it. Section 4 discusses ambiguity diversification and joint ambiguity of prospect portfolios. Section 5 pushes this theory further and uses it to incorporate ambiguity into asset pricing theory. Section 6 discusses our theoretical results with respect to the related literature and Section 7 concludes.

2 Shadow probability theory

The decision-making model presented in this paper relies on the foundation of cumulative prospect theory (CPT) for uncertainty, introduced by Tversky and Kahneman (1992)[96] and its axiomatization proposed by Wakker (2010)[103]. In order to provide some background, we first review the basic principles of this theory and its relevant preliminary definitions. Next, we generalize this theory to multidimensional uncertainty by adding an additional tier of prospects, this time with respect to the capacities which govern the outcomes in the original CPT. To simplify our exposition, whenever possible our results are proven in static discrete settings; however all of the presented results can be easily applied to dynamic continuous settings.

2.1 Preliminaries

Prospect theory assumes two differentiated phases in the decision-making process: framing and valuation. In the framing phase, based on the information she has, the DM constructs a representation of the possible acts, contingencies and consequences, called prospects, which are relevant to her decision. In the valuation phase, based on her preferences, the DM assesses the value of each prospect and chooses accordingly. This paper concentrates on the valuation process assuming well framed prospects. We also assume that DMs do not exhibit unawareness, i.e., they are aware of all states of nature.

The main idea of cumulative prospect theory is that the utility of an uncertain prospect is the sum of the utilities of the outcomes, each weighted by a transformation of the cumulative probabil-

\footnotetext{15}{This axiomatization, Theorem 12.3.5 in Wakker (2010)[103], is based on the tradeoff consistency approach developed by Tversky et al. (1988)[98] and Wakker (1989)[102].}

\footnotetext{16}{Tversky and Kahneman (1986)[95] study the principles that govern the representation of acts, contingencies and consequences, as prospects.}
ity distribution function, while the transformation is applied separately to losses and gains. This theory assumes that probabilities (weighting or capacities in terms of CPT) are subjective and not necessarily additive, i.e., do not add to unity. To formulate different preferences toward gains and losses, CPT assumes that the utility arising from the outcomes are determined with respect to a fixed reference consequence (status quo). Outcomes greater than the reference are considered as a gain, while outcomes lower than the reference are considered as a loss.

Formally, let \( S \) be a (finite or infinite) state space, endowed with an algebra of subsets of \( S \). Later, when generalizing the CPT, we refer to this space as the subordinated space. It is assumed that exactly one state can be realized, which is unknown to the DM while making her choice. Subsets of the state space are called events and are denoted by \( \mathcal{E} \). The set of events, \( \Xi \), is a \( \sigma \)-algebra of subsets of \( S \). The complementary event, \( \mathcal{E}^c \), consists of all states \( s \in S \) not contained in \( \mathcal{E} \). We define \( X \) to be the set of consequences, also called outcomes. Since our paper mostly deals with financial aspects, consequences are confined to real numbers, \( X \subseteq \mathbb{R} \), thus any consequence \( x \in \mathbb{R} \). We assume that there is a neutral consequence, denoted \( x_k \in X \), with respect to which all other elements of \( X \) are interpreted either as a gain or a loss.

A prospect is a function from states into consequences, \( f : S \to X \), describing the resulting consequence associated with each state \( s \in S \). We denote \( \mathcal{F} \) to be the set of all possible prospects. If no confusion arises, we often suppress function notations and write \( f_j \) for \( f(s_j) \). When possible, for simplicity’s sake, we assume that prospects take a finite number number of values. A prospect \( f \in \mathcal{F} \) is represented as a sequence of \( j = 1, \ldots, n \) pairs

\[
  f = (\mathcal{E}_1 : x_1, \cdots, \mathcal{E}_j : x_j, \cdots, \mathcal{E}_n : x_n),
\]

where \( x_j \) is the consequence if event \( \mathcal{E}_j \) occurs, i.e., \( x_j \) is the outcome under each state \( s \in \mathcal{E}_j \), and \( (\mathcal{E}_1, \cdots, \mathcal{E}_n) \) is a partition of the state space \( S \). Prospects, designating state-consignment consequences, are assumed to be equipped with a complete sign-ranking with respect to outcomes,

\[
  x_1 \leq \cdots \leq x_{k-1} \leq x_k \leq x_{k+1} \leq \cdots \leq x_n,
\]

for some \( 1 \leq k \leq n \). That is, all consequences are ranked not only with respect to each other, but also with respect to the neutral consequence \( x_k \). We refer the neutral point \( k \) with the neutral value \( x_k \) as the reference point and the reference consequence, respectively. It is important to note that the reference consequence \( x_k \) can be any value and is not necessarily required to be \( x_k = 0 \).

\footnote{In the original prospect theory, Kahneman and Tversky (1979)[60] encounter two problems: the theory does not always satisfy stochastic dominance, and prospects taking a large number of outcomes are not supported. In addition to the generalization for risk to uncertainty, CPT addresses these two issues.}

\footnote{Following Wakker and Tversky (1993)[104]. Wakker (2010, Appendix G)[103] and Kothiyal et al. (2011)[66], the state space, \( S \), can consist of an infinite number of states. Therefore, all results presented in this paper are valid as regards an infinite state space.}

\footnote{It is common to assume that the reference point is the status quo, exogenously given. In the financial world, when consequences are returns rather than quantitative outcomes, a natural objective reference consequences can possibly be 0 or the risk-free rate. However, any other consequences can also be selected.}
Sometimes, when it is clear from the context, we refer to the prospect \( f \in \mathcal{F} \) with a vector of outcomes \( X = (x_1, \ldots, x_n) \) simply as a random variable, but possibly without specified probabilities.

All consequences \( x \in X \) are defined with respect to the designated reference consequence \( x_k \in X \). Any outcome \( x_j \in X \) is a loss if \( x_j < x_k \), a nongain if \( x_j \leq x_k \), a gain if \( x_k < x_j \), a nonloss if \( x_k \leq x_j \) and a neutral consequence if \( x_j = x_k \). The cumulative event of loss is, thus, defined by \( \mathcal{E}_L = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_k \), while the cumulative event of gain by \( \mathcal{E}_G = \mathcal{E}_{k+1} \cup \cdots \cup \mathcal{E}_n \), denoted \( L \) and \( G \), respectively. A prospect \( f \in \mathcal{F} \) is called positive or strictly positive if its consequences are all nonnegative or positive, respectively. Similarly, a prospect is called negative or strictly negative if its consequences are all nonpositive or negative, respectively. All other prospects are called mixed prospects.

For any prospect \( f \in \mathcal{F} \) let \( f^- \) denote the loss part of \( f \) obtained by assigning \( f^-_j = f_j \) if \( f_j \leq x_k \) and \( f^-_j = x_k \) if \( x_k < f_j \). Similarly, the gain part of any prospect \( f \in \mathcal{F} \) is denoted by \( f^+ \) and obtained by assigning \( f^+_j = f_j \) if \( x_k < f_j \) and \( f^+_j = x_k \) if \( f_j \leq x_k \).

The domain of preference relation, \( \succ \), is the nondegenerated set of all prospects. The relations \( \prec, \succ, \sim, \preccurlyeq, \text{ and } \succcurlyeq \) are defined as usual. Prospects yielding the same consequence \( x \in X \) for each state \( s_j \in S \) are called constant prospects, and are designated by their constant consequences, \( x \). A certainty equivalent (CE) of a prospect \( f \) is a constant prospect, \( x \), such that \( f \sim x \). A real function \( V : \mathcal{F} \rightarrow \mathbb{R} \) assigns to each prospect \( f \in \mathcal{F} \) a value \( V(f) \) such that \( V(f) \geq V(g) \) iff \( f \succ g \). A capacity \( W \) is a function on \( 2^S \) that assigns to each event \( A \subseteq S \) a number \( W(A) \) satisfying \( W(\emptyset) = 0 \), \( W(S) = 1 \) and if \( A \subset B \subset S \) then \( 0 \leq W(A) \leq W(B) \); that is, the capacities are monotonic. Since capacities are decision weights that represent the DM’s subjective probabilities, sometimes we simply refer to them as probabilities.

CPT holds if there is a strictly increasing continuous utility function, \( U : X \rightarrow \mathbb{R} \), satisfying \( U(x_k) = 0 \), and the two weighting functions \( W^- \) and \( W^+ \) for losses and for gains, respectively. The utility function, \( U(\cdot) \), characterizes the DM’s preferences toward risk. As usual, a concave function indicates risk aversion, while a convex function indicates risk seeking. As in classical expected utility theory, the utility function \( U(\cdot) \), representing the DM’s preferences, is unique up to affine transformation.\(^{20}\) CPT assumes that the DM might exhibit loss aversion, risk aversion for gains, and risk seeking for losses at the same time.\(^{21}\) Confining attention to uncertainty, our analysis assumes that neither of these preferences hold. We simply assume risk aversion for gains and losses; however, such preferences can be easily incorporated into our model.

The CPT value of a prospect \( f \in \mathcal{F} \) is evaluated by

\[
V(f) = V(f^-) + V(f^+),
\]

\(^{20}\)This property is also referred to as unique up to unit and level or unique up to scale and location.

\(^{21}\)Loss aversion in terms of CPT, is modeled by a steeper utility function for loss, i.e., \( x \leq x_k \), compared to the utility function for gains, i.e., \( x_k < x \). Most prior studies report risk aversion for gains, while some behavioral works document risk seeking for losses, for example Shefrin and Statman (1985)[91], Wang and Fischbeck (2004)[105], and Abdellaoui et al. (2008)[1], while others document risk aversion for losses, for example Cohen et al. (1987)[21], Shead and Hodgins (2009)[90]. Loss aversion, in the context of optimal portfolio selection, has also been studied, for example, by Barberis and Huang (2001)[7].
such that

\[ V(f^-) = \sum_{j=1}^{k} \pi_j^- U(x_j) \quad \text{and} \quad V(f^+) = \sum_{j=k+1}^{n} \pi_j^+ U(x_j), \]

where the decision weights, \( \pi^- \) and \( \pi^+ \), satisfy \( \pi^-(f^-) = (\pi_1^-, \ldots, \pi_k^-) \) and \( \pi^+(f^+) = (\pi_{k+1}^+, \ldots, \pi_n^+) \).

The decision weights for losses, \( 1 \leq j \leq k \), are defined by

\[ \pi_j^- = W^- (\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_j) - W^- (\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_{j-1}), \quad 2 \leq j \leq k, \]

and \( \pi_1^- = W^- (\mathcal{E}_1) \), \( j = 1 \).

The decision weights for gains, \( k + 1 \leq j \leq n \), are defined by

\[ \pi_j^+ = W^+ (\mathcal{E}_j \cup \cdots \cup \mathcal{E}_n) - W^+ (\mathcal{E}_{j+1} \cup \cdots \cup \mathcal{E}_n), \quad k + 1 \leq j \leq n - 1, \]

and \( \pi_n^+ = W^+ (\mathcal{E}_n), \quad j = n \).

Practically, \( \pi \) obtains the meaning of a marginal decision weight: For any \( j \leq k \) the expression \( \pi_j^- = W^- (\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_j) - W^- (\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_{j-1}) \) is the event’s \( j \) marginal decision weight for losses, and for any \( k < j \) the expression \( \pi_j^+ = W^+ (\mathcal{E}_j \cup \cdots \cup \mathcal{E}_n) - W^+ (\mathcal{E}_{j+1} \cup \cdots \cup \mathcal{E}_n) \) is the event’s \( j \) marginal decision weight for gains. It is important note that the weights, \( W \), does not necessarily add to unity.\(^{22}\)

The weighting functions \( W^- \) and \( W^+ \) are also written \( W^- (\mathcal{E}) = w^- (P(\mathcal{E})) \) and \( W^+ (\mathcal{E}) = w^+ (P(\mathcal{E})) \), where \( P : \Xi \to [0, 1] \) is a probability (possibly subjective) measure, i.e., \( W \) is the composition of \( w \) and \( P \).\(^{23}\)

The weighting function \( w (\cdot) \) is assumed to be a strictly increasing function from the unit interval to itself, satisfying \( w^- (0) = w^+ (0) = 0 \) and \( w^- (1) = w^+ (1) = 1 \). Because \( w (\cdot) \) can be nonlinear, \( W \) need not be additive. That is, if \( A \cap B = \emptyset \) then \( W(A \cup B) \neq W(A) + W(B) \).

The weighting functions capture the DM’s perception of likelihood and can be considered as representing pessimism or optimism. The utility function, \( U \), characterizes her preferences toward risk, i.e., risk aversion or risk seeking. The generalized version of CPT, presented in this paper, proves that the weighting functions \( W^- \) and \( W^+ \) are determined by the presence of ambiguity and the DM’s attitude toward it. This proves that subadditive weighting functions are obtained by aversion to ambiguity.

The functional representation of a prospect \( f \in \mathcal{F} \) is formulated by

\[
V(f) = \sum_{j=1}^{k} \left[ w^- (P(\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_j)) - w^- (P(\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_{j-1})) \right] U(x_j) + \\
\sum_{j=k+1}^{n} \left[ w^+ (P(\mathcal{E}_j \cup \cdots \cup \mathcal{E}_n)) - w^+ (P(\mathcal{E}_{j+1} \cup \cdots \cup \mathcal{E}_n)) \right] U(x_j).
\]

Our decision-making model uses the basic structure of CPT and its axiomatic foundation. The next

\(^{22}\)To simplify notation, the weighting function \( W \) is used when referring to both \( W^- \) and \( W^+ \). The same conventions are used for \( w \) and \( \pi \).

\(^{23}\)Similarly, the weighting functions can be applied directly to consequences, without specifying an underlying state space; see Jaffray (1989)[57].
Theorem 2.1. [Wakker, 2010, Theorem 12.3.5] Under the structural assumption above and the assumption that the preference relation $\succ$ is truly mixed, the following two statements are equivalent for $\succ$ on the set of prospects $\mathcal{F}$.

(i) CPT holds, where utility is continuous and strictly increasing.

(ii) $\succ$ satisfies: weak ordering; monotonicity; continuity; gain-loss consistency; sign-tradeoff consistency.

The preference $\succ$ is truly mixed if it satisfies $(\mathcal{E} : x_1, \mathcal{E}^c : x_k) \prec (\mathcal{E} : x_1, \mathcal{E}^c : x_2) \prec (\mathcal{E} : x_k, \mathcal{E}^c : x_2)$ for some event $\mathcal{E} \in \Xi$ and consequences $x_1, x_2 \in X$, such that $x_1 < x_k < x_2$. This requirement is imposed to ensure nondegeneracy. Weak ordering implies that the preference relation $\succ$ is complete and transitive. Continuity means that $\succ$ is continuous in the product topology on $\mathbb{R}^n$. That is, the subset of prospects $\{f \in \mathcal{F} \mid f \succ g\}$ and $\{f \in \mathcal{F} \mid f \preceq g\}$ are closed subsets of $\mathbb{R}^n$ for any $g \in \mathcal{F}$. Gain-loss consistency assures a separable evaluation of gains and losses that holds if for all $f, g \in \mathcal{F}$ if $f^- \sim g^-$ and $f^+ \sim g^+$ then $f \sim g$. Sign-tradeoff consistency is sustained if improving consequence in any indifference relation $\sim$ breaks the relation. Formally, if $(\mathcal{E} : a, \mathcal{E}^c : x) \sim (\mathcal{E} : b, \mathcal{E}^c : y), (\mathcal{E} : c, \mathcal{E}^c : x) \sim (\mathcal{E} : d, \mathcal{E}^c : y)$ and $(\mathcal{E} : a', \mathcal{E}^c : x) \sim (\mathcal{E} : b, \mathcal{E}^c : y)$ then $a' = a$.

We conclude this brief review by presenting the definition of CPT for an infinite state space, in which prospects take an infinite number of outcomes. Analogously to the discrete version of CPT, the value of a prospect with an infinite support is defined by

$$V(f) = -\int_{-\infty}^{k} w^-(P(\{s \in S \mid U(f_s) < t\})) dt + \int_{k}^{\infty} w^+(P(\{s \in S \mid U(f_s) > t\})) dt.$$ 

Since $f$ can be considered as a probability measure, this equation can be written as follows

$$V(f) = -\int_{-\infty}^{k} W^-(f^{-1}[U^-(\infty, t)]) dt + \int_{k}^{\infty} W^+(f^{-1}[U^-(t, \infty)]) dt.$$ 

If the probability measure $P$ has a density function $P' = p$, then the CPT value function can be defined

---

24 Other CPT axiomatizations for general utility form have appeared in Luce and Fishburn (1991)[71], Luce (1991)[70], Tversky and Kahneman (1992)[96], Wakker and Tversky (1993)[104], Chew and Wakker (1996)[52] and Kothiyal et al. (2011)[66].

25 Considering a finite state space, if strong monotonicity holds, gain-loss consistency and true mixedness can be dropped (Wakker and Tversky (1993)[104]). Enforcing comonotonicity instead of ranks, sign-tradeoff consistency can be weakened to sign-tradeoff consistency (Wakker and Tversky (1993)[104] and Köhlerling and Wakker (2003)[64]).

26 Tversky and Kahneman (1992)[96] refer to this requirement as double matching.

27 In order to maintain CPT, the main theorem of Tversky and Kahneman (1992)[96] imposes comonotonic independence instead of sign-tradeoff consistency.

28 For further discussion on the conditions imposed on the preference relation $\succ$, such that CPT holds, the reader is referred to Tversky and Kahneman (1992)[96] and Wakker (2010)[103].
presented by

\[ V(f) = \int_{-\infty}^{k} U(x) w^{-f} (1 - P(x)) p(x) \, dx + \int_{k}^{\infty} U(x) w^{+f} (P(x)) p(x) \, dx, \]  \tag{2}  

where here \( x \) is simply referred to as a random variable and the probability measure \( P \) is not necessarily additive.

2.2 The directing space

At this point our model branches off CPT by incorporating an additional layer of uncertainty, this time with respect to probabilities. This model, called *Shadow probability theory* (SPT), generalizes CPT by assuming that probabilities (capacities) are random. The main idea behind this extension is that there exists an additional separate latent space, called a *directing-space*, which dominates the probability of observable consequences in the *subordinated-space*, also referred to as *outcome-space*.\(^{29}\) While making her choice, the DM does not know which event will be realized, either in the subordinated space or in the directing-space, which means that she knows neither the realized outcome nor the realized probability of outcomes.

