

ECONOMIC AND FINANCIAL DECISIONS UNDER UNCERTAINTY

Louis Eeckhoudt
Catholic University of Mons

Christian Gollier
University of Toulouse

Harris Schlesinger
University of Alabama

February 7, 2004

Contents

0.1	Preface	7
I	Decision theory	11
1	Risk aversion	13
1.1	An historical perspective on risk aversion	14
1.2	Definition and characterization of risk aversion	18
1.3	Risk premium and certainty equivalent	21
1.4	Degree of risk aversion	25
1.5	Decreasing absolute risk aversion and prudence	28
1.6	Relative risk aversion	30
1.7	Some classical utility functions	32
1.8	Bibliographical references and extensions	35
2	The measures of risk	39
2.1	Increases in risk	40
2.1.1	Adding noise	40
2.1.2	Mean-preserving spreads in probability	42
2.1.3	The integral condition and risk-averse preferences	44
2.1.4	Preference for diversification	46
2.1.5	And the variance?	47
2.2	Aversion to downside risk	48
2.3	First-degree stochastic dominance	49
2.4	Bibliographical references and extensions	51

II	Risk management	53
3	Insurance decisions	55
3.1	Optimal insurance: An illustration	58
3.2	Optimal coinsurance	59
3.3	Comparative statics in the coinsurance problem	64
3.4	The optimality of deductible insurance	67
3.5	Bibliographical references and extensions	71
4	Static portfolio choices	77
4.1	The one-risky-one-riskfree-asset model	78
4.1.1	Description of the model	78
4.1.2	The equity premium and the demand for stocks	80
4.2	The effect of background risk	81
4.3	Portfolios of risky assets	83
4.3.1	Diversification in the expected utility model	83
4.3.2	Diversification in the mean-variance model	85
4.4	Bibliographical references and extensions	87
5	Static portfolio choices in an Arrow-Debreu economy	91
5.1	Arrow-Debreu securities and arbitrage pricing	92
5.2	Optimal portfolios of Arrow-Debreu securities	95
5.3	A simple graphical illustration	97
5.4	Bibliographical references and extensions	99
6	Consumption and saving	101
6.1	Consumption and saving under certainty	101
6.1.1	Aversion to consumption fluctuations over time	104
6.1.2	Optimal consumption growth under certainty	106
6.2	Uncertainty and precautionary savings	108
6.3	Risky savings and precautionary demand	111
6.4	Time consistency	113
6.5	Bibliographical references and extensions	115
7	Dynamic portfolio management	119
7.1	Backward induction	120
7.2	The dynamic investment problem	121
7.3	Time diversification	126

7.4	Portfolio management with predictable returns	128
7.5	Bibliographical references and extensions	132
8	Risk and information	135
8.1	The value of information	136
8.1.1	An example	136
8.1.2	A general model	138
8.1.3	Value of information and risk aversion	141
8.2	Comparative statics analysis	142
8.2.1	Real-option value and irreversibility	144
8.2.2	Savings and the early resolution of uncertainty	146
8.3	The Hirshleifer effect	147
8.4	Bibliographical references and extensions	149
9	Optimal prevention	153
9.1	Prevention under risk neutrality	154
9.2	Risk aversion and optimal prevention	155
9.3	Prudence and optimal prevention	157
9.4	Bibliographical references and extensions	158
III	Risk sharing	161
10	Efficient allocations of risks	163
10.1	Risk sharing: An illustration	164
10.2	Description of the economy and definition	166
10.3	Characterization of efficient allocations of risk	168
10.3.1	The mutuality principle	169
10.3.2	The sharing of the macroeconomic risk	171
10.4	Aggregation of preferences	174
10.5	Bibliographical references and extensions	176
11	Asset pricing	179
11.1	Competitive markets for Arrow-Debreu securities	179
11.2	The first theorem of welfare economics	181
11.3	The equity premium	182
11.4	The capital asset pricing model	185
11.5	Two fund separation theorem	188

11.6	Bond pricing	189
11.6.1	The risk-free rate	189
11.6.2	Factors affecting the interest rate	191
11.6.3	The yield curve	194
11.7	Bibliographical references and extensions	195
IV Extensions		197
12	Asymmetric Information	199
12.1	Adverse selection	200
12.1.1	Full insurance	201
12.1.2	Pooling contracts	203
12.1.3	Separating contracts	205
12.2	Moral hazard	207
12.3	The principal-agent problem	211
12.3.1	Binary effort with a risk-neutral principal	211
12.3.2	Continuous effort with a risk-averse principal	215
12.4	Bibliographical references and extensions	217
13	Alternative decision criteria	221
13.1	The independence axiom and the Allais' paradox	224
13.2	Rank-dependent expected utility	226
13.3	Ambiguity aversion	230
13.4	Prospect theory and loss aversion	233
13.5	Some Concluding Thoughts	235
13.6	Bibliographical references and extensions	236

0.1 Preface

Risk is an ever-prevalent challenge to both individuals and society. When you dress yourself every morning, do you not ask what the weather will be like today? And after recalling the latest weather forecast, do you not wonder whether or not the forecast for today will be accurate? The weather forecast itself relies heavily on the rules of probability theory, as does the fact that today's weather might not behave as the forecast predicts. How you react to the uncertain weather ahead says something about your so-called "risk preferences." If the forecast calls for a 10% chance of rain, do you carry your umbrella when you walk to a restaurant for lunch? How about with a 50% chance of rain? Obviously the answer will not be the same for each individual.

