

NHH



# Optimal Risk Sharing: Expected Utility Theory Versus the Dual Theory of Choice Under Risk

**Thomas Michael Fjærвик**

**Supervisor: Knut Kristian Aase**

MSc Economics and Business Administration, Economic Analysis

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

## **Abstract**

An introduction to the dual theory of choice under risk is given. Optimal risk sharing under both expected utility theory and the dual theory of choice under risk is reviewed. Central results to insurance in pure demand theory is found to be very similar under both theories. The exception is optimal coinsurance. Central results are also found to be similar concerning Pareto optimal risk sharing between an insurer and a potential policyholder, but some differences arise. The general structure of Pareto optimal risk sharing is affected by the underlying choice theory. For both Pareto optimal risk sharing and pure demand theory similarities/differences are attempted explained by properties underlying the respective choice theories. A brief introduction to distortion risk measures and their relation to the dual theory of choice under risk is given. Before the concluding remarks, a brief discussion concerning the normative and descriptive validity of each choice theory is presented. In general it seems that the dual theory of choice models risk sharing between firms well, while expected utility theory models risk sharing concerning individuals well. This seems to be a result of agents' attitudes towards wealth under the different theories.

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>3</b>  |
| <b>2</b> | <b>The Dual Theory of Choice Under Risk</b>                              | <b>5</b>  |
| 2.1      | The Axioms . . . . .   | 6         |
| 2.2      | Representation Theorems . . . . .  | 9         |
| 2.3      | Marginal Utility of Wealth vs. Risk Aversion . . . . .                   | 11        |
| 2.4      | Characterization of Risk Aversion . . . . .                              | 13        |
| <b>3</b> | <b>Optimal Insurance: Pure Demand Theory</b>                             | <b>17</b> |
| 3.1      | Optimal Coinsurance: Expected Utility Theory . . . . .                   | 17        |
| 3.2      | Optimal Coinsurance: The Dual Theory of Choice Under Risk . . . . .      | 19        |
| 3.3      | Optimality of Deductibles . . . . .                                      | 22        |
| 3.4      | Optimal Reinsurance . . . . .  | 24        |
| <b>4</b> | <b>Pareto Optimal Risk Sharing: Expected Utility Theory</b>              | <b>25</b> |
| 4.1      | The Benefits of Risk Sharing . . . . .                                   | 25        |
| 4.2      | Risk Sharing in the Absence of Costs . . . . .                           | 26        |
| 4.3      | Risk Sharing in the Presence of Costs . . . . .                          | 33        |
| <b>5</b> | <b>Distortion Risk Measures</b>  | <b>37</b> |
| 5.1      | Introduction to Distortion Risk Measures . . . . .                       | 37        |
| 5.2      | Properties of Distortion Risk Measures . . . . .                         | 39        |
| <b>6</b> | <b>Pareto Optimal Risk Sharing: The Dual Theory of Choice Under Risk</b> | <b>41</b> |
| 6.1      | Introducing the Problem . . . . .  | 41        |
| 6.2      | Characterization of Pareto Optimal Allocations . . . . .                 | 42        |

|          |  |           |
|----------|--|-----------|
| 6.3      | The Case of One Policyholder and One Insurer . . . . . | 49        |
| 6.4      | Some Points About Pareto Optimal Allocations . . . . . | 51        |
| <b>7</b> | <b>Discussion</b>                                      | <b>55</b> |
| <b>8</b> | <b>Conclusion</b>                                      | <b>59</b> |
|          | <b>References</b>                                      | <b>61</b> |

# 1. Introduction

This thesis is concerned with the structure of optimal risk sharing. There has been extensive amounts of articles written on the subject of risk sharing. Most of these seem to be restricted to the confines of expected utility theory. This is perhaps not a surprise, as many people consider it the standard workhorse for analysis of decisions under risk/uncertainty. It is perhaps so commonly used that many take it for granted as the only theory for analysis of decisions under risk/uncertainty. As a master student at the Norwegian School of Economics I never once encountered another framework for decision making, not counting brief introductions to prospect theory or brief mentions of other theories. I was therefore intrigued when my supervisor introduced me to the dual theory of choice under risk developed by Yaari (1987). Taking into account that my favorite subjects at the Norwegian School of Economics concerned optimal insurance decisions and risk sharing in the expected utility framework, it seemed natural to explore these subjects under the dual theory. This provided the motivation for this thesis.

The main goal of this thesis is to present implications of both theories for risk sharing, and to present some central results in the area of optimal risk sharing under both theories. A comparison of results with explanations and discussions naturally follow. Another goal of the thesis is to present the subject matter in a comprehensible way, not only to mathematically well-versed readers, but to a broader audience with a certain competence in economics/finance. For the most part, complicated mathematics and derivations are avoided in order to facilitate economic clarity. However, in order to include the reader in certain points of the analysis, some mathematics will be required. Sometimes this serves the purpose of preparing the reader for further study of other articles, and sometimes it facilitates explanation of results and their origin in the different theoretical frameworks.

The main structure of the thesis is as follows. Section two presents the dual theory of choice under risk. The most central aspects of the dual theory is presented, and at some points more detailed explanations are given when deemed necessary. This overview of the dual theory includes comparisons with the expected utility framework. The choice theories are quite similar, as the reader shall see, but there is however a noteworthy difference. Section

three presents optimal insurance decisions under both theories. This section is limited to pure demand analysis, i.e, we only consider the demand for insurance by a potential policyholder and take the insurer as given. Section four reviews the case of Pareto optimal risk sharing under the expected utility theory. Initially costs are excluded from the analysis and the Pareto optimal risk sharing allocations are presented. Thereafter follows a presentation of Pareto optimal risk sharing arrangements in the presence of costs. Results are specialized to one insurer and one potential policyholder for clarity and ease of exposition. Section five gives a brief introduction to distortion risk measures. They represent an insurance premium principle and it is explained how maximizing dual utility is equivalent to minimizing a distortion risk measure. Some properties of distortion risk measures are also presented. Section six deals with Pareto optimal risk sharing under the dual theory and makes use of distortion risk measures in the analysis. The main structure of section six is centered around an article by Ludkovski and Young (2009). The reason for this is that the general structure of this article is reoccurring in the literature, as is explained in section six. This section is somewhat technical to start with. However, the diligent reader will be rewarded, as there is insight to be gained from this. Technicalities and rigour is sacrificed for clarity later. Section seven discusses briefly both decision theories as normative and descriptive theories, using insights gained from the prior sections. Concluding remarks are given in section eight.

I would like to thank my supervisor Knut Kristian Aase for very insightful comments and suggestions. Beyond that, his extensive knowledge and grasp of the subject is what sparked my interest in risk sharing when I first started my master's degree. I am grateful that he took the time, in his last semester before retirement, to supervise this thesis.

## 2. The Dual Theory of Choice Under Risk

Proposed by Yaari (1987), the dual theory of choice under risk is a modification of expected utility theory. In fact, its axiomatic foundation is the same, except for the so-called independence axiom. This last axiom is, as Yaari puts it, "laid on its side." Whereas in expected utility theory one requires independence with respect to convex combinations of risky prospects formed along the probability axis, in the dual theory one requires independence with respect to convex combinations formed along the payment axis. Before discussing this any further, the axioms will be stated as presented in Yaari (1987). The presentation will also include comments to the axioms, as well as some necessary details preceding the axioms themselves. Details from Yaari (1987) which are not central to this exposition are omitted.

$V$  is the set of all random variables defined on a given probability space, with values in the unit interval. Note that the values of all  $v \in V$  will be interpreted as payments. Also note that expected utility theory is not limited to variables with support in the unit interval. Having mentioned this, we adhere to the exposition below mostly as published by Yaari (1987). For each  $v \in V$ , define the decumulative distribution function (DDF) of  $v$ , denoted  $G_v$ , by

$$G_v(t) = Pr(v > t), \quad 0 \leq t \leq 1.$$

$G_v$  is nonincreasing, right-continuous and satisfies  $G_v(1) = 0$ . For all  $v \in V$ , one can confirm using integration by parts that

$$\int_0^1 G_v(t) dt = Ev,$$

where  $E$  is the expectation operator.

We assume that a preference relation  $\succeq$  is defined on  $V$ . Also, we let the symbols  $\succ$  and  $\sim$  stand for strict preference and indifference. We will now define a family of functions  $\Gamma$ , which will be referred to in the axioms.

$$\Gamma = \{G : [0, 1] \mapsto [0, 1] \mid G \text{ is nonincreasing, right continuous and satisfies } G(1) = 0\}$$

Using the notation from Yaari (1987), we are now ready to first present the axioms of expected utility theory.

## 2.1 The Axioms

**Axiom A1:** Let  $u$  and  $v$  belong to  $V$ , with respective DDF's  $G_u$  and  $G_v$ . If  $G_u = G_v$ , then  $u \sim v$ . This axiom says that if two risky prospects have the same decumulative distribution functions, then they are equally risky (S. S. Wang & Young, 1998). Put differently, the agent is indifferent between two prospects with the same decumulative distribution functions.

**Axiom A2:**  $\succeq$  is reflexive, transitive and connected. This axiom says that for any risky prospects  $x, y, z \in X$ , the following holds (Kreps, 2018)

1.  $x \succeq x$  for all  $x \in X$
2. If  $x \succeq y$  and  $y \succeq z$  then  $x \succeq z$  for all  $x, y, z \in X$
3. for all  $x, y \in X$ ,  $x \succeq y$  or  $y \succeq x$  or both.

In words, we could say that: 1) all risky prospects are weakly preferred to themselves. 2) If  $x$  is weakly preferred to  $y$  and  $y$  is weakly preferred to  $z$  then  $x$  is weakly preferred to  $z$ . 3) The preference relation orders all pairs of risky prospects.

**Axiom A3:** Let  $G, G', H, H'$ , belong to  $\Gamma$ ; assume that  $G \succ G'$ . Then, there exists an  $\varepsilon > 0$  such that  $\|G - H\| < \varepsilon$  and  $\|G' - H'\| < \varepsilon$  imply  $H \succ H'$ , where  $\|\cdot\|$  is the  $L_1$ -norm, i.e.  $\|m\| = \int |m(t)|dt$ . This continuity axiom is stronger than that required for expected utility theory. As an example, one could consider the "standard" continuity axiom of expected utility theory (Levin, 2006). It simply says that if you have three risky prospects  $a, b$ , and  $c$ , and  $a$  is weakly preferred to  $b$  is weakly preferred to  $c$ , then there exists a constant  $\alpha \in (0, 1)$  such that the agent is indifferent between a convex mixture between  $a$  and  $c$  and the prospect  $b$ . A standard example is letting  $a$  be the prospect of receiving \$10,  $b$  the prospect of receiving nothing and  $c$  the prospect of being killed. Then there exists an  $\alpha$ , however close to 1, such that the agent is indifferent between receiving \$10 with probability  $\alpha$  and getting killed with

probability  $(1 - \alpha)$ . Some would argue that this should not be the case realistically, while others would make the argument that it is entirely rational, considering there is always a positive probability of dying in any instant.

**Axiom A4:** If  $G_u(t) \geq G_v(t)$  for all  $t, 0 \leq t \leq 1$ , then  $G_u \succeq G_v$ . This axiom simply says that if the probability that the value  $t$  of a random variable  $u$  is always greater than or equal to the probability of the same value  $t$  of a random variable  $v$ , then the distribution of  $u$  is weakly preferred to the distribution of  $v$ . In other words the preferences are monotone with respect to first-order stochastic dominance.

**Axiom A5EU (independence):** If  $G, G'$ , and  $H$  belong to  $\Gamma$  and  $\alpha$  is a real number satisfying  $0 \leq \alpha \leq 1$ , then  $G \succeq G'$  implies  $\alpha G + (1 - \alpha)H \succeq \alpha G' + (1 - \alpha)H$ . This axiom states that, if you weakly prefer one risky prospect  $G$  to another risky prospect  $G'$ , then you will also weakly prefer a convex combination of  $G$  and some other risky prospect  $H$  to a convex combination of  $G'$  and the same risky prospect  $H$ . Put differently, preferences between risky prospects does not change by introducing another risky prospect.

In order to state the axioms for the dual theory of choice under risk we need only modify the independence axiom. However, in order to do this we follow Yaari (1987) and first define inverse decumulative distribution functions. In his article, Yaari gives a very general definition of the inverse DDFs. To hopefully ease the exposition of his paper, we will now adopt the definition of inverse DDFs given by S. S. Wang and Young (1998), with notation adapted to our setting.

**Definition:** Let  $G_x(t) \in \Gamma$ . Then the inverse function  $G_x^{-1}$  is defined by

$$G_x^{-1}(q) = \inf\{t \geq 0 : G_x(t) \leq q\}, \quad 0 \leq q \leq 1$$

with  $G_x^{-1}(0) = 1$ , if  $G_x(t) \geq 0$  for all  $t \geq 0$ .

Using this definition of the inverse, if the DDF  $G$  is invertible then  $G^{-1}$  is just the usual inverse of  $G$ , which is also the case for the definition given by Yaari (1987). The dual independence axiom will now be stated. Again, for ease of exposition it will be formulated as in (S. S. Wang & Young, 1998). In the original paper by Yaari the axiom is stated by first defining a mixture operation on  $\Gamma$  which makes it a mixture space, as in Herstein and

Milnor (1953).

**Axiom A5 (Dual independence):** If  $X \prec Y$ , if  $Z$  is any risk, and if  $p$  is any number in  $[0, 1]$ , then  $W \prec V$ , in which  $W$  and  $V$  are the random variables with inverse DDF's given by  $pG_X^{-1} + (1 - p)G_Z^{-1}$  and  $pG_Y^{-1} + (1 - p)G_Z^{-1}$ , respectively.

In his paper, Yaari (1987) gives another statement of the dual independence axiom that is perhaps more suited for economic interpretation. He also proves that these two statements are equivalent, something that will not be repeated here. We will, however, now repeat a definition needed to interpret this statement of the axiom.

**Definition:** Let  $u$  and  $v$  belong to  $V$ . We say that  $u$  and  $v$  are comonotonic if, and only if, for every  $s$  and  $s'$  in  $S$ , the inequality

$$(u(s) - u(s'))(v(s) - v(s')) \geq 0$$

is true.

**Axiom A5 (Direct Dual Independence):** Let  $u, v$  and  $w$  belong to  $V$  and assume that  $u, v$  and  $w$  are pairwise comonotonic. Then, for every real number  $\alpha$  satisfying  $0 \leq \alpha \leq 1$ ,  $u \succeq v$  implies  $\alpha u + (1 - \alpha)w \succeq \alpha v + (1 - \alpha)w$ .

As Yaari (1987) points out, this is a convex combination of real functions, and not a probability mixture. It could be thought of as taking pointwise averages of the values of the random variables. The economic interpretation here follows from the concept of hedging. If the random variables are, as stated in the axiom, pairwise comonotonic, then no mixing of the random variables will serve as a hedge. Put differently, one cannot influence how a random variable varies, by mixing with another random variable whose values always vary in the same direction across states.

An example could be in place here. Consider two states,  $s$  and  $s'$  and two random variables  $u$  and  $w$ . To be more concrete suppose state  $s$  is sunny weather and state  $s'$  is rainy weather, and suppose  $u$  and  $w$  represent sale of ice cream and sale of sunscreen, respectively, and that they can take on values "high" corresponding to state  $s$  and "low" corresponding to state  $s'$ . One could not reduce the variance of the total sales of ice cream and sunscreen by mixing the two because the change in sales across states always moves in the same direction. Had instead a third random variable  $v$  representing sales of umbrellas

been mixed with  $u$ , the total variability of sales across states could probably be reduced. This is due to the fact that the number of umbrellas sold would be distributed "low" in state  $s$  and "high" in state  $s'$ , and thus always cancelling out the "high" and "low" sales of ice cream in the corresponding states. Going beyond this very simplified example, it is easy to imagine how preferences between random variables could change when introducing mixing with a third random variable that only serves as a hedge against one of the other random variables. This should make the economic interpretation of the axiom quite clear. The following is an informal verbalization of the axiom: "Let  $u$ ,  $v$  and  $w$  be such that none is a hedge against the other. Then, a convex mixing of the values of the random variables does not change the order of preference between them."

## 2.2 Representation Theorems

At this point, we are ready to state two representation theorems. The first one will be the widely known expected utility theorem. The other one will be a representation theorem for the dual theory of choice under risk. Both theorems will be presented exactly as in Yaari (1987). Again, note that the expected utility theorem as presented here is defined on the unit interval, which need not be the case in general. Also note the following notation: If  $x$  and  $p$  are between 0 and 1,  $[x; p]$  represents a random variable with support  $x$  and 0 with probabilities  $p$  and  $1 - p$ , respectively.