Taking Ellsberg’s three-color urn experiment as an example, the subordinated space and its prospects are defined by the prizes attached to the four possible events, i.e., \((R)\), \((B)\), \((RY)\) and \((BY)\).\(^{30}\) The probabilities of these events are determined by unobservable events in the directing space. A directing space can, for example, be defined by the amount of black balls and the amount of yellow balls that the experimenter put in the urn. An event in the directing space, for instance, can be ”The experimenter put 30 red balls, 10 black balls, and 50 yellow balls in the urn.” The DM, of course, does not know how many balls of each type the experimenter has put in the urn.

Another example could be a farmer who has to decide which seeds to sow. Her income is determined by the amount of her harvest, which in turn is determined by the amount of rain. A specific amount of rain does not guarantee a specific harvest quantities, but only dominates the probabilities of reaching various harvest quantities. The amount of rain, however, is random and unknown at the point of making the decision. In terms of SPT, the farmer’s income defines the subordinated space, while the weather defines the directing space. Each type of seed is modeled by a different prospect over the subordinated space, while the amount of rain is modeled by a prospect in the directing space. The probabilities of the various incomes generated by each type of seed are governed by events (amount of rain) in the directing space.

The state space, \( S \), and its associated prospects, \( \mathcal{F} \), which have already been defined in the previous subsection, now take on the meaning of *subordinated space*. Probabilities of events \( \mathcal{E} \in \Xi \), occurring in the subordinated space, are governed by events in the *directing state space* (or *directing space* for short), \( \Omega \), consists of a finite or an infinite number of *directing states* \( \omega \). In other words, the probability \( P \) of each of the events \( \mathcal{E}_1, \ldots, \mathcal{E}_n \in \Xi \) is random and determined by finite subsets \( \varepsilon_i = \{\omega_1, \ldots, \omega_l\} \)

\(^{29}\)We deviate from the mathematical meaning of *space* and use this term in its broader meaning to describe a universe.

\(^{30}\)A description of this experiment is presented in the introduction.
of the directing space $\Omega$, where $\omega_1, \ldots, \omega_t \in \Omega$. A subset $\epsilon_i$ is called a directing event and the set of directing events is denoted $\Sigma$, i.e., a $\sigma$-algebra of subsets of $\Omega$. The subset of events $(\epsilon_1, \ldots, \epsilon_m)$ is assumed to be a partition of the directing state space $\Omega$, i.e., $\epsilon_i \cap \epsilon_l = \emptyset$, $\forall i \neq l$.

A directing consequence is a probability measure, possibly nonadditive, $P_i$, over the subordinated space $S$, where $P_i$ stands for $P(\cdot | \epsilon_i)$. $P$ denotes the set of consequential probability measures. In a finite state space the probability measure $P_i$ takes the form of probability vector $q_i = (q_{i,1}, \ldots, q_{i,j}, \ldots, q_{i,n})$, assigning to each event its probability, where $q_{i,j}$ stands for $q(\epsilon_j | \epsilon_i)$. Sometimes, when the context is clear, we omit the index of the subordinated event $j$ and simply write $q_i$.

Without loss of generality, we assume that $q : \Omega \to [0, 1]$ is additive.

A directing prospect, $g$, is a function from the directing space into the set of probability measures, $g : \Omega \to P$, describing the resulting probability measure associated with each directing state $\omega \in \Omega$. For simplicity’s sake, when possible we assume that directing prospects take a finite number of values. The set of directing prospects is denoted $\mathcal{G}$, where for simplicity’s sake, at this point we assume that $\mathcal{G}$ contains a single prospect. The DM does not have a direct choice over directing-prospects; however, her subjective perception of likelihoods, resulting in aversion or seeking ambiguity, is inspired by the nature of the directing prospects. A directing prospect $g \in \mathcal{G}$ can be defined by a sequence $i = 1, \ldots, m$ of pairs

$$g = (\epsilon_1 : q_1, \ldots, \epsilon_i : q_i, \ldots, \epsilon_m : q_m),$$

where the outcome of the prospect, $q_i$, is a probability measure. It is important to note that the directing prospects are degenerating in the sense that they need not be equipped with either a ranking or a reference point.\(^{31}\)

We define a second-order weighting function (second-order capacity), $\chi(\cdot)$, on $2^\Omega$, which assigns to each directing event $\epsilon \subset \Omega$ a number $\chi(\epsilon)$, satisfying $\chi(\emptyset) = 0$, $\chi(\Omega) = 1$, and if $A \subset A \subset \Omega$ then $0 \leq \chi(A) \leq \chi(B)$. We refer to $\chi(\epsilon_i)$ as the directing weight of the directing event $\epsilon_i$ and use the notation $\chi_i$ for short. It is important to note that the directing weight function $\chi(\cdot)$ need not be additive.

Figure 1 gives a diagrammatic representation of the two spaces: the outcome space and the directing space, and the relation between them.

We conclude this subsection by defining the following notational conventions $\epsilon_j = \epsilon_1 \cup \cdots \cup \epsilon_j$ and $\epsilon_{K \cdots J} = \epsilon_k \cup \cdots \cup \epsilon_j$ to denote cumulative events. The same notational conventions are applies to all other variables, for example $P_j = P(\epsilon_1 \cup \cdots \cup \epsilon_j)$ and $P_{K \cdots J} = P(\epsilon_k \cup \cdots \cup \epsilon_j)$.

### 3 Ambiguity

The main at the base of our model is that risk and risk attitude apply to the subordinated space, as in classical expected utility theory, whereas the level of ambiguity and the DM’s preferences toward it

\(^{31}\)In contrast to prospects in the outcome space, in which we enforce the notion that the rank and sign of outcomes of events agree for all prospects considered, such that the same decision weights can be used for all prospects. As regards directing prospects, this restriction can also be canceled.
apply to the directing space. This section concentrates on the implications of the directing space for ambiguity and the DM’s preferences toward it.

3.1 Ambiguity and attitudes toward it

The DM is assumed to have a (second-order) preference relation, $\succ$, over the set of all directing prospects, $\mathcal{G}$, satisfying the conditions of Theorem 2.1. Thus, the functional form of this preference is characterized by a continuous and strictly increasing function $\psi(\cdot)$. This function frames the DM’s attitude toward ambiguity and is referred to as probability-utility or $p$-utility for short. Preferences and their utility representation over the directing space are degenerated in the sense that a distinction of losses from gains is not required, such that the directing weights satisfy $\chi^{-}(\cdot) = \chi^{+}(\cdot) = \chi(\cdot)$ and the $p$-utility, $\psi(\cdot)$, is the same for both losses and gains. Therefore, true mixing is not imposed on the second-order preference, $\succ$. Applying Theorem 2.1 to the directing space proves that the functional representation $\psi(\cdot)$ coincides with the second-order preference relation, $\succ$.

Similarly to preferences toward risk, we identify three types of preferences toward ambiguity: aversion to ambiguity, molded by a concave $\psi(\cdot)$; loving ambiguity, modeled by a convex $\psi(\cdot)$; and neutrality to ambiguity, modeled by a linear $\psi(\cdot)$. For future examples we also define two special types of preference toward ambiguity: constant relative ambiguity aversion (CRAA), which takes the functional form $\psi(p) = \frac{1^\frac{1}{1-\eta}}{1-\eta}$ and constant absolute ambiguity aversion (CAAA), which takes the functional form $\psi(p) = -\frac{1}{\frac{1}{1-\eta} + \eta}$, where $\eta$ designates the coefficient of ambiguity aversion.

Using this notion of ambiguity and the DM’s preferences toward it, we suggest the following definition of capacities.

Figure 1: The two spaces
**Definition 3.1.** The *subjective probability* \( Q(\mathcal{E}_j) \) of an event \( \mathcal{E}_j \in \Xi \) is determined by

\[
Q(\mathcal{E}_j) = \psi^{-1} \left( \frac{\chi(\xi_j)}{\sum_{i=1}^{m} \chi(\xi_i)} \psi(\xi_j|\xi_i) \right),
\]

where \( \psi : \mathbb{R} \to \mathbb{R} \).\(^{32} \)

In Shadow theory, while estimating the value of any prospect \( f \in \mathcal{F} \) in the subordinated space, \( \mathcal{S} \), the DM’s preferences over the directing space, \( \Omega \), are served to shape her subjective probabilities, \( Q \), associated with each event, \( \mathcal{E}_j \in \Xi \), in the subordinated space. Formally, Equation (3) is integrated into Equation (1) to obtain the value of prospect \( f \).

**Model 3.2.** In the Shadow theory framework, the value of a prospect \( f \in \mathcal{F} \) is

\[
V(f) = \sum_{j=1}^{k} \left[ w^{-} \left( \psi^{-1} \left( \frac{\sum_{i=1}^{m} \chi(\xi_i)}{\sum_{i=1}^{m} \chi(\xi_i)} \psi(q_{j,1-\ldots-J}) \right) \right) \right] U(x_j) + \sum_{j=k+1}^{n} \left[ w^{+} \left( \psi^{-1} \left( \frac{\sum_{i=1}^{m} \chi(\xi_i)}{\sum_{i=1}^{m} \chi(\xi_i)} \psi(q_{j,J+1-\ldots-N}) \right) \right) \right] U(x_j).
\]

The functional representation of the DM’s aggregate preferences, presented in Model 3.2, makes a complete distinction between beliefs and preferences and between risk and ambiguity. First-order beliefs are assessed by the random probability measures \( q_1, \ldots, q_i, \ldots, q_m \); second-order beliefs are assessed by the directing capacities \( \chi_1, \ldots, \chi_i, \ldots, \chi_m \); preferences toward risk are represented by the utility function \( U(\cdot) \); and preferences toward ambiguity are represented by the p-utility function \( \psi(\cdot) \). The functions \( w^{-}(\cdot) \) and \( w^{+}(\cdot) \) can be interpreted as the DM’s functional representation of pessimism or optimism. *Pessimism* holds if worsening the rank increases the weighting assigned by \( w(\cdot) \), i.e., bigger weights are assigned for worse ranks. *Optimism* holds if improving the rank increases the weighting assigned by \( w(\cdot) \), i.e., bigger weights are assigned for better ranks. The separation obtained by Model 3.2 allows us to isolate and investigate the distinct impact of each component on prospects’ values. Yet, even more important, it enables the simplification of the model to an applicative form as, in fact, the rest of the paper does. This simplification paves the way for empirical research that tests the impact of ambiguity using data.

In CPT, if the weighting function \( W = w(P(\mathcal{E})) \) is additive, and hence so is the probability measure, then \( \pi_j \) is simply the probability \( P(\mathcal{E}_j) \) of event \( \mathcal{E}_j \). It follows readily from the definition of \( \pi(\cdot) \) and \( W(\cdot) \) that for both positive and negative prospects the decision weights add to unity. For mixed prospects, however, the sum can be either smaller or greater than 1. In SPT, adding the element of ambiguity aversion allows the probability to be nonadditive, even when the weighting functions are additive. Our model distinguishes between two sources of nonadditivity: ambiguity

---

\(^{32}\) Considering a directing space with an infinite support, the subjective probability function can be constructed in a manner similar to Equation (2).
through the subjective probabilities \( Q(\cdot) \) and pessimism, through the weighting function \( w(\cdot) \). In principle, a DM can exhibit ambiguity aversion while still having optimistic perceptions of likelihoods.

Our next goal is to simplify Equation (4) in Model 3.2 into a friendly applicative form, but first we need additional notations. Let

\[
p_j = E[P_j] = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} q_{i,j}
\]

be the expected probability of event \( \mathcal{E}_j \) and

\[
\xi_j^2 = \text{Var}(P_j) = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_{i,j} - p_j)^2
\]

be its variance. The covariance of the probability of two events \( \mathcal{E}_j \) and \( \mathcal{E}_l \) is defined by

\[
\xi_{j,l} = \text{Cov}(P_j, P_l) = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_{i,j} - p_j) (q_{i,l} - p_l)
\]

Now, to simplify our exposition and make the model more applicable, we approximate the subjective probability of event, \( P(\mathcal{E}_j) \), by taking a second-order Taylor approximation around \( p_j \). We apply the same method as that used by Arrow (1965)[5] and Pratt (1964)[78] within the expected utility framework, but this time we apply it to the directing space with respect to the consequential probabilities. Arrow-Pratt restrict their results to the case of small consequences. In our case, this restriction is not required since consequences are probabilities ranging between 0 and 1. However, since we are dealing with probabilities, a restriction on the p-utility function \( \psi(\cdot) \) is enforced to assure a nonnegative approximation of the subjective probabilities, \( Q(\cdot) \).

**Theorem 3.3.** Assume a continuous and twice-deferential p-utility \( \psi(\cdot) \), satisfying

\[
\frac{1}{2} \left( \frac{\psi''(p_{\mathcal{E}_j})}{\psi'(p_{\mathcal{E}_j})} \xi_{j}^2 - \frac{\psi''(p_{\mathcal{E}_j \cup \mathcal{E}_l})}{\psi'(p_{\mathcal{E}_j \cup \mathcal{E}_l})} \xi_{j \cup l}^2 \right) \leq p_{\mathcal{E}_j},
\]

for any events \( \mathcal{E}_j, \mathcal{E}_l \in \Xi \), then the subjective probability of an event \( \mathcal{E}_j \in \Xi \) can be approximated by

\[
Q(\mathcal{E}_j) = p_j + \frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} \xi_j^2.
\]

In a similar way, the subjective probability of a cumulative event \( \mathcal{E}_k \cdots \mathcal{E}_j = \mathcal{E}_k \cup \cdots \cup \mathcal{E}_j \) can be approximated by

\[
Q(\mathcal{E}_k \cup \cdots \cup \mathcal{E}_j) = p_{K \cdots J} + \frac{1}{2} \frac{\psi''(p_{K \cdots J})}{\psi'(p_{K \cdots J})} \xi_{K \cdots J}^2.
\]

Theorem 3.3 characterizes the DM’s subjective probabilities, which satisfy \( Q(\emptyset) = 0 \) and \( Q(S) = 1 \). Lemma A.2 proves that if \( A \subset B \subset S \) then \( Q(A) \leq Q(B) \) and, thus, \( Q(\cdot) \) is a capacity. The
under these technical conditions. \( p \) is bounded by \( \xi \) probabilities defined in Equation (3). Lemma A.1 proves that the variance of each probability, \( \xi \), is bounded by \( p_j (1 - p_j) \). Henceforth we assume that the \( p \)-utility and variance of probabilities fall under these technical conditions.

To establish a new terminology that will serve for the reminder of the work, we define the following.

**Definition 3.4.** We refer to the expression

\[
\varphi_J = -\frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} \xi_j^2
\]

as the **probability premium** of event \( E_J \) and \( \xi_j^2 \) is referred to as the **ambiguity of event \( J \) (e-ambiguity for short).** We call the expression \(-\frac{\psi''(p_j)}{\psi'(p_j)}\), the coefficient of absolute ambiguity aversion while the expression \(-p_j \frac{\psi''(p_j)}{\psi'(p_j)}\), is referred to as the coefficient of relative ambiguity aversion.\(^{33}\)

The probability premium is composed of two components: the DM’s preferences toward ambiguity, framed by \(-\frac{\psi''(\cdot)}{\psi'(\cdot)}\), and the level of e-ambiguity, measured by \( \xi^2 \). Preferences toward ambiguity can be aversion, \(-\frac{\psi''(\cdot)}{\psi'(\cdot)} > 0\), loving, \(-\frac{\psi''(\cdot)}{\psi'(\cdot)} < 0\), or neutrality, \(-\frac{\psi''(\cdot)}{\psi'(\cdot)} = 0\). These preferences can possibly be different for gains and for losses. Clearly a higher ambiguity aversion or a higher level of e-ambiguity implies a greater probability premium and a lower subjective probability. When the probability of any event \( E_j \in \Xi \) is perfectly known, its variance equals zero, its e-ambiguity is zero, i.e., \( \xi^2 = 0 \), and therefore the probability premium also equals zero. When the DM is ambiguity-neutral then \(-\frac{\psi''(\cdot)}{\psi'(\cdot)} = 0\), which implies that the probability premium equals zero. Subadditive probabilities are obtained when the DM exhibits ambiguity aversion, i.e., \(-\frac{\psi''(\cdot)}{\psi'(\cdot)} < 0\). Superadditive probabilities are obtained when she exhibits ambiguity loving, i.e., \(-\frac{\psi''(\cdot)}{\psi'(\cdot)} > 0\).\(^{34}\) It is important to note that according to Condition (5) the coefficient of absolute ambiguity aversion is bounded, such that for a high enough level of ambiguity aversion subjective probabilities tend to zero, while for a high enough level of ambiguity loving subjective probabilities tend to unity.

In general, the subjective probability measure, \( Q(\cdot) \), is nonadditive, that is

\[
Q(E_1 \cup \cdots \cup E_k \cup E_{k+1} \cup \cdots \cup E_j) \neq Q(E_1 \cup \cdots \cup E_k) + Q(E_{k+1} \cup \cdots \cup E_j).
\]

\(^{33}\)These definitions are equivalent to Arrow-Pratt’s coefficients of absolute risk aversion and relative risk aversion, respectively.

\(^{34}\)These results coincide with the nonadditive priors of Gilboa (1987)[38] and the Choquet expected utility (CEU) model of Schmeidler (1989)[86].
The expected probability of event or cumulative event, however, is additive:

\[ p(\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_k \cup \mathcal{E}_{k+1} \cup \cdots \cup \mathcal{E}_j) = p(\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_k) + p(\mathcal{E}_{k+1} \cup \cdots \cup \mathcal{E}_j). \]

Since subjective probability has an additive component (the expected probability) with an additional nonadditive component (the probability premium), we may refer to it as *semiadditive*. The source of the nonadditivity is the e-ambiguity, measured by the variance of probabilities, \( \xi^2_{1 \ldots j} \), which in most cases satisfies

\[ \xi^2_{1 \ldots j} \neq \xi^2_{1 \ldots K-1} + \xi^2_{K \ldots J}. \]

The ambiguities of two events (e-ambiguities) are not independent.\(^{35}\) The source of nonadditivity is the correlation between the probabilities of every two pairs of events. Since our measure of e-ambiguity is the variance of the possible probabilities of the event, the e-ambiguity of a union of events should also include the covariances between the probabilities of all the pairs of sub-events. In fact, the e-ambiguity of a union of two events is the sum of the e-ambiguity of each event in the union, plus twice the covariance between the probabilities of these events, as the next proposition proves.