Likewise, an individual may react differently to different consequences from the same risk. A person who decides she does not need to carry her umbrella with such a small risk of rain, may decide nonetheless to stop by the parking lot on the way to the restaurant to put the top up on her new cabriolet automobile. To quote from Peter Bernstein, "The ability to define what may happen in the future and to choose among alternatives lies at the heart of contemporary societies." (Bernstein, 1998) An understanding of risk and how to deal with it is an essential part of modern economies. Recognizing risks, quantifying risks, analyzing them, treating them and incorporating risks into our decision-making processes is the focus of this book.

Of course attempting to model human behavior is never easy. People may behave slightly differently from day to day. They also like to experiment in order to learn about their own tastes and preferences. Still, there are many basic principles that hold with much regularity. For the most part, this book models behavior using the expected utility model as developed in its modern form by von Neumann and Morgenstern (1948). While this basic approach is generally well accepted, it is not without its detractors. We discuss many of the major criticisms in the last chapter of this book.

It is important when reading this book to keep in mind that we are deriving models that help us to understand behavior towards risk. It is not assumed that people actually solve the mathematical problems that we present here. Indeed, most readers probably have a relative who cannot solve an optimization problem, yet decide every year to purchase an automobile insurance policy.

We also confine ourselves to risks that involve economic and financial

decisions. Obviously there are many other risks that one must deal with in everyday life, such as whether or not to take a new medication with potential untoward side effects, or which scientific journal provides the best publication outlet for a newly written research paper.

This book is designed for use in advanced undergraduate and beginning doctoral courses. We cover a broad array of topics in enough detail so that the book may be used as a self-contained text. Alternatively, one can use the first two “basics” chapters, together with a selection of later chapters, as a basis for courses in macroeconomics, insurance, portfolio choice and asset pricing. Such courses can easily adapt the book for the intended use, and supplement it with additional readings or projects.

The book starts by introducing the basic concepts of risk and risk aversion that are crucial throughout the rest of the text. Part two of the text applies these basic concepts to a multitude of personal decisions under risk. Part 3 uses the results about personal decision making to show how markets for risk are organized and how risky assets are priced. Our final part introduces two important points of departure: decision making under imperfect information and alternatives to the expected utility framework.

Each chapter of the book concludes with a discussion of the relevant literature, together with some suggestions for readers who would like to read more on the topic. We also provide an appendix that contains many problems related to each of the thirteen chapters.

The only mathematics contained in this book is calculus and simple algebra. We use discrete examples for time and for probabilities throughout the text. Although the mathematics is important, the logic and intuition are more important and this is stressed throughout the book. Many of the concepts that are derived here might not be easy to understand upon a first reading. We urge the readers to take the time to re-read difficult parts of the book and to work on the related problems in the Appendix.

The book’s three authors have spent collectively more than 60 years working on research projects related to the topics we present here. We each learned many new things while writing this book. And we continue to be curious, as we still have much to learn. We will feel that this book has been a “success,” if some of our curiosity transfers to the reader.

References

Bernstein, P. L., (1998), *Against the Gods*, John Wiley and Sons.

von Neumann, J. and O. Morgenstern, (1948), *Theory of Games and Economic Behavior*, Princeton University Press.

Louis Eeckhoudt, Mons (Belgium)

Christian Gollier, Toulouse (France)

Harris Schlesinger, Tuscaloosa, Alabama (USA)

Part I
Decision theory

Chapter 1

Risk aversion

Risk is an ever-prevalent challenge to both individuals and society. When you dress yourself every morning, do you not ask what the weather will be like today? And after recalling the latest weather forecast, do you not wonder whether or not the forecast for today will be accurate? The weather forecast itself relies heavily on the rules of probability theory, as does the fact that today's weather might not behave as the forecast predicts. How you react to the uncertain weather ahead says something about your so-called "risk preferences." If the forecast calls for a 10% chance of rain, do you carry your umbrella when you walk to a restaurant for lunch? How about with a 50% chance of rain? Obviously the answer will not be the same for each individual. Likewise, an individual may react differently to different consequences from the same risk. A person who decides she does not need to carry her umbrella with such a small risk of rain, may decide nonetheless to stop by the parking lot on the way to the restaurant to put the top up on her new cabriolet automobile. To quote from Peter Bernstein, "The ability to define what may happen in the future and to choose among alternatives lies at the heart of contemporary societies." (Bernstein, 1998) An understanding of risk and how to deal with it is an essential part of modern economies. Recognizing risks, quantifying risks, analyzing them, treating them and incorporating risks into our decision-making processes is the focus of this book.

As the reader can guess from its title, this chapter looks at a basic concept behind modeling individual preferences in the face of risk. As with any social science, we of course are fallible and susceptible to second-guessing in our theories. It is nearly impossible to model many natural human tendencies such as "playing a hunch" or "being superstitious." However, we can

develop a systematic way to view choices made under uncertainty. Hopefully, our models can capture the basic human tendencies enough to be useful in understanding market behavior towards risk. In other words, even if we are not correct in predicting behavior under risk for every individual in every circumstance, we can still make general claims about such behavior and can still make market predictions, which after all are based on the “marginal consumer.”