**Theorem 0 (EU):** A preference relation  $\succeq$  satisfies Axioms A1-A4 and A5EU if, and only if, there exists a continuous and nondecreasing real function  $\phi$ , defined on the unit interval, such that, for all  $u$  and  $v$  belonging to  $V$ ,

$$u \succeq v \Leftrightarrow E\phi(u) \geq E\phi(v).$$

Moreover, the function  $\phi$ , which is unique up to a positive affine transformation, can be selected in such a way that, for all  $t$  satisfying  $0 \leq t \leq 1$ ,  $\phi(t)$  solves the preference equation

$$[1; \phi(t)] \sim [t; 1].$$

It is worthwhile to talk about this theorem. It says that if and only if one can represent preferences omitted by a preference relation by a function satisfying certain properties, does the preference relation satisfy axioms A1-A4 and A5EU. This is quite a strong result as it allows us to make use of the big toolbox that is mathematics to analyze problems where an agent has preferences satisfying axioms A1-A4 and A5EU. In addition, the theorem also tells us the construction of such a function. The direct interpretation of the preference equation is that an agent is indifferent between receiving 1 with probability  $\phi(t)$  and receiving  $t$  with probability 1. The function  $\phi(t)$  must satisfy this preference equation for all values of  $t$ , where  $0 \leq t \leq 1$ . Having found such a function, we may define a new function by applying a transformation  $f(t) := a\phi(t) + b$ , where  $a > 0$ , and this new function still represents the same preferences as before.

**Theorem 1 (Dual theory):** A preference relation  $\succeq$  satisfies Axioms A1-A5 if, and only if, there exists a continuous and nondecreasing real function  $f$ , defined on the unit interval, such that, for all  $u$  and  $v$  belonging to  $V$ ,

$$u \succeq v \Leftrightarrow \int_0^1 f(G_u(t))dt \geq \int_0^1 f(G_v(t))dt.$$

Moreover, the function  $f$ , which is unique up to a positive affine transformation, can be selected in such a way that, for all  $p$  satisfying  $0 \leq p \leq 1$ ,  $f(p)$  solves the preference equation

$$[1; p] \sim [f(p); 1].$$

Much of what we mentioned after the EU theorem also applies here. The interpretation of the indifference equation is somewhat less abstract. It says that the function  $f(p)$  must be such that the agent is indifferent between receiving 1 with probability  $p$  and receiving  $f(p)$  with probability 1. Say that we have a lottery that takes the values 0 and 1 with probabilities  $\frac{1}{2}$  and  $\frac{1}{2}$ . If the agent is indifferent between facing this lottery and receiving say  $\frac{3}{4}$  with certainty, then the function  $f(\frac{1}{2})$  must assign the value  $\frac{3}{4}$ . It could be noted for the interested reader that Yaari (1987) gives a proof of theorem 1 which is quite elegant and which largely makes use of the previously proven theorem 0, i.e. the expected utility theorem. In other

words, if the reader is familiar with a proof of the expected utility theorem, the proof of theorem 1 should be quite understandable.

Immediately one can note two properties of the utility function  $U(v) := \int_0^1 f(G_v(t))dt$ . The first is that  $U(v)$  assigns to each risky prospect, i.e. to each random variable  $v \in V$  its certainty equivalent. Put differently, the agent values receiving an amount of money  $U(v)$  just as much as receiving the uncertain prospect  $v$ . For comparison purposes, within the framework of expected utility theory, the certainty equivalent  $CE$  of an uncertain prospect  $v$  is implicitly defined as  $u(CE) = Eu(v)$ , where  $E$  is the expectation operator and  $u$  is the function representing the preferences of the agent. Another noteworthy property of the utility function  $U(v)$  is that it is linear in payments. This simply means that if one applies a fixed positive affine transformation to the values of a random variable  $v$ , then the value  $U(v)$  is transformed in the same manner. This is in contrast to expected utility theory, which is linear in probabilities. Intuitively this difference between the theories makes sense, since the only difference in their axiomatic foundation is the independence axiom. Moreover, the independence axiom for expected utility states independence with respect to probabilities of outcomes while the dual independence axiom states independence with respect to outcomes of probability measures/distributions.

## 2.3 Marginal Utility of Wealth vs. Risk Aversion

Equipped with some properties of the dual utility function  $U(v)$  we now present what is arguably a very intuitive appeal of the dual theory, namely that diminishing marginal utility of wealth and risk aversion are not entwined as in expected utility theory. In order to explain this, we first state what defines marginal utility of wealth and what defines risk aversion in expected utility theory. Diminishing marginal utility of wealth in expected utility theory stems from the agent's preferences being modeled by a concave utility function. Consequently, an agent's utility function exhibits the following properties: 1)  $u'(\cdot) > 0$  and 2)  $u''(\cdot) \leq 0$  (weak concavity). In order to be precise, one should say that properties 1) and 2) models an agent with non-increasing marginal utility of wealth. Substituting property 2) with 3)  $u''(\cdot) < 0$  (strict concavity) yields diminishing marginal utility of wealth. In plain english, gaining x dollars gives less added utility the more initial wealth the agent is endowed

with. Risk aversion could within the framework of expected utility be defined by saying that a risk averse agent is someone who is not willing to accept a fair gamble, with a fair gamble being defined as having expected value 0. Let  $w$  be the risk-free initial wealth of an agent and let a lottery  $z$  be such that  $Ez = 0$ . If  $Eu(w+z) \leq Eu(w)$  then the agent is risk averse. But this expression can be recognized as simply being Jensen's inequality, which tells us that in order for the inequality to hold then the function  $u(\cdot)$  must be concave. It is thus evident that an agent whose preferences are represented by an expected utility representation must be either both risk averse and have diminishing marginal utility, or exhibit none of those traits.

As stated in Safra and Segal (1998, p. 1) "Constant risk aversion means that adding the same constant to all outcomes of two distributions, or multiplying all their outcomes by the same positive constant, will not change the preference relation between them." Put differently, if an agent has preferences represented by a preference relation  $\succeq$  and if that agent has preference  $u \succeq v$  for any  $u$  and any  $v$ , then prospects  $\hat{u} = au + b$  and  $\hat{v} = av + b$  will be ordered  $\hat{u} \succeq \hat{v}$ . The original ordering of the prospects is preserved by the preference relation. Taking this to be the definition of both constant absolute risk aversion and constant relative risk aversion, and recognizing that the dual utility function is linear in payments, one can prove the following corollary in Yaari (1987): If the preference relation  $\succeq$  satisfies A1-A5, then, for all  $u$  and  $v$  belonging to  $V$ , we have

$$u \succeq v \Leftrightarrow au + b \succeq av + b,$$

provided  $a > 0$  and provided  $au + b$  and  $av + b$  both belong to  $V$ .

In words, under A1-A5, agents always display constant absolute risk aversion as well as constant relative risk aversion. This is not true for expected utility if we assume risk averse agents (and if their wealth is different from 1), as will now be shown. We will here make use of the Arrow-Pratt measures of absolute and relative risk aversion (Pratt, 1975), denoted  $A(w)$  and  $R(w)$ , respectively.

$$A(w) := \frac{-u''(w)}{u'(w)} \text{ and } R(w) := \frac{-u''(w)w}{u'(w)} = A(w)w$$

From the definitions of the coefficient of absolute and relative risk aversion, it is evident

that constant relative risk aversion under expected utility implies decreasing absolute risk aversion. The reason for this is that, under the assumption of a risk averse agent,  $A(w)$  is positive (we here implicitly assume  $w \geq 0$ ). If  $R(w) = c$  where  $c$  is a constant, then we get that  $A(w) = \frac{1}{w}R(w) = \frac{c}{w}$ . Taking the derivative of  $A(w)$  we get that  $A'(w) = -\frac{c}{w^2} < 0$  for all  $w$ . This shows that, assuming a risk averse agent, constant relative risk aversion under expected utility implies decreasing absolute risk aversion. The same argument could be made had we assumed risk seeking agents. The only remaining possibility would be that agents are risk neutral, i.e.  $u''(w) = 0$ . We can verify that under risk neutrality, agents display both constant absolute risk aversion and constant relative risk aversion and that their coefficients are given by 0. Put differently, when displaying constant absolute and relative risk aversion, agents with preferences represented by expected utility will rank random variables by comparing their mean. As is pointed out in Yaari (1987), by remembering the convenient relationship  $Ev = \int_0^1 G_v(t)dt$  and by considering Theorem 1, one can verify that under the dual theory, agents rank random variables by comparing their means if, and only if, the function  $f$  in Theorem 1 is the identity function. Since the allowable constructions of  $f$  for theorem 1 to hold is not limited to the identity function, risk neutrality need not be the case. The agent's attitude towards risk in the dual theory is thus not connected with his attitude towards wealth in the manner of expected utility theory. In fact, the preference functional of the dual theory does not exhibit diminishing marginal utility, but rather constant marginal utility of money, regardless of the risk preferences of the agent. For readers familiar with rank-dependent utility developed by John Quiggin, (interested readers may look up for instance Quiggin (1993)), the dual theory can be recognized as a special case where the transformation of wealth in the preference functional is  $u(x) = x$ . Remember that in rank-dependent utility there is both a transformation of wealth and a transformation of probabilities, while in the dual theory this transformation of wealth is the identity function so in practice there is only a transformation of probabilities.

## 2.4 Characterization of Risk Aversion

Next we wish to consider how risk aversion is characterized under the dual theory. Yaari (1987) defines a preference relation  $\succeq$  to be risk averse if  $u \succeq v$  as long as the following holds

for all  $T$  satisfying  $0 \leq T \leq 1$ , with equality for  $T = 1$ :

$$\int_0^T G_u(t)dt \geq \int_0^T G_v(t)dt.$$

This expression can be rewritten in terms of the cumulative distribution functions  $F_u(t)$  and  $F_v(t)$ , which yields a familiar form

$$\int_0^T [F_v(t) - F_u(t)]dt \geq 0.$$

This can be recognized as second order stochastic dominance. In other words, a preference relation is risk averse if  $u \succeq v$  as long as  $u$  second order stochastically dominates  $v$ . This is an appropriate time to clear up a possible confusion. One can read in much of the literature concerning expected utility theory, for example Eeckhoudt, Gollier, and Schlesinger (2011), that this last integral definition is equivalent to other conditions that are dependent upon the expected utility representation to hold. One may then think that taking this condition to define risk aversion implies that the conditions for the expected utility representation to exist holds, which is not quite the case for the dual theory. Guriev (2001) mentions that the neutrality axiom (A1) is sufficient for writing a dislike towards mean-preserving spreads in terms of distribution functions, which then yields the above integral condition. Röll (1987) shows that defining a dual mean preserving spread in terms of the inverse distribution functions, is no different than the conventional definition of a mean preserving spread in terms of distribution functions. Also note that in expected utility theory, defining risk aversion as a dislike to mean preserving spreads is equivalent to defining risk aversion to be such that the certainty equivalent of any prospect is less than or equal to the expectation of the prospect. The two definitions both follow from the concavity of the utility function in expected utility theory. The definitions are not equivalent in the dual theory. The interested reader is referred to Röll (1987). Hopefully, this clarifies any confusion. In his article, Yaari (1987) proves that an agent exhibits risk aversion if, and only if, the function  $f$  from his representation theorem representing the preference relation  $\succeq$  is nondecreasing (stated in the representation theorem) and convex. The proof will not be repeated here. However, it is instructive to rewrite the representation of an agent's preferences in order to gain intuition about how it works. As is suggested by Yaari (1987), one can use integration by parts to obtain the following:

$$U(v) = \int_0^1 f(G_v(t))dt = f(G_v(t))t \Big|_0^1 - \int_0^1 t f'(G_v(t))G'_v(t)dt.$$

Using the fact that  $f(0) = 0$  and  $f(1) = 1$  and also that  $G_v(t) = 1 - F_v(t)$ , where  $F_v(t)$  is the cumulative distribution function of  $v$ , the expression reduces to

$$U(v) = \int_0^1 t f'(G_v(t))dF_v(t).$$

It may now become clear why  $f$  is often referred to as a distortion function. Firstly one may note that

$$\int_0^1 f'(G_v(t))dF_v(t) = \int_0^1 \frac{d}{dt}[-f(G_v(t))]dt = g(1) - g(0) = 1.$$

In words, the expression  $\int_0^1 f'(G_v(t))dF_v(t)$  are nonnegative weights that sum to 1. By then noticing that  $\int_0^1 t dF_v(t)$  is the mean of  $v$ , it becomes clear that  $U(v)$  is a distorted mean of  $v$ . When calculating the distorted mean, each  $t$  is given a nonnegative weight. We may also note that a convex  $f$ , i.e. a risk averse agent, means that  $f'' \geq 0$ . By this we know that the derivative of  $f$  is nondecreasing. This means that for values of  $t$  generating small values of  $G_v(t)$ , the weights will be low relative to the weights of values of  $t$  generating high values of  $G_v(t)$  (of course as long as the function  $f''$  is not constant for all values of its argument). Since  $G_v(t)$  is a decumulative distribution function, this means that the distorted probability of bad outcomes is higher than the undistorted probability and vice versa. Intuitively, this could be stated as saying that a risk averse agent values uncertain prospects by behaving pessimistically. We may also offer some intuition here in relation to the definition of risk aversion through a dislike for mean preserving spreads. Consider any lottery  $u$  and consider a lottery  $v$  that is a mean preserving spread of  $u$ . We can think intuitively as follows: Since  $v$  is a mean preserving spread of  $u$ , it pays the same on average (by average we mean the expected value), but in some states it pays less and in some states it pays more. For a risk averse agent to dislike  $v$  in relation to  $u$ , he must be more affected by getting less in the bad states than by getting more in the good states. This is exactly what happen when the agent has a convex distortion function. We now illustrate how the distortion function works by a simple example. Imagine a person facing the following prospect:

1. Event A: Lose \$1000 with probability 0.2
2. Event B: Gain \$ 10 with probability 0.3

### 3. Event C: Gain \$100 with probability 0.5

Assuming that this person is risk averse, the weights generated by the expression

$$f'(G_v(t))dF_v(t) := w_i$$

for  $i = A, B, C$  could be  $w_A = 0.5$ ,  $w_B = 0.4$ ,  $w_C = 0.1$ , making him evaluate the prospect lower than if he had been risk neutral, in which case he would have just considered it's expected value.

One could ask the question: Is there a simple way to measure an agent's degree of risk aversion under the dual theory? Put differently, is there a simple way to measure who is the more risk averse of two (or more) agents, similar to the Arrow-Pratt coefficient of absolute risk aversion? In Yaari (1986) this question is raised. He suggests 5 ways to define how comparisons of risk aversion might be carried out. It is beyond the scope of this thesis to present them all, but we mention two of the more obvious definitions. Definition 1 suggests that agent one is more risk averse than agent 2 if there exists a convex function  $g$ , defined on the interval  $[0, 1]$ , such that  $f_1 = g(f_2(p))$  for all  $p$ . In other words, agent 1 is more risk averse than agent 2 if agent 1's distortion function  $f_1$  can be obtained as a convex transformation of agent 2's distortion function  $f_2$ . In practice, finding a function  $g$  that gives  $f_1$  as a transformation of  $f_2$  and checking if it is convex or not could hardly be called a simple way of comparing agents' risk aversion. The fourth definition given in Yaari (1986) is analogous to the Arrow-Pratt coefficient of absolute risk aversion. Under this definition, agent 1 is more risk averse than agent 2 if  $\frac{f_1''(p)}{f_1'(p)} \geq \frac{f_2''(p)}{f_2'(p)}$  holds for all  $0 < p < 1$ . This definition assumes twice differentiable and strictly increasing  $f_1$  and  $f_2$ . It is shown that these two definitions are equivalent when the functions satisfy the above mentioned differentiability conditions. In fact, the theorem also establishes equivalence with definition 2 and 3 in Yaari (1986), as well as establishing that definition 1 implying definition 5. It thus seems that under differentiability conditions, there exists an analogous way to compare risk aversion under the dual theory to the Arrow-Pratt coefficient of absolute risk aversion under expected utility theory. We will however not explore the usefulness of such a measure under the dual theory any further. Definition 1 and 4 mentioned corresponds to the dual case of proposition 1.5 in Eeckhoudt et al. (2011).