**Proposition 3.5.** E-ambiguity satisfies

\[ \xi^2_{1 \ldots J} = \xi^2_{1 \ldots K} + \xi^2_{K+1 \ldots J} + 2 \xi_{1 \ldots K,K+1 \ldots J}, \]

where \( \xi_{1 \ldots K,K+1 \ldots J} \) stands for the covariance between the probabilities of the two subsets of events, i.e.,

\[ \text{Cov} [P(\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_k), P(\mathcal{E}_{k+1} \cup \cdots \cup \mathcal{E}_j)]. \]

A special case is the e-ambiguity that arises when considering an event \( \mathcal{E} \in \Xi \) and its complement event \( \mathcal{E}^c \in \Xi \), since the e-ambiguity of their union equals zero, i.e., \( \xi^2_S = 0 \). The next lemma applies to this case.

**Lemma 3.6.** The covariance between event \( \mathcal{E} \in \Xi \) and its complement event \( \mathcal{E}^c \in \Xi \) satisfies

\[ \text{Cov}[P(\mathcal{E}), P(\mathcal{E}^c)] = -\text{Var}[P(\mathcal{E})] = -\text{Var}[P(\mathcal{E}^c)] \]

and thus \( \xi^2_S = 0 \).

The results of Proposition 3.5 and Lemma 3.6 coincide with the findings of the *support theory* of Tversky and Koehler (1994)[97] and that of Rottenstreich and Tversky (1997)[82]. Support theory demonstrates that the judged probability of an event generally increases when its description is unpacked into disjoint components, and decreases by unpacking the alternative description. Furthermore, this theory states that judged probabilities are complimentary in the binary case, and subadditive in the general case. One can easily see from Proposition 3.5 that, when

\(^{35}\)One can write \( \phi(\mathcal{E}_j, \mathcal{E}_1) = \xi^2_{j,1} - \xi^2_{j} - \xi^2_{1} \) to obtain the M"obius transform. See, for example, Chateauneuf and Jaffray (1989)[17] and Grabisch et al. (2000)[44].
of loss, and 

\[ V(\xi) = -\frac{1}{2} \psi''(p_{1\cdots j}) \xi^2_{1\cdots j} - \frac{1}{2} \psi''(p_{1\cdots j-1}) \xi^2_{1\cdots j-1} \] 

\( V(\xi) \) is the disutility caused to an ambiguity-averse DM by ambiguity. If the DM’s preferences toward ambiguity are CRAA, then the value of a prospect is

\[ W(\xi) = \sum_{j=1}^{n} \left[ w^+(p_{1\cdots j-1} - \varphi_{1\cdots j-1}) - w^- (p_{1\cdots j} - \varphi_{1\cdots j}) \right] U(x_j) + \sum_{j=k+1}^{n} \left[ w^+(p_{1\cdots j-1} - \varphi_{1\cdots j-1}) - w^- (p_{1\cdots j} - \varphi_{1\cdots j}) \right] U(x_j) \]

Recall that attitude toward ambiguity can possibly be different for gains and for losses. In this case, the coefficient of ambiguity aversion takes the form 

\[ -\frac{\psi''(p_{1\cdots j})}{\psi'(p_{1\cdots j})} \] 

for any event \( 1 \leq j < k \) of loss, and 

\[ -\frac{\psi''(p_{1\cdots j})}{\psi'(p_{1\cdots j})} \] 

for any event \( k \leq j \leq n \) of gain.

**Example 3.7.** As an example, assume an ambiguity-averse DM who exhibits constant absolute ambiguity aversion (CAAA). Considering the simple case where the probability weighting functions are linear of the type \( w(P) = w^- (P) = w^+(P) = P \), since \( p \) is additive and 

\[ -\frac{\psi''(\xi)}{\psi'(\xi)} = \eta, \] 

one can easily verify that the value of a prospect \( f \) in Equation (6) is simplified to

\[ V(f) = \sum_{j=1}^{n} p_j U(x_j) - \frac{1}{2} \eta \sum_{j=1}^{k} \left[ \xi^2_{1\cdots j} - \xi^2_{1\cdots j-1} \right] U(x_j) - \frac{1}{2} \eta \sum_{j=k+1}^{n} \left[ \xi^2_{1\cdots j-1} - \xi^2_{1\cdots j-1} \right] U(x_j). \]

The first component of the value function is the classical expected utility. The second two components are the disutility caused to an ambiguity-averse DM by ambiguity. If the DM’s preferences toward ambiguity are CRAA, then the value of a prospect is

\[ V(f) = \sum_{j=1}^{n} p_j U(x_j) - \frac{1}{2} \eta \sum_{j=1}^{k} \left[ \frac{\xi^2_{1\cdots j}}{p_{1\cdots j}} - \frac{\xi^2_{1\cdots j-1}}{p_{1\cdots j-1}} \right] U(x_j) - \frac{1}{2} \eta \sum_{j=k+1}^{n} \left[ \frac{\xi^2_{1\cdots j}}{p_{1\cdots j}} - \frac{\xi^2_{1\cdots j}}{p_{1\cdots j}} \right] U(x_j). \]

In both cases, if the DM’s ambiguity preferences are different for losses and for gains, then \( \eta = \eta_L \) for \( 1 \leq j < k \) and \( \eta = \eta_G \) for \( k \leq j \leq n \).
When the weighting functions, $w^- (\cdot)$ and $w^+ (\cdot)$, are linear we use the following notational conventions
\[
\hat{\varphi}_j = \begin{cases} 
\varphi_{1\ldots J} - \varphi_{1\ldots J-1}, & j \leq k \\
\varphi_{J\ldots N} - \varphi_{J+1\ldots M}, & j > k 
\end{cases}
\]
to denote the marginal probability premium. The subjective probability of a subordinated event $\mathcal{E}_j \in \Xi$, thus, takes the form
\[
Q (\mathcal{E}_j) = p (\mathcal{E}_j) - \hat{\varphi} (\mathcal{E}_j).
\]

3.2 The ambiguity measure

An important research question arises when decisions are involved with ambiguity: How should prospects be ranked according to the criteria of ambiguity? The key way to address this question, however, is by determining a well-defined measure of ambiguity. The main goal of this subsection is to provide such a measure. This section uses the properties of the e-ambiguity, identified in the previous subsection, to construct an ambiguity measure for prospects. It opens with a supporting lemma and then constructs the ambiguity measure.

Lemma 3.8. The variance of the aggregate probability of loss is equal to the variance of the aggregate probability of gain, that is
\[
\text{Var} [P_L] = \text{Var} [P_G].
\]

We now turn to prove one of the main statements of this paper: an ambiguity measure.\(^{36}\)

Theorem 3.9. The degree of ambiguity (ambiguity for short) of any prospect $f \in \mathcal{F}$, denoted $\aleph^2$, is defined by
\[
\aleph^2 [f] = 2 \text{Var} [P_L] + 2 \text{Var} [P_G] = 4 \text{Var} [P_L]. \tag{7}
\]

In other words, twice the sum of the variance of the aggregate probability of loss and the variance of the aggregate probability of gain is the measure of the level of ambiguity embedded in a prospect.\(^{37}\)

To normalize our measure to units of probability, we sometimes use
\[
\aleph = 2 \sqrt{\text{Var} [P_L]},
\]
\(^{36}\)The following statement is defined as a theorem, due to its importance in the overall framework of this paper.

\(^{37}\)One may consider
\[
\sum_{j=1}^{k} \begin{bmatrix} \xi_{1\ldots J}^2 & \xi_{1\ldots J-1}^2 \\ p_{1\ldots J} & p_{1\ldots J-1} \end{bmatrix} + \sum_{j=k+1}^{n} \begin{bmatrix} \xi_{J\ldots N}^2 & \xi_{J+1\ldots N}^2 \\ p_{J\ldots N} & p_{J+1\ldots N} \end{bmatrix},
\]
as a more accurate measure of ambiguity. However, we show that $\aleph^2$, which is a much more simple expression, is a good approximation for the level of ambiguity.
as the measure for ambiguity.

The heart of the idea, which lies at the base of the ambiguity measure, ℵ^2, is the notion that probabilities in our framework are random and fluctuate around some reference point. The selected reference point, however, must be meaningful to the DM when she makes decisions. Since decisions involve potential loss, with a non-zero probability (truly mixed), the native reference point is the consequence which distinguishes between losses and gains. This selection is motivated by the insight that the DM cares about the magnitude of the perturbation of the probability of loss and gain; hence, she takes this parameter into account when making choices.

Our measure of ambiguity, ℵ^2, allows for the ranking of prospects according to their level of ambiguity. Theorem 3.13 below proves that, indeed, the ranking of prospects done by an ambiguity-averse DM coincides with the ranking according to ℵ^2. This tool is an important component in the extended toolbox suggested by this paper. It paves the way for financial and economic models to formulate the ”price” of ambiguity, as suggested in Section 5, and to study other financial implications of ambiguity.

The minimal possible degree of ambiguity, zero degree of ambiguity, is obtained when the probabilities are perfectly known. The maximum possible level of ambiguity, ℵ^2 = 1, is obtained when the probability of gains is either 0 or 1 with equal odds. In this most extreme case, the variance of probability of gain and the variance of probability of loss, attains its maximal possible value, 1/4. We normalize this variance of probability by 4 to obtain an ambiguity measure bounded between 0 and 1. It is important to note that the measure of ambiguity ℵ^2 depends on a reference point, x_k, which determines the set of gain-outcomes and the set of loss-outcomes. If, for instance x_k = x_1 or x_k = x_n, i.e., the DM sees all outcomes either as gains or all outcomes as losses, respectively, then the measure of ambiguity equals zero. These are the only two cases in which a prospect is not truly mixed, and therefore does not satisfy the conditions of CPT (Theorem 2.1).

When there is no objective reference point agreed upon by all DMs, which makes a clear differentiation between losses and gains, the ambiguity measure ℵ^2 should be considered as a subjective measure of ambiguity. Concerning financial assets, for example, a zero return or the return on the risk-free asset can possibly be an objective reference point agreed upon by all financial DMs.

Example 3.10. Taking the first part of the Ellsberg experiment as an example, the likelihood of drawing a black (B) ball can result in one of the possible values $0, 1, 2, \ldots, 60$. Considering only strictly positive outcomes as gains ($8$ in this case), the ambiguity degree (twice the standard deviation of the probability of loss) of this gamble is ℵ[B] = 0.3912. In the second part of the experiment, while betting on red or yellow (RY), the likelihoods of gain can result in one of the possible values $30, 31, \ldots, 90$, which in turn also implies an ambiguity level of ℵ[RY] = 0.3912. Table 2 is a stylized description of this example, where E[x] and Var[x] are computed using expected probabilities, p, of the subordinated events.

Now, let us assume that instead of 60 black and yellow balls with an unknown proposition, the urn

---

38 The experiment is described in detail in the Introduction section.
Table 2: The Ellsberg example

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>Y</th>
<th>B</th>
<th>P_G</th>
<th>E[P_G]</th>
<th>E[x]</th>
<th>Var[x]</th>
<th>( \aleph )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
<td>3</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>(B)</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>( \frac{2}{3}, \frac{1}{3}, \frac{1}{3}, \ldots, \frac{40}{90}, \frac{50}{90} )</td>
<td>( \frac{1}{3} )</td>
<td>3</td>
<td>18</td>
<td>0.3912</td>
</tr>
<tr>
<td>(RY)</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>( \frac{2}{3}, \frac{1}{3}, \frac{1}{3}, \ldots, \frac{40}{90}, \frac{50}{90} )</td>
<td>( \frac{2}{3} )</td>
<td>6</td>
<td>18</td>
<td>0.3912</td>
</tr>
<tr>
<td>(BY)</td>
<td>0</td>
<td>9</td>
<td>9</td>
<td>( \frac{2}{3} )</td>
<td>( \frac{2}{3} )</td>
<td>6</td>
<td>18</td>
<td>0</td>
</tr>
</tbody>
</table>

contains only 30 black and yellow balls with an unknown proposition in addition to the 30 red balls. The degree of ambiguity regarding the bet on the black ball decreases to \( \aleph[B] = 0.2981 \). If the amount of unknown black and yellow balls is 90, then the degree of ambiguity regarding the bet on the black ball increases to \( \aleph[B] = 0.4377 \). Finally, if the urn contains only 60 balls and all of them are black or yellow with an unknown proportion, then the level of ambiguity, which in this case is identical for the black and yellow balls, is \( \aleph = 0.5868 \). If there are only 10 balls (again with an unknown proportion) in the urn, then \( \aleph = 0.6324 \).

Our next step would be to show that indeed an ambiguity-averse DM prefers the prospects with the relatively low \( \aleph^2 \). For this purpose, we suggest the following definition:

**Definition 3.11.** A true mixed prospect \( f \in \mathcal{F} \) is more ambiguous than a prospect \( g \in \mathcal{F} \) if there exists a random variable \( \epsilon \in [0,1] \) such that

\[
P(f_j) - p(f_j) =_d P(g_j) - p(g_j) + \epsilon,
\]

for at least one outcome \( f_j \), where \( =_d \) means equal in distribution, \( p(f_j) = E[P(f_j)] \) and \( E[\epsilon \mid P(f_j)] = E[\epsilon] = 0 \).

In other words, Definition 3.11 states that the probabilities of a more ambiguous event fluctuate more, which implies a greater \( \epsilon \)-ambiguity. This definition coincides with our measure of ambiguity, \( \aleph^2 \). In particular, this implies that the event of loss satisfies

\[
P(L_f) - E[P(L_f)] =_d P(L_g) - E[P(L_g)] + \epsilon,
\]

which in turn implies that the variance of the probability of loss on \( f \) is higher than the variance of the probability of loss on \( g \). Thus, the level of ambiguity of \( f \) is at least as high as the ambiguity of \( g \), i.e., \( \aleph^2[g] \leq \aleph^2[f] \). The opposite is also true: by definition, a higher \( \aleph^2 \) means a higher variance of the probability of loss, and hence a stronger perturbation of probabilities as suggested by Definition 3.11.

The proof that ranking prospects by \( \aleph^2 \) coincides with a ranking made by an ambiguity-averse DM, considers symmetric prospects. Financial-economic theory usually concentrates on normal probability

\footnote{The condition \( E[\epsilon \mid P(f_j)] = E[\epsilon] \) means that \( \epsilon \) is mean-independent of \( P(f_j) \).}
distribution, (or quadratic utility function) to prove that risk can be measured by variance: see for example, Markowitz (1952)[75] and LeRoy and Werner (2001)[67]. Our proof relaxes the requirement of a normally distributed prospect in the sense that normal distributions are a subset of symmetric distributions.

**Definition 3.12.** A prospect \( f = (E_1 : x_1, \cdots, E_j : x_j, \cdots, E_n : x_n) \), is symmetric around the point of symmetry, \( x_s \), if \( x_{s-j} = x_{s+j} \) and \( P_i,s-j = P_i,s+j \), \( \forall j = 1, \ldots, s \) and \( \forall i = 1, \ldots, m \). We call such a prospect a **symmetric prospect**. If, however, only the first condition holds, we refer to it as an **outcome-symmetric**.

The next theorem ties our ambiguity measure to the DM’s preferences toward ambiguity and proves that ranking prospects according to the ambiguity measure, \( \mathbb{A}^2 \), is equivalent to the ranking made by an ambiguity-averse DM. Thus, in order to isolate the impact of ambiguity, we must compare two prospects that have identical properties, except ambiguity. Therefore, the comparison is between two prospects that have an identical set of outcomes and an identical expected probability for every outcome, such that the only difference between them is the dispersion of probabilities around their expected probabilities, event-wise. To eliminate the impact of preferences toward risk, we also assume a risk-neutral DM with a reference point \( x_k \leq x_s \). The reason for this assumption is that in symmetric prospects the point of symmetry equals the expected outcome and the reference point must be lower than the expectation; otherwise, the DM will not choose this prospect.

**Theorem 3.13.** Assume symmetric prospects \( f, g \in \mathbb{F} \) sharing the same set of outcomes, \( X \), and having the same expected probability for each outcome, i.e., \( E[P(f_j)] = E[P(g_j)], \forall j = 1, \ldots, n \).

Prospect \( f \) is more ambiguous than prospect \( g \), i.e., \( \mathbb{A}^2[f] \leq \mathbb{A}^2[g] \), if and only if any ambiguity-averse DM, with a reference value \( x_k \), prefers \( g \) to \( f \); that is, if she prefers the prospect with the lower \( \mathbb{A}^2 \) over the prospect with the higher \( \mathbb{A}^2 \).

Theorem 3.13 proves that ranking a prospect according to the measure of ambiguity, \( \mathbb{A}^2 \), is identical to the ranking of a DM who exhibits aversion toward ambiguity. Surprisingly, ranking prospects by \( \mathbb{A}^2 \) also coincides with the order of a DM who is ambiguity-averse to gains, but an ambiguity-lover as regards losses. This is true when the DM’s reference point is below the distribution’s point of symmetry. This is proved in the following theorem.

**Theorem 3.14.** Assuming symmetric prospects, \( f, g \in \mathbb{F} \), sharing the same set of outcomes, \( X \), and having the same expected probability of each outcome, i.e., \( E[P(f_j)] = E[P(g_j)], \forall j = 1, \ldots, n \). Let the prospect \( f \) be more ambiguous than prospect \( g \), i.e., \( \mathbb{A}^2[g] \leq \mathbb{A}^2[f] \). A DM with a reference value of \( x_k \leq x_s \), who is ambiguity-averse as regards gains but an ambiguity-lover as regards losses, prefers the less ambiguous prospect over the prospect with the relatively high ambiguity, \( g \succ f \).

\[\text{40Here, } f \text{ and } g \text{ are referred to as random variables; thus, the index } j \text{ designates outcome rather than event.}\]
3.3 Attributes of $\aleph^2$

Interesting points arise about the nature of ambiguity, especially regarding the relationship between $\aleph^2$ and variance. This subsection studies the properties of ambiguity, paying special attention to the normal distribution, yet with unknown random parameters. We start with some technical properties and then proceed to discuss the implications of variance for ambiguity.

The measure of ambiguity, $\aleph^2$, is invariant to monotone changes of outcomes, while adjusting the reference point accordingly. A monotonic chance of the outcome of each event does not affect its probability; thus, the prospect’s level of ambiguity remains unchanged. The next lemma proves this property.