To use (vaguely) mathematical language, the understanding of this chapter is a *necessary* but not *sufficient* condition to go further into the analysis. Because of the importance of risk aversion in decision making under uncertainty, it is worthwhile to first take an “historical” perspective about its development and to indicate how economists and decision scientists progressively have elaborated upon the tools and concepts we now use to analyze risky choices. In addition, this “history” has some surprising aspects that are interesting in themselves. To this end, our first section in this chapter broadly covers these retrospective topics. Subsequent sections are more “modern” and they represent an intuitive introduction to the central contribution to our field, that of Pratt (1964).

1.1 An historical perspective on risk aversion

As it is now widely acknowledged, an important breakthrough in the analysis of decisions under risk was achieved when Daniel Bernoulli, a distinguished Swiss mathematician, wrote in Latin in St. Petersburg in 1738 a paper entitled : “*Specimen Theoriae Novae de Mensura Sortis*”, or “Exposition of a New Theory on the Measurement of Risk”. Bernoulli’s paper, translated in English in Bernoulli (1954), is essentially nontechnical. Its main purpose is to show that two people facing the same lottery may value it differently because of a difference in their psychology. This idea was quite novel at the time, since famous scientists before Bernoulli (among them Pascal and Fermat) had argued that the value of a lottery should be equal to its mathematical expectation and hence identical for all people, independently of their risk attitude.

In order to justify his ideas Bernoulli uses three examples. One of them, the “St. Petersburg paradox” is quite famous and it is still debated today in scientific circles. It is described in most recent texts of finance and microeconomics and for this reason we do not discuss it in detail here. Peter tosses

a fair coin repetitively until the coin lands head for the first time. Peter accepts to give to Paul 1 ducat if head appears on the first toss, 2 ducats if head appears only on the second toss, 4 ducats if head appears for the first time on the third toss, and so on in order to double the reward to Paul to each additional toss necessary to see head for the first time. The question raised by Bernoulli is to determine how much Paul would be ready to pay to Peter to accept to play this game.

Unfortunately, the celebrity of the paradox has overshadowed the other two examples given by Bernoulli that show that, most of the time, the value of a lottery is not equal to its mathematical expectation. One of these two examples, which presents the case of an individual named "Sempronius," wonderfully anticipates the central contributions that will be made to risk theory about 230 years later by Arrow, Pratt and others.

Let us quote Bernoulli:¹

"Sempronius owns goods at home worth a total of 4000 ducats and in addition possesses 8000 ducats worth of commodities in foreign countries from where they can only be transported by sea. However our daily experience teaches us that of [two] ships one perishes".

In modern-day language, we would say that Sempronius faces a risk on his wealth. This wealth may be represented by a lottery \tilde{x} , which takes on a value of 4000 ducats with probability $1/2$ (if his ship is sunk), or 12000 ducats with probability $1/2$. We will denote such a lottery \tilde{x} as being distributed as $(4000, 1/2; 12000, 1/2)$. Its mathematical expectation is given by:

$$E\tilde{x} \equiv \frac{1}{2}4000 + \frac{1}{2}12000 = 8000 \text{ ducats.}$$

Now Sempronius has an ingenious idea. Instead of "trusting all his 8000 ducats of goods to one ship" he now "trusts equal portions of these commodities to two ships". Assuming that the ships follow independent but equally dangerous routes, Sempronius now faces a more diversified lottery \tilde{y} distributed as $(4000, 1/4; 8000, 1/2; 12000, 1/4)$. Indeed, if both ships perish, he would end up with his sure wealth of 4000 ducats. Because the two risks are independent, the probability of these joint events equals the product of

¹We altered Bernoulli's probabilities to simplify the computations. In particular, Bernoulli's original example had one ship in ten perish.

the individual events, i.e., $(1/2)^2 = 1/4$. Similarly, both ships will succeed with probability $1/4$, in which case his final wealth amounts to 12000 ducats. Finally, there is the possibility that only one ship succeeds to download the commodities safely, in which case only half of the profit is obtained. The final wealth of Sempronius would then just amount to 8000 ducats. The probability of this event is $1/2$ because it is the complement of the other two events which have each a probability of $1/4$.

Since common wisdom suggests that diversification is a good idea, we would expect that the value attached to \tilde{y} exceeds that attributed to \tilde{x} . However, if we compute the expected profit, we obtain that

$$E\tilde{y} = \frac{1}{4}4000 + \frac{1}{2}8000 + \frac{1}{4}12000 = 8000 \text{ ducats,}$$

the same value as for $E\tilde{x}$! If Sempronius would measure his well-being ex ante by his expected future wealth, he should be indifferent to diversify or not. In Bernoulli's example, we obtain the same expected future wealth for both lotteries, even though most people would find \tilde{y} more attractive than \tilde{x} . Hence, according to Bernoulli and to modern risk theory, the mathematical expectation of a lottery is not an adequate measure of its value. Bernoulli suggests a way to express the fact that most people prefer \tilde{y} to \tilde{x} : a lottery should be valued according to the "expected utility" that it provides. Instead of computing the expectation of the monetary outcomes, we should use the expectation of the utility of the wealth. Notice that most human beings do not extract utility from wealth. Rather, they extract utility from consuming goods that can be purchased with this wealth. The main insight of Bernoulli is to suggest that there is a nonlinear relationship between wealth and the utility of consuming this wealth.