# 3. Optimal Insurance: Pure Demand Theory

In this section some well known results regarding insurance contracts will be presented. In its entirety, the section is devoted to pure demand theory, i.e, an insurer is taken as given. The main aim of the section is to compare results developed under expected utility theory, with results developed under the dual theory of choice under risk. Another aim is to give some understanding as to why any differences or similarities of results under the different theories may arise. In addition, following this exposition will hopefully serve any readers unfamiliar with the dual theory well, in that analyses of some simple insurance decisions may provide a better understanding of the dual theory's connection to, as well as its differences from, expected utility. Firstly we will consider the optimal amount of coinsurance that a policyholder demands from an insurer. In this setting, supply of insurance is taken as given, i.e. the models do not include the optimality of the coinsurance contracts from the viewpoint of the insurer. Mossin (1968) shows that a risk averse agent whose preferences are represented by an expected utility function will buy full insurance when the insurance premium is actuarially fair, i.e. the premium is equal to the expected loss. If the insurance premium includes a loading such that the premium is higher than the expected loss, it will not be optimal to buy full insurance. In contrast to this, Doherty and Eeckhoudt (1995) carried out a similar analysis using the dual theory of choice under risk to represent the policyholder's preferences. They found that with no loading, or with a sufficiently small loading, it is optimal to buy full insurance. However, as soon as the loading reaches a threshold, no coverage is optimal.

## 3.1 Optimal Coinsurance: Expected Utility Theory

In order to compare these results more closely, we now introduce some common notation. Let  $W$  be the agent's initial wealth and let the random variable  $X$  represent a loss. The

agent may buy an insurance contract that in the case of a loss pays out  $\alpha X$  and with a premium  $P(\alpha) = \alpha(1 + \lambda)E(X)$ . The loading  $\lambda$  represents a risk premium demanded by the insurer. One may also think of it as cover for the insurer's costs. For simplicity, we denote  $P(\alpha) = \alpha P^*$ , where  $P^* = (1 + \lambda)E(X)$ . To some extent the following derivation of Mossin's theorem is based on the development given in Eeckhoudt et al. (2011). If the agent buys insurance, the utility of his final wealth will be  $u(W - X + \alpha X - \alpha P^*)$ . The objective of the agent seeking insurance is to choose the optimal rate of coinsurance, i.e, to choose the  $\alpha$  that maximizes his expected utility. Since his objective is to maximize his expected utility, we may define a function

$$F(\alpha) := Eu(W - x + \alpha X - \alpha P^*)$$

and maximize this function with respect to the rate of insurance coverage. For simplicity and expositional clarity we assume that certain conditions are fulfilled such that it is possible to differentiate inside the expectation operator. Differentiating  $F(\alpha)$  twice, we get that

$$F''(\alpha) = E[(X - P^*)^2 u''(W - X + \alpha X - \alpha P^*)].$$

Inspecting this we may note that the squared term is always positive, and by the assumption of risk aversion so is  $u''$ . We may conclude that demand for coinsurance is a concave function of the rate of coinsurance  $\alpha$ . This ensures that the optimal rate of coinsurance will be found by considering the first order condition, which is  $E[(X - P^*)u'(W - X + \alpha X - \alpha P^*)] = 0$ . One can easily verify that full insurance is optimal when the loading factor  $\lambda = 0$ , by evaluating  $F'(1)$  which is then equal to 0, i.e, the first order condition for optimality is fulfilled. The aware reader would point out that this is no surprise at all, since with the loading factor  $\lambda = 0$ , the expression for expected utility with full insurance reduces to  $Eu(W - E(X)) \geq Eu(W - X + \alpha X - \alpha E(X))$  when  $\alpha < 1$  by risk aversion. However, with a positive loading factor  $\lambda$ , i.e, with a positive risk premium, one can easily verify that  $F'(1) < 0$ , meaning that it would be optimal to reduce the rate of coinsurance  $\alpha$ . The policyholder is then left better off by keeping some risk, rather than getting rid of all risk. For the sake of comparison with the dual theory, note that, as long as there are positive loading costs/risk premium (which are not too large), there may be interior solutions to the problem of optimal coinsurance. Intuitively, it could be argued that such a solution makes sense. It is

not difficult to imagine being faced with the prospect of a possible loss on some wealth, let's say a boathouse. If one considers the possibility of a devastating storm or perhaps a fire as very unlikely, one would probably not want to pay a high insurance premium (well above the actuarial value of the boathouse), even if it was worth quite a bit. However, one might still not want to gamble with the prospect of losing a valuable asset, regardless of the magnitude of the probability of a loss. In such a case, paying a lower premium to insure at least some of the possible losses would make sense. How an agent's risk aversion varies with wealth should also play a role in determining the agent's willingness to pay for insurance. In the case of decreasing absolute risk aversion, which could for example be represented by a utility function of the form  $u(x) = \log(x)$ , the agent is less willing to pay for insurance the more wealthy that agent is. In fact, as his wealth increases his willingness to pay asymptotically approaches the actuarial value of the loss. Again, this makes intuitive sense. Who would be willing to pay much more than they expect to lose to rid themselves of the risk of losing a small amount of their wealth? If one scales up that possible loss, it does not seem farfetched that faced with transaction costs, one would only be willing to insure some of the possible loss. To not let intuition create confusion, it should be emphasized again that we are now talking about whether or not partly insuring risks through coinsurance makes sense.

## 3.2 Optimal Coinsurance: The Dual Theory of Choice Under Risk

Next we turn to the dual theory. Recall that earlier, we expressed the utility function in the dual theory as  $U(v) = \int w f'(G_v(w)) F'_v(w) dw$ , where  $G_v$  is the decumulative distribution function of  $v$ ,  $F'_v$  is the derivative of the cumulative distribution function of  $v$  (the probability density function of  $v$ ) and  $f$  is a distortion function, which is not to be mistaken for the probability density function of  $v$ . Assuming a risk averse agent, this distortion function  $f$  is convex. However, in order to present the optimal rate of coinsurance we shall adopt the approach of Doherty and Eeckhoudt (1995) which first presents an equivalent formulation of  $U(v)$ . The reason for this is threefold: 1) this formulation is arguably more intuitive than the one first developed by Yaari (1987), as it only makes use of the more familiar concept of the cumulative distribution function of a random variable, rather than the decumulative distribution function. 2) For readers interested in the more detailed exposition given in the

original article, the material should be easier to follow with this consistent approach. 3) Having the approach of the dual theory presented with more familiar concepts might make it more clear why the original formulation of Yaari (1987) is convenient in this case. The equivalent formulation is

$$U(v) = \int wg'(K(w))k(w)dw,$$

where  $K(w)$  is the cumulative distribution function of  $v$ ,  $k(w)$  is the probability density function of  $v$  and  $g$  is increasing and concave, i.e,  $g' > 0$  and  $g'' < 0$ . As before, the agent faces a possible loss  $X$  on his initial wealth  $W = w_0$ . As is pointed out by Doherty and Eeckhoudt (1995), when modeling this situation it will be convenient to use the loss distribution. Consequently, we could formulate this problem as

$$U(v) = \int (w_0 - x)h'(F(x))f(x)dx.$$

With this formulation, risk aversion in the dual theory is present through the convexity of  $h$ . Just as before, the final wealth of the policyholder with insurance is  $w_0 - X + \alpha X - P(\alpha)$  and just as before  $P(\alpha) = \alpha(1 + \lambda)E(X)$ . Omitting some details, and calling this wealth prospect  $v$ , we get that

$$U(v) = w_0 - \alpha(1 + \lambda)E(X) - (1 - \alpha) \int xh'(F(x))f(x)dx.$$

Writing it in this form makes it easy to see that the value of the prospect is linear in  $\alpha$ , something that should be noted. Let us now examine how the value of the prospect changes when we vary  $\alpha$ . Taking the derivative of  $U(v)$  with respect to  $\alpha$ , we get that

$$\frac{\partial U(v)}{\partial \alpha} = \int xh'(F(x))f(x)dx - (1 + \lambda)E(X).$$

The first thing to note is that the first order condition for optimality is independent of the proportion  $\alpha$  of optimal coinsurance. Now we must remember that the the distortion of the probabilities in the first term are such that the distorted expectation will always be larger than the expectation of  $X$ . This will in turn imply that, dependig upon the value of  $\lambda$ , it is either optimal with full coverage or with no coverage. For  $\lambda = 0$  or smaller than some critical value  $c$ , full insurance, ie.  $\alpha = 1$ , is optimal. Once the threshold  $c$  is reached, it becomes optimal to buy no insurance coverage. This is a so-called bang-bang result. Interestingly,

this result is quite different from that which was obtained under expected utility. The reason for this difference is the following. A coinsurance contract takes a linear form, i.e. the wealth level of the policyholder varies linearly with the proportion of coinsurance. But remember that the utility function under the dual theory is linear in payments, meaning that  $U(av + b) = aU(v) + b$ . Before buying insurance, the prospective policyholder is facing a random distribution of wealth. Buying coinsurance transforms this wealth distribution, but in a linear fashion. Thus, since the policyholder is assumed risk averse, when the loading  $\lambda = 0$ , due to risk aversion the policyholder prefers full insurance to no insurance. This also holds when one reduces the level of wealth by a small amount by increasing the price of insurance, i.e. by a loading  $\lambda > 0$ . However, when the price of insurance becomes too high, no matter the proportion  $\alpha$  of coinsurance, the linear transformation of wealth induced by purchasing coverage, becomes less preferred to purchasing no coverage. In contrast, within the expected utility framework wealth is transformed in a nonlinear fashion. Since the insurance contract is linear, this makes it possible to obtain interior solutions, i.e. solutions in which risk sharing between an insurer and a policyholder occurs. As under expected utility, one could make a case for the intuitive appeal of the optimal coinsurance contract under the dual theory. Say for instance that one was considering insuring something that either works perfectly, or does not work at all. If this asset was something crucial to a prospective policyholder, it might not be rational to purchase partial coverage, since in any event as long as a loss occurs, one must replace the asset. Also, consider contemplating purchasing insurance coverage for something of little value (imagine a very risk averse agent). If an agent has risk preferences like that, it might be tedious to be very detail oriented as far as degree of coverage for different items goes. A quick and less tedious solution to get rid of such small risks would be to either insure an item fully, or not insure at all. In any event, one could make a case for both types of preferences. Nevertheless, it is interesting to note that the solutions are quite different, even though the theories share almost exactly the same axiomatic system. Also interesting is the fact that Mossin's theorem, with a small modification when the loading is positive but sufficiently small, holds for the dual theory.

Schlesinger (1997) writes about extending Mossin's theorem to incorporate non-expected utility models such as for instance the dual theory of choice under risk. In his article, which builds on work done on risk aversion by Segal and Spivak (1990), he mentions that Machina (1995) has already extended Mossin's theorem to nonexpected utility models. However, in extending it he implicitly assumed risk aversion of order 2. For definitions of risk aversion

of different orders, see for instance Segal and Spivak (1990). Sufficient for our purpose is to say the following: If an agent exhibits risk aversion of order 2 then, at the limit and when the amount of risk is infinitesimal, the behavior of the agent will be in a risk neutral manner (Schlesinger, 1997). A risk averse agent whose preferences can be represented by expected utility and whose utility function is differentiable satisfy second order risk aversion. If an agent exhibits such preferences, then he will always accept a positive fraction, albeit possibly very small, of any gamble with an expected payoff greater than zero. In contrast, individuals who exhibit risk aversion of order 1 will find some such gambles unacceptable, no matter how small the fraction of the risk they are offered. They will find the expected payoff too small relative to the risk, even though it is positive. Schlesinger (1997) presents a modified version of Mossin's theorem, which holds for risk aversion of any order (and so includes the dual theory):

**Modified Mossin's theorem:** A risk averse individual (not necessarily an expected utility maximizer) buying proportional insurance coverage will choose

1. full coverage ( $\alpha^* = 1$ ) if  $\lambda = 0$
2. full coverage ( $\alpha^* = 1$ ) or partial coverage ( $\alpha^* < 1$ ) if  $\lambda > 0$ .

Note that an agent whose preferences are represented by the dual theory is covered, since the optimal coverage  $\alpha^*$  is not restricted to being larger than zero.

### 3.3 Optimality of Deductibles

Another important result, originally due to Arrow (1974), is that a policyholder that is also an expected utility maximizer always prefers an insurance contract with a straight deductible to a coinsurance contract with the same premium. Mathematically speaking, a contract with a straight deductible exhibits the following indemnity schedule:

$$I(x) = \begin{cases} 0 & x \leq D \\ x & x - D > 0 \end{cases}$$

In words, the insurer pays out nothing if the loss is less than or equal to an amount  $D$ , the deductible. Once the loss exceeds the deductible, the insurer pays out the portion of the

loss which exceeds this limit. It should be noted here that the optimality of deductibles is not inconsistent with Mossin's theorem. Mossin's theorem deals with risk averse individuals buying proportional insurance coverage. It does not consider optimality over the whole set of possible insurance contracts, only over the set of possible coinsurance contracts. Doherty and Eeckhoudt (1995) talks about the optimality of deductibles as well as claiming that the result also holds under the dual theory, but they do not go as far as demonstrating it in a formal manner. Gollier and Schlesinger (1996) provides a very neat proof of the optimality of deductible insurance contracts, which is entirely independent of the agents being expected utility maximizers. Contrary to other proofs of Arrow's theorem, for instance those made under the assumption of expected utility maximizing agents, this proof makes no use of optimal control theory or the calculus of variations. The proof is thus both more general, more accessible and arguably it builds much more intuition for why straight deductibles are optimal compared to the more mathematically advanced proofs. Gollier and Schlesinger (1996) show that policies with a deductible second-degree stochastically dominates any other insurance policy with the same premium. The proof thus only depends on a preference functional to be such that a risk averse agent dislikes mean preserving spreads in the sense of Stiglitz and Rothschild (1970), which is equivalent to second order stochastic dominance. In other words, the preference functional must satisfy second degree stochastic dominance. The reader may recall that Yaari (1987) defines risk aversion in terms of second order stochastic dominance when proving that a risk averse agent has a convex distortion function. It can also be seen in Guriev (2001) how an agent that dislikes mean preserving spreads under the dual theory is consistent with a risk averse agent under the dual theory. S. S. Wang and Young (1998) in their article about the ordering of risks in the dual theory versus in expected utility theory, demonstrate that the preference functional admitted from the dual theory satisfies second degree stochastic dominance. They develop second degree dual stochastic dominance, and then show it's equivalence to second degree stochastic dominance. The equivalence holds also for first order stochastic dominance, but not for higher orders than two. A more direct proof that risk averse agents with dual theory preference functionals dislikes mean preserving spreads can be found in Demers and Demers (1990). On the basis of this we conclude that Arrow's theorem is robust enough to be directly extended to the dual theory. Again, it is interesting to note that a theorem developed originally under the strict assumptions of expected utility theory is robust enough to be extended beyond the confines of it's original assumptions.

### 3.4 Optimal Reinsurance

Lastly, we will present an interesting result from S. S. Wang and Young (1998). The article in question is about the ordering of risks in the dual theory of choice under risk compared to the ordering of risks under expected utility. We shall not stray to far away from the subject matter of this section, but the result in itself bears great similarities to previous results presented, and as we shall explain, these similarities arise from the same background. The result is about the optimal reinsurance contract under the dual theory of choice. To reduce the need for introducing new notation and to avoid reiteration of material, we ask the reader to first read the section below about distortion risk measures and then return to this result.

**Definition:** (S. S. Wang & Young, 1998) The reinsurance contract  $I^* \in \mathbb{I}_p$  is the optimal reinsurance contract with respect to  $H_g[X - I^*(X)] < H_g[X - I(X)]$  for all  $I \in \mathbb{I}_p$ . That is, the valuation operator applied to the retained claims,  $X - I(X)$ , is minimum for the optimal reinsurance contract.

$\mathbb{I}_p$  is the set of all indemnity schedules satisfying usual assumptions, namely that  $I(0) = 0$ ,  $0 \leq I' \leq 1$  and  $E[I(X)] = P$ . We now present the result, which is also from S. S. Wang and Young (1998):

**Result:** The optimal reinsurance contract  $I^* \in \mathbb{I}_p$  is of the form  $I^*(X) = (x - d)_+$ , in which  $d$  is defined by  $E[(X - d)_+] = P$  and in which  $g$  is a concave distortion.

As is mentioned in the section about distortion risk measures, a concave distortion function of the valuation operator is equivalent to a risk averse agent.