**Lemma 3.15.** Applying a monotonic increasing transformation to a prospect, with respect to its consequences, while adjusting the reference value accordingly, leaves the ambiguity level unchanged, that is

$$\aleph^2 [T(f)] = \aleph^2 [f],$$

for any monotonic function $T(\cdot)$.

Assume that a monotonic transformation is applied to a prospect, while the reference value, $x_k$, is not adjusted accordingly. In this case, the difference between the ambiguity of $f$ and the ambiguity of $T(f)$ cannot be determined without additional information about the nature of the probability measure, defined by its density function. In general, $\aleph^2 [T(f)]$ can be higher, lower or equal to $\aleph^2 [f]$.

We may sometimes assume that the type of probability distribution of events in the subordinated space, $S$, is known; however, the parameters governing this distribution are unknown and dictated by the directing space, $\Omega$. We use the following notational conventions to designate these random parameters. Greek letters denote probability-moments, conditional upon a directing event $\varepsilon_i \in \Sigma$. The conditional mean and variance of a prospect $f_{|\varepsilon_i} = X_i$ are defined by

$$\mu_{X_i} = \sum_{j=1}^{n} P_i(x_j) x_j,$$

and

$$\sigma^2_{X_i} = \sum_{j=1}^{n} P_i(x_j) (x_j - \mu_{X_i})^2,$$

respectively. The conditional covariance between two prospects $f_{|\varepsilon_i} = X_i$ and $g_{|\varepsilon_i} = Y_i$ is defined by

$$\sigma_{X_i, Y_i} = \sum_{j=1}^{n} P_i(x_j \cap y_j) (x_j - \mu_{X_i})(y_j - \mu_{Y_i}).$$

The conditional correlation coefficient, $\rho_{X_i, Y_i}$, is thus $\rho_{X_i, Y_i} = \frac{\sigma_{X_i, Y_i}}{\sigma_{X_i} \sigma_{Y_i}}$. When the context is clear, to save notations, the index $i$, designating the directing event, is omitted.

The mean of a prospect, denoted $E[f] = E[X]$, is defined by

$$E[X] = \sum_{j=1}^{n} E[P_i(x_j)] x_j = \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} P_{i,j} x_j = \sum_{i=1}^{m} \frac{\sum_{j=1}^{n} \chi_i x_j}{\sum_{i=1}^{m} \chi_i} \mu_{X_i},$$

24
while its variance is defined by

\[ \text{Var}[X] = \sum_{j=1}^{n} \mathbb{E}[P_j(x_j)] (x_j - \mathbb{E}[X])^2 = \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{\chi_i}{\sum_{j=1}^{m} \chi_i} P_{i,j} (x_j - \mathbb{E}[X])^2. \]

Similarly, the covariance between two prospects \( f = X \) and \( g = Y \) is defined by

\[ \text{Cov}[X,Y] = \sum_{j=1}^{n} \mathbb{E}[P_j(x_j \cap y_j)] (x_j - \mathbb{E}[X]) (y_j - \mathbb{E}[Y]) = \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{\chi_i}{\sum_{j=1}^{m} \chi_i} P_{i,j} (x_j - \mathbb{E}[X]) (y_j - \mathbb{E}[Y]), \]

and the correlation coefficient, \( \text{Cor}[X,Y] \), is defined by

\[ \text{Cor}[X,Y] = \frac{\text{Cov}[X,Y]}{\text{std}[X] \text{std}[Y]}. \]

A prospect is said to be normally distributed if all possible probability distributions of events in the subordinated space, \( S \), are normal, whereas the parameters governing the distribution are random and dominated by events in the directing space, \( \Omega \). As regards the matter of completion, and as a building block for future usage in our asset pricing model, we also prove that ambiguity is invariant to increasing affine transformation in the case of prospects that have an infinite support (normally distributed).

**Lemma 3.16.** Assume a normally distributed prospect. Applying an affine transformation \( c + hf \), where \( h > 0 \), to prospect \( f \in \mathcal{F} \), while the reference value, \( x_k \), is adjusted accordingly, i.e., \( c + hx_k \), leaves the level of ambiguity unchanged.\(^{41}\) That is, the ambiguity measure satisfies the following relation

\[ \mathcal{A}^2[c + hf] = \mathcal{A}^2[f] = 4 \text{Var} \left[ \int_{-\infty}^{x_k} \frac{1}{\sqrt{2\pi \sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \, dx \right] \]

\( \forall \, c, h \in \mathbb{R}. \)

It is important to note that if a prospect is limited to having only uniform distributions then, by construction, the set of directing consequences, \( \mathcal{P} \), consists of a single probability measure. Hence, its level of ambiguity is always zero, regardless of the reference point.

We now turn to study the interaction between ambiguity and variance. This interesting point has significant implications for a wide range of financial issues. In general, if the variance of all possible probability measures, \( \mathbb{P}(\cdot) \), increases, then all probability mass functions, in a finite support framework, and all probability density functions, in an infinite support framework, become flatter. A flatter density function infers a lower variance of probabilities and thus, conditional upon the reference point, in most cases a lower ambiguity. One can take, for example, the normal distribution, in which when increasing the variance to infinity the probability distribution tends to the uniform distribution, which in turn implies a zero degree of ambiguity. This important insight indicates a negative relation between risk, measured by variance, and ambiguity, measured by \( \mathcal{A}^2 \). The implications of this result for finance are enormous, mainly concerning portfolio selection. It infers that diversifying ambiguity by a

\(^{41}\)Explicitly, the notation \( c + hf \) applies to the prospect’s consequences. In the case of a finite support, for example, \( c + hf = (\xi_1 : c + hx_1, \ldots, \xi_j : c + hx_j, \ldots, \xi_n : c + hx_n) \).
portfolio of prospects (assets) in most cases is not feasible. Subsection 4 elaborates on the implication of ambiguity for diversification.

The relation between the ambiguity of a prospect and its variance is not trivial and contingent on the DM’s reference point. Increasing the outcomes of a prospect by applying a monotonic transformation and adjusting the reference point accordingly, Lemma 3.15 shows that the level of ambiguity remains unchanged, whereas the variance increases. The same is true when decreasing the prospect’s outcomes, yet preserving its outcomes’ ranking: ambiguity remains unchanged, but the variance decreases. The following theorem defines the conditions under which an increasing variance leads to a lower ambiguity for the case of normally distributed prospects.

**Theorem 3.17.** Assume a normally distributed prospect and a reference point, $x_k$, satisfying

$$
\mu_i - \sigma_i \leq x_k \leq \mu_i + \sigma_i,
$$

\( \forall i = 1, \ldots, m, \) then a higher variance implies a lower ambiguity.

It is reasonable to assume that the DM holds a reference point that is around the expected outcome of the prospect and is always lower than this value. Considering the return on the risky asset, for example, we may expect that the reference point will be lower than the expected return, but higher than the return on a risk-free asset. In this case, Theorem 3.17 proves that a uniform increment of the variance across all possible probability measures has a negative effect on the level of ambiguity.\[42\] Considering normally distributed prospects, the ambiguity measure can be established using the standard normal probability distribution and takes the form

$$
\mathcal{K}^2[f] = 4 \text{Var} \left[ \int_{-\infty}^{x_k} \frac{1}{\sqrt{2\pi \sigma_X^2}} e^{-\frac{(x-x^*)^2}{2\sigma_X^2}} dx \right] = 4 \text{Var} \left[ \Phi \left( \frac{x_k - \mu_X}{\sigma_X} \right) \right],
$$

(9)

where $\Phi(\cdot)$ stands for the standard normal cumulative probability distribution. One can see that since $\Phi(\cdot)$ is a monotonic increasing function, when the reference point, $x_k$, is bounded to the interval $[\max(\mu_i - \sigma_i), \min(\mu_i + \sigma_i)]$, a higher variance implies less centralized probabilities of loss; the variance of the probably of loss is then low, which in turn implies a lower level of ambiguity.

Let us now consider the case where the reference point lies in the interval $[\max(\mu_i - \sigma_i), \min(\mu_i + \sigma_i)]$ and the variance decreases. Decreasing the variance makes the probability density function steeper; the variance of probability is higher and the ambiguity is also higher. However, at some point, decreasing variance, while keeping all other parameters unchanged, causes the reference point to deviate from the interval $[\max(\mu_i - \sigma_i), \min(\mu_i + \sigma_i)]$. Then, decreasing variance results in a lower ambiguity level. When the variance tends to zero, the ambiguity level also tends zero, unless the select reference point is equal to the expected outcome under all possible priors, i.e., all probability distributions have an identical mean.

\[42\] A further elaboration on this issue is presented in Section 4, which analyzes the joint ambiguity of prospects.
Figure 2 demonstrates the effect of variance on the level of ambiguity. In this figure, the prospect is normally distributed with a known mean $\mu_X = 0.1$, but with random standard deviation that can take one of the two possible values, $\sigma_{X_1} = 0.05$ or $\sigma_{X_2} = 0.1$, with equal likelihoods. The selected reference point is $x_k = 0.09$. The x-axes describes the increment factor of the variance, varying from 0 to 2. One can see that an increasing variance results in a higher ambiguity only for a narrow range of factors, between 0 and 0.2. For this range of factors, the variances are 0.01 and 0.02, while the reference point is 0.09; thus, Condition (8) is violated. For any factor larger than 0.2, Condition (8) is satisfied; therefore, a higher variance implies a lower ambiguity.

![Figure 2: Ambiguity and variance](image)

If all distributions are symmetric and they all share the same point of symmetry, $x_s$, which is equal to the reference point, $x_k$, then $P_{i,L} = P_{i,G} = \frac{1}{2}$, $\forall i = 1, \ldots, m$, which implies a zero level of ambiguity, i.e., $\aleph^2 = 0$. Consider a normally distributed prospect, for example, if all possible probability measures have an identical mean equal to the reference point, then their variances play no role as regards ambiguity which equals zero. If, however, the reference point is not equal to the common mean, then a strictly positive ambiguity level, dominated by the possible variances, is observed.

Concerning normally distributed prospects, Figures 3 and 4 give a diagrammatic representation of the degree of ambiguity. Figure 3 describes the standard normal probability density function, while Figure 4 describes its associated cumulative probability distribution. The values of x-axes are the upper integration limit $\frac{x_k - \mu_X}{\sigma_X}$ of the probability of loss (Equation (9)). These values differentiate between losses and gains. The y-axes in Figure 3 describes the amplitude of the probability density function. The y-axes in Figure 4 describes the cumulative probability of loss. The bright shadow region depicts the range of the integration boundary, $\frac{x_k - \mu_X}{\sigma_X}$, when the variance is relatively low, while the dark region depicts the range of the integration boundary when the variance is relatively high. In Figure 4, one can observe that decreasing the variance causes a shift to the left of the region of the possible probabilities of loss to the range of a flatter cumulative probability over which the variance is computed to obtain the ambiguity measure. However, the greater spread of the integration boundary $\frac{x_k - \mu_X}{\sigma_X}$, when the variance is relatively low, has a stronger impact, which results in a higher ambiguity.
When the integration boundary \( \frac{x_k - \mu X}{\sigma_X} \) is higher than 1 or lower than \(-1\), decreasing variance shifts the range to the left, where the cumulative probability distribution is flat enough, such that it decreases the ambiguity.

Figure 3: Ambiguity over a normally distributed prospect: A density function representation of the standard normal distribution.

Figure 4: Ambiguity over a normally distributed prospect: A cumulative representation of the standard normal distribution.

Figure 5 provides an additional way to observe the impact of variance on the level of ambiguity. The curves describe normal probability density functions that have identical means, but different random variances. The shadow area around each curve describes the possible probability density functions, resulting in a fluctuation of the variance. The bright shadow region depicts the range of the density functions when the variance is relatively low, while the dark shadow region describes the case
of relatively high variance. Figure 5 shows that if the reference point is within one standard deviation from the mean, then a lower variance, designated by the bright area, implies a sharper curve compared to the dark area with the higher variance. Clearly, sharper density functions have a positive effect on the variance of probabilities and, thus, on the level of ambiguity.

![Figure 5: Ambiguity over normally distributed prospect with known mean and unknown (random) variance.](image)

The impact of changes in regard to the distributions’ mean can also be studied from Figure 5. Let us now assume that the mean is also random around the reference point. Changing the possible means while keeping the reference point, \( x_k \), unchanged, such that the means become closer to the reference point, increases the ambiguity level. The maximal level of ambiguity is obtained when the reference point is located as close as possible to the means. This insight can be explained by the following intuition. Recall that the reference point is subjectively selected by the DM. When the reference point is close to the random means, either through the DM’s preliminary selection or because the circumstances were changed, she is considered more sensitive to changes in the probability of loss, such that her estimated level of ambiguity is now higher.

Considering an empirical examination of \( \mathbb{N}^2 \), while assuming that outcomes are normally distributed and the reference point is close to the possible means, the variance of the expression \( \frac{2x - \mu_X}{\sigma_X} \) can serve as a proxy for the variance of probability of loss.\(^{43}\) This expression is the upper integration boundary of the standard normal cumulative distribution function, \( \Phi (\cdot) \), which is used to estimate the expected probability of loss (Equation (9)). Since near its mean the normal cumulative distribution function is almost linear, taking the variance of the integration boundaries is a reasonable approximation to the variance of cumulative probabilities. This approximation significantly simplifies computing the level of ambiguity from the data.

To limit the length of this paper, we leave the study of the relation between variance and ambiguity over other types of probability measures for future research.

\(^{43}\)See Brenner and Izhakian (2011)[11].
4 Joint ambiguity and diversification

Can ambiguity be diversified in a similar manner to risk diversification by an asset portfolio? In other words, are two assets, when combined, less ambiguous than each asset taken separately? Now that we have a well-defined ambiguity measure, this subsection addresses this critical question and studies other characteristics of the ambiguity of prospect portfolios. For simplicity’s sake, most of the results in this subsection are demonstrated for a portfolio that consists of two prospects. However, all the presented results can be easily extended to portfolios that consist of any number of prospects. Let us begin with a very simple example that practically demonstrates the main theoretical results of this subsection.

Example 4.1. Assume two states of nature and two prospects: A and B. The outcomes of prospects A and B are $X_A = (-2, 1, 1, 2)$ and $X_B = (2, 1, 1, -2)$, respectively. Their possible probabilities are

$$
\begin{bmatrix}
P_1 \\
P_2
\end{bmatrix} =
\begin{bmatrix}
0.1 & 0.4 & 0.4 & 0.1 \\
0.2 & 0.3 & 0.3 & 0.2
\end{bmatrix},
$$

where a row, $i = 1, 2$, indicates a probability measure and a column, $j = 1, \ldots, 4$, indicates a state of nature. The probability measures are assigned with equal (second-order) capacities, i.e., $\chi = (0.5, 0.5)$, and the reference point is $x_k = 0.5$. The expected probabilities of the subordinated events are, thus,

$$p = \chi_1 P_1 + \chi_2 P_2 = (0.15, 0.35, 0.35, 0.15).$$

Using $p$, one can easily calculate the means $E[A] = 0.7$ and $E[B] = 0.7$, of prospects A and B, respectively. The variances of these two prospects are also identical: $\text{Var}[A] = \text{Var}[B] = 1.41$. Since the reference point is $x_k = 0.5$, prospect A faces a loss only in state $j = 1$ and prospect B faces a loss only in state $j = 4$:

$$X_A = \begin{bmatrix}
-2 & 1 & 1 & 2 \\
L & G & | & L
\end{bmatrix}$$

$$X_B = \begin{bmatrix}
2 & 1 & 1 & -2 \\
G & | & L
\end{bmatrix}$$

Concerning prospect A, conditional upon the directing events $i = [1, 2]$, its probabilities of loss are $P_{A,L} = (0.1, 0.2)$, and the same is true for prospect B, $P_{B,L} = (0.1, 0.2)$. Thus, the variance of the probability of loss is $\text{Var}[P_{A,L}] = \text{Var}[P_{B,L}] = 0.0025$, which in turn implies an ambiguity level of $\mathbb{A}[A] = \mathbb{A}[B] = 0.1$.

Now let us assume an equally weighted portfolio consisting of prospects A and B, denoted $h$. Its random outcomes can be

$$X_h = \begin{bmatrix}
0 & 1 & 0 \\
L & G & L
\end{bmatrix}.$$
The expected outcome of this portfolio is $E[X_h] = 0.7$ and its variance is $\text{Var}[X_h] = 0.21$. Portfolio $h$ faces a loss in states $j = 1, 4$; therefore, the probabilities of loss are $P_{h,L} = (0.2, 0.4)$, which implies $\aleph[h] = 0.2$. Portfolio $h$ maintains the same expected outcome as prospects $A$ and $B$, i.e., $E[h] = E[A] = E[B] = 0.7$, but with a lower variance, i.e., $\text{Var}[h] < \text{Var}[A] = \text{Var}[B]$. The level of ambiguity, however, is doubled $\aleph[h] = 2\aleph[A] = 2\aleph[B] = 0.2$, which demonstrates that while risk can be diversified, in general, ambiguity cannot.

This example demonstrates that while risk can be reduced by a diversified prospect portfolio, generally this is not true for ambiguity. Example 4.1 illustrates the nature of ambiguity and the possibility of diversifying it. In most cases, when ambiguity is present, adding assets to a portfolio decreases its variance; thus, increases its ambiguity (see Theorem 3.13). The rest of this subsection proves this notion analytically.

Let us assume a portfolio $h$ consists of $t$ prospects $f_1, \ldots, f_t \in \mathcal{F}$, labeled $l = 1, \ldots, t$. All the prospects share the same subordinated space, $S$, and the same directing space, $\Omega$. Since consequences of events are different across prospects and the order is determined by consequences, we refer to any prospect $l$ simply as a random variable

$$X_l = (x_{l,1}, \ldots, x_{l,j}, \ldots, x_{l,n}),$$

where the index $j$ designates the $j$-ordered outcome. It is important to note that using this representation, the probabilities of outcomes across prospects can possibly be correlated. For example, the probabilities of two outcomes of two different prospects, which occur for the same event, are correlated. We use the following notational convention to designate a prospect portfolio

$$f_h = h_1f_1 + \cdots + h_t f_t,$$

where $f_1, \ldots, f_t \in \mathcal{F}$, and $h_l$ is a nonnegative proportion of prospect $l$ in the portfolio. The aggregate consequences, $Z$, of $h$ are ranked with respect to an aggregate reference point, $z_k$.