What ultimately matters for the decision maker ex post is how much satisfaction he or she can achieve with the monetary outcome rather than the monetary outcome itself. Of course, there must be a relationship between the monetary outcome and the degree of satisfaction. This relationship is characterized by a utility function u , which for every wealth level x tells us the level of "satisfaction" or "utility" $u(x)$ attained by the agent with this wealth. Of course, this level of satisfaction derives from the goods and services that the decision maker can purchase with a wealth level x . While the outcomes themselves are "objective", their utility is "subjective" and specific to each decision-maker depending upon his or her tastes and preferences. Although the function u transforms the objective result x into a perception

$u(x)$ by the individual, this transformation is assumed to exhibit some basic properties of rational behavior. For example, a higher level of x (more wealth) should induce a higher level of utility: the function should be increasing in x . Even for someone who is very altruistic, a higher x will allow them to be more philanthropic. Readers familiar with indirect utility functions from microeconomics (essentially utility over budget sets, rather than over bundles of goods and services) can think of $u(x)$ as essentially an indirect utility of wealth, where we assume that prices for goods and services are fixed. In other words, we may think of $u(x)$ as the highest achievable level of utility from bundles of goods that are affordable when our income is x .

Bernoulli argues that if the utility u is not only increasing but also concave in the outcome x , then the lottery \tilde{y} will have a higher value than the lottery \tilde{x} , in accordance with intuition. A twice-differentiable function u is concave if and only if its second derivative is negative, i.e. if the marginal utility $u'(x)$ is decreasing in x .² In order to illustrate this point let us consider a specific example of a utility function, such as $u(x) = \sqrt{x}$, which is an increasing and concave function of x . Using these preferences in Sempronius' problem, we can determine the expectation of $u(x)$:

$$Eu(\tilde{x}) = \frac{1}{2}\sqrt{4000} + \frac{1}{2}\sqrt{12000} = 86.4$$

$$Eu(\tilde{y}) = \frac{1}{4}\sqrt{4000} + \frac{1}{2}\sqrt{8000} + \frac{1}{4}\sqrt{12000} = 87.9.$$

Because lottery \tilde{y} generates a larger expected utility than lottery \tilde{x} , the former is preferred by Sempronius. The reader can try using concave utility functions other than the square root function to obtain the same type of result. In the next section, we formalize this result.

Notice that the concavity of the relationship between wealth x and satisfaction/utility u is quite a natural assumption. It simply implies that the marginal utility of wealth is decreasing with wealth: one values a one-ducats increase in wealth more when one is poorer than when one is richer. Observe that in Bernoulli's example, diversification generates a mean-preserving

²For simplicity, we maintain the assumption that u is twice differentiable throughout the book. However, a function need not be differentiable to be concave. More generally, a function u is concave if and only if $\lambda u(a) + (1 - \lambda)u(b)$ is smaller than $u(\lambda a + (1 - \lambda)b)$ for all (a, b) in the domain of u and all scalars λ in $[0, 1]$. A function must, however, be continuous to be concave.

transfer of wealth from the extreme events to the mean. Transferring some probability weight from $x = 4000$ to $x = 8000$ increases expected utility. Each probability unit transferred yields an increase in expected utility equaling to $u(8000) - u(4000)$. On the contrary, transferring some probability weight from $x = 12000$ to $x = 8000$ reduces expected utility. Each probability unit transferred yields a reduction in expected utility equaling to $u(12000) - u(8000)$. But the concavity of u implies that

$$u(8000) - u(4000) > u(12000) - u(8000), \quad (1.1)$$

i.e., that the positive effect of these combined mean-preserving transfers must dominate the negative effect. This is why all investors with a concave utility would support Sempronius' strategy to diversify risks.

1.2 Definition and characterization of risk aversion

We assume that the decision maker lives for only one period, which implies that he immediately uses all his final wealth to purchase and to consume goods and services. Later in this book, we will disentangle wealth and consumption by allowing the agent to live for more than one period. Final wealth comes from initial wealth w plus the outcome of any risk borne during the period.

Definition 1 *An agent is risk-averse if, at any wealth level w , he or she dislikes every lottery with an expected payoff of zero: $\forall w, \forall \tilde{z}$ with $E\tilde{z} = 0$: $Eu(w + \tilde{z}) \leq u(w)$.*

Observe that any lottery \tilde{z} with a non-zero expected payoff can be decomposed into its expected payoff $E\tilde{z}$ and a zero-mean lottery $\tilde{z} - E\tilde{z}$. Thus, from our definition, a risk-averse agent always prefers receiving the expected outcome of a lottery with certainty, rather than the lottery itself. For an expected-utility maximizer with a utility function u , this implies that, for any lottery \tilde{z} and for any initial wealth w ,

$$Eu(w + \tilde{z}) \leq u(w + E\tilde{z}). \quad (1.2)$$

If we consider the simple example from Sempronius' problem, with only one ship the initial wealth w equals 4000, and the profit \tilde{z} takes value 8000 or 0