A short proof in S. S. Wang and Young (1998) demonstrates why so called stop-loss contracts (straight deductibles as defined above) are optimal reinsurance contracts. The proof is based on the fact that the agent in question satisfies second degree stochastic dominance, which we now know to be true. The main takeaway here is that the optimal reinsurance contract is of the same type as the optimal insurance contract, and these results both follow from risk aversion. It should also be noted that optimal reinsurance contracts under expected utility theory are also stop-loss contracts. A simple and straightforward proof of the optimality of stop-loss reinsurance contracts under expected utility theory is found in Gerber and Pafum (1998).

## 4. Pareto Optimal Risk Sharing: Expected Utility Theory

This section will deal with Pareto optimal allocations of risk. It will review some classical results of Pareto optimal risk sharing derived under the expected utility framework. As in the previous section, the counterparts to these results will be presented, the counterparts here being results of Pareto optimal risk sharing derived under the dual theory. However, these counterparts will be presented in section six. As a reminder to the reader, here is the verbal definition of a Pareto optimal sharing rule given in Wilson (1968): "A sharing rule is Pareto optimal if there is no alternative sharing rule which would increase the utility of some member(s) without decreasing the (expected) utility of any other members." His use of the word members is due to the fact that he wrote about risk sharing in syndicates. For the interested reader, Wilson developed his theory under the axioms of Savage (1972), published originally in 1954. These axioms made it mandatory for agents to have only subjective probability measures. Wilson showed that individual risks do not matter for expected utility maximizers. The optimal consumption of each individual is given by a nondecreasing function of the total risk. In our exposition we only consider risk sharing under one objective probability measure. Note that we put "expected" in parenthesis as Wilson worked with expected utilities. Replacing "expected" with "dual" still gives a correct definition. As we shall see, these sharing rules will be a set of functions.

### 4.1 The Benefits of Risk Sharing

Consider figure 1, which illustrates the benefits of risk sharing.  $U_1(W_1)$  and  $U_2(W_2)$  represents the utility to the respective agents 1 and 2. Note that the utility functions are not restricted to the expected utility representation, they could also for example take the form of a dual utility, i.e  $U(v)$  as presented earlier. The arc connecting the x-axis and the y-axis is called the Pareto frontier. It represents all Pareto optimal allocations of risk between agent 1 and

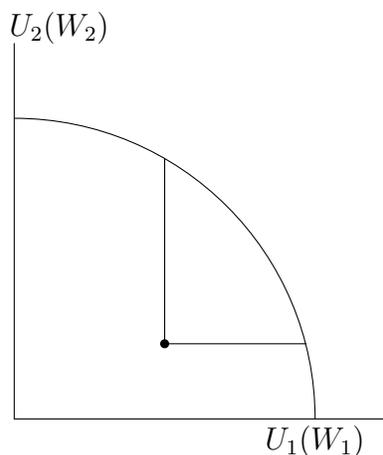


Figure 4.1: The benefits of risk sharing

agent 2. The initial allocation of risk is given by the black dot inside the Pareto frontier. For a Pareto optimal allocation to be individually rational, the agents must be at least as well off after the reallocation of risk as they were in the initial situation. This condition is graphically represented by the two lines originating from the initial allocation of risk. It should be clear that all solutions to the right of and above the black dot are utility improving for both agents and as such individually rational. However, remember from our definition of Pareto optimality given above, that for an allocation to be Pareto optimal there should not exist another allocation that is utility improving for at least one agent, without decreasing the utility of any other agent. It should then be clear that the individually rational Pareto optimal allocations of risk are given by the part of the Pareto frontier which is limited by both individual rationality constraints. For interested readers, Aase (2008) gives a nice overview of optimal risk sharing under expected utility theory. He presents much of what is presented here, but in a more modern setting, using amongst others results from functional analysis to present the material.

## 4.2 Risk Sharing in the Absence of Costs

In the following, the set up of Borch (1962) will be briefly presented. It is worth noting that it was stated by Borch (1962, p. 426) "there has been considerable controversy over the plausibility of the various formulations which can be given to these axioms." By this

he was of course referring to the axioms of the expected utility theorem. However, he next dismissed these controversies in the context of reinsurance, stating: "there is no need to take up this question here, since it is almost trivial that the Bernoulli hypothesis must hold for a company in the insurance business." This illustrates how expected utility theory may be well suited to analyze problems of optimal risk sharing, and especially in the context of insurance. Even though this statement is given without a convincing argument, it is the author's impression that similar thoughts are prevalent today.

1. There are  $n$  insurance companies, each having their own portfolio of insurance contracts.
2. There are two elements defining each company's risk situation
  - $F_i(x_i)$ , the probability that the total amount of claims to be paid in the company's portfolio do not exceed  $x_i$
  - The funds,  $S_i$ , which the company has available to pay claims

$x_1, \dots, x_n$  are assumed stochastically independent. Now, Borch (1962) turns to what he calls the "Bernoulli hypothesis", what might be better known to a younger audience as the expected utility theorem. Prior to negotiations, a company will have attached to its risk situation the following utility

$$U_i(S_i, F_i(x_i)) = \int_0^\infty u_i(S_i - x_i) dF_i(x_i).$$

Initially, company  $i$  must pay  $x_i$ , i.e., its portfolio of claims. In the reinsurance markets, treaties can be negotiated which redistributes the amount of claims the firms are obligated to pay initially. Such treaties could be represented as a set of functions:

$$y_i(x_1, x_2, \dots, x_n), \quad (i = 1, 2, \dots, n).$$

$y_i(x_1, x_2, \dots, x_n)$  is the amount company  $i$  must pay if claims in all the initial portfolios amount to  $x_1, x_2, \dots, x_n$ . These new treaties must satisfy

$$\sum_{i=1}^n y_i(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i.$$

Each company will then have their utility changed to

$$U_i(y) = \int_R u_i(S_i - y_i(x))dF(x),$$

where  $F(x)$  is the joint probability distribution of  $x_1, x_2, \dots, x_n$ , and  $R$  is the positive orthant in  $n$ -dimensional  $x$ -space.

What Borch (1962) derives in his paper are Pareto optimal sharing rules of risk (of course the broader scope of the paper is an attempt to derive equilibrium in a reinsurance market, as the title of the paper states). If we have one set of treaties represented by a vector  $y$  and there exists another set of treaties represented by  $\bar{y}$ , such that

$$U_i(y) \leq U_i(\bar{y})$$

with at least one inequality, it is obvious that  $\bar{y}$  is preferred to  $y$ . If no such  $\bar{y}$  exists, we say that the set of treaties represented by  $y$  is Pareto optimal. Borch goes on to derive a differential equation which shows that the amount  $y_i(z)$  which each company must pay will depend only on  $z = x_1 + x_2 + \dots + x_n$ . This is a noteworthy property. It says that each firm's Pareto optimal allocation of risk only depends on the total amount of claims from all insurers. This effectively shows that any Pareto optimal set of treaties is equivalent to a pool arrangement. In other words, all the companies' portfolios are handed over to a pool where the claims against the pool are distributed among the members according to some Pareto optimal sharing rule. For the record, as well as for comparison purposes later on, the differential equation is stated here:

$$\frac{dy_i(z)}{dz} = \frac{\frac{k_i}{u_i''(S_i - y_i(z))}}{\sum_{j=1}^n \frac{k_j}{u_j''(S_j - y_j(z))}}$$

Moffet (1979) was, perhaps, the first to use Borch's theorem from Borch (1960) to illustrate Pareto optimal risk sharing between a policyholder and an insurer. His statement of the theorem is as follows (specialized to two agents: "If  $u_i'(\cdot) > 0$  and  $u_i''(\cdot) < 0$ , for  $i = 1, 2$ , then an arrangement  $\{y_1, y_2\}$  is Pareto optimal if and only if

$$u_i'[W_i - y_i(x)] = k_i u_1'[W_1 - y_1(x)]$$

where  $k_2$  is a positive constant chosen arbitrarily and  $k_1 = 1$ . We also remind the reader that the treaties depend only upon the total amount of claims amongst all firms, i.e.

$$y_i(x) = f_i\left(\sum_{i=1}^n x_i\right) \text{ such that } f_i : \mathbb{R} \mapsto \mathbb{R} \text{ is nondecreasing,}$$

$i = 1, 2$  if the arrangement is Pareto optimal. As we did in the section of optimal insurance, we will now consider optimal insurance contracts, but in a broader setting. There are two important differences to this analysis from those in the preceding section: 1) We now consider the optimality of contracts from the perspective of both supply and demand simultaneously. This is in contrast to the preceding section, where we only considered the optimality of contracts for a prospective policyholder, i.e, we only considered insurance demand with supply taken as given. 2) Our criterion for the optimality of contracts is Pareto optimality, which is a natural criterion with more than one agent. In what follows, some outtakes of Moffet (1979) will be presented.

Consider two individuals, individual 1 is a potential policyholder and individual 2 is an insurer. Suppose they have initial wealth  $W_1$  and  $W_2$  respectively, and that individual 1 faces a random loss  $X$ . Then if individual 1 pays a premium  $P$ , he will receive an indemnity  $I(x)$  should a loss  $x$  occur. Applying Borch's theorem, a Pareto optimal insurance scheme is then given by

$$u_2'[W_2 + P - I(x)] = k u_1'[W_1 - P - x + I(x)],$$

where we skip the subscript on  $k_2$  for notational convenience. To obtain a more informative solution, one must reduce the set of feasible solutions. A natural initial condition is  $I(0) = 0$ . The condition simply states that there is no payouts from the insurer if the policyholder does not incur a loss. This initial condition makes it possible to determine  $k$  by evaluating the Pareto optimal insurance scheme at the initial condition and solving for  $k$ . Doing this, and substituting back for  $k$  in the Pareto optimal insurance scheme yields

$$\frac{u_2'[W_2 + P - I(x)]}{u_2'[W_2 + P]} = \frac{u_1'[W_1 - P - x + I(x)]}{u_1'[W_1 - P]}$$

From this, one can solve for the change in the optimal indemnity given an increase in the

loss  $x$ . The expression, which looks a little bit messy at first glance, is

$$\frac{\partial I(x)}{\partial x} = \frac{u_2'[W_2 + P]u_1''[W_1 - P - x + I(x)]}{u_1'[W_1 - P]u_2''[W_2 + P - I(x)] + u_2'[W_2 + P]u_1''[W_1 - P - x + I(x)]}$$

After inspecting the expression, it should be quite clear that  $0 < \frac{\partial I(x)}{\partial x} < 1$ . It also follows, since  $I(0) = 0$ , and by the mean-value theorem that  $0 < I(x) < x$  for  $x > 0$ . In other words, under the assumption that both the insurer and the policyholder are risk averse, coinsurance is the Pareto optimal insurance scheme. One can recall from the previous section that for the policyholder, taking the insurance sector as given, the optimal insurance policy is one with a straight deductible. It is perhaps not surprising that this contract is not Pareto optimal when both agents are risk averse, as one would intuitively expect that what is the most optimal for agent 1 is generally not optimal for agent 2. Mathematically, one can easily verify that insurance with a straight deductible is not Pareto optimal when both agents are risk averse. Recall that a deductible insurance contract is given by

$$I(x) = \begin{cases} 0 & x \leq D \\ x & x - D > 0 \end{cases}$$

Thus,  $\frac{\partial I(x)}{\partial x} = 0$  for  $x < D$  and  $\frac{\partial I(x)}{\partial x} = 1$  for  $x > D$ , which violates  $0 < \frac{\partial I(x)}{\partial x} < 1$ . As is mentioned by Moffet (1979), insurance policies with a deductible are often observed in practice, so this raises questions about whether or not the theory is flawed and whether or not people act irrationally. At first glance one can observe that the theory does not take into account factors such as different costs and their structures. As this is of course prevalent in practice, it would seem overly confident to make statements about people's rationality from such a simplified analysis. We will return to a discussion of such questions in section 7.

We round off our review of Moffet (1979) by presenting an approximation of the Pareto optimal  $I(x)$ , as well as presenting a closed form solution to  $I(x)$  for a specific utility function that corresponds exactly to the approximation. The approximation is given by

$$I(x) = \frac{A_1(W_1 - P)}{A_1(W_1 - P) + A_2(W_2 + P)}x,$$

where  $A_i(\cdot)$  is the Arrow-Pratt measure of absolute risk aversion as defined in section 1. As with other HARA utility functions, exponential utility of the form  $u_1(x) = 1 - e^{-ax}$  and

$u_2(x) = 1 - e^{-bx}$ , where  $a, b > 0$ , yields a very nice expression for  $I(x)$ , namely

$$I(x) = \frac{a}{a+b}x$$

We see that the Pareto optimal insurance contract when both the insurer's and the policyholder's preferences are represented by exponential utility functions are completely determined by the agents' constant coefficient of absolute risk aversion. The Pareto optimal contract is then coinsurance, where the rate of coinsurance is equal to the policyholder's share of the sum of the absolute risk aversion between the two agents.

We now introduce the concept of risk tolerance. An agent's risk tolerance is the inverse of an agent's risk aversion. Mathematically,

$$T(w) = \frac{1}{A(w)} = \frac{-u'(w)}{u''(w)}.$$

In the case of infinitesimal risks, Wilson (1968, p. 121) offers a verbal explanation (paraphrased): "We can interpret the risk tolerance  $T(w)$  as half the tolerable variance per unit of compensating risk premium for infinitesimal risks when the income of agent  $i$  is  $x_i$ ". One can derive a general expression that is exact for Pareto optimal indemnity schedules for the class of HARA utility functions with equal degree of cautiousness. The agents' risk tolerances are then linear in wealth  $w$ , i.e, they are on a form  $t_i = \beta_i + \alpha w$ , where  $\alpha$  is the cautiousness parameter and  $\alpha, \beta_i \geq 0$ . We then get the following general expression for the optimal indemnity schedule:

$$I_p(x) = \frac{\beta_{PH} + \alpha(w_{PH} + p)}{\beta_I + \beta_{PH} + \alpha(w_{PH} + w_I)}x,$$

where  $p$  is the insurance premium. It is thus clear that for general HARA utility with equal cautiousness, Pareto optimal indemnity schedules are linear. Below is an illustration of different insurance contracts. The horizontal axis measures the loss while the vertical axis measures the indemnity paid out.

One should note that the policies in figure 4.2 and figure 4.3 will never be Pareto optimal under expected utility theory in the absence of costs. As was shown above in the case of a deductible insurance policy (figure 4.2), this is because the Pareto optimal indemnity

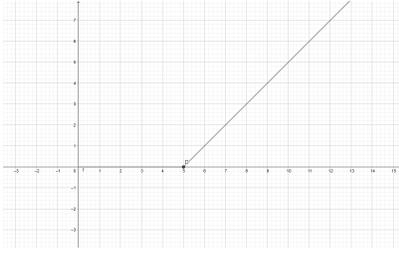


Figure 4.2: Policy With a Deductible

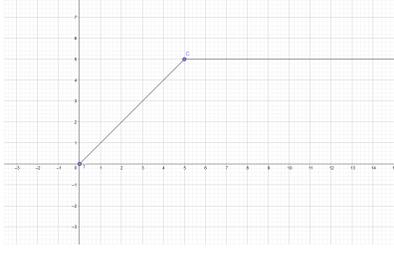


Figure 4.3: Policy With a Ceiling

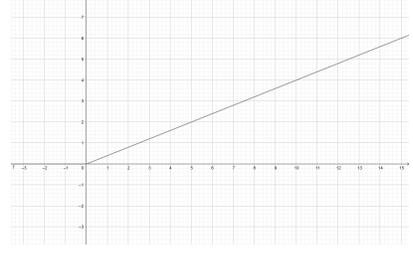


Figure 4.4: Pareto Optimal with HARA Utility

schedule has the following characteristic:

$$0 < \frac{\partial I(x)}{\partial x} < 1,$$

and the argument for the non-optimality of contracts such as the one in figure 4.3 is analogous to the argument for the non-optimality of contracts with a deductible.