Since prospects can be represented as random variables, the probability of gain and the probability of loss on a prospect portfolio can be represented by a convolution of probabilities. The ambiguity of a portfolio, $h = (h_1, h_2)$, consisting of two prospects, $X, Y \in \mathcal{F}$, with a finite support is, thus, defined by the following convolution

$$\aleph^2[hX + h_2 Y] = 4\text{Var} \left[ \sum_{j=1}^{k} \sum_{l=1}^{n} P_1(\lfloor h_2 y_l \rfloor) P_2(\lfloor h_1 x_j - h_2 y_l \rfloor) \right],$$

where $P_1$ and $P_2$ are the (random) probability measures associated with the outcomes of prospects $X$ and $Y$, respectively. The reference point $x_k$ is adjusted to $\lfloor h_1 x_k \rfloor$, where $\lfloor \cdot \rfloor$ stands for the lower

---

44If prospects are related to two different subordinated spaces with two sets of events $\Xi_1$ and $\Xi_2$, we can construct a new set of events $\Xi$ containing all the events in $\Xi_1 \times \Xi_2$. The same principle can be applied to the directing space; in the case of two different directing spaces we can construct a directing space $\Sigma = \Sigma_1 \times \Sigma_2$.
Considering prospects taking the real numbers, \( \mathbb{R} \), as a support, the ambiguity of a portfolio consisting of two prospects is

\[
\mathbb{N}^2 [h_1 X + h_2 Y] = 2 \text{Var} \left[ \int_{-\infty}^{z_k} \int_{-\infty}^{\infty} P(x,y) \, dy \, dx \right],
\]

where \( z_k = h_1 k_1 + h_2 k_2 \), and \( P(x,y) \) is the joint density of the random variables \( X \) and \( Y \). If the probabilities of subordinated events are not independent, then we cannot extract the expression further without additional assumptions about the probability distribution. If, however, the random probabilities of the two random variables are independent, then the ambiguity measure obtains the form

\[
\mathbb{N}^2 [h_1 X + h_2 Y] = 2 \text{Var} \left[ \int_{-\infty}^{z_k} \int_{-\infty}^{\infty} P_1(h_2 y) P_2(h_1 x - h_2 y) \, dy \, dx \right].
\]

When the type of probability distribution is known, although the parameters governing it are unknown, we can typify the level of ambiguity. Considering, for example, normally distributed random variables, then a closed form solution of the level of ambiguity can be achieved as the following proposition suggests.

**Proposition 4.2.** Assume two normally distributed prospects, \( X, Y \in \mathcal{F} \), with random means, \( \mu_X \) and \( \mu_Y \), and random standard deviations, \( \sigma_X \) and \( \sigma_Y \), respectively. The ambiguity of the compounded prospect \( Z = h_1 X + h_2 Y \) is then

\[
\mathbb{N}^2 [Z] = 4 \text{Var} \left[ \int_{-\infty}^{z_k} \frac{1}{\sqrt{2\pi \sigma_Z^2}} e^{-\frac{(x-\mu_Z)^2}{2\sigma_Z^2}} \, dx \right],
\]

where \( z_k = h_1 x_k + h_2 y_k \), \( \mu_Z = h_1 \mu_Z + h_2 \mu_Y \) and \( \sigma_Z^2 = h_1^2 \sigma_X^2 + h_2^2 \sigma_Y^2 + 2 h_1 h_2 \rho_X Y \sigma_X \sigma_Y \).

Next, using Proposition 4.2, we prove that an increasing correlation between assets reduces the level of ambiguity. In some sense, this result is a bit counterintuitive, since we are tempted to wrongly conclude that, like risk which increases with the correlation between assets, so too does ambiguity. However, this conclusion is incorrect, due to the fact that when the correlation increases, the variance of the portfolio also increases and (as we prove in Theorem 4.3) a higher variance usually implies lower ambiguity.

**Theorem 4.3.** Assume a portfolio \( h \) consists of \( t \) normally distributed prospects \( X_1, \ldots, X_t \in \mathcal{F} \), whose random outcome \( Z = \sum_t h_t X_t \) is characterized by the random mean \( \mu_Z = \sum_t h_t \mu_{X_t} \) and the random variance \( \sigma_Z^2 = \sum_t h_t^2 \sigma_{X_t}^2 + \sum_{t \neq i} h_t h_d \rho_{X_t X_d} \sigma_{X_t} \sigma_{X_d} \). If the reference point, \( z_k = \sum_t h_t x_{1,k} \),

45 The lower bound over the reference point is applied to the index of the outcome. In other words, the new reference point is the index of the maximal outcome \( x_j \); yet satisfies \( x_k \leq h_1 x_j \).

46 Recall that \( \mu_Z \) and \( \sigma_Z \) are random variables subordinated by the directing event \( \varepsilon_1, \ldots, \varepsilon_t, \ldots, \varepsilon_m \).
satisfies
\[ \mu_{Z,i} - \sigma_{Z,i} \leq z_k \leq \mu_{Z,i} + \sigma_{Z,i}, \]

(10)

\( \forall i = 1, \ldots, m \), then an increasing correlation between any pair of prospects in portfolio \( h \) decreases its degree of ambiguity.

Theorem 4.3 proves that when the reference point is in the range of less than one standard deviation from the possible means, then the correlation between assets has a negative effect on the level of ambiguity. The variance of the portfolio, however, is positively affected by the correlation. The opposite, a decreasing correlation, has a positive impact on ambiguity. This is true as long as the reference point, \( z_k \), satisfies \( \max (\mu_{Z,i} - \sigma_{Z,i}) \leq z_k \leq \min (\mu_{Z,i} + \sigma_{Z,i}) \). Decreasing the correlation to a low enough level, while keeping the reference point unchanged, causes the reference point to violate Condition (10); that is, the reference point deviates from the required range \( (\mu_{Z,i} - \sigma_{Z,i}) \leq z_k \leq (\mu_{Z,i} + \sigma_{Z,i}) \). In that case, a decreasing correlation has a negative impact on ambiguity. However, since while finalizing her reference point the DM takes into account the variance of the prospect portfolio, the reference point is assumed to be adjusted to the level of variance (risk). Hence, while the correlation changes, the reference point is dynamically adjusted accordingly.

Theorem 4.3 infers that since the maximal correlation is obtained when \( \rho_{X_l,X_d} = 1, \forall i,d \), any portfolio with non-perfectly-positive correlated prospects is more ambiguous than a single prospect. The next theorem studies this property.

**Theorem 4.4.** Assume that the conditions of Theorem 4.3 hold and \( \frac{\mu_{X_l,i}}{\sigma_{X_l,i}} = \frac{\mu_{X_d,i}}{\sigma_{X_d,i}} \), \( \forall i = 1, \ldots, m \) and \( \forall l,d = 1, \ldots, t \). Then, ambiguity is not diversifiable. The ambiguity of a prospect portfolio is higher than the accumulated ambiguity of all prospects included in the portfolio, when considered separately. That is,

\[ \sum_{l=1}^{t} h_l [X_1 + \cdots + X_l] \geq h_1[X_1] + \cdots + h_t[X_t], \]

where \( \sum_{l=1}^{t} h_l = 1 \) and \( 0 \leq h_t, \forall t = 1, \ldots, t \).

Even if we relax the assumptions of Theorem 4.4 to allow for a variability of \( \frac{\mu_X}{\sigma_X} \) across prospects conditional upon a directing event, in many cases ambiguity still cannot be diversified. Using numerical simulations, one can see that the ability to diversify ambiguity is a matter of the relations between the prospects’ random means and variances across prospects and across directing events. For example, Figure 6 simulates the ambiguity of a two-prospect portfolio as a function of the proportion of prospects in the portfolio. It assumes two identical, non-correlated, normally distributed prospects with a known mean, \( \mu_X = 0.1 \), and a random standard deviation that can take one of the two possible values \( \sigma_{X_1} = 0.05 \) or \( \sigma_{X_2} = 0.1 \), with equal likelihoods. The selected reference point is \( x_k = 0.09 \). The values between 0 to 1 on the x-axes are the positive proportion of the first asset in the portfolio. The y-axes shows the level of ambiguity and standard deviation of the portfolio. The solid curve depicts the level of ambiguity, while the dashed curve depicts the expected standard deviation.
Let us now assume an equally weighted portfolio, consisting of perfectly uncorrelated and normally distributed prospects, all having identical means and identical variances. Increasing the number of prospects in the portfolio decreases its variance. As the number of prospects tends to infinity, the variance tends to zero. In this case, either the reference point deviates by more than one standard deviation from the mean or it equals the mean. Both cases imply a perfectly diversified ambiguity, i.e., $\mathbb{R}^2 = 0$.

The results of this section have considerable implications as regards common financial theory. From a paretoical point of view, the existing financial literature asserts that investors should aim to minimize the variance by holding a fully risk-diversified portfolio. Our results, however, imply that this is usually not true for ambiguity. Increasing the number of assets in the portfolio decreases its variance, but at the same time it usually increases its level of ambiguity, such that holding a fully diversified portfolio is not necessarily optimal for an ambiguity-averse DM. It has been documented that individuals tend to hold under-diversified portfolios (3-4 stocks); see for example Barber and Odean (2000) and Goetzmann and Kumar (2008). This phenomena can possibly be explained by the nature of ambiguity. Our results also coincide with the behavioral findings of Bossaerts et al. (2010) who demonstrate that in the presence of ambiguity, ambiguity-averse investors are reluctant to hold ambiguous assets.

The relation between ambiguity and under-diversification has also been discussed by Uppal and Wang (2003). Using a calibration, they demonstrate that small differences in the ambiguity of the distribution of marginal returns will result in a portfolio that is significantly under-diversified, relative to the standard mean-variance portfolio. Our model provides the theoretical explanation for these results. "Own-company stock", "home-country equity" and "limited stock market participation" are other financial puzzles upon which STP might help to shed new light.
5 Asset pricing

Classical asset pricing theory assumes that the probabilities of outcomes are known and agreed upon by all the investors (DMs). This section relaxes this assumption and instead assumes that probability measures are random, and therefore cannot be known to the investors. The main goal of this section is to investigate the investors’ saving behavior with respect to pricing and allocations while facing ambiguity. Special attention is given to characterizing the uncertainty premium and differentiating it into two separate premiums of assets (prospects): risk premium and ambiguity premium, resulting from risk aversion and ambiguity aversion, respectively.

For simplicity’s sake, it is henceforth assumed that the probability weighting functions are linear of the form \( w^- (Q) = Q \) and \( w^+ (Q) = Q \), and that capacities over the directing space are uniformly distributed.\(^{47}\) Recall that the expected probabilities, \( p \), are additive, but the probability premiums, \( \varphi \), are not; the value of a prospect \( f \in \mathcal{F} \) is therefore

\[
V (f) = \sum_{j=1}^{k} (p_j - (\varphi_{1...J} - \varphi_{1...J-1})) U (x_j) + \sum_{j=k+1}^{n} (p_j - (\varphi_{j...N} - \varphi_{J+1...N})) U (x_j).
\]

Reasonably, we assume that the reference point \( x_k \) is relatively close to zero and satisfies \( x_k \leq E [x] \).

Considering financial decisions, outcomes, \( X \), can possibly be the rate of returns on an asset. That is, state’s \( j = 1, \ldots, n \) outcomes are \( x_j = 1 + r_j \), where \( r \) designates net return. Under these settings, we assume that the risk-free rate, denoted \( r_f \), is the objective reference point agreed upon by all investors.

5.1 The ambiguity premium

The risk premium is the premium that a DM is willing to pay for replacing a risky bet with its expected outcome. The ambiguity premium is the premium that a DM is willing to pay for replacing an ambiguous bet with a risky bet with an identical expected outcome. The uncertainty premium is the total premium that a DM is willing to pay for replacing an ambiguous bet with its expected outcome, i.e., the accumulation of risk premium and ambiguity premium. Formally, the uncertainty premium, denoted \( K \), is defined by

\[
U (E [x] - K) = \sum_{j=1}^{k} [p_j - (\varphi_{1...J} - \varphi_{1...J-1})] U (x_j) + \sum_{j=k+1}^{n} [p_j - (\varphi_{j...N} - \varphi_{J+1...N})] U (x_j). \tag{11}
\]

The certainty equivalent, \( CE = E [x] - K \), of a prospect \( f \in \mathcal{F} \) is a constant prospect \( CE \in \mathcal{F} \) satisfying \( V (CE) = V (f) \). In other words, the certainty equivalent is the constant prospect the DM is willing to exchange a risky-ambiguous (uncertain) prospect for. The next theorem approximates the uncertainty premium and differentiates it into risk premium and ambiguity premium.

\(^{47}\)Assuming that the random parameters governing first-order distribution are continuously distributed over a close interval, i.e., continuous second-order capacities, and that investors have no additional information regarding these second-order capacities, then this assumption is supported by maximum entropy considerations.
Theorem 5.1. Assume a DM whose preferences are characterized by a twice differentiable utility function, \( U(\cdot) \), and a twice differentiable p-utility function, \( \psi(\cdot) \). The uncertainty premium is then assessed by

\[
K = -\frac{1}{2} \frac{U''(E[x])}{U'(E[x])} \operatorname{Var}[x] - \frac{1}{8} \left[ \frac{\psi''(E[P_L])}{\psi'(E[P_L])} + \frac{\psi''(E[P_G])}{\psi'(E[P_G])} \right] E[x] \mathbb{N}^2[x],
\]

where \( R \) is the risk premium and \( A \) is the ambiguity premium.\(^{48}\)

Theorem 5.1 provides a complete differentiation of the two aspects. First, a distinction between risk premium and ambiguity premium is obtained. Second, within each premium, a distinction between the sources of premiums, i.e., preference and beliefs, is achieved. The obtained risk premium,

\[
R = -\frac{1}{2} \frac{U''(E[x])}{U'(E[x])} \operatorname{Var}[x],
\]

is exactly the Arrow-Pratt risk premium. Higher risk, measured by \( \operatorname{Var}[x] \), or higher risk aversion, measured by the Arrow-Pratt coincident of absolute risk aversion, \( \frac{U''(\cdot)}{U'(\cdot)} \), both result in a higher risk premium. The level of risk, \( \operatorname{Var}[x] \), is a matter of the DM’s beliefs, while risk aversion is a matter of her preferences toward it.\(^{49}\) Independently, each one of the components has a positive impact on the risk premium.

Recall that the ambiguity premium is the premium that a DM is willing to pay, in terms of return, for exchanging a risky-ambiguous prospect with a risky, but non-ambiguous one. The ambiguity premium,

\[
A = -\frac{1}{8} \left[ \frac{\psi''(E[P_L])}{\psi'(E[P_L])} + \frac{\psi''(E[P_G])}{\psi'(E[P_G])} \right] E[x] \mathbb{N}^2[x],
\]

possesses attributes resembling those of risk premium, but with respect to probabilities rather than to consequences. A separation between beliefs about probabilities, measured by \( \mathbb{N}^2 \), and preferences toward it, measured by the coefficient of absolute ambiguity aversion, \( \frac{\psi''(\cdot)}{\psi'(\cdot)} \), is obtained. A DM who exhibits ambiguity aversion has a concave p-utility function and thus \( \frac{\psi''(\cdot)}{\psi'(\cdot)} > 0 \), which implies a positive ambiguity premium. If, however, the DM exhibits ambiguity-seeking behavior, then the ambiguity premium is negative, since \( \frac{\psi''(\cdot)}{\psi'(\cdot)} < 0 \). Clearly, when the DM is ambiguity-neutral, \( \frac{\psi''(\cdot)}{\psi'(\cdot)} = 0 \), which implies a zero ambiguity premium. A zero ambiguity premium is also obtained when the probabilities of all events are perfectly known, i.e., \( \mathbb{N}^2 = 0 \). Higher ambiguity aversion implies a higher \( \frac{\psi''(\cdot)}{\psi'(\cdot)} \) and thus a higher ambiguity premium, \( A \). A higher level of ambiguity, \( \mathbb{N}^2 \), has the same effect

\(^{48}\)The proof of Theorem 5.1 applies the same method as that of Arrow (1965)\(^5\) and Pratt (1964)\(^{78}\), while we also restrict ourselves to relatively small consequences. Izhakian and Benninga (2011)\(^{54}\) and Nau (2006)\(^{77}\) also use the same method.

\(^{49}\)In fact, the variance of a lottery, which serves as a common measure for risk, doesn’t always coincides with risk level. In other words, a higher variance does not always indicate a riskier asset, and a risk-averse DM may possibly prefer a lottery with a higher variance over a lottery with a lower variance, but with an identical expected payoff (LeRoy and Werner (2001)\(^{67}\)).
on \( \mathcal{A} \).

Together, theorems 5.1 and 3.17 imply that when the volatility changes, the changes in risk premium and the changes in ambiguity premium may have opposite signs. This result coincides with those of Izhakian and Benninga (2010)[54] and Ui (2011)[99]; however, they explain it by changes in risk aversion preferences and by limited market participation, due to ambiguity preferences, respectively, while in our model this result emerges from the nature of ambiguity, independently of preferences.

Recall that the level of ambiguity, \( \mathbb{N}^2 \), is determined with respect to the DM’s reference point, which distinguishes between gains and losses. If the reference point is selected, such that all outcomes are considered gains or all outcomes are considered losses, i.e., true mixing does not hold, then obviously the ambiguity level is zero and so is the ambiguity premium.

As an example, the next corollary shows the different premiums in the case of a DM typified by constant relative risk aversion (CRRA) and constant absolute ambiguity aversion (CAAA).

**Corollary 5.2.** Assume a DM who is characterized by constant relative risk aversion (CRRA),

\[
U(x_j) = \begin{cases} \frac{x_j^{1-\gamma} - x_k^{1-\gamma}}{1-\gamma}, & \gamma \neq 0 \\ \ln(x_j) - \ln(x_k), & \gamma = 0 \end{cases},
\]

(12)

and constant absolute ambiguity aversion (CAAA)

\[
\psi(q_i) = -\frac{e^{-\eta q_i}}{\eta},
\]

then the uncertainty premium is approximately

\[
\mathcal{K} \approx \gamma \frac{1}{2} \frac{\text{Var}[x]}{\mathcal{R}} + \eta \frac{1}{4} \frac{E[x] \mathbb{N}^2[X]}{\mathcal{A}}.
\]

For the sake of completeness, the next corollary shows the different premiums in the case of a DM typified by CRRA and constant relative ambiguity aversion (CRAA).