1.2. DEFINITION AND CHARACTERIZATION OF RISK AVERSION 19

with equal probabilities. Because our intuition is that Sempronius must be risk averse, it must follow that

$$\frac{1}{2}u(12000) + \frac{1}{2}u(4000) \leq u(8000). \quad (1.3)$$

If Sempronius could find an insurance company that would offer full insurance at an actuarially fair price of $E\tilde{z} = 4000$ ducats, Sempronius would be better off by purchasing the insurance policy. We can check whether inequality (1.3) is verified in Figure 1.1. The right-hand side of the inequality is represented by point f on the utility curve u . The left-hand side of the inequality is represented by the middle point on the arc ae , i.e., by point c . This can immediately be checked by observing that the two triangles abc and cde are equivalent, since they have the same base and the same angles. We observe that f is above c : ex-ante, the welfare derived from lottery \tilde{z} is smaller than the welfare obtained if one were to receive its expected payoff $E\tilde{z}$ with certainty. In short, Sempronius is risk-averse. From this Figure, we see that this is true whenever the utility function is concave. The intuition of the result is very simple: if marginal utility is decreasing, then the potential loss of 4000 reduces utility more than the increase in utility generated by the potential gain of 4000. Seen ex-ante, the expected utility is reduced by these equally weighted potential outcomes.

INSERT FIGURE 1.1 ABOUT HERE

It is noteworthy that equations (1.1) and (1.3) are exactly the same. The preference for diversification is intrinsically equivalent to risk aversion, at least under the Bernoullian expected utility model.

Using exactly the opposite argument, it can easily be shown that if u is convex, the inequality in (1.2) will be reversed. Therefore, the decision maker prefers the lottery to its mathematical expectation and he reveals in this way his inclination for taking risk. Such individual behavior will be referred to as risk loving. Finally, if u is linear, then the welfare Eu is linear in the expected payoff of lotteries. Indeed, if $u(x) = a + bx$ for all x , then we have

$$Eu(w + \tilde{z}) = E[a + b(w + \tilde{z})] = a + b(w + E\tilde{z}) = u(w + E\tilde{z}),$$

which implies that the decision maker ranks lotteries according to their expected outcome. The behavior of this individual is called risk-neutral.

In the next Proposition, we formally prove that inequality (1.2) holds for any lottery \tilde{z} and any initial wealth w if and only if u is concave.

Proposition 2 *A decision maker with utility function u is risk-averse, i.e. inequality (1.2) holds for all w and \tilde{z} , if and only if u is concave.*

Proof: The proof of sufficiency is based on a second-order Taylor expansion of $u(w + z)$ around $w + E\tilde{z}$. For any z , this yields

$$u(w + z) = u(w + E\tilde{z}) + (z - E\tilde{z})u'(w + E\tilde{z}) + 0.5(z - E\tilde{z})^2u''(\xi(z))$$

for some $\xi(z)$ in between z and $E\tilde{z}$. Because this must be true for all z , it follows that the expectation of $u(w + \tilde{z})$ is equal to

$$Eu(w + \tilde{z}) = u(w + E\tilde{z}) + u'(w + E\tilde{z})E(\tilde{z} - E\tilde{z}) + 0.5E[(\tilde{z} - E\tilde{z})^2u''(\xi(\tilde{z}))].$$

Observe now that the second term of the right-hand side above is zero, since $E(\tilde{z} - E\tilde{z}) = E\tilde{z} - E\tilde{z} = 0$. In addition, if u'' is uniformly negative, then the third term takes the expectation of a random variable $(\tilde{z} - E\tilde{z})^2u''(\xi(\tilde{z}))$ that is always negative, as it is the product of a squared scalar and negative u'' . Hence, the sum of these three terms is less than $u(w + E\tilde{z})$. This proves sufficiency.

Necessity is proven by contradiction. Suppose that u is not concave. Then, there must exist some w and some $\delta > 0$ for which $u''(x)$ is positive in the interval $[w - \delta, w + \delta]$. Now take a small zero-mean risk $\tilde{\varepsilon}$ such that the support of final wealth $w + \tilde{\varepsilon}$ is entirely contained in $(w - \delta, w + \delta)$. Using the same Taylor expansion as above yields

$$Eu(w + \tilde{\varepsilon}) = u(w) + 0.5E[\tilde{\varepsilon}^2u''(\xi(\tilde{\varepsilon}))].$$

Because $\xi(\tilde{\varepsilon})$ has a support that is contained in $[w - \delta, w + \delta]$ where u is locally convex, $u''(\xi(\tilde{\varepsilon}))$ is positive for all realizations of $\tilde{\varepsilon}$. Consequently, it follows that $E[\tilde{\varepsilon}^2u''(\xi(\tilde{\varepsilon}))]$ is positive, and $Eu(w + \tilde{\varepsilon})$ is larger than $u(w)$. Thus, accepting the zero-mean lottery $\tilde{\varepsilon}$ raises welfare and the decision maker is not risk-averse. This is a contradiction. ■