Lemaire (1991), originally published in 1979, calculated Pareto optimal treaties in reinsurance using exponential utility functions of the form  $u_j(x) = \frac{1}{c_j}(1 - e^{-c_j x})$ . Using Borch's theorem under the constraint that the sum of all treaties must equal the sum of all claims, yields Pareto optimal sharing rules of the following interesting form:

$$y_j(x) = q_j z + y_j(0),$$

where

$$q_j = \frac{1/c_j}{\sum_{i=1}^n \frac{1}{c_i}} \text{ and } y_j(0) = W_j - q_j \sum_{i=1}^n \left( W_i + \frac{1}{c_i} \log \frac{k_i}{k_j} \right),$$

where  $W_i$  is agent  $i$ 's initial wealth for  $i = 1, 2, \dots, n$ . Sufficive for gaining intuition of the above Pareto optimal sharing rule, remember that an agent's risk tolerance function is a measure of how much risk the agent is willing to cope with. The reader is reminded that the agents' risk tolerance here is given by  $\frac{1}{c_i}$ , for  $i = 1, 2, \dots, n$ . We then see that each agent pays a share  $q_j$  of each claim. The size of this share is determined by how well that agent copes with risk. The higher the risk tolerance, the higher the share of each claim paid by the agent. In addition to this, each agent may pay side-payments to other agents. This compensates the least risk averse companies for paying greater amounts of the claims. Note that these

side payments sum to zero, i.e.,  $\sum_{j=1}^n y_j(0) = 0$ . This should be expected, since the optimal sharing rules are derived under the constraint that the sum of all treaties equals the sum of all wealth. Nothing disappears in pure exchange. All risk is still there, it is just redistributed. For a nice overview of the structure of the Pareto optimal risk exchanges with proofs, and well explained illustrations of the same point made in Lemaire (1991) and more, the reader is referred to Gerber and Pafum (1998). One thing that should be noted, which is quite easy to see from example 7 in Gerber and Pafum (1998), is that under the assumption that all agents have an exponential utility function, the only thing that differs between Pareto optimal risk exchanges are the side payments. All Pareto optimal risk exchanges are characterized by a fixed quota given by each agent's risk tolerance divided by the sum of all risk tolerances, and side payments that sum to 0. Another set of side payments that sum to 0 gives a new Pareto optimal risk exchange with the same fixed quota.

There is a nice result which is a direct consequence of Borch's theorem (Aase, 2008), that first appeared in Wilson (1968). In a Pareto optimum, the sum of the risk tolerances of each agent equals the risk tolerance of the market (which can be modeled by a representative agent). Mathematically

$$T_M = \sum_i^n t_i.$$

An interesting and immediate consequence of this is that a risk neutral agent will carry all available risk in a Pareto optimum. The intuition for this is that he is completely indifferent to risk, so it will always be utility improving for non risk neutral agents to transfer risk to him, without his utility being reduced (as long as his expected value after the risk sharing arrangement remains the same or higher).

### 4.3 Risk Sharing in the Presence of Costs

As we pointed to when reviewing Moffet (1979), the Pareto optimal sharing of risk between an insurer and a policyholder did not include any mentioning of costs. Raviv (1979) extends his analysis to include a cost function  $c(I)$ , which is a function of the indemnity schedule in the insurance contract. This cost function is meant to capture both fixed and variable cost. It is assumed that  $c(0) = a \geq 0$ ,  $c'(\cdot) \geq 0$  and  $c''(\cdot) \geq 0$ . In words, it is assumed that the cost function includes fixed costs that may equal zero, and that costs may increase

disproportionately (at least proportionately) with the size of the indemnity. It may be noted from Aase (2017) that this assumption of a convex cost function permits the optimization problem below to be convex which will lead to a maximum. This assumption, which may in certain situations not be intuitive, is thus made to simplify the mathematical problem. Concave costs would not guarantee the existence of a maximum. Similar to before, the insurer (the supply side) is assumed to maximize his expected utility, which is given by  $E\{V[W_0 + P - I(x) - c(I(x))]\}$  and which is subject to an individual rationality constraint, i.e  $E\{V[W_0 + P - I(x) - c(I(x))]\} \geq V(W_0)$ . For clarity,  $W_0$  represents the insurer's initial wealth. The policyholder is assumed to be strictly risk averse ( $U'(w) > 0, U''(w) < 0$ ) with initial wealth  $w$ , and subject to a potential loss  $x$ . He also faces an individual rationality constraint, namely  $E\{U[w - P - x + I(x)]\} \geq E\{U[w - x]\}$ . The derivation of Pareto optimal contracts make use of optimal control theory. Proofs and detailed derivations will not be repeated here as the economic intuition that is to be gained from those are small at best, unless the reader has a thorough understanding of the mathematical framework. However, the general outline of the approach given in Raviv (1979) could be instructive to note. The problem is stated as follows:

$$\max_{P, I(x)} \bar{U}(P, I) := \int_0^t U[w - P - x + I(x)]f(x)dx$$

subject to

$$0 \leq I(x) \leq x$$

and

$$\bar{V}(P, I) := \int_0^t V[W_0 + P - I(x) - c(I(x))]f(x)dx \geq k$$

where  $k$  is a constant and  $k \geq V(W_0)$ .

Verbally, the problem is to find the premium  $P$  and indemnity schedule  $I(x)$  that maximize the final wealth of the policyholder. The first constraint says that the indemnity schedule cannot exceed the loss and the second constraint says that the expected utility of final wealth of the insurer must be at least as big as the insurer's utility of initial wealth. In other words, we seek insurance contracts described by the pair  $(P, I(x))$  that gives the policyholder the maximum attainable utility while not making the insurer worse off in any way, and without the possibility of unreasonable indemnity schedules. It should be noted that varying the value of  $k$  will generate the Pareto optimal frontier, yielding a situation as

in figure 1 above. Raviv (1979) solves the problem in two steps. First he assumes a fixed  $P$  and optimal insurance is derived as a function of  $P$ . Thereafter, the optimal  $P$  is found, and thereby the problem is solved.

When  $P$  is fixed, Raviv (1979) finds two possible forms of the optimal indemnity schedule  $I^*(x)$ :

$$1. \begin{cases} I^*(x) = 0 & x \leq \bar{x}_1 \\ 0 \leq I^*(x) \leq x & x > \bar{x}_1 \end{cases}$$

$$2. \begin{cases} I^*(x) = x & x \leq \bar{x}_2 \\ 0 \leq I^*(x) \leq x & x > \bar{x}_2 \end{cases}$$

1) is a deductible insurance policy, with  $\bar{x}_1 = D$  as the deductible and coinsurance above the deductible. It thus may differ from the optimal contract for the policyholder, which consists of a straight deductible  $x - \bar{x}_1$  for losses exceeding  $\bar{x}_1$ . 2) is called a policy with an upper limit of full insurance. Below the loss  $\bar{x}_2$ , the indemnity schedule equals the loss. Above  $\bar{x}_2$ , the indemnity schedule takes the form of coinsurance. One may also observe that in both cases, when  $0 < I^*(x) < x$ , we have that

$$I^{*'}(x) = \frac{A_{PH}(Y)}{A_{PH}(Y) + A_I(Z)(1 + c') + c''/(1 + c')}$$

where  $Y = w - P - x + I^*(x)$  and  $Z = W_0 + P - I^*(x) - c(I^*(x))$ . As before  $A_i(\cdot)$  denotes agent  $i$ 's index of absolute risk aversion. One should note the similarity to the approximation for  $I(x)$  derived in Moffet (1979). One should also note that this expression is not linear in  $x$ , as was the case for general HARA utility with equal cautiousness.

Raviv (1979) goes on to show that policies with upper limits are dominated by policies without upper limits. In other words, he shows that the Pareto optimal policies are not of the upper limit type. From this we can infer that Pareto optimal contracts must be of type 1 as defined above. The question still remains: Under what conditions do we get Pareto optimal policies with a deductible greater than 0? The answer, also given by Raviv (1979, p. 90) is: "A necessary and sufficient condition for the Pareto optimal deductible to be equal to zero is  $c'(\cdot) = 0$  (i.e.  $c(I) = a$  for all  $I$ )." This means that a Pareto optimal contract does not include a deductible if, and only if, the cost function is constant. This is in line with what we saw earlier, where policies with a deductible were never Pareto optimal when costs were not included. It is therefore noteworthy that costs may be present and the results

from Moffet (1979) still holds, however only if costs are fixed. As soon as the costs depend on the indemnity schedule, Pareto optimal contracts will include a deductible. It should be mentioned here that Aase (2017) shows that Pareto optimal contracts contains a deductible also in the presence of quasi-fixed costs, regardless of whether or not actual variable costs are zero. A quasi-fixed costs in such that, whenever a claim is made, regardless of it's size, a cost  $a > 0$  is incurred. Also noteworthy is the fact that the result of Raviv (1979) does not depend on risk preferences of either the insurer or the policyholder. This observation sheds light on the robust result of Arrow (1974), which was presented in the section about optimal insurance contracts. In that section, insurers were taken to be risk neutral and costs were assumed to be proportional to the indemnity schedule. One could also interpret the insurers to have been risk averse, with the risk premium included in the loading. It is now possible to deduce, at least in the case of agents being expected utility maximizers, that the optimality of deductibles in that section was due to the cost structure and not due to other assumptions such as risk neutrality (one could view insurers in that section as risk neutral if the loading only includes costs, and not a risk premium). However, the assumption of risk neutrality yielded straight deductibles. A risk averse agent with proportional costs  $\lambda$ , would obtain a coinsurance for losses above the deductible. It can be seen that the level of coinsurance is given by

$$I^{*'}(x) = \frac{A_{PH}(Y)}{A_{PH}(Y) + A_I(Z)(1 + \lambda)} < 1$$

in accordance with the expression for  $I^{*'}(x)$  presented above. It should also be mentioned that coinsurance above the deductibles may arise not just because of risk aversion, but because of strictly convex cost functions. This may happen even if the agent is risk neutral. Intuitively this makes sense by noticing that while risk aversion implies that an agent's utility function is nonlinear, strictly convex cost functions mean cost function nonlinearity. In the case of risk neutral agents and coinsurance above the deductible, one can thus think about the nonlinearity of the cost function acting in place of the nonlinearity that would arise with a risk averse agent.

# 5. Distortion Risk Measures

## 5.1 Introduction to Distortion Risk Measures

To make this thesis self contained, it is necessary with a brief introduction to distortion risk measures. Firstly, we briefly mention what a risk measure is. To do this, we quote Denuit, Dhaene, Goovaerts, Kaas, and Laeven (2006, p. 1): "Mathematically, a risk measure is a mapping from a class of random variables to the real line. Economically, a risk measure should capture the preferences of the decision maker." Distortion risk measures are risk measures that were first introduced to insurance pricing by S. Wang (1996). However, the reader has already been introduced to this risk measure in some sense. Distortion risk measures are more or less the preference relation that Yaari (1987) developed, which has already been presented. In fact, they coincide when agents are risk averse, which is pointed out in T. Boonen (2013) and which will hopefully be clear shortly. S. Wang (1996) has extended Yaari's preference relation to hold for unbounded random variables and has also provided an axiomatic foundation in S. S. Wang, Young, and Panjer (1997) that justifies a proposed insurance premium principle given in S. Wang (1996). It follows from these axioms that one can describe the price of an insurance risk by a Choquet integral representation with respect to a distorted probability. For readers unfamiliar with the Choquet integral, a short introduction and a review of some interesting applications can be found in Z. Wang and Yan (2006).

The axioms of S. S. Wang et al. (1997) will not be presented here. Properties of distortion risk measures and their connection to the dual theory is the subject matter of interest. The curious, and mathematically well oriented reader, is referred to S. S. Wang et al. (1997) for details. Hopefully, this brief exposition will make the unfamiliar reader capable of following the coming analysis of Pareto optimal risk sharing under the dual theory. We start out by recalling that under the dual theory, one can represent the certainty equivalent of a bounded

economic prospect as a Choquet integral:

$$\int_0^1 g[G_v(t)]dt$$

where  $G_v(t)$  is the decumulative distribution function of a random variable  $v$  defined on the unit interval,  $g$  is increasing,  $g(0) = 0$  and  $g(1) = 1$ . It can then be shown that the axioms postulated in S. S. Wang et al. (1997) are both necessary and sufficient conditions for a premium functional  $H$  to have a Choquet integral representation as follows:

$$H[X] = \int Xd(g \circ P) = \int_0^\infty g[G_X(t)]dt,$$

where  $g$  is increasing with  $g = 0$  and  $g = 1$ . The choice of words "premium functional" may be misleading. We choose to take this up here since the term was used in S. S. Wang et al. (1997). This is not to be confused with premium functionals that price insurance contracts in financial markets with symmetric information and no arbitrage. Such premium functionals must be linear (Albrecht, 1992), i.e,  $H[aX + bY] = aH[X] + bh[Y]$ . As will be seen later in this section, this is generally not the case for  $H[X]$ , which is normally referred to as a premium principle. To avoid confusion we thus refrain from calling  $H[X]$  a premium functional, and adhere to the expression premium principle. Albrecht (1992) talks about the implications for insurance pricing when restricted to linear premium functionals. Properties for our insurance premium principle will for the most part not be mentioned here as the properties central to our later exposition may be presented when necessary. We will however take note of the following property: If  $g$  is concave, then  $H$  preserves second order stochastic dominance, i.e,

$$\int_x^\infty G_X(t)dt \leq \int_x^\infty G_Y(t)dt \text{ for all } x \geq 0 \implies H[X] \leq H[Y].$$

This is an appropriate time to clear up a possible confusion. The attentive reader may remember that risk aversion under the dual theory is represented by a convex distortion function. But distortion risk measures are characterized by a concave distortion function. To clear up this confusion, we follow S. Wang (1996). Assume an agent with initial wealth 1 facing a random loss  $X$  on the interval  $[0, 1]$ . He thus faces the following prospect  $V = 1 - X$ . Denote by  $P$  the insurance premium for risk  $X$  (remember that distortion risk measures

could be interpreted as insurance premiums). Assume a risk averse agent, i.e. an agent whose distortion function  $h$  is convex. The aim is now to show that the insurance premium could be represented by a concave distortion function, without violating risk aversion. We have the following relation:

$$1 - P = \int_0^1 h[G_{1-x}(t)]dt.$$

We also have that  $G_{1-x}(t) = \Pr\{X < 1 - t\}$ , which yields

$$1 - P = \int_0^1 h[1 - G_X(1 - t)]dt = \int_0^1 h[1 - G_X(z)]dz$$

Solving for  $P$  yields

$$P = \int_0^1 g[G_X(z)]dz, \text{ where } g(x) = 1 - h(1 - x).$$

One can also verify easily that  $g$  is concave if and only if  $h$  is convex:  $g'(x) = h'(1 - x)$  and  $g''(x) = -h''(1 - x)$  which means that in order for  $g''(x) \leq 0$  which is weak concavity, we must have that  $h''(1 - x) \geq 0$  which is weak convexity. Finally, note that distortion risk measures can be extended to unbounded real-valued random variables, in which case the mathematical formulation is

$$H[Y] = \int Xd(g \circ P) = \int_{-\infty}^0 (g[G_X(t)] - 1)dt + \int_0^{\infty} g[G_X(t)]dt.$$

## 5.2 Properties of Distortion Risk Measures

We round of this brief introduction of distortion risk measures by listing some properties. Most of these are formulated as in S. S. Wang and Young (1998), but for clarity with later material we also refer to Ludkovski and Young (2009):

1.  $H_g[-X] = -H_h[X]$ , where  $h(p) = 1 - g(1 - p), 0 \leq p \leq 1$ . We recognize this from above, where we showed that  $g$  is concave if and only if  $h$  is convex.
2. If  $g(p) \geq p$  for all  $p \in [0, 1]$ , then  $H_g[X] \geq E[X]$ . With a concave distortion function the distortion risk measure takes a value at least as high as the expected value of the random variable  $X$ .

3. Also if  $g$  is concave and  $a, b \geq 0$ , we have  $H_g[aX + b] = aH_g[X] + b$ . Verbally, the distortion risk measure is translation invariant and exhibits positive homogeneity.
4. For concave  $g$ :  $H_g[X + Y] \leq H_g[X] + H_g[Y]$ . For concave distortion function distortion risk measures are subadditive. In connection to what we mentioned about the term premium functional, we may now note that the insurance premium principle is not generally linear.
5. For convex  $g$ :  $H_g[X + Y] \geq H_g[X] + H_g[Y]$ . For convex distortion function distortion risk measures are superadditive.
6. For comonotonic  $X$  and  $Y$ :  $H_g[X + Y] = H_g[X] + H_g[Y]$ . Distortion risk measures are comonotonic additive.
7. Distortion risk measures preserves convex ordering. Ordering by a convex order is equivalent to second degree stochastic dominance with equal means. Thus, for readers familiar with second order stochastic dominance and expected utility (as mentioned before) one can think of distortion risk measures preserving second degree stochastic dominance in the sense of Stiglitz and Rothschild (1970).