**Corollary 5.3.** Assume a DM who is characterized by CRRA (Equation (12)) and constant relative ambiguity aversion (CRAA)

\[
\psi(q) = \frac{q^{1-\eta}}{1-\eta},
\]

then the uncertainty premium is approximately

\[
\mathcal{K} \approx \gamma \frac{1}{2} \frac{\text{Var}[x]}{\mathcal{R}} + \eta \frac{1}{8} \left( \frac{1}{E[P_L]} + \frac{1}{E[P_G]} \right) E[x] \mathbb{N}^2[X].
\]

\( ^{50} \) A more standard formulation of CRRA, \( U(x) = \frac{x^{1-\gamma}}{1-\gamma} \) for \( \gamma \neq 1 \) and \( U(x) = \ln(x) \) for \( \gamma = 1 \) otherwise, is not normalized to \( U(x_k) = 0 \) as required by Theorem 2.1.
The financial literature usually considers the returns on assets as the consequences. When the net return on asset, \( r \), is relatively small and the probability of loss and gain are close to \( \frac{1}{2} \), the approximated ambiguity premium can be simplified to
\[
\mathcal{A} \approx -\frac{1}{4} \frac{\psi''(E[P_L])}{\psi'(E[P_L])} N^2[r],
\]
and the expected return is thus
\[
E[r] \approx r_f - \frac{1}{2} \frac{U''(E[r])}{U'(E[r])} \text{Var}[r] - \frac{1}{4} \frac{\psi''(E[P_L])}{\psi'(E[P_L])} N^2[r],
\]
(13)
where \( r_f \) stands for the return on the risk-free asset. Brenner and Izhakian (2011)[11] test this equation empirically and show that ambiguity has a significant explanatory power for the return.

5.2 The pricing kernel and the discount factor

To better clarify the impact of ambiguity on the prices assigned to assets, we construct the state prices in an economy with a finite support. For this purpose, we assume a complete market with a risk- and ambiguity-averse pricing representative DM, who has a time and state separable utility function.

The DM's objective is to maximize her two periods' consumption, \( c_0 \) and \( c_1 \), given her initial wealth \( w_0 \). Define the state price of state of nature \( j = 1, \ldots, n \) by \( \Psi_j \), and the contingent claim on state \( j \) by \( x_j \). The DM's problem is then defined by
\[
\max_{x_1, \ldots, x_n} U(c_0) + \sum_{j=1}^n (p_j - \hat{\varphi}_j) U(c_j),
\]
subject to the consumption constraint at time \( t = 0 \):
\[
c_0 = w_0 - \sum_{j=1}^n x_j \Psi_j,
\]
and the consumption constraint of every state of nature \( j \) at time \( t = 1 \):
\[
c_j = x_j.
\]

Solving the DM’s maximization problem, in equilibrium the state’s \( j \leq k \) price is
\[
\Psi_j = (p_j - \hat{\varphi}_j) \frac{U'(x_j)}{U'(x_0)} = \left[ p_j - \left( -\frac{1}{2} \frac{\psi''(p_j)}{\psi'(p_j)} \xi_j + \frac{1}{2} \frac{\psi''(p_{j-1})}{\psi'(p_{j-1})} \xi_{j-1} \right) \right] \frac{U'(x_j)}{U'(x_0)},
\]
(14)
and the state’s \( k < j \) price is
\[
\Psi_j = (p_j - \hat{\varphi}_j) \frac{U'(x_j)}{U'(x_0)} = \left[ p_j - \left( -\frac{1}{2} \frac{\psi''(p_{j+1})}{\psi'(p_{j+1})} \xi_{j+1} + \frac{1}{2} \frac{\psi''(p_{j+1})}{\psi'(p_{j+1})} \xi_{j} \right) \right] \frac{U'(x_j)}{U'(x_0)}.
\]

The obtained state price, \( \Psi_j \), is a function of four factors. The first factor is the marginal rate of substitution (MRS), \( \frac{U'(x_j)}{U'(x_0)} \), which in turn is a function of risk aversion. As a consequence of risk aversion, modeled by a concave utility function, the price of a state of nature with relatively low

\[51\] A pricing representative DM is a single artificial DM in an economy with heterogeneous agents, where prices in equilibrium could be rationalized as if originating from the preferences of that artificial agent. However, we do not consider the question of aggregation of preferences.
outcome, $x_j$, is relatively high. The intuition for this result is clear: prices are high in states of nature with relatively low supply. The higher the risk aversion the lower the MRS, which results in a relatively low state price.

The second factor is the absolute ambiguity aversion, formulated by the coefficient $-\frac{\psi''(j)}{\psi'(j)}$. High ambiguity aversion implies a high probability premium, which in turn implies a low subjective probability, and therefore low prices of all states of nature faced with ambiguity. The third factor is the level of ambiguity of each state of nature $j$, measured by $\xi_j^2 - \xi_{j-1}^2$. Ambiguity negatively affects the state price; when the variance of the probability of state $j$ is high, i.e., a high e-ambiguity level, the price of this state is low. The expected probability of the specific state of nature, $p_j$, is the fourth factor. The price of the contingent claim on state $j$ is positively affected by the expected probability that this specific state will be realized.

Let us now assume that the risk-free security, paying a constant rate of return, $r_f$, is traded and the market is in equilibrium. Using the state prices, defined by Equation (14), one can also extract the risk free-rate, $r_f$. The risk-free security satisfies

$$1 = \sum_{j=1}^{n} r_f (p_j - \hat{\varphi}_j) \frac{U'(x_j)}{U'(x_0)},$$

and therefore

$$r_f = \frac{U'(x_0)}{\mathbb{E}[U'(x_j)] - \sum_{j=1}^{n} \hat{\varphi}_j U'(x_j)}$$

One can easily see that as in classical financial-economic theory, the higher the MRS or the higher the risk aversion, the lower the risk-free rate. In equilibrium, a high MRS or a high risk-aversion implies a higher demand for the safe asset; its price is higher and therefore its return is lower. Adding ambiguity to the economy has an opposite effect; when ambiguity is present, the risk-free rate is higher compared to conditions without ambiguity. This conclusion seems counter intuitive, since we are tempted to believe that, independently, ambiguity and risk have the same affect on pricing. However, as we have proved, ambiguity and risk are not independent. There is usually a negative relation between ambiguity and risk, i.e., increasing risk decreases ambiguity, and increasing ambiguity decreases risk. This notion can explain the positive impact ambiguity has on the risk-free rate. These results might help explain the documented puzzles of high equity returns together with low risk-free rates, as in the studies of Mehra and Prescott (1985)[76] and Weil (1989)[106]. These results also coincide with

\[52\] The equity premium puzzle asserts that risk aversion is too low to explain the observed high returns on risky assets. The risk-free rate puzzle asserts that the risk aversion, which explains the equity premium, is too high to coincide with the observed low risk-free rate.
Abel’s (2002) insights about the risk-free rate puzzle.\footnote{Abel (2002)\cite{Abel2002} proposes to model the DM’s pessimism by subjective probabilistic beliefs that are stochastically dominated by the objective distribution.}

The state prices, defined by Equation (14), can also be used to evaluate the expected return on uncertain assets. In equilibrium, the return \( r \) on an uncertain asset \( X \) must satisfy

\[
1 = \sum_{j=1}^{n} \left[ r_j (p_j - \hat{\gamma}_j) \frac{U'(x_j)}{U'(x_0)} \right].
\]

Using the covariance rule, \( E [x r] = E [x] E [r] + \text{Cov} [x, r] \), the expected return can then be formulated

\[
E [r_j] \approx \frac{U' (x_0)}{E [U' (x_j)]} - \frac{\text{Cov} [U' (x_j), r_j]}{E [U' (x_j)]} + \frac{1}{2E [U' (x_j)]} \left[ \frac{1}{2} \frac{\psi'' (p_L)}{\psi' (p_L)} U' (x_k) r_k - \frac{1}{2} \frac{\psi'' (p_G)}{\psi' (p_G)} U' (x_{k+1}) r_{k+1} \right] \mathbb{N}^2 [x].
\]

Assuming a quadratic utility function, Equation (16) paves the way towards incorporating ambiguity into the consumption capital asset pricing model (CCAPM). Such a generalization of the CAPM allows for the refinement of ambiguity to systematic ambiguity and idiosyncratic ambiguity; systematic ambiguity is driven by wide economy shocks, while idiosyncratic ambiguity is driven by firm-specific shocks. However, in this case the shocks occur to the probability distributions, rather than to consequences. Izhakian (2011)\cite{Izhakian2011} uses a different approach to make the distinction between systematic and idiosyncratic risk. He extends the mean-variance framework to the mean-uncertainty framework and uses this extended framework to generalize the CAPM to incorporate ambiguity.

### 5.3 Optimal portfolio

Given the amount allocated to saving out of her total wealth, the DM faces the decision of how much to allocate to the risk-free asset and how much to allocate to the uncertain asset. In other words, given a single uncertain asset and a risk-free asset, the question is: what is the optimal saving portfolio for a DM with risk-aversion and ambiguity-aversion preferences? The next theorem characterizes the optimal proportions of the two assets.

**Theorem 5.4.** For a sufficiently small risk premium and a sufficiently small ambiguity premium, the optimal fraction, \( z \), allocated to the uncertain asset is

\[
z \approx \frac{E [r] - r_f}{2 Var [r] \left[ - \frac{1}{2} \frac{\psi'' (p_L)}{\psi' (p_L)} (r_k - r_f)^2 - \frac{1}{2} \frac{\psi'' (p_G)}{\psi' (p_G)} (r_{k+1} - r_f)^2 \right] \mathbb{N}^2 [r]},
\]

where \( r \) is the return on the uncertain asset and \( r_k \) is the reference point in terms of return.
and the variance have on the proportion invested in an uncertain asset, this is not obvious for ambiguity. The direct impact of an increasing expected return or decreasing variance is a higher optimal proportion allocated to the uncertain asset. Variance, however, also has an indirect impact on the allocation to the uncertain asset through ambiguity; usually a negative impact, as Theorem 3.17 demonstrates. This implies that a higher ambiguity does not necessarily enforce a smaller allocation to the uncertain asset; see for example Abel (2002)[2]. The next proposition characterizes the conditions under which ambiguity has a negative impact on the allocation to the uncertain asset.

**Proposition 5.5.** If $(E[r] - r_f) < \text{Var}[r]$, then increasing the ambiguity decreases the allocation to the uncertain asset and increases the allocation to the risk-free asset.

Proposition 5.5 implies that for a high enough variance increasing ambiguity decreases the allocation to the uncertain asset. When the return on the uncertain asset is relatively small, Equation (13) indicates that if the condition $\frac{\psi'(E[P_L])}{\psi'(E[P_C])} N^2 [r] \leq - \frac{U''(E[r])}{U'(E[r])} \text{Var}[r]$ holds, an increasing ambiguity decreases the allocation to the uncertain asset. These results pave the way for further studies about the consequences of ambiguity to asset pricing and related topics. However, to limit the scope of this paper, we leave this interesting study to the attention of future research.

### 6 Related Literature

The fundamental assumption of expected utility theory (von Neumann-Morgenstern (vNM, 1944)[101] and Savage (1954)[83]) paradigms is that a decision maker, whose preferences have an expected utility representation, knows—or acts as if she knows—the probabilities of all states of nature. Critics of these models mainly attack their ability to model decisions made in reality: they are not realistic and do not reflect the full picture of uncertainty. In the real world, uncertainty has no numerical form agreed upon by all DMs; for example, real investments, financial assets and insurance have no common prospects for financial decision making.

Since Knight (1921)[63] and Ellsberg’s (1961)[26] seminal works, utility theory research has been making a concerted effort to treat decision processes under uncertainty and explain the violation of expected utility theory. These attempts have generated several new ideas. One notion is that, while making decisions, individuals consider outcomes in terms of gains and losses, relative to a neutral reference point, rather than final asset positions, i.e., objects of choice are prospects rather than acts. This notion, first introduced by Markowitz (1952)[75], is the cornerstone of the original prospect theory suggested by Kahneman and Tversky (1979)[60]. Previous literature mainly concentrates on the implications of the reference point for risk and shows that individuals are generally risk-averse as regards gains, risk-seeking as regards losses, and that losses loom larger than gains. This paper concentrates on the role of the reference point for decisions made under ambiguity. It uses the reference point and the distinction between losses and gains to assess ambiguity. Relying on CPT, our

---

54 For surveys, see Gilboa (2009)[39] and Wakker (2010)[103].
55 The reference point is also used to incorporate into prospect theory the tendency to overweight small probabilities and to underweight high probabilities.
model shows that decision weights can be obtained from a nonadditive transformation of the random probabilities scales.

Our Shadow theory generalizes Tversky and Kahneman’s (1992)[96] CPT. Shadow theory can be viewed as emerging from two different well known disciplines modeling ambiguity: multiple-prior and non-additive probability, to suggest a general unified model. The multiple-prior discipline, suggested by Gilboa and Schmeidler (1989)[40], sets aside the assumption that DMs have a unique prior, and presumes they have a set of priors. In their max-min expected utility with multiple priors (MEU) model, Gilboa and Schmeidler assert that an ambiguity-averse DM evaluates her ex-ante welfare by computing the expected utility, conditional upon the worst prior. The subjective non-additive probabilities of Gilboa (1987)[38] and the Choquet expected utility (CEU) of Schmeidler (1989)[86] state that belief can be represented by a single, but non-additive, prior. In the CEU model, uncertainty aversion results in a sub-additive prior.\textsuperscript{56,57}

CPT generalizes the CEU by adding reference-dependence and sign-dependence to the rank-dependence condition, i.e., it applies a cumulative functional separately to gains and to losses.\textsuperscript{58} Practically, it implements a two-sided variant of Choquet expected utility with possibly different capacities for gains and for losses. In SPT, capacities are generated by the randomness of probability distribution, such that nonadditive subjective-probabilities are obtained by the DM’s preferences over the directing-space. One may perceive this randomness as a random selection of one probability (not necessarily additive) measure out of a set of possible measures. Adding a directing-space that dominates the probabilities in the outcome-space (MEU) equipped with second-order capacities creates this mechanism.

The concept of modeling an attitude toward ambiguity by relaxing the reduction between first- and second-order probabilities was first suggested by Segal (1987)[87].\textsuperscript{59} Other models inspired by this idea include: Klibanoff et al.’s (2005)[61], smooth model of ambiguity, its recursive form, also proposed by Klibanoff et al. (2009)[62], and its generalization to include intertemporal substitution, proposed by Ju and Miao (2011)[59] and Hayashi and Miao (2011)[50]; the second-order probability sophistication of Ergin and Gul (2009)[32], Nau (2006)[77], and Chew and Sagi (2008)[19]. In contrast to our model, preferences toward ambiguity in these works are taken with respect to the possible expected utilities or with respect to the possible certainty equivalents, such that a complete distinction between the impact of risk preferences and ambiguity preferences on decisions is not trivial. Applying preferences toward ambiguity to a separate space—the directing-space in our model—provides a convenient mechanism with which to achieve a complete distinction between the impact of the two different preferences, thus,\textsuperscript{56}

\textsuperscript{56}See also Wakker (1989)[102], Gilboa and Schmeidler (1993)[41], Epstein (1999)[28] and Zhang (2002)[109].

\textsuperscript{57}For a convex nonadditive measure, i.e., $W (A \cup B) \geq W (A) + W (B) - W (A \cap B)$, the MEU and the CEU give the same decision rule.

\textsuperscript{58}Our results are sustained without enforcing sign-dependence.

\textsuperscript{59}Other studies that used the multiple-prior concept include: Segal and Spivak (1990)[88], Loomes and Segal (1994)[69], Epstein and Wang (1994)[30], Epstein and Schneider (2003)[29], Casadesus et al. (2000)[14] Siniscalchi (2006)[93] and Seo (2009)[89], to name a few. The weighted MEU ($\alpha$-MEU) model also assumes multiple priors and extends the MEU by also considering the best priors: Ghirardato et al. (1998)[35], Ghirardato et al. (2004)[36] and Marinacci (2002)[74], for example.
allowing for the measurement of ambiguity.

STP can be interpreted as a model of robustness in the presence of model uncertainty. This class of multiple-prior models assumes an uncertainty of the true probability law governing the realization of states and DMs’ concern with model misclassification and, thus, look for robustness decision-making; see, for example, Anderson et al. (1999)[4], Hansen et al. (1999)[49], Hansen and Srage (2001)[46], Maccheroni et al. (2006)[72] and Hansen and Sarg (2007)[48]. This literature investigates the robustness of decision rules in modeling misspecification with respect to the underlying probability. Generally speaking, this line of models assumes that the DM suffers from model-misspecification, formulated by a disutility function that penalizes all possible probability measures for deviating from a reference probability measure (reference model), where the deviation from the reference model is measured by relative entropy. SPT relaxes the requirement of having a reference probability measure with respect to which the relative entropy of density functions is calculated. Instead, all that is required is a reference outcome that distinguishes losses from gains, such that only cumulative probabilities are considered. In this sense, our model simplifies the misspecification/robustness models.

SPT is also related to Siniscalchi’s (2009)[94] vector expected utility (VEU), which assumes a baseline probability and a collection of random variables, called adjusted factors, which depict exposure to different sources of ambiguity. The value of a prospect considers an anchor-von Neumann-Morgenstern-expected-utility, with respect to the baseline probability, and a functional over the covariances between adjusted factors and utility, also with respect to the baseline probability. Other models that consider reference expected utility include those of Roberts (1980)[81], Quiggin et al. (2004)[80] and Grant and Polak (2007)[45], for example; or reference prior: Einhorn and Hogarth (1986)[25], Gajdos et al. (2004)[34], Gajdos et al. (2008)[33]. An element of confidence added to the generated set of priors is suggested by Kopylov’s (2006)[65] ϵ-contamination. In his model, the set of priors is generated around a reference prior with respect to a ”contaminating” ϵ. Chateauneuf et al. (2007)[16] suggest new capacities (neo-additive) obtained from an α-maxmin expected utility with a set of priors generated by ϵ-contamination. All these models require the identification of a reference prior, which, if at all possible, practically is not a trivial task.