The above Proposition is in fact nothing more than a rewriting of the famous Jensen's inequality. Consider any real-valued function ϕ . Jensen's inequality states that $E\phi(\tilde{y})$ is smaller than $\phi(E\tilde{y})$ for any random variable \tilde{y} if and only if ϕ is a concave function. It builds a bridge between two alternative definitions of the concavity of u : the negativity of u'' and the property that any arc linking two points on curve u must lie below this curve. Figure 1.1 illustrates this point. It is intuitive that decreasing marginal utility

($u'' < 0$) means risk aversion. In a certainty world, decreasing marginal utility means that an increase in wealth by 100 dollars has a positive effect on utility that is smaller than the effect of a reduction in wealth by 100 dollars. Then, in an uncertain world, introducing the risk to gain or to lose 100 dollars with equal probability will have a negative net impact on expected utility. In expectation, the benefit of the prospect to gain 100 dollars is outweighed by the cost of the prospect to lose 100 dollars with the same probability. Over the last two decades, many prominent researchers in the field challenged the idea that risk aversion comes only from decreasing marginal utility. Some even challenged the idea itself that there should be any link between the two.³

1.3 Risk premium and certainty equivalent

A risk-averse agent is an agent who dislikes zero-mean risks. The qualifier "zero-mean" is very important. A risk-averse agent may like risky lotteries if the expected payoffs that they yield is large enough. Risk-averse agents may want to purchase risky assets if their expected returns exceed the risk free rate. Risk-averse agents may dislike purchasing insurance if it is too costly to acquire. In order to determine the optimal trade-off between the expected gain and the degree of risk, it is useful to quantify the effect of risk on welfare. This is particularly useful when the agent subrogates the risky decision to others, as is the case when we consider public safety policy or portfolio management by pension funds for example. It is important to quantify the degree of risk aversion to help people to know themselves better, and to help them making better decisions in the face of uncertainty. Most of this book is precisely about this problem. Clearly, people have different attitudes towards risks. Some are ready to spend more money than others to get rid of a specific risk. One way to measure the degree of risk aversion of an agent is to ask her how much she is ready to pay to get rid of a *zero mean* risk \tilde{z} . The answer to this question will be referred to as the risk premium Π associated to that risk. For an agent with utility function u and initial

³This question will be discussed in the last chapter of this book. Yaari (1987) provides a model that is dual to expected utility where agents may be risk-averse in spite of the fact that their utility is linear in wealth.

wealth w , the risk premium must satisfy the following condition:

$$Eu(w + \tilde{z}) = u(w - \Pi). \quad (1.4)$$

The agent ends up with the same welfare either by accepting the risk or by paying the risk premium Π . When risk \tilde{z} has an expectation that differs from zero, we usually use the concept of the certainty equivalent. The certainty equivalent e of risk \tilde{z} is the sure increase in wealth that has the same effect on welfare as having to bear risk \tilde{z} , i.e.,

$$Eu(w + \tilde{z}) = u(w + e). \quad (1.5)$$

When \tilde{z} has a zero mean, comparing (1.4) and (1.5) implies that the certainty equivalent e of \tilde{z} equals minus its risk premium Π .

A direct consequence of Proposition 2 is that the risk premium Π is nonnegative when u is concave, i.e. when she is risk-averse. In Figure 1.2, we measure Π for the risk $(-4000, 1/2; 4000, 1/2)$ for initial wealth $w = 8000$. Notice first that the risk premium is zero when u is linear, and it is nonpositive when u is convex.

INSERT FIG 1.2 ABOUT HERE

One very convenient property of the risk premium, is that it is measured in the same units as wealth, e.g. we can measure Sempronius' risk premium in ducats. Although the measure of satisfaction or utility is hard to compare between different individuals – what would it mean to say Sempronius was "happier" than Alexander? – the risk premium is not. We can easily determine whether Sempronius or Alexander is more affected by risk \tilde{z} by comparing their two risk premia.

The risk premium is a complex function of the distribution of \tilde{z} , of initial wealth w and of the utility function u . We can estimate the amount that the agent is ready to pay for the elimination of this zero-mean risk by considering small risks. Assume that $E\tilde{z} = 0$. Using a second-order and a first-order Taylor approximation for the left-hand side and the right-hand side of equation (1.4) respectively, we obtain that

$$u(w - \Pi) \simeq u(w) - \Pi u'(w)$$

and

$$\begin{aligned} Eu(w + \tilde{z}) &\simeq E[u(w) + zu'(w) + 0.5\tilde{z}^2u''(z)] \\ &= u(w) + u'(w)E\tilde{z} + 0.5u''(w)E\tilde{z}^2 \\ &= u(w) + 0.5\sigma^2u''(w), \end{aligned}$$

where $E\tilde{z} = 0$ and $\sigma^2 = E\tilde{z}^2$ is the variance of the outcome of the lottery. Replacing these two approximations in equation (1.4) yields

$$\Pi \simeq \frac{1}{2}\sigma^2 A(w), \quad (1.6)$$

where the function A is defined as

$$A(w) = \frac{-u''(w)}{u'(w)}. \quad (1.7)$$

Under risk aversion, function A is positive. It would be zero or negative respectively for a risk-neutral or risk-loving agent. $A(\cdot)$ is hereafter referred to as the degree of absolute risk aversion of the agent. From (1.6), we see that the risk premium associated with risk $\tilde{\varepsilon}$ for an agent with wealth w is approximately equal to one-half the product of the variance of \tilde{z} and the degree of absolute risk aversion of the agent evaluated at w . Equation (1.6) is known as the Arrow-Pratt approximation, as it was developed independently by Arrow (1963) and Pratt (1964).