Wirch and Hardy (2001) shows that distortion risk measures strictly preserves second order stochastic dominance if and only if the distortion function is strictly concave. In math we have that if  $X \succ_{2nd} Y$ , then  $H_g[X] > H_g[Y]$  if, and only if, the distortion function  $g$  is strictly concave. If the distortion function is not strictly concave, i.e,  $g''(t) = 0$  for some  $t$  then we may have a situation where  $X \succ_{2nd} Y$  and  $H_g[X] = H_g[Y]$ . For convenience we remind the reader of the definition for second order stochastic dominance. The following definition is from Wirch and Hardy (2001):  $X$  second order stochastically dominates  $Y$ , if

$$\int_x^\infty S_X(t)dt \leq \int_x^\infty S_Y(t)dt,$$

for all  $x \geq 0$ , with strict inequality for some  $x \in (0, \infty)$ . This can of course be rewritten in terms of cumulative distribution functions, yielding

$$\int_x^\infty [F_Y(t) - F_X(t)] \geq 0.$$

# 6. Pareto Optimal Risk Sharing: The Dual Theory of Choice Under Risk

## 6.1 Introducing the Problem

Ludkovski and Young (2009) considers Pareto optimal risk sharing using distortion risk measures. We shall start by introducing the model in question, thereafter presenting some results that will eventually lead to a result about Pareto optimal insurance contracts between one insurer and a policyholder. The first part of this section may be somewhat technical for some readers. For readers uninterested in mathematical derivations and technicalities it is sufficient to read the introduction and then skip straight to Theorem 2, where the structure of Pareto optimal allocations of risk is described, followed by an illustrative example. Mathematically well-versed readers are encouraged to follow the material as presented for additional insight into how properties of distortion risk measures affect the structure of Pareto optimal allocations of risk. We shall start with a definition.

**Definition:**((Ludkovski & Young, 2009))  $Y$  is said to precede (or be preferred to)  $Z$  in convex order if  $\int_0^q S_Y^{-1}(p)dp \leq \int_0^q S_Z^{-1}(p)dp$  for all  $q \in [0, 1]$  with equality at  $q = 1$ . We write  $Y \leq_{cx} Z$ .

We can first think of a setting with two or more agents that are facing a random loss  $X_i$  which summed over all agents equals a total loss of  $X$ . Their preference functionals are given by  $V_i$ . Now consider the following set:

$$\mathbb{A}(X) := \left\{ Y := (Y_1, Y_2, \dots, Y_n) : X = \sum_{i=1}^n Y_i, V_i(Y_i) \text{ finite} \right\}.$$

This set consists of all the allocations of the loss  $X$  that are such that the sum of the individual allocations sum to the total loss and are such that the preference functionals of

all agents are finite. The problem here is to find Pareto optimal allocations of risk when all agents are subject to individual rationality constraints. This is exactly what we did before under the expected utility theory. If agents exchange risk from the original allocation  $X$  to the allocation  $Y \in \mathbb{A}(X)$ , each agent  $i$  then face a random loss (or payout) denoted:

$$Z_i = Y_i + (a_i + b_i Y_i + c_i EY_i) = (1 + b_i)Y_i + a_i + c_i EY_i.$$

$a_i \geq 0$  and is interpreted as a fixed cost incurred upon transferring individual risk  $X_i$  to a group of agents.  $b_i$  is interpreted as cost arising from the size of the random loss  $Y_i$ . This makes intuitive sense for instance in the insurance industry, where it seems natural that some costs may increase proportionally with the size of the loss. One could for instance think of costs increasing the more resources spent to investigate the validity of arising claims.  $c_i \in \mathbb{R}$  are costs reflecting the expected size of the payout. For comparison purposes, one may note that  $a_i + b_i Y_i$  is a special case of the cost function in Raviv (1979), which the reader might recall was increasing and convex.

All agents seek to minimize  $H_{g_i}(Z_i)$ , so each agent seeks to minimize

$$H_{g_i}(Z_i) = H_{g_i}[(1 + b_i)Y_i + a_i + c_i EY_i] = (1 + b_i)H_{g_i}(Y_i) + a_i + c_i EY_i,$$

where the second equality follows from translation invariance and positive homogeneity. We may note that the fixed cost  $a_i$  will not alter the minimum of  $H_{g_i}(Z_i)$ , so that minimizing  $H_{g_i}(Z_i)$  is equivalent to minimizing

$$V_i(Y_i) := (1 + b_i)H_{g_i}(Y_i) + c_i EY_i.$$

## 6.2 Characterization of Pareto Optimal Allocations

We are now ready to describe Pareto optimal allocations. To do this, we find it instructive to present a lemma from Ludkovski and Young (2009). For readers interested in the proof of the theorem about Pareto optimal allocations that is to follow, an understanding of this lemma is important. This lemma also reminds the reader about an important property of distortion risk measures.

**Lemma 1:** If  $X^* = (X_1^*, X_2^*, \dots, X_n^*) \in \mathbb{A}(X)$  is Pareto optimal, then so is  $(X_1^*, X_2^*, \dots, X_j^* +$

$\beta, \dots, X_k^* - \beta, \dots, X_n^*)$  for any  $\beta \in \mathbb{R}$  and any  $j, k = 1, 2, \dots, n$ .

As straightforward as it may seem to some, this deserves some elaboration. Verbally lemma 1 says that if we have an allocation of risk such that all of the risk is allocated and all of the agents have finite risk measures, and if this allocation is Pareto optimal, then we can construct another Pareto optimal allocation with the same properties by adding a constant to one or more agents and subtracting a constant from one or more agents, while all these constants sum to zero. Why is this possible? Since the distortion risk measures are cash equivariant, these constants do not matter in the calculation of risk. Thus when we want to calculate the minimum value an agent's preference functional, constants do not matter. Only the allocation of risk matters. For the experienced reader this next comment will seem unnecessarily basic. However, if one finds this lemma confusing it may help to remember that a Pareto optimal allocation is not the same as the best allocation each agent can achieve. It is fully possible to find a Pareto optimal allocation that gives lower aggregated utility to a group as a whole. It will still be Pareto optimal as long as at least one agent is made worse off by changing the allocation. One interesting thing may be noted about Lemma 1: A Pareto optimal allocation is determined up to side-payments that sum to 0. If we remember back to the section about Pareto optimal risk sharing under expected utility theory, the example from Lemaire (1991) showed a Pareto optimal allocation that included side payments which sum to 0. The reader may recall that when all agents have utility functions given by exponential utility, all Pareto optimal risk exchanges differ only by side payments that sum to 0. A consequence of Lemma 1 is that all Pareto optimal allocations under the dual theory may be altered by side payment which sum to 0. T. Boonen (2013) states what we have just explained, namely that Lemma 1 also holds when agents are represented by exponential (CARA) utility functions. We shall now present a theorem. The proof will not be presented, but a part of the proof will be instructive to discuss to show the usefulness of lemma 1. As mentioned in Ludkovski and Young (2009, p. 92): "We use Lemma 1 to characterise the set of Pareto optimal allocations when we view them as points in  $\mathbb{R}^n$  via the mapping  $F : \mathbb{A}(X) \mapsto \mathbb{R}^n$  given by  $F(Y) = (V_1(Y_1), V_2(Y_2), \dots, V_n(Y_n))$ ."

**Theorem 1** (Ludkovski & Young, 2009): We have the following two alternatives:

1. If there exist  $i, j \in \{1, 2, \dots, n\}$ , such that  $1 + b_i + c_i \neq 0$  and  $(1 + b_i + c_i)(1 + b_j + c_j) \leq 0$ , then no Pareto optimal allocation in  $\mathbb{A}(X)$  exists.

2. Otherwise, the image of the set of Pareto optimal allocations in  $\mathbb{A}(X)$  under the mapping  $F$  is a hyperplane in  $\mathbb{R}^n$  given by

$$\left\{x \in \mathbb{R}^n : \sum_{i=1}^n \frac{(V_i(X_i^*) - x_i)}{1 + b_i + c_i} = 0\right\},$$

in which  $X^* \in \mathbb{A}(X)$  is any Pareto optimal allocation. Furthermore, one obtains such a Pareto optimal allocation  $X^*$  by minimizing

$$\sum_{i=1}^n \frac{V_i(Y_i)}{|1 + b_i + c_i|}$$

over  $Y \in \mathbb{A}(X)$ .

To note the usefulness of the Lemma 1, we shall demonstrate how it is used to prove the first part of the theorem. One may assume that  $1 + b_1 + c_1 < 0$  and  $1 + b_2 + c_2 \geq 0$ . This is for convenience and without any loss of generality as we could have chosen any numbers  $i$  and  $j$ . Next we may consider any allocation  $Y \in \mathbb{A}(X)$ . Choosing  $Z = (Y_1 + 1, Y_2 - 1, Y_3, \dots, Y_n)$  will then be a strict improvement of  $Y$ . This can be seen as follows:

$$V_1(Z_1) = (1+b_1)H_{g_1}(Y_1+1)+c_1E(Y_1+1) = 1+b_1+(1+b_1)H_{g_1}(Y_1)+c_1+c_1EY_1 = V(Y_1)+1+b_1+c_1 < V(Y_1)$$

Notice that the only property we used here was translation invariance. By the same reasoning  $V_2(Z_2) = V(Y_2) - (1 + b_2 + c_2) \leq V_2(Y_2)$ . Since we have improved the allocation for agent 1 without making it worse for any other agents, the first allocation cannot have been Pareto optimal. But we could start out in our final allocation reverse the process and end up with a worse allocation. But this allocation would clearly not be Pareto optimal, and according to Lemma 1 it should be. Thus, no Pareto optimal allocation can exist in  $\mathbb{A}(X)$ . For the full proof of the theorem, readers are referred to the original article by Ludkovski and Young (2009).

Theorem 1 tells us how the Pareto optimal allocations look like as points in a hyperplane in  $\mathbb{R}^n$  and in addition it tells us how to obtain such a Pareto optimal allocation. One may choose an allocation  $Y \in \mathbb{A}(X)$  and proceed to minimize a sum of preference functionals given that allocation. Doing that for any allocation  $Y \in \mathbb{A}(X)$  will yield a Pareto optimal allocation. Omitting the details, it is next shown that one may restrict attention to the

following set of allocations when seeking Pareto optimal allocations:

$$\mathbb{C}(X) := \left\{ (f_1(X), f_2(X), \dots, f_n(X)) \in \mathbb{A}(X) : f_i \text{ continuous, non-decreasing, } \sum_{i=1}^n f_i(x) = x \text{ for } x \in \mathbb{R} \right\}.$$

What this really means is that we may restrict attention to comonotone allocations when searching for a Pareto optimal allocation of risk. It is interesting to note that if we look for optimal allocations in a set of comonotone allocations, the share of risk borne by each agent depends only on the total risk  $X$ . This is similar to what Borch (1962) noted when analyzing Pareto optimal allocations under expected utilities. We will now present Theorem 2 in Ludkovski and Young (2009), which will tell us about the structure of Pareto optimal risk allocations under the dual theory. It will also be of use when determining the form of optimal insurance contracts between one insurer and a policyholder.

**Theorem 2** (Ludkovski & Young, 2009): Suppose  $(1 + b_i + c_i)(1 + b_j + c_j) > 0$  for all  $i, j = 1, 2, \dots, n$ . Then,  $X^* = (f_1^*(X), f_2^*(X), \dots, f_n^*(X)) \in \mathbb{C}(X)$  is a Pareto optimal allocation if and only if

$$\sum_{i \in \mathbb{I}} (f_i^*)'(t) = 1 \text{ for } \mathbb{I} = \operatorname{argmin}_{k=1,2,\dots,n} \frac{(1 + b_k)g_k(S_X(t)) + c_k S_X(t)}{|1 + b_k + c_k|},$$

and  $(f_i^*)'(t) = 0$  otherwise. Note that  $\mathbb{I}$  is the set of quantiles of the decumulative distribution function that minimizes the expression given after argmin for each agent.

This theorem may appear hard to interpret, but the reader may gain from studying the proof in the original article. The general outline of the proof is to show that minimizing the expression in theorem 1 that gives a Pareto optimal allocation is equivalent to minimizing the following expression:

$$\sum_{i=1}^n \int_0^\infty \frac{(1 + b_i)g_i(S_X(t)) + c_i(S_X(t))}{|1 + b_i + c_i|} df_i(t).$$

They then proceed to argue that this expression is minimized when

$$\sum_{i \in \mathbb{I}} (f_i^*)'(t) = 1 \text{ for } \mathbb{I} = \operatorname{argmin}_{k=1,2,\dots,n} \frac{(1 + b_k)g_k(S_X(t)) + c_k S_X(t)}{|1 + b_k + c_k|},$$

and  $(f_i^*)'(t) = 0$  otherwise. We provide the following intuition for why the expression is minimized by the above-mentioned. We know that we have to be in the set  $\mathbb{C}(X)$ , which means that  $\sum_{i=1}^n f_i(x) = x$  must be satisfied. Differentiating the expression on both sides, we get that  $\sum_{i=1}^n f_i'(x) = 1$ . We then say to minimize the integral, the agent whose distortion function  $g_i$  minimizes the expression in the integral for a given quantile, must have his optimal function  $(f_i^*)'(t) = 1$  in that quantile (ignoring equality in the argmin for clarity of intuition). The reason why it must be one is that all the risk must be distributed for all quantiles of the risk. For all quantiles where an agent does not minimize the expression inside the integral, the optimal function  $(f_i^*)'(t) = 0$ . It should then seem quite clear that the sum is minimized by

$$\sum_{i \in \mathbb{I}} (f_i^*)'(t) = 1 \text{ for } \mathbb{I} = \operatorname{argmin}_{k=1,2,\dots,n} \frac{(1 + b_k)g_k(S_X(t)) + c_k S_X(t)}{|1 + b_k + c_k|}.$$

This theorem tells us that a Pareto optimal allocation is split up into tranches, i.e, the risk is sliced up. If no agents share the same argmin, then each agent covers a slice of the risk alone. This, however, also means that for all other parts of the risk, that same agent is fully insured. This holds for all agents.

Theorem 2 might seem very unclear to some readers. To make the theorem a little bit clearer, it is instructive consider it in the absence of costs. The set  $\mathbb{I}$  then simplifies to

$$\mathbb{I} = \operatorname{argmin}_{k=1,2,\dots,n} g_k(S_X(t)).$$

In this simplified setting, this seems like an appropriate time for an example.

**Example 6.1:** Consider two agents that are going to share the following risk. In state 1 the combined loss is  $\frac{7}{15}$  and in state 2 the combined loss is 1. The agents preferences are given by the distortion functions  $g_1(x) = \min(\frac{3}{2}x, 1)$  and  $g_2(x) = \sqrt{x}$ . The probability of state 1 occurring is  $\frac{4}{9}$  and the probability of state 2 occurring is  $\frac{5}{9}$ . We must first determine which agent's function is minimized by the quantiles  $\frac{7}{15}$  and 1. The reader can verify that  $g_1$  minimizes the first quantile while  $g_2$  minimizes the second quantile. We then know by Theorem 2 the construction of the derivatives of each agents' optimal function  $f^*$ . We get that

$$(f_1^*)'(x) = \begin{cases} 1 & 0 \leq x \leq \frac{7}{15} \\ 0 & \text{otherwise} \end{cases}$$

and similarly

$$(f_2^*)'(x) = \begin{cases} 0 & 0 \leq x \leq \frac{7}{15} \\ 1 & t > \frac{7}{15}. \end{cases}$$

The optimal functions are then given by

$$(f_1^*)(x) = \begin{cases} t & 0 \leq x \leq \frac{7}{15} \\ \frac{7}{15} & otherwise \end{cases}$$

and

$$(f_2^*)(x) = \begin{cases} 0 & 0 \leq x \leq \frac{7}{15} \\ t - \frac{7}{15} & t > \frac{7}{15}. \end{cases}$$

The optimal functions are piecewise linear and continuous. They are also nondecreasing. We only have to verify that all the risk is distributed and that the derivatives sum to 1 at each quantile. The total risk in state 1 is  $\frac{7}{15}$ . We have that  $f_1^*(\frac{7}{15}) = \frac{7}{15}$  and that  $f_2^*(\frac{7}{15}) = 0$ . All the risk is allocated in state 1. Similarly, for state 2 we have that  $f_1^*(1) = \frac{7}{15}$  and  $f_2^*(1) = 1 - \frac{7}{15} = \frac{8}{15}$ . All the risk is allocated in state 2. It's easy to see that the derivatives sum to 1 at each quantile. Thus, we have found a Pareto optimal distribution of risk. Had the initial allocation of the agents been for example  $X_1 = X_2 = \frac{1}{2}X$  one could check whether or not the Pareto optimal allocation satisfied both agents' individual rationality constraints. If this had not been the case one could have altered the Pareto optimal allocation by side-payments to obtain one that satisfied both agents individual rationality constraints.