Izhakian and Izhakian (2009a)[55] suggest a new mathematical formulation for multidimensional uncertainty and Izhakian and Izhakian (2009b)[56] show its implementation as regards decision theory. They suggest that observed real outcomes and real probabilities are projections of consequences in a generalized space called phantom space. The norm over this space, which maps phantom probabilities to real probabilities, can be considered as a subjective probability weighting function. In their model, the norm is applied to single events, possibly both to consequences and to their probabilities, while in our model capacities are applied to cumulative events.

This is the point to emphasize; compared to all of the models mentioned above, SPT is different as regards one major aspect: ambiguity is applied to a separate space—the directing space—and not

60Relative entropy is the expected log Radon-Nikodym derivative. Technically, all alternative models have to be absolutely continuous with respect to the reference model so that an entropy measure exists. Others studies that use relative entropy models includes those of Hansen and Sarg (2003)[47], Uppal and Wang (2003)[100], Maenhout (2004)[73], among others.
over any element of utility, i.e., expected utility, certainty equivalent or event-wise utility, which in turn are driven by preference toward risk. This structure enables a complete distinction between preferences and beliefs and between risk and ambiguity, such that ambiguity can be measured empirically. Furthermore, in this model no need arises to identify reference probability distribution. SPT provides a formal way to compare the choices of two DMs who have different attitudes toward ambiguity, or different levels of ambiguity: for example, two DMs who share the same information and the same attitude toward risk but have different levels of ambiguity sensitivity, or two DMs who share the same attitude toward risk and ambiguity but possess different information (different levels of ambiguity). The ability to conduct this type of comparative static is of primary importance, as it allows for the identification of the pure effect of introducing ambiguity and attitude toward ambiguity into a model, for both losses and for gains.

The behavioral decision literature documents different preferences toward ambiguity when individuals face gains, as compares to their preferences when facing losses - usually ambiguity aversion for gains and ambiguity seeking for losses (Einhorn and Hogarth (1986)[25], Ho et al. (2009)[51], and Chakravarty and Roy (2009)[15], among other). Our model supports different preferences for ambiguity concerning losses and gains. Interestingly, overconfidence is also linked to ambiguity. Brenner et al. (2011)[12] show that while being exposed to ambiguity individuals are less overconfident about the likelihoods to outperform a benchmark portfolio. In terms of SPT, these results imply that individuals assign lower subjective probabilities to gains, i.e., a higher subadditivity, when exposed to ambiguity, compared to the probabilities assigned when ambiguity is not present. In particular, their results provide experimental evidence supporting our model.

A few methods with which to measure ambiguity have been suggested in the literature. Dow and Werlang (1992)[23] measure uncertainty as the sum of the probability of an event and the probability of its complement event. They show that probabilities are additive, if and only if, the uncertainty aversion equals zero for any event. Assuming a normal distribution with an unknown mean, Ui (2011)[99] suggests measuring ambiguity according to the difference between the minimal possible mean and the true mean. However, the true mean is unknown. In their empirical analysis, Anderson et al. (2009)[3] measure uncertainty via the degree of disagreement of professional forecasters, attributing different weights to each forecaster. However, their results can also be attributed to heterogeneous beliefs across analysts. Jewitt and Mukerji (2011)[58] investigate the ranking of ambiguous acts as revealed by the DM’s preferences.

Bewley (2011)[8] assumes a multiple-prior setting and suggests measuring the level of ambiguity according to the critical confidence interval. Boyle et al. (2011)[10] assume mean-variance preferences, with known variances but possibly ambiguous means; DMs are characterized by max-min preferences. As in Bewley (2011)[8], they measure ambiguity using the confidence interval, α, of the expected excess return on assets, defined by

$$
\alpha - \text{confidence interval} = \left\{ \mu : \frac{(\mu - \hat{\mu})^2}{\sigma^2_{\mu}} \leq \alpha^2 \right\},
$$
where $\hat{\mu}$ is the common estimated value of the excess mean return, $\mu$, and $\sigma_{\hat{\mu}}$ is its standard error. Our measure of ambiguity, $\aleph^2$, is broader in the sense that also it takes into account the perturbation of variance. As we have shown in Section 4, the characteristics of variance play an important role when it comes to ambiguity, especially as regards the diversification prospect portfolios.

Boyle et al. (2011)[10] show that, when the level of risk and the level of ambiguity are both low, increasing the correlation between assets results in a flight to familiarity, or in other words, an escaping of ambiguity. When $(E[r] - r_f) < \text{Var}[r]$, these results coincide with our results (Proposition 5.5), only in the sense that increasing ambiguity decreases the allocation to the uncertain asset. In our model, increasing the correlation between assets increases the variance of the portfolio, and conditional upon the reference point, this usually results in a lower level of ambiguity. When $(E[r] - r_f) < \text{Var}[r]$ a higher level of ambiguity implies a smaller allocation to the uncertain asset; when $(E[r] - r_f) \geq \text{Var}[r]$ an increasing level of ambiguity has the opposite effect, i.e., greater allocation to the uncertain asset. The reason that our results are partially different than those of Boyle et al. (2011)[10] is that the ambiguity measure, $\aleph^2$, takes into consideration the impact of ambiguous variance, while the confidence interval in Boyle et al. (2011)[10] ignores it. In our model, the criteria determining the impact of ambiguity on the allocation to uncertain assets is endogenously determined, while in Boyle et al. (2011)[10] the critical confidence interval is subjective and assumed to be given exogenously.

The implication of ambiguity as regards asset pricing has been studied mainly through focusing on the theoretical aspects. Based on Klibanoff et al.’s (2005)[61] smooth model of ambiguity, Izhakian and Benninga (2011)[54] added an ambiguity premium to the conventional risk premium showing that ambiguity is negatively affected by increasing risk aversion, such that increasing risk aversion might result in a lower uncertainty premium. In contrast to our model, Izhakian and Benninga (2011)[54] use the variance of expected utility to measured ambiguity, such that a complete separation between risk and ambiguity is not obtained. Ui (2011)[99] also breaks down the equity premium into risk premium and ambiguity premium and proves that, due to changes in market participation, changes in risk premium and changes in ambiguity premium may have opposite signs. Chen and Epstein (2002)[18] extend the MEU model to continuous-time recursive multiple-priors utility and demonstrate a separation between ambiguity premium and risk premium. Epstein and Schneider (2008)[27] also use the max-min model to show that ambiguity premium depends on idiosyncratic risk in fundamentals. These works use the max-min model, such that a separation between preferences and beliefs is not feasible.

Dow and Werlang (1992)[23] use a nonadditive probability measure (CEU) to explore optimal investment decisions under ambiguity. Other works that study the implication of ambiguity as it relates to asset pricing include those of Epstein and Wang (1994)[30], Chen and Epstein, (2002)[18], Cao et al. (2005)[13], Easley and O’Hara (2009)[24] explain limited market participation by ambiguity.

---

61 Based on the Klibanoff et al.’s (2005)[61] model, Gollier (2006)[43] showed that, generally, it is not true that ambiguity aversion has the same effect as risk aversion, such that increasing ambiguity aversion may increase the demand for uncertain assets.


63 Segal and Spivak (1990)[88] also analyze the ambiguity premium, which they call a premium of order 2.
Levy et al. (2005)[13], Ju and Miao (2011)[59], among others. Levy et al. (2003)[68] suggest a functional form consisting of both CPT and financial market equilibria. Izahkain (2011)[53] uses the SPT framework to generalize mean-variance preference to mean-uncertainty preference. He incorporates ambiguity into the capital asset pricing model (CAPM) and proves a simple formalization of the beta-ambiguity, in addition to the beta-risk, such that a distinction between systematic ambiguity and idiosyncratic ambiguity is demonstrated.

7 Conclusion

This paper introduces a novel model of decision making under ambiguity, called Shadow probability theory, generalizing the the Choquet expected utility with non-additive priors model of Schmeidler (1989)[86], and the cumulative prospect theory of Tversky and Kahneman (1992)[96]. This model assumes that probabilities (capacities) of observable events in a subordinated outcome-space are random and dominated by second-order concealed prospects in a directing probability-space. The structure of two separate spaces allows for a complete distinction between risk and ambiguity and between preferences and beliefs. The level of ambiguity and the decision maker’s attitude toward it are then measured with respect to the directing space. Risk and risk attitude, on the other hand, apply to the subordinated space, as in classical expected utility theory. Ambiguous probabilities can be considered as a new dimension of risk attitude: probabilistic sensitivity, i.e., the nonlinear ways in which individuals may process probabilities. Perceived probabilities are nonadditive, ambiguity-aversion results in a sub-additive subjective probability measure, while ambiguity seeking results in a super-additive measure. Ambiguity in our model takes the form of probability perturbation (volatility) with respect to a reference point that distinguishes losses from gains. We show that this structure allows for the construction of a natural ambiguity measure, $\aleph^2$, which proves to be empirically testable.

Analyzing the nature of ambiguity, measured by $\aleph^2$, for prospect portfolios, proves that in most cases ambiguity cannot be diversified. Counterintuitively, adding a prospect to a portfolio of prospects usually increases its ambiguity. The intuition for this result is that ambiguity in our model is measured by the variance of probability of loss and, thus, is positively affected by the amplitude of probability density functions. Adding a prospect to a portfolio of prospects decreases its variance, the probability density functions becomes steeper, the incline of the cumulative probability also becomes steeper; thus, the level of ambiguity is higher.

Using this model of choice, the paper generalizes the classical asset pricing theory to incorporate ambiguous probabilities. It shows the conditions under which a higher ambiguity has a negative impact on the allocation to uncertain assets. It generalizes the Arrow-Pratt theory and clearly differentiates between ambiguity premium and risk premium. The ambiguity premium suggested by this paper can be measured empirically. To the best of our knowledge, this is the first suggested model which has been tested empirically with data, rather than in laboratory experiments or calibrations. Brenner and Izhakian (2011)[11] empirically prove that ambiguity, measured by $\aleph^2$, has a significant negative impact on traded stocks’ return.
We are convinced that this innovative model can pave the way and shed new light on many puzzling financial phenomena. For example, the existing financial literature basically asserts that investors should hold a fully diversified portfolio (while in reality individuals usually hold under-diversified portfolios, 3-4 stocks). Our results challenge this notion and demonstrate that diversification is not always optimal. Incorporating ambiguity into the decision environment indicates a need to reassess this notion.
References


[65] I. Kopylov.


A  Appendix

A.1  Supporting Lemmata

Lemma A.1. The subordinated probability, P_j, satisfies \( \xi_j^2 \leq p_j (1 - p_j) \), \( \forall j = 1, \ldots, n \), where \( p_j = E[P_j] \).

Proof of Lemma A.1. Given the expected probability \( p \), the maximal possible variance is obtained for the binary case where \( P_i \) can be either 0 or \( y \), such that the variance is

\[
\text{Var}[P] = \alpha (0 - p)^2 + (1 - \alpha) (y - p)^2,
\]

for some \( \alpha \in [0, 1] \). Since \( p = \alpha 0 + (1 - \alpha) y \), which implies \( y = p_1 - \alpha \), \( \text{Var}[P] = \alpha p^2 + (1 - \alpha)^2 = \frac{\alpha}{(1 - \alpha)^2} p^2 \).

The parameter \( y \) is a probability and must satisfy \( y = p_1 - \alpha \leq 1 \), thus \( \alpha \leq (1 - p) \), which implies that the maximal possible variance is \( \text{Var}[P] = p (1 - p) \).

Lemma A.2. Assume that the p-utility \( \psi(\cdot) \), satisfies,

\[
\frac{1}{2} \left( \frac{\psi''(p_{E_j})}{\psi'(p_{E_j})} \xi_{E_i}^2 - \frac{\psi''(p_{E_i \cup E_j})}{\psi'(p_{E_i \cup E_j})} \xi_{E_i \cup E_j}^2 \right) \leq p_{E_j},
\]

for any events \( E_j, E_i \in \Xi \). If \( A \subset B \subset S \) then \( Q(A) \leq Q(B) \).

Proof of Lemma A.2. Writing \( B = A \cup Y \), then

\[
Q(B) - Q(A) = p_A + p_Y + \frac{1}{2} \frac{\psi''(p_B)}{\psi'(p_B)} \xi_B^2 - p_A - \frac{1}{2} \frac{\psi''(p_A)}{\psi'(p_A)} \xi_A^2
\]

\[
= p_Y + \frac{1}{2} \frac{\psi''(p_{A \cup Y})}{\psi'(p_{A \cup Y})} \xi_{A \cup Y}^2 - \frac{1}{2} \frac{\psi''(p_A)}{\psi'(p_A)} \xi_A^2,
\]

which is non-negative according to the assumption of the Lemma.

A.2  Proofs of Section 2

Proof of Theorem 3.3. The subjective probability \( Q(E_j) \) of event \( E_j \in \Xi \) can explicitly be written

\[
Q(E_j) = \psi^{-1}(\psi(p_j - \varphi_j)) = \psi^{-1}\left( \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} \psi(q_{i,j}) \right),
\]

for some \( \varphi_j \). Taking the first-order Taylor approximation of \( \psi(p_j - \varphi_j) \) around \( p_j \) yields

\[
\psi(p_j - \varphi_j) \approx \psi(p_j) + \psi'(p_j) (p_j - \varphi_j - p_j) = \psi(p_j) - \varphi_j \psi'(p_j).
\]
Ignoring the weighted summation in RHS of Equation (18) for the moment, the second-order Taylor approximation of \( \psi(q_i) \) around \( p_j \) is

\[
\psi(q_{i,j}) \approx \psi(p_j) + \psi'(p_j)(q_{i,j} - p_j) + \frac{1}{2}\psi''(p_j)(q_{i,j} - p_j)^2.
\]

Since \( \psi(p_j) \), \( \psi'(p_j) \) and \( \psi''(p_j) \) are constants, applying the weighted sum yields

\[
\sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} \psi(q_i) \approx \psi(p_j) + \frac{1}{2}\psi''(p_j) \xi_j^2.
\]

Equating (19) to (20) and organizing terms yields

\[
\varphi_j = -\frac{1}{2}\frac{\psi''(p_j)}{\psi'(p_j)} \xi_j^2.
\]

Substituting \( \varphi_j \) into Equation (18) proves the theorem.

**Proof of Proposition 3.5.** By definition

\[
\xi_j^2 = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_{i,J} - p_{J})^2.
\]

Since \( q \) is additive and, thus, \( p \) is also additive

\[
\xi_j^2 = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} [(q_{i,1\ldots K} - p_{1\ldots K}) + (q_{i,K+1\ldots J} - p_{K+1\ldots J})]^2.
\]

Therefore

\[
\xi_j^2 = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_{i,1\ldots K} - p_{1\ldots K})^2 + \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_{i,K+1\ldots J} - p_{K+1\ldots J})^2 + \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_{i,1\ldots K} - p_{1\ldots K})(q_{i,K+1\ldots J} - p_{K+1\ldots J})
\]

\[
= \xi_{1\ldots K}^2 + \xi_{K+1\ldots J}^2 + 2\xi_{1\ldots K,K+1\ldots J}.
\]

**Proof of Lemma 3.6.** Let \( P_i(\mathcal{E}) = q_i \), thus, since \( q_i \) is additive \( P_i(\mathcal{E}^c) = 1 - q_i \). The covariance takes the form

\[
\text{Cov}[P(\mathcal{E}), P(\mathcal{E}^c)] = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_i - p)(q_i - p) = \sum_{i=1}^{m} \frac{\chi_i}{\sum_{i=1}^{m} \chi_i} (q_i - p)(p - q_i),
\]

and therefore

\[
\text{Cov}[P(\mathcal{E}), P(\mathcal{E}^c)] = -\text{Var}[P(\mathcal{E})].
\]
The second equality is obtained by

\[
\text{Var} \left[ P \left( \mathcal{E} \right) \right] = \sum_{i=1}^{m} \frac{X_i}{\sum_{j=1}^{m} X_i} (q_i - P)^2 = \sum_{i=1}^{m} \frac{X_i}{\sum_{j=1}^{m} X_i} (q_i^e - P^e) = \text{Var} \left[ P \left( \mathcal{E}^e \right) \right].
\]

\[
\square
\]

**Proof of Lemma 3.8.** We can write

\[
\text{Var} \left[ P_L \right] = \sum_{i=1}^{m} \frac{X_i}{\sum_{j=1}^{m} X_i} (q_{i,1} - E[q_{1\ldots K}])^2 = \sum_{i=1}^{m} \frac{X_i}{\sum_{j=1}^{m} X_i} ((1 - q_{i,1\ldots K}) - (1 - E[q_{1\ldots K}])^2,
\]

where the second equality is obtained since \( q \) is additive. Thus,

\[
\text{Var} \left[ P_L \right] = \sum_{i=1}^{m} \frac{X_i}{\sum_{j=1}^{m} X_i} (q_{i,K+1\ldots N} - E[q_{K+1\ldots N}])^2 = \text{Var} \left[ P_G \right] .
\]

\[
\square
\]

**Proof of Theorem 3.9.** Given two prospects with an identical expected outcome, the first equality, derived by Theorem 3.13, proves that an ambiguity-averse DM prefers the prospect with a lower \( \aleph^2 \) over the prospect with a relatively high \( \aleph^2 \). The second equality is obtained by the fact that the variance of the probability of loss equals the variance of the probability of gain (Lemma 3.8).

\[
\square
\]

**Proof of Theorem 3.13.** Without loss of generality, we prove this theorem for the case of a simple linear weighting function \( w \left( P \right) = P \).\(^{64}\) According to equation (6) the value of prospect \( g \in \mathcal{F} \) is

\[
V(g) = \sum_{j=1}^{k} \left[ \left( p_{1\ldots J} + \frac{1}{2} \psi^r \left( p_{1\ldots J} \right) \xi^2_{1\ldots J} \right) - \left( p_{1\ldots J-1} + \frac{1}{2} \psi^r \left( p_{1\ldots J-1} \right) \xi^2_{1\ldots J-1} \right) \right] x_j + \sum_{j=k+1}^{n} \left[ \left( p_{J+1\ldots N} + \frac{1}{2} \psi^r \left( p_{J+1\ldots N} \right) \xi^2_{J+1\ldots N} \right) - \left( p_{J+1\ldots N} + \frac{1}{2} \psi^r \left( p_{J+1\ldots N} \right) \xi^2_{J+1\ldots N} \right) \right] x_j
\]

\[
= \sum_{j=k+1}^{n} p_j x_j + \frac{1}{2} \sum_{j=1}^{k} \left[ \psi^r \left( p_{1\ldots J} \right) \xi^2_{1\ldots J} - \psi^r \left( p_{1\ldots J-1} \right) \xi^2_{1\ldots J-1} \right] x_j + \frac{1}{2} \sum_{j=k+1}^{n} \left[ \psi^r \left( p_{J+1\ldots N} \right) \xi^2_{J+1\ldots N} - \psi^r \left( p_{J+1\ldots N} \right) \xi^2_{J+1\ldots N} \right] x_j.
\]