The cost of risk, as measured by the risk premium, is approximately proportional to the variance of its payoffs. Thus, the variance might appear to be a good measure of the degree of riskiness of a lottery. This observation induced many authors to use a mean-variance decision criterion for modeling behavior under risk. In a mean-variance model, we assume that individual risk attitudes only depend upon the mean and the variance of the underlying risks. However, the validity of these models is dependent on the degree of accuracy of the approximation in (1.6), which can be considered accurate only when the risk is small or in very special cases. In such cases, the mean-variance approach for decisions under risk, which has historically played a very important role in the development of the theory of finance, can be seen as a special case of the expected utility theory. In most cases however, the risk premium associated with any (large) risk will also depend upon the other moments of the distribution of the risk, not just its mean and variance. For example, it seems intuitive that whether or not \tilde{x} is symmetrically distributed about its mean matters for determining the risk premium. The degree of skewness (i.e. third moment) might very well affect the desirability of a risk. Hence, two risks with the same mean and variance, but one with a distribution that is skewed to the right and the other with a distribution that is skewed to the left, should not be expected to necessarily have the same risk premium. A similar argument can be made about the kurtosis

(fourth moment), which is linked to the probability mass in the tails of the distribution.

At this stage, it is worth noting that, at least for small risks, the risk premium increases with the size of the risk proportionately to the square of this size. To see this, let us assume that $\tilde{z} = k\tilde{\varepsilon}$, with $E\tilde{\varepsilon} = 0$. Parameter k can be interpreted as the size of the risk. When k tends to zero, the risk becomes very small. Of course, the risk premium is a function of the size of the risk. We may expect that this function $\Pi(k)$ is increasing in k . We are interested in describing the functional form linking the risk premium Π to the size k of the risk. Because the variance of \tilde{z} equals k^2 times the variance of $\tilde{\varepsilon}$,⁴ we obtain that

$$\Pi(k) \simeq \frac{1}{2}k^2\sigma_{\tilde{\varepsilon}}^2A(w),$$

i.e., the risk premium is approximately proportional to the square of the size of the risk. From this observation, we can observe directly that, not only does $\Pi(k)$ approach zero as k approaches zero, but also $\Pi'(0) = 0$. This is an important property of expected utility theory. At the margin, accepting a small zero-mean risk has *no* effect on the welfare of risk-averse agents! We say that risk aversion is a second-order phenomenon.⁵ In the small, we – the expected-utility maximizers – are all risk neutral.

Proposition 3 *If the utility function is differentiable, the risk premium tends to zero as the square of the size of the risk.*

Proof: In the following, we prove formally that $\Pi'(0) = 0$, as suggested by the Arrow-Pratt approximation in our comments above. The relationship between Π and k can be obtained by fully differentiating the equation $Eu(w + k\tilde{\varepsilon}) = u(w - \Pi(k))$ with respect to k . This yields

$$\Pi'(k) = \frac{-E\tilde{\varepsilon}u'(w + k\tilde{\varepsilon})}{u'(w - \Pi(k))}. \quad (1.8)$$

We directly infer that $\Pi'(0) = 0$, since by assumption $E\tilde{\varepsilon} = 0$. ■

⁴The general formula is

$$\text{Var}(a\tilde{x} + b\tilde{y}) = a^2\text{Var}(\tilde{x}) + b^2\text{Var}(\tilde{y}) + 2ab\text{Cov}(\tilde{x}, \tilde{y}).$$

⁵This property in general models, not restricted to expected utility, is called "second-order risk aversion." Within the expected-utility model, this property relies on the assumption that the utility function is differentiable.

1.4 Degree of risk aversion

Let us consider the following simple decision problem. An agent is offered a take-it-or-leave-it offer to accept lottery \tilde{z} with mean μ and variance σ^2 . Of course, the optimal decision is to accept the lottery if

$$Eu(w + \tilde{z}) \geq u(w), \quad (1.9)$$

or, equivalently, if the certainty equivalent e of \tilde{z} is positive. In the following, we examine how this decision is affected by a change in the utility function.

Notice at this stage that an increasing linear transformation of u has no effect on the decision maker's choice, and on certainty equivalents. Indeed, consider a function $v(\cdot)$ such that $v(x) = a + bu(x)$ for all x , for some pair of scalars a and b , where $b > 0$. Then, obviously $Ev(w + \tilde{z}) \geq v(w)$ yields exactly the same restrictions on the distribution of \tilde{z} as condition (1.9). The same analysis can be done on equation (1.5) defining certainty equivalents. The neutrality of certainty equivalents to linear transformations of the utility function can be verified in the case of small risks by using the Arrow-Pratt approximation. If $v \equiv a + bu$, it is obvious that $A(x) = -v''(x)/v'(x) = -bu''(x)/bu'(x) = -u''(x)/u'(x)$ for all x . Thus, by (1.6), risk premia for small risks are not affected by the linear transformation. Because the certainty equivalent equals the mean payoff of the risk minus the risk premium, the same neutrality property holds for certainty equivalents.