Example 6.1 also illustrates another property of Pareto optimal allocations under the dual theory. One can construct a representative agent from all agents' distortion functions. This representative agent will then have the following distortion function:  $g_R^*(x) = \min[g_i(x) : i \in N]$ , where  $N$  is the set of all agents. In words, for all values of  $x$  the distortion function of the representative agent takes the form of the lowest distortion function of all agents. This is what we did implicitly in example 6.1 (readers may verify this by calculating example 6.1 themselves) and follows quite intuitively from Theorem 2. A formal justification of this property of Pareto optimal allocations is given in T. Boonen (2013). The same article also covers conditions such that a Pareto optimal allocation of risk is unique up to side-payments. For our purposes, this is not necessary to bring up so the interested reader is referred to T. Boonen (2013).

T. Boonen (2013) in proposition 3.6 states the following noteworthy property of Pareto

optimal allocations with distortion risk measures. It says that if an agent has a distortion function that is smaller than all other agents' distortion functions, then it is optimal to shift all risk to this agent. Having followed example 6.1 and noted the construction of the representative agent's distortion function, this may already seem clear. This is similar to what we saw under expected utility theory, where we saw that if one agent is risk neutral while the others are risk averse, then it is Pareto optimal for that agent to bear all the risk. This seems intuitive, as we know that this agent may be compensated by side-payments and that the only thing that matters for risk allocation are attitudes to risk, not side-payments. In this case, the side-payments merely function as a way for the agent assuming all the risk to satisfy his individual rationality constraint. The reader may recall that a risk neutral agent under the dual theory will value prospects after the prospect's expected value. From this one may observe that a risk neutral preference relation under the dual theory is also an expected utility function. It is in fact the only expected utility function that is also a dual utility function. This might strengthen the readers intuition for why the risk allocations are so similar under both choice theories in the case of one risk neutral agent. However, we may note a difference in Pareto optimal allocations between our two choice theories. Under expected utility, one does not obtain a risk sharing arrangement such as in Theorem 2. As an example, we saw how Pareto optimal risk sharing would look like under HARA-utility with equal cautiousness. Such a risk sharing arrangement was characterized by each agent assuming a fixed proportion of the risk. On the contrary, theorem 2 says that agents assume total responsibility for some parts of the risk, and full insurance otherwise. As we observe risk sharing as postulated by Theorem 2 in practice (T. Boonen, 2013), the dual theory seems fitting as a descriptive theory in this case. We wish to point out that such risk sharing arrangements can also be constructed under expected utility theory, but that would require another set of assumptions and/or optimality criteria, which will not be explained here.

We have now reached a suitable point of this section to reflect a little on what is to be achieved by all these derivations. For readers well-versed in mathematics, a deep understanding may be obtained. For readers not quite as well-versed in mathematics, frustration and a possible increase in lack of understanding may result. The author, which is not trained extensively in mathematics, sympathize with the latter group of people. It does not appear controversial to assume that the majority of economists are not extensively trained in mathematics, and in any event potential readers of this thesis probably values economic clarity more than mathematical finesse. Until this point of the section, one objective of presenting

somewhat detailed derivations of the theory has been to increase the readers understanding of distortion risk measures, some key properties of distortion risk measures for our analysis and how such properties of distortion risk measures directly affect the structure of Pareto optimal risk sharing. In addition, many articles on the subject follow a similar buildup where they introduce the problem, reduce the set of allocations to only include comonotone allocations and then present a general theorem related to their specific subject of study before presenting specializations. A satisfactory understanding of this first part of the section thus serves as a good preparation for the interested reader. Hopefully, the readers who followed the development of the first part of this section could appreciate the exposition. We will now sacrifice mathematical rigour in the hope that it will facilitate increased economic clarity.

### 6.3 The Case of One Policyholder and One Insurer

For comparison purposes with the case of Pareto optimal insurance between one insurer and a policyholder in the case of expected utility maximizers, we now direct attention to the special case of  $n = 2$  agents. We consider the case where  $(1 + b_1 + c_1)(1 + b_2 + c_2) > 0$ . Suppose a potential policyholder is faced with a random loss  $X$ . We choose this to be agent 2. Agent 2 wants to purchase insurance  $f(X)$  from an insurer, which is then agent 1. The agents may thus split the risk between them, in which case  $f(X)$  is the insurer's part of the risk  $X$ , and  $X - f(X)$  is the policyholder's retained part of the risk. We suppose that the insurer receives  $(1 + \lambda)E[f(X)]$ , with  $\lambda > 0$ , as a premium from the policyholder. We choose some parameter values for the agents which arguably makes sense:  $a_1 = 0$ ,  $b_1 > 0$ ,  $c_1 = -(1 + \lambda)$ ,  $a_2 = (1 + \lambda)E[X]$ ,  $b_2 = 0$  and  $c_2 = -(1 + \lambda)$ . This means that the insurer has no fixed costs with the risk transfer, while the policyholder has a fixed cost equal to a risk-adjusted premium to rid himself of the expected loss. These specific values for the cost structure means that  $(1 + b_1 + c_1)(1 + b_2 + c_2) > 0$  is reduced to  $b_1 < \lambda$ . They also yield individual rationality constraints

$$(1 + \lambda)E[f(X)] \geq (1 + b_1)H_{g_1}(f(X))$$

for the insurer and

$$H_{g_2}(f(X)) \geq (1 + \lambda)E[f(X)]$$

for the policyholder. If we interpret the right hand side of the insurers rationality constraint to be a risk-adjusted cost of entering into a contract, we see that the insurer only enters into a contract if the premium is at least as high as this risk-adjusted cost. The risk adjustment is done via the distortion risk measure and the cost  $b_1$ . Similarly for the policyholder, his risk-adjusted benefit from receiving  $f(X)$  exceeds the cost  $(1 + \lambda)E[f(X)]$ . It is usual to assume that the policyholder is more risk averse than the insurer. Since an agents' risk aversion is captured through the concavity of  $g$ , we may then assume that the policyholder's distortion function can be expressed as a concave transformation of the insurer's distortion function (the reader might remember the discussion of this in section 2). An equivalent statement is that  $g_2 \geq g_1$ . By Theorem 2, the optimal function  $f^*$  is

$$(f^*)'(t) = \begin{cases} 1, & \text{if } g_1(S_X(t)) - S_X(t) < \frac{\lambda - b_1}{\lambda(1 + b_1)}[g_2(S_X(t)) - S_X(t)] \\ \beta & \text{if } g_1(S_X(t)) - S_X(t) = \frac{\lambda - b_1}{\lambda(1 + b_1)}[g_2(S_X(t)) - S_X(t)] \\ 0, & \text{otherwise.} \end{cases}$$

where  $\beta \in [0, 1]$  is arbitrary. Having this expression for the optimal  $f^*$ , we can perform some basic comparative statics to see how the Pareto optimal insurance coverage is affected by parameter changes. We may note that an increase in  $\lambda$  will lead to an increase in the optimal insurance  $f^*$ . In other words, the indemnity that the insurer pays to the policyholder increases when the loading increases. When the insurers's cost increases, the optimal indemnity that the insurer pays to the policyholder decreases. These effects both makes sense. The policyholder demands a higher indemnity when the insurer's risk premium increases, and the insurer is less willing to supply insurance when his own costs increases. One may also note that a more concave  $g_2$  will lead to a higher optimal indemnity schedule. The policyholder wishes to buy more insurance at a given price the more risk averse he is.

For comparison purposes with the expected utility case, we include the following proposition from Ludkovski and Young (2009):

**Proposition 1:** "If  $\frac{g_1(p) - p}{g_2(p) - p}$  increases for  $p \in (0, 1)$ , then deductible insurance is optimal, that is,

$$f^*(x) = (x - d)_+$$

is optimal with the deductible given by

$$d = \inf \left\{ t : \frac{g_1(S_X(t)) - S_X(t)}{g_2(S_X(t)) - S_X(t)} \leq \frac{\lambda - b_1}{\lambda(1 + b_1)} \right\}.$$

If no such  $d$  exists, then  $f^* \equiv 0$ ."

Since our argument  $p$  in the functions  $g_1(p)$  and  $g_2(p)$  is  $p = S_X(t)$ , which we recall is the decumulative distribution function of the loss  $X$ , we may note that an increase in  $\frac{g_1(p)-p}{g_2(p)-p}$  for  $p \in (0, 1)$  means that  $\frac{g_1(S_X(t))-S_X(t)}{g_2(S_X(t))-S_X(t)}$  decreases for values of  $t \geq 0$ . Looking at the expression for the optimal deductible then it is clear that the optimal deductible is given by the first quantile of  $X$  that makes the inequality  $\frac{g_1(S_X(t))-S_X(t)}{g_2(S_X(t))-S_X(t)} \leq \frac{\lambda-b_1}{\lambda(1+b_1)}$  hold. It may be instructive to check whether or not this matches our intuition. We define a function  $h(b_1) := \frac{\lambda-b_1}{\lambda(1+b_1)}$  and recall that  $\lambda > 0$ . By differentiating this expression, we can check how the optimal deductible changes with a change in  $b_1$ . Doing this, we get

$$h'(b_1) = -\frac{\lambda + 1}{\lambda(1 + b_1)^2},$$

which is always negative. An increase in  $b_1$  reduces  $h(b_1)$ . In other words, when the cost of the insurer increases, the optimal deductible is higher, which makes sense. The converse is also true. We remind the reader that this result is derived under some specialized assumptions on the parameters. However, as we mentioned before, these assumptions about the parameters coincide with common sense. We thus see that under some assumptions, we get a result similar to the one obtained by Raviv (1979), namely that in the presence of costs (that are not fixed), deductible insurance is Pareto optimal. There is however a notable difference. The type of contract with a deductible admitted by dual utility maximizers are so-called stop-loss contracts. All losses over the deductible is covered. In the expected utility case, losses above the deductible are given by a coinsurance which yields indemnities with less than full coverage. It should also be noted that it is possible to obtain contracts under the dual theory which includes a deductible without having full coverage above the deductible.

## 6.4 Some Points About Pareto Optimal Allocations

Asimit and Boonen (2018) studies Pareto optimal insurance contracts where there are one potential policyholder and multiple insurers. Each insurer is willing to insure a part of or

all of the risk that the potential policyholder is facing initially. As before, the necessary components of the analysis will be presented here, while the interested reader is referred to Asimit and Boonen (2018) for further details. Some brief assumptions are made about a set of random variables defined on a probability space given by  $(\Omega, \mathcal{F}, \mathbb{P})$ . Define  $L^p(\mathbb{P})$  to be the set of p-integrable random variables and  $L_+^p(\mathbb{P})$  the set of non-negative p-integrable random variables. Adopting the notation of (Asimit & Boonen, 2018) we recall from the last section that distortion risk measures are represented mathematically as:

$$\rho[Y] = \int_0^\infty g(S_Y(z))dz - \int_{-\infty}^0 [1 - g(S_Y(z))]dz$$

where  $S_Y$  is the usual decumulative distribution function of the random variable  $Y$  and  $g$  is as described previously. Boonen proceeds to assume that there is one potential policyholder seeking to share his initial risk with multiple insurers. The initial risk is a potential loss  $X \in L_+^p(\mathbb{P})$ , i.e a nonnegative random variable. The possible insurers that the policyholder trades with is given by a set  $S \subseteq N$ , where  $N = \{1, 2, \dots, n\}$ . Each insurer accepts part of the risk  $X_i \in L_+^p(\mathbb{P})$  where naturally  $X_i \leq X$ . Upon accepting a share  $X_i$  of the total risk they receive an insurance premium  $\pi_i \geq 0$ . After completing the trading of risk the policyholder then retains a share  $X_{PH} = X - \sum_{i \in S} X_i$  and pays  $\sum_{i \in S} \pi_i$ .

A contract  $(\pi^S, X^S)$  is called feasible if

$$\rho_i(X_i - \pi_i) \leq 0 \text{ for all } i \in S \text{ and } \rho_{PH}(X_{PH} + \sum_{i \in S} \pi_i) \leq \rho_{PH}(X).$$

These are individual rationality constraints. All insurers must have no more risk than they had before the trade adjusted by their premium, and their initial risk  $\rho(0) = 0$ . Similarly, the policyholder must have no more risk than before the trade. To be clear, when we talk about levels of risk we are talking about the values of the distortion risk measures. Recall that minimizing a distortion risk measure is equivalent to maximizing utility under the dual theory.

In short, Asimit and Boonen (2018) considers a very similar situation to Ludkovski and Young (2009), and as one might expect the results are very similar. A result similar to Theorem 2 shows that the optimal indemnity schedule is layered, where reinsurers cover different parts of the risk of the insurer. For reference, this result is proposition 4.1 in (Asimit & Boonen, 2018). A more detailed derivation and explanation of the result can be

found in (T. J. Boonen, Tan, & Zhuang, 2016) section 4. Since this result is so similar to what we just presented, we do not present it here. However, we wish to make a small digression, which will serve two purposes. The first is to make a point about the general structure of Pareto optimal solutions under the dual theory versus expected utility theory. The other is to explain where this difference stems from. This result is based on the following risk measure: Let an exponential utility function be given by  $u_i(w) = -\gamma_i e^{-\frac{w}{\gamma_i}}$ . We then recognize that an agent's risk tolerance  $t_i$  is given by  $t_i = \gamma_i$ . The aim is to minimize the expected loss, and thus the aim is to minimize  $-E[u_i(-X)]$ . We will now assume that our risk measure is given by

$$\rho_i(X) - u_i^{-1}(E[u_i(-X)]) = \gamma_i \ln(E[e^{\frac{X}{\gamma_i}}]),$$

which is also known as the entropic risk measure. Note that this is not a distortion risk measure, but it shares some properties with distortion risk measures. By using a risk measure which shares just some of the properties that characterizes distortion risk measures, we shall highlight how these properties affect the Pareto optimal allocation of risk. More on the entropic risk measure can be found in Barriau and El Karoui (2005). It is straightforward to verify that this risk measure satisfies translation invariance. One may also note the similarity of this risk measure and the utility function used by Lemaire (1991) in the section about Pareto optimality under expected utility theory. Not surprisingly, the result is also very similar and is given by proposition 4.2 in (Asimit & Boonen, 2018): When  $X \in L_+^\infty(\mathbb{P})$  and  $\rho_i$ , where  $i \in S \cup PH$ , is as described above, then a risk allocation  $x^S$  is Pareto optimal if, for any  $i \in S$ ,

$$X_i = \frac{\gamma_i}{\sum_{j \in S \cup PH} \gamma_j} X.$$

In words, if  $X$  is a non-negative (a negative loss cannot occur) and integrable (a technical condition) random variable, and all agents have entropic risk measures as defined above, then the optimal risk allocation is such that all insurers covers a share of the total risk given by their risk tolerance relative to the sum of all other insurers' and the policyholder's risk tolerance. This is very similar to what we saw earlier under expected utility theory. However, there are no explicit side-payments mentioned here. The attentive reader will not be surprised by this, as we have determined earlier that Pareto optimal allocations under the dual theory is determined up to side-payments. This is a straightforward implication of translation invariance, as was explained when we presented Lemma 1. We may thus add and subtract constants that sum to 0 and obtain a new Pareto optimal allocation. These

side-payments are captured by the premiums  $\pi_i$  in this model. This result highlights the role of the property translation invariance on the structure of Pareto optimal allocations. Under expected utility theory, the explicit examples we presented made use of utility functions that does not satisfy translation invariance. However, in the case of utility functions of the following form  $u_i(x) = 1 - e^{-a_i x}$ , we found that one may alter Pareto optimal allocations by side-payments that sum to 0, and that these side-payments vary with the agent weights  $k_i$  in the optimization problem.

Lastly in this section, we wish to briefly make a comment. Since expected utility maximizers exhibit diminishing marginal utility, while dual utility maximizers exhibit constant marginal utility, one could argue on the basis of that (*ceteris paribus*) that individuals should have preferences represented by expected utility while firms should have preferences represented by dual utility. T. J. Boonen (2017) performs an analysis with two sets of agents. The first set is made up of expected utility maximizers while the second set is made up of dual utility maximizers. He gives a representation of Pareto optimal contracts in such a market, and he finds that both results of Pareto optimal risk sharing under expected utility and Pareto optimal risk sharing under the dual theory can be extended to encapsulate both. For details about this and more, the interested reader is referred to (T. J. Boonen, 2017).