Considering prospect \( f \in \mathcal{F} \), since \( \epsilon \) and \( P_j \) are mean-independent, the subjective probability of an event \( \mathcal{E}_j \in \Xi \) is

\[
Q \left( \mathcal{E}_j \right) = p_j + \frac{1}{2} \frac{\psi^r \left( P_j \right)}{\psi^r \left( p_j \right)} \left( \text{Var} \left[ P_j \right] + \text{Var} \left[ \epsilon \right] \right).
\]

\[^{64}\text{LeRoy and Werner (2001, Chapter 10)}[67] \text{use the same method but with respect to risk.}\]
Therefore, its value is
\[
V(f) = \sum_{j=1}^{n} p_j x_j + \frac{1}{2} \sum_{j=1}^{k} \left[ \frac{\psi''(p_{1-j})}{\psi'(p_{1-j})} (\xi^2_{1-j} + \text{Var}[\epsilon_{1-j}]) - \frac{\psi''(p_{j-1})}{\psi'(p_{j-1})} (\xi^2_{j-1} + \text{Var}[\epsilon_{j-1}]) \right] x_j
+ \frac{1}{2} \sum_{j=k+1}^{n} \left[ \frac{\psi''(p_{j-N})}{\psi'(p_{j-N})} (\xi^2_{j-N} + \text{Var}[\epsilon_{j-N}]) - \frac{\psi''(p_{j+1-N})}{\psi'(p_{j+1-N})} (\xi^2_{j+1-N} + \text{Var}[\epsilon_{j+1-N}]) \right] x_j.
\]

Thus,
\[
V(f) - V(g) = \frac{1}{2} \sum_{j=1}^{k} \left[ \frac{\psi''(p_{1-j})}{\psi'(p_{1-j})} \text{Var}[\epsilon_{1-j}] - \frac{\psi''(p_{j-1})}{\psi'(p_{j-1})} \text{Var}[\epsilon_{j-1}] \right] x_j + \frac{1}{2} \sum_{j=k+1}^{n} \left[ \frac{\psi''(p_{j-N})}{\psi'(p_{j-N})} \text{Var}[\epsilon_{j-N}] - \frac{\psi''(p_{j+1-N})}{\psi'(p_{j+1-N})} \text{Var}[\epsilon_{j+1-N}] \right] x_j.
\]

The organizing terms of Equation (21) yield
\[
V(f) - V(g) = \frac{1}{2} \frac{\psi''(p_{1-K})}{\psi'(p_{1-K})} \text{Var}[\epsilon_{1-K}] x_k + \frac{1}{2} \sum_{j=1}^{k-1} \frac{\psi''(p_{1-j})}{\psi'(p_{1-j})} \text{Var}[\epsilon_{1-j}] (x_j - x_{j+1}) + \frac{1}{2} \sum_{j=k+1}^{n} \frac{\psi''(p_{j-N})}{\psi'(p_{j-N})} \text{Var}[\epsilon_{j-N}] (x_{j+1} - x_j).
\]

Since the DM is ambiguity-averse, i.e., \( \frac{\psi''(p_{j-N})}{\psi'(p_{j-N})} < 0 \) for \( 1 \leq j \leq n \), and \( 0 < (x_{j+1} - x_j) \), the second component in the first line of (22) is positive, while the second component in the second line of Equation (22) is negative. However, because prospects are symmetric with \( x_k \leq x_s \), the absolute value of the negative component is greater than the positive component and, thus, their sum is negative. The first component in the first line and the second line of (22) are both negative; therefore, \( V(f) - V(g) \leq 0 \), which implies \( g \succ f \).

For the opposite direction, let us assume \( V(f) \geq V(g) \). Since all the parameters in the value functions \( V(f) \) and \( V(g) \) of prospects \( f \) and \( g \) are identical, except \( \text{Var}[P_L] \), Equation (21) implies that \( \text{Var}[\epsilon] \geq 0 \) and, thus, \( \text{Var}[P_L] \) of prospect \( f \) is higher than \( \text{Var}[P_L] \) of prospect \( g \). According to Lemma 3.8 this is also true for the probability of gain \( P_L \); therefore, \( R^2[f] > R^2[g] \).

**Proof of Theorem 3.14.** Considering Equation (22), since the DM is ambiguity-averse for gains, i.e., \( \frac{\psi''(p_{j-N})}{\psi'(p_{j-N})} < 0 \) for \( k+1 \leq j \leq n \), and \( 0 < (x_{j+1} - x_j) \), all the components in the second line of Equation (22) are negative. Because the DM is an ambiguity-lover as regards losses, \( 0 < \frac{\psi''(p_{j-N})}{\psi'(p_{j-N})} \), and \( (x_j - x_{j+1}) < 0 \), the second element of the first line of (22) is also negative. The first element of the first line of (22) is positive. However, it’s smaller than the absolute value of the first element in the second line, since outcomes are symmetric and the reference point is lower than the symmetry point. In this case, \( V(f) - V(g) \leq 0 \), which implies \( g \succ f \).

**Proof of Lemma 3.15.** Since the function \( T(\cdot) \) is an increasing transformation, applying it to a
prospect $f$, i.e., applying the transformation to all the consequences of the prospect, does not change the ranking of these consequences. Because the references point $x_k$ is also adjusted accordingly, the event of loss, $E_i \cup \cdots \cup E_k$, remains unchanged. Its probabilities $P_L = P(E_K)$ also remain unchanged as well as $N^2$.

**Proof of Lemma 3.16.** Writing the ambiguity measure explicitly,

$$N^2 [c + hf] = 2 \text{Var} [P_L] = 2 \text{Var} \left[ \int_{-\infty}^{c+hx_k} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-c-h\mu X)^2}{2\sigma X^2}} \, dx \right].$$

Changing the integration variable to $x = c + hy$ yields the required result.

**Proof of Theorem 3.17.** Writing the ambiguity measure explicitly

$$N^2 [f] = 4 \text{Var} [P_i (x \leq x_k)] = 4 \text{Var} \left[ \int_{-\infty}^{x_k} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \right] = 4 \text{Var} \left[ \int_{-\infty}^{x_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dx \right],$$

then

$$\frac{\partial N^2}{\partial \sigma} = 2 E \left[ (P - E[P]) (P' - E[P']) \right].$$

(23)

Using Leibniz’s integral rule yields

$$P_i' = \frac{dP_i}{d\sigma_i} = -\frac{(x_k - \mu_i)}{\sigma_i^2 \sqrt{2\pi}} e^{-\frac{(x_k-\mu_i)^2}{2\sigma_i^2}} = -\frac{z}{\sigma_i \sqrt{2\pi}} e^{-\frac{(z)^2}{2}},$$

where $z = \frac{x_k - \mu_i}{\sigma_i}$. Differentiating with respect to $z$ implies

$$\frac{\partial P_i'}{\partial z} = \frac{z^2 - 1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(z)^2}{2}} \leq 0,$$

for $z^2 = \left(\frac{x_k - \mu_i}{\sigma_i}\right)^2 \leq 1$, which is satisfied for $\mu_i - \sigma_i \leq x_k \leq \mu_i + \sigma_i$. Therefore, $P_i'$ is a decreasing function of $\frac{x_k - \mu_i}{\sigma_i}$ in the range $\mu_i - \sigma_i \leq x_k \leq \mu_i + \sigma_i$ for any $i = 1, \ldots, m$. However, the cumulative probability distribution, $P_i$, is an increasing function of $\frac{x_k - \mu_i}{\sigma_i}$. Therefore, the covariance between $P$ and $P'$ is negative and thus is $\frac{\partial N^2}{\partial \sigma}$ in Equation (23).

**A.3 Proofs of Section 4**

**Proof of Proposition 4.2.** Using the convolution theory, it can be shown that the sum of two normally distributed variables is also normally distributed with mean $\mu_Z = h_1 \mu_X + h_2 \mu_Y$ and variance $\sigma_Z^2 = h_1^2 \sigma_X^2 + h_2^2 \sigma_Y^2 + 2h_1 h_2 \sigma_{XY}$. Theorem 3.9 then proves the proposition.
Proof of Theorem 4.4. Writing the ambiguity of $Z$ explicitly

$$N^2[Z] = 4 \text{Var} \left[ \int_{-\infty}^{z_k} \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{(z - \mu_Z)^2}{2\sigma_Z^2}} \, dz \right] = 4 \text{Var} \left[ \int_{-\infty}^{z_k} \frac{z_k - \mu_Z}{\sqrt{\sum_{i=1}^{k} h_i x_i + \sum_{i=1}^{k} \sum_{j \neq i} h_i a_{ij} x_i x_j}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz \right].$$

Since any prospect has a nonnegative weight in the portfolio, a higher correlation implies a greater $\sigma_Z$, which in turn, according to Theorem 3.13, implies a lower ambiguity.

Proof of Theorem 4.4. We prove this statement for two prospects $X, Y \in \mathcal{F}$; the proof then can easily be extended by induction to any number of prospects. Since $\frac{\mu_{X,i}}{\sigma_{X,i}} = \frac{\mu_{Y,i}}{\sigma_{Y,i}}$, writing $\mu = a \mu_X$ and $\sigma = a \sigma_X$, the ambiguity level of the two prospects is identical

$$4 \text{Var} \left[ \Phi \left( \frac{x_k - \mu}{\sigma} \right) \right] = 4 \text{Var} \left[ \Phi \left( \frac{a x_k - a \mu X}{a \sigma_X} \right) \right] = 4 \text{Var} \left[ \Phi \left( \frac{x_k - \mu X}{\sigma_X} \right) \right].$$

Hence, $h N^2[X] + (1 - h) N^2[Y] = 4 \text{Var} \left[ \Phi \left( \frac{x_k - \mu_X}{\sigma_X} \right) \right]$. The ambiguity level of the portfolio is

$$I = 4 \text{Var} \left[ \Phi \left( \frac{z_k - h \mu X - (1 - h) \mu Y}{\sqrt{h^2 \sigma_X^2 + (1 - h)^2 \sigma_Y^2 + 2h(1 - h) \rho_{X,Y} \sigma_X \sigma_Y}} \right) \right] = 4 \text{Var} \left[ \Phi \left( \frac{z_k - h \mu X - a (1 - h) \mu_X}{\sqrt{h^2 \sigma_X^2 + (1 - h)^2 a^2 \sigma_X^2 + 2ah(1 - h) \rho_{X,Y} \sigma_X^2}} \right) \right],$$

where $z_k = h x_k + (1 - h) y_k$. According to Theorem 4.3, $I$ attains its minimal value when the correlation is maximal, i.e., $\rho_{X,Y} = 1$; thus,

$$I \geq 4 \text{Var} \left[ \Phi \left( \frac{h x_k + a (1 - h) x_k - h \mu X - a (1 - h) \mu_X}{\sqrt{h^2 \sigma_X^2 + (1 - h)^2 a^2 \sigma_X^2 + 2ah(1 - h) \rho_{X,Y} \sigma_X^2}} \right) \right] = 4 \text{Var} \left[ \Phi \left( \frac{x_k - \mu X}{\sigma_X} \right) \right].$$

A.4 Proofs of Section 5

Proof of Theorem 5.1. The first-order Taylor approximation of the LHS of Equation (11) with respect to $E[x]$ is

$$LHS = U(E[x] - K) = \sum_{j=1}^{n} p_j U(E[x] - K) \approx \sum_{j=k}^{n} p_j (U(E[x]) - KU'(E[x])).$$
Writing the RHS of Equation (11) as

\[
RHS = \sum_{j=1}^{n} p_j U(x_j) - \left( \sum_{j=1}^{k} (\varphi_{1\ldots j} - \varphi_{1\ldots j-1}) U(x_j) + \sum_{j=k+1}^{n} (\varphi_{j\ldots N} - \varphi_{j+1\ldots N}) U(x_j) \right),
\]

the second-order Taylor approximation of I around \( E[x] \) is

\[
I \approx \sum_{j=1}^{n} p_j \left( U(E[x]) + U'(E[x]) (x_j - E[x]) + \frac{1}{2} U''(E[x]) (x_j - E[x])^2 \right) = U(E[x]) + \frac{1}{2} U''(E[x]) \text{Var}[x].
\]

Taking the first-order Taylor approximation of \( II \) around \( E[x] \) yields\(^{65}\)

\[
II \approx \sum_{j=1}^{k} (\varphi_{1\ldots j} - \varphi_{1\ldots j+1}) (U(E[x]) + U'(E[x]) (x_j - E[x])) + \sum_{j=k+1}^{n} (\varphi_{j\ldots N} - \varphi_{j+1\ldots N}) (U(E[x]) + U'(E[x]) (x_j - E[x])).
\]

Let \( x_{j+1} - x_j \approx \Delta \), then

\[
II \approx \varphi_{1\ldots K} (U(E[x]) + U'(E[x]) (x_k - E[x])) + \varphi_{K+1\ldots N} (U(E[x]) + U'(E[x]) (x_{k+1} - E[x])) - \Delta \sum_{j=1}^{k} \varphi_{1\ldots j} U'(E[x]) + \Delta \sum_{j=1+1}^{n} \varphi_{j\ldots N} U'(E[x]).
\]

Since \( \varphi_1 \) and \( \varphi_n \) are relatively small and \( \Delta \sum_{j=1}^{k} \varphi_{1\ldots j} \approx \Delta \sum_{j=1+1}^{n} \varphi_{j\ldots N} \), then

\[
II \approx \varphi_{1\ldots K} U(x_k) + \varphi_{K+1\ldots N} U(x_{k+1}),
\]

where the approximation is obtained by going from the Taylor approximation backward to the accurate values of the function at outcomes \( x_k \) and \( x_{k+1} \). Because \( x_k \) is relatively close to zero, \( U(x_k) = 0 \), and \( U(\cdot) \) is almost linear around the reference point, \( x_k \), then \( U(x_k) \approx U'(E[x]) E[x] \). Therefore,\(^{66}\)

\[
II \approx \frac{1}{4} \left[ -\frac{1}{2} \frac{\psi''(p_L)}{\psi'(p_L)} - \frac{1}{2} \frac{\psi''(p_G)}{\psi'(p_G)} \right] \text{Var}[x] U'(E[x]) E[x].
\]

Combining equations LHS, RHS, I and II, the uncertainty premium can be approximated by

\[
\mathcal{K} \approx -\frac{1}{2} \frac{U''(E[x])}{U'(E[x])} \text{Var}[x] - \frac{1}{8} \left[ \frac{\psi''(p_L)}{\psi'(p_L)} + \frac{\psi''(p_G)}{\psi'(p_G)} \right] \text{Var}[x] E[x].
\]

\(^{65}\)Recall that \( \varphi_j \) is the probability premium, holding an order of magnitude of the variance of probability. Thus, \( \varphi_j \) is smaller by one order of magnitude than probabilities.\(^{66}\)In CPT the utility function around zero is almost linear; see for example Segal and Spivak (1990)[88] and Levy at al. (2003)[68].
Proof of Corollary 5.2. CRRA implies $U'(x) = x^{-\gamma}$ and $U''(x) = -\gamma x^{-\gamma - 1}$. CAAA implies $\psi'(q_i) = e^{-\eta q_i}$ and $\psi''(q_i) = -\eta e^{-\eta q_i}$. Substituting in Theorem 5.1 yields the result.

Proof of Corollary 5.3. CRRA implies $U'(x) = x^{-\gamma}$ and $U''(x) = -\gamma x^{-\gamma - 1}$. CAAA implies $\psi'(q_i) = q_i^{-\eta}$ and $\psi''(q_i) = q_i^{-1-\eta}$. Substituting for the e-premium of the event of loss yields $\varphi_L = \frac{1}{2} \frac{1}{E[P_G]} \xi_L^2$ and similarly for the event of gains $\varphi_G = \frac{1}{2} \frac{1}{E[P_G]} \xi_G^2$. Then, according to Theorem 5.1, the result is obtained.

Proof of Theorem 5.4. Given the allocated wealth to saving, the DM optimal portfolio allocation problem is defined by

$$\max_z \mathbb{E}[U(z(r - r_f) + r_f)] = \max_z \sum_{j=1}^{n} Q(r_j) U(z(r_j - r_f) + r_f),$$

where $z$ is the fraction allocated to the uncertain asset. Thus, the first-order condition is

$$\sum_{j=1}^{n} Q(r_j) \left[ U'(z(r_j - r_f) + r_f) (r_j - r_f) \right] = 0.$$

Taking the first-order Taylor approximation of the marginal utility with respect to $r$ around $r_f$ yields

$$\sum_{j=1}^{n} Q(r_j) \left[ U'(r_f) (r_j - r_f) + z U''(r_f) (r_j - r_f)^2 \right] \approx 0.$$

Solving for $z$ yields

$$z \approx \frac{\sum_{j=1}^{n} Q(r_j) (r_j - r_f)}{-U''(r_f) \sum_{j=1}^{n} Q(r_j) (r_j - r_f)^2}.$$

Since for $j \leq k$, the subjective probability takes the form $Q(r_j) = p_j - (\varphi_{1...J} - \varphi_{1...J-1})$ and for $j > k$ it takes the form $Q(r_j) = p_j - (\varphi_{J+1...N} - \varphi_{J+1...N-1})$, then from the same consideration of Theorem 5.1

$$z \approx \frac{[\mathbb{E}[r] - r_f] - \frac{1}{2} \left[ \frac{1}{2} \psi''(p_L) (r_k - r_f) - \frac{1}{2} \psi''(p_C) (r_{k+1} - r_f) \right] \mathbb{N}^2 [r]}{-U''(r_f) \left( \mathbb{Var}[r] - \frac{1}{2} \left[ -\frac{1}{2} \psi''(p_L) (r_k - r_f)^2 - \frac{1}{2} \psi''(p_C) (r_{k+1} - r_f)^2 \right] \mathbb{N}^2 [r] \right)}.$$

Proof of Proposition 5.5. Differentiating Equation (17) with respect to $\mathbb{N}^2$ proves the proposition.