## 7. Discussion

This section introduces some new material, but is largely based upon insights from the preceding sections. It will consider the descriptive and normative validity of the different choice theories. One could ask the question: are the expected utility theory and the dual theory normative or descriptive theories? Starting with expected utility theory, we could note the following. As a descriptive theory it has been widely criticized. Tversky (1975) describes several experiments where expected utility fails. The experiments described mostly amounts to violations of the independence axiom. It should be noted that Tversky (1975) does not describe these experiments in detail, so any potential framing effects etc. that could be prevalent in the experiments is not discussed and unknown. The following example may be known to the reader: A person may choose between two menus when ordering dinner. The first menu consists of steak for the main course and ice cream for dessert. The other menu consists of halibut for the main course and ice cream for dessert. Let's say the agent prefers the second menu to the first menu. If the independence axiom holds, that means he prefers halibut to steak, since ice cream is a common alternative. However, imagine introducing the same appetizer to both menus, namely scallops. In real life, it is easy to imagine someone switching preference from menu 2 to menu 1, to not have two dishes in a row consisting of seafood (preference for variation). However, this is not in line with the independence axiom. Many would accept this anecdotal evidence as a basis for rejecting the independence axiom. As we know, modifying the independence axiom yields the dual theory of choice. Does the dual theory outperform expected utility theory as a descriptive theory? As is pointed out by Guriev (2001), several studies has shown that the dual theory does not perform well as a descriptive theory. He cites both Harless and Camerer (1994) and Hey and Orme (1994) to back up his claims. It should be noted that the experiments performed in the above-mentioned studies were done on individuals. It seems natural that individuals exhibit diminishing marginal utility of money. It would arguably be more natural for firms to have constant marginal utility of money. Perhaps the results would have been different had the experiments been carried out on firms (assuming that to be possible).

As we know, risk aversion in expected utility theory is equivalent to concavity of the

utility function. As we mentioned earlier, one cannot separate the notion of diminishing marginal utility of money and risk aversion in expected utility theory. One could say, as a counterargument to the normative validity of expected utility theory, that it is not very satisfactory to postulate risk aversion as a normative decision criterion as a consequence of an agent exhibiting diminishing marginal utility of money. Put differently, if one finds it reasonable to postulate diminishing marginal utility of money, then it seems unsatisfactory as a normative theory to require risk averse behavior on the basis of diminishing marginal utility of money. Tversky (1975) calls risk aversion in expected utility theory an epiphenomenon. As far as risk aversion goes, one could argue that the dual theory is more flexible, since it not only separates the notion from diminishing marginal utility, but also separates the notion of weak and strong risk aversion (Gollier & Machina, 2013), as we pointed to when describing risk aversion under the dual theory. However, as a normative theory, exhibiting constant marginal utility of money seems to be at odds with common sense. Intuitively, it is difficult to convince oneself that a billionaire and a poor person has the same marginal utility of money. For that reason alone, some might be tempted to dismiss the normative validity of the dual theory when applied to individuals. Another thing that may seem at odds with the normative validity of the dual theory is the distortion of the probabilities. At first glance at the theory, subjectively distorting the probabilities of events occurring and using this as a basis to calculate the utility of the prospect seems irrational, meant in the colloquial sense of the word. However, it would only be irrational if this was done because the true probabilities was distorted in the agent's perception, i.e, if the agent perceived the probabilities of events occurring differently to what they actually were. As is pointed out by Yaari (1987) this is not the case, and neither can be the case due to the neutrality axiom.

For our purposes, when judging the validity of the different theories as descriptive, we should see whether or not they describe risk sharing as it is observed in practice. We have seen that when it comes to pure demand theory, both expected utility theory and the dual theory postulate that the optimal contracts have indemnity schedules with a straight deductible. Contracts with a deductible is very common in practice. This result is also the case for optimal reinsurance contracts, when limited to pure demand theory. However, when considering optimal coinsurance (so limiting the set of admissible indemnity schedules), the two theories differ. Expected utility theory predicts full coverage with no loading, and less than full coverage with positive loading. The optimal coverage is decreasing in the loading. On the other hand, the dual theory predicts full coverage up to a certain amount of

loading, and then no coverage at all, a so-called bang-bang solution. Doherty and Eeckhoudt (1995) points out that this result has a certain intuitive appeal, as such a solution appears to be encountered frequently in markets. We would argue that from a normative point of view, such solutions seem justified only if the negative utility of researching and purchasing insurance outweighs the benefits of risk sharing. This is not captured directly by the model, rather it follows from a property of the underlying choice theory, but if by descriptive we mean describing actual behavior (without regard for causality), then the model could still be called descriptive. On the basis of these remarks, intuition would suggest that expected utility theory seems better adapted as a normative theory than the dual theory in the case of pure demand theory. On the point where they differ in results, i.e, optimal coinsurance, intuition can easily agree with reducing coverage somewhat in proportion to increasing costs, i.e, intuition can easily agree with expected utility theory. The intuition does not seem quite as clear cut from the dual theory.

The more realistic case is of course to model both supply and demand in the insurance markets. In this setting, the models differ somewhat. The dual theory predicts that Pareto optimal risk allocation is divided into tranches, where each agent assumes responsibility for a tranche of the risk. Under expected utility theory, such a solution does not occur. One of the closed-form solutions exemplified under expected utility theory yielded proportional risk sharing contracts. As is mentioned in Ludkovski and Young (2009), tranching of risk is observed in practice. It seems that the dual theory has more descriptive power in this setting. However, we should not forget that these results are derived under a set of assumptions and optimality criteria. Relaxing and/or changing assumptions and optimality criteria makes it possible to explain tranching also under expected utility theory. Having mentioned this, adhering to the model settings of this thesis, the dual theory seems to better describe Pareto optimal risk sharing. In the case of 1 insurer and one potential policyholder negotiating Pareto optimal insurance contracts, we found both similarities and differences. Under expected utility theory we found that deductibles, with coinsurance for losses over the deductible, occur as long as the cost function of the insurer is not constant. It seems reasonable to assume that an insurer's costs related to an insurance contract are not entirely fixed. Since deductibles are observed in practice we conclude that expected utility theory can explain this well. It is possible to obtain deductibles also under the dual theory, but under a more restrictive set of assumptions. The contract then exhibits a deductible with full coinsurance over the deductible. It may seem that expected utility theory better explains

the simple case of one insurer and one potential policyholder, while the dual theory better explains things such as tranching in credit markets. This seems to fit well with intuition, as we mentioned earlier that one could make a case for modeling individuals with expected utility theory (because of diminishing marginal utility of wealth under the assumption of risk aversion), while modeling firms under the dual theory (because of constant marginal utility of wealth under the assumption of risk aversion).

## 8. Conclusion

As we have seen, there are some notable similarities and some notable differences between expected utility theory and the dual theory of choice under risk. When it comes to the structure of Pareto optimal allocations of risk, it is interesting that both theories have solutions that are, in effect, pool arrangements. It is also interesting to note that for insurance decisions, prices may be determined independently of the risk allocation under the dual theory. First, the optimal allocation of risk is determined. Thereafter, prices are negotiated so that both agents' individual rationality constraints are satisfied. This is generally not the case under expected utility theory. In the special case of one risk neutral agent, both theories give the same structure on Pareto optimal risk allocations, namely that the risk neutral agent carries all the risk. This is very much in line with intuition as the special case of risk neutral agents gives the only dual theory preference functional that is also an expected utility functional.

As for the more general structure of Pareto optimal allocations of risk, we get tranching of the risk under the dual theory. This is an interesting result as it conforms to what is observed in credit markets. The structure of Pareto optimal allocations of risk under the dual theory is highly impacted by the property translation invariance shared by all distortion risk measures. This property stems from linearity in payments under the dual theory. We have mentioned how this property is connected to constant marginal utility of money, which we have argued is a desirable property when modeling firms. It is therefore interesting that the dual theory seems to explain risk allocations in credit markets so well. Such risk sharing arrangements is not explained by expected utility theory. We exemplified a Pareto optimal risk sharing arrangement under expected utility theory by assuming agents with exponential utility. The closed form solution obtained exhibited indemnity schedules that are linear in the loss. Adhering to the notion that it is desirable with constant marginal utility of money when modeling firms, it fits well with intuition that the dual theory seems to outperform expected utility theory when modeling risk sharing arrangements between firms, since expected utility maximizers exhibit diminishing marginal utility of money.

On the other hand, when describing insurance decisions between one insurer and one po-

tential policyholder, we have seen that, in the presence of costs that are not fixed, expected utility theory always predicts insurance policies with a deductible, and a form of coinsurance over the deductible (which may also be full coverage). As such contracts are prevalent in markets, expected utility theory seems well adapted to this setting. The dual theory may yield similar results, but may also give rise to other indemnity schedules. The setting in the dual theory that yields an indemnity schedule with a straight deductible is fairly restricted. We have argued earlier that diminishing marginal utility is a desirable property when modeling individuals. The results from the theories thus seems to conform with intuition, as the predictions from expected utility theory seems more descriptive in the case of one insurer and one potential policyholder than the predictions from the dual theory.

When it comes to pure demand theory, a noteworthy finding was that the optimality of deductibles could be directly extended to include the dual theory. With small modifications, a theorem of optimal coinsurance contracts first derived under expected utility theory could also be extended to include the dual theory. These results are thus robust. We presented arguments in support of the normative and intuitive validity of both theories. However, the arguments in support of the normative and intuitive validity of expected utility theory was, in the author's opinion, more convincing. Again, this fits well with intuition as we modeled individual demand for insurance, and it should be uncontroversial to postulate diminishing marginal utility of money for individuals. The descriptive validity of expected utility theory also seems higher when modeling individuals, as mentioned in the previous section. However, as far as normative validity goes, it is arguably better to be able to separate an agent's attitude towards risk from an agent's attitude towards wealth, as is the case under the dual theory.

# References

- Aase, K. K. (2008). Optimal risk-sharing and deductibles in insurance. *Encyclopedia of Quantitative Risk Analysis and Assessment*.
- Aase, K. K. (2017). Optimal insurance policies in the presence of costs. *Risks*, 5(3), 46.
- Albrecht, P. (1992). Premium calculation without arbitrage? A note on a contribution by G. Venter. *ASTIN Bulletin: The Journal of the IAA*, 22(2), 247–254.
- Arrow, K. J. (1974). Optimal insurance and generalized deductibles. *Scandinavian Actuarial Journal*, 1974(1), 1–42.
- Asimit, V., & Boonen, T. J. (2018). Insurance with multiple insurers: A game-theoretic approach. *European Journal of Operational Research*, 267, 778–790.
- Barrieu, P., & El Karoui, N. (2005). Inf-convolution of risk measures and optimal risk transfer. *Finance and stochastics*, 9(2), 269–298.
- Boonen, T. (2013). Risk redistribution with distortion risk measures. *Work in progress*.
- Boonen, T. J. (2017). Risk sharing with expected and dual utilities. *ASTIN Bulletin: The Journal of the IAA*, 47(2), 391–415.
- Boonen, T. J., Tan, K. S., & Zhuang, S. C. (2016). The role of a representative reinsurer in optimal reinsurance. *Insurance: Mathematics and Economics*, 70, 196–204.
- Borch, K. (1960). The safety loading of reinsurance premiums. *Scandinavian Actuarial Journal*, 1960(3-4), 163–184.
- Borch, K. (1962). Equilibrium in a reinsurance market. *Econometrica*, 30(3), 424–444.  
Retrieved from <http://www.jstor.org/stable/1909887>
- Demers, F., & Demers, M. (1990). Price uncertainty, the competitive firm and the dual theory of choice under risk. *European Economic Review*, 34(6), 1181–1199.
- Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R., & Laeven, R. (2006). Risk measurement with equivalent utility principles. *Statistics & Decisions*, 24(1/2006), 1–25.
- Doherty, N. A., & Eeckhoudt, L. (1995). Optimal insurance without expected utility: The

- dual theory and the linearity of insurance contracts. *Journal of Risk and Uncertainty*, 10(2), 157–179.
- Eeckhoudt, L., Gollier, C., & Schlesinger, H. (2011). *Economic and financial decisions under risk*. Princeton University Press.
- Gerber, H. U., & Pafum, G. (1998). Utility functions: from risk theory to finance. *North American Actuarial Journal*, 2(3), 74–91.
- Gollier, C., & Machina, M. J. (2013). *Non-expected utility and risk management: A special issue of the geneva papers on risk and insurance theory*. Springer Science & Business Media.
- Gollier, C., & Schlesinger, H. (1996). Arrow’s theorem on the optimality of deductibles: a stochastic dominance approach. *Economic Theory*, 7(2), 359–363.
- Guriev, S. (2001). On microfoundations of the dual theory of choice. *The Geneva papers on risk and insurance theory*, 26(2), 117–137.
- Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6), 1251–1289. Retrieved from <http://www.jstor.org/stable/2951749>
- Herstein, I. N., & Milnor, J. (1953). An axiomatic approach to measurable utility. *Econometrica*, 21(2), 291–297. Retrieved from <http://www.jstor.org/stable/1905540>
- Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6), 1291–1326. Retrieved from <http://www.jstor.org/stable/2951750>
- Kreps, D. (2018). *Notes on the theory of choice*. Routledge.
- Lemaire, J. (1991). Borch’s theorem: A historical survey of applications. In *Risk, Information and Insurance* (pp. 15–37). Springer.
- Levin, J. (2006). Choice under uncertainty. *Unpublished material*. Retrieved from <https://web.stanford.edu/~jdlevin/Econ%20202/Uncertainty.pdf>
- Ludkovski, M., & Young, V. R. (2009). Optimal risk sharing under distorted probabilities. *Mathematics and Financial Economics*, 2(2), 87–105.
- Machina, M. J. (1995). Non-expected utility and the robustness of the classical insurance paradigm. In *Non-expected utility and risk management* (pp. 9–50). Springer.
- Moffet, D. (1979). The risk sharing problem. *Geneva Papers on Risk and Insurance*, 11, 5–13.
- Mossin, J. (1968). Aspects of rational insurance purchasing. *Journal of Political Economy*,

- 76(4, Part 1), 553–568.
- Pratt, J. W. (1975). Risk aversion in the small and in the large. In *Stochastic Optimization Models in Finance* (pp. 115–130). Elsevier.
- Quiggin, J. (1993). *Generalized expected utility theory: The rank dependent model*. Springer Science & Business Media.
- Raviv, A. (1979). The design of an optimal insurance policy. *The American Economic Review*, 69(1), 84–96.
- Röell, A. (1987). Risk aversion in Quiggin and Yaari’s rank-order model of choice under uncertainty. *The Economic Journal*, 97, 143–159.
- Safra, Z., & Segal, U. (1998). Constant risk aversion. *Journal of Economic Theory*, 83(1), 19–42.
- Savage, L. J. (1972). *The Foundations of Statistics*. Dover Publications, Inc.
- Schlesinger, H. (1997). Insurance demand without the expected-utility paradigm. *The Journal of Risk and Insurance*, 64(1), 19–39. Retrieved from <http://www.jstor.org/stable/253910>
- Segal, U., & Spivak, A. (1990). First order versus second order risk aversion. *Journal of Economic Theory*, 51(1), 111–125.
- Stiglitz, J. E., & Rothschild, M. (1970). Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3), 225–243.
- Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis*, 163–173.
- Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1), 71–92.
- Wang, S. S., & Young, V. R. (1998). Ordering risks: Expected utility theory versus yaari’s dual theory of risk. *Insurance: Mathematics and Economics*, 22(2), 145–161.
- Wang, S. S., Young, V. R., & Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2), 173–183.
- Wang, Z., & Yan, J.-a. (2006). Unpublished. *Choquet Integral and Its Applications: A Survey [slides]*. Retrieved from <http://159.226.47.19/academic/workshop/workshop10/%D1%CF%BC%D3%B0%B2.pdf>
- Wilson, R. (1968). The theory of syndicates. *Econometrica*, 36(1), 119–132. Retrieved from <http://www.jstor.org/stable/1909607>

- Wirch, J. L., & Hardy, M. R. (2001). Distortion risk measures: coherence and stochastic dominance. In *International congress on insurance: Mathematics and economics* (pp. 15–17).
- Yaari, M. E. (1986). Univariate and multivariate comparisons of risk aversion: a new approach. In W. Heller, R. Starr, & D. Starret (Eds.), *Essays in Honor of Kenneth J. Arrow, Volume 3* (pp. 173–187). Cambridge University Press Cambridge, UK.
- Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1), 95–115.  
Retrieved from <http://www.jstor.org/stable/1911158>