Limiting the analysis to small risks, we see from this analysis that agents with a larger absolute risk aversion $A(w)$ will be more reluctant to accept small risks. The minimum expected payoff that makes the risk acceptable for them will be larger. This is why we say that A is a measure of the degree of risk aversion of the decision maker. From a more technical viewpoint, $A = -u''/u'$ is a measure of the degree of concavity of the utility function. It measures the speed at which marginal utility is decreasing.

We are now interested to extend these observations to any risk, not only small risks. We consider the following definition for comparative risk aversion.

Definition 4 *Suppose that agents u and v have the same wealth w that is arbitrary. An agent v is more risk-averse than another agent u with the same initial wealth if any risk that is undesirable for agent u is also undesirable for agent v . In other words, the risk premium of any risk is larger for agent v than for agent u .*

This must be true independent of the common initial wealth level w of the two agents. If this definition were restricted to small risks, we know from the above analysis that this would be equivalent to requiring that

$$A_v(w) = \frac{-v''(w)}{v'(w)} \geq \frac{-u''(w)}{u'(w)} = A_u(w),$$

for all w . If limited to small risks, v is more risk-averse than u if function A_v is uniformly larger than A_u . We say in this case that v is more concave than u in the sense of Arrow-Pratt. It is important to observe that this is equivalent to the condition that v is a *concave* transformation of u , i.e., that there exists an increasing and concave function ϕ such that $v(w) = \phi(u(w))$ for all w . Indeed, we have that $v'(w) = \phi'(u(w))u'(w)$ and

$$v''(w) = \phi''(u(w)) (u'(w))^2 + \phi'(u(w))u''(w),$$

which implies that

$$A_v(w) = A_u(w) + \frac{-\phi''(u(w))u'(w)}{\phi'(u(w))}.$$

Thus, A_v is uniformly larger than A_u if and only if ϕ is concave. This is equivalent to requiring that A_v be uniformly larger than A_u or that v be a concave transformation of u . It yields that agent v values small risks less than agent u . Do we need to impose more restrictions to guarantee that agent v values any risk less than agent u , i.e., that v is more risk-averse than u ? The following Proposition, which is due to Pratt (1964), indicates that no additional restriction is required.

Proposition 5 *The following three conditions are equivalent:*

- a) *Agent v is more risk-averse than agent u , i.e., the risk premium of any risk is larger for agent v than for agent u .*
- b) *For all w , $A_v(w) \geq A_u(w)$.*
- c) *Function v is a concave transformation of function u : $\exists \phi(\cdot)$ with $\phi' > 0$ and $\phi'' \leq 0$ such that $v(w) = \phi(u(w))$ for all w .*

Proof: We have already shown that (b) and (c) are equivalent. That (a) implies (b) follows directly from the Arrow-Pratt approximation. We now prove that (c) implies (a). Consider any lottery \tilde{z} . Let Π_u and Π_v denote the risk premium for zero-mean lottery \tilde{z} of agent u and agent v respectively. By definition, we have that

$$v(w - \Pi_v) = Ev(w + \tilde{z}) = E\phi(u(w + \tilde{z})).$$

Define random variable \tilde{y} as $\tilde{y} = u(w + \tilde{z})$. Because ϕ is concave, $E\phi(\tilde{y})$ is smaller than $\phi(E\tilde{y})$ by Jensen's inequality. It thus follows that

$$v(w - \Pi_v) \leq \phi(Eu(w + \tilde{z})) = \phi(u(w - \Pi_u)) = v(w - \Pi_u).$$

Because v is increasing, this implies that Π_v is larger than Π_u . ■

In the case of small risks, the only thing that we need to know to determine whether a risk is desirable is the degree of concavity of u locally at the current wealth level w . For larger risks, the Proposition above shows that we need to know much more to take a decision. Namely, we need to know the degree of concavity of u at all wealth levels. The degree of concavity must be increased at all wealth levels to guarantee that a change in u makes the decision maker more reluctant to accept risks. If v is locally more concave at some wealth levels and is less concave at other wealth levels, the comparative analysis is intrinsically ambiguous.

To illustrate the Proposition, let us go back to the example of Sempronius' single ship yielding outcome $\tilde{z} = (0, 1/2; 8000, 1/2)$, with a initial wealth $w_0 = 4000$ ducats. If Sempronius' utility function is $u(w) = \sqrt{w}$, his certainty equivalent of \tilde{z} equals $e_u = 3464.1$, since

$$\frac{1}{2}\sqrt{4000} + \frac{1}{2}\sqrt{12000} = 86.395 = \sqrt{7464.1}$$

Alternatively, suppose that Sempronius' utility function is $v(w) = \ln(w)$, which is also increasing and concave. It is easy to check that v is more concave than u in the sense of Arrow-Pratt. Indeed, these functions yield

$$A_v(w) = \frac{1}{w} \geq \frac{1}{2w} = A_u(w)$$

for all w . From the above Proposition, this change in utility should reduce the certainty equivalent of any risk. In the case of $w_0 = 4000$ and $\tilde{z} \sim$

$(0, 1/2; 8000, 1/2)$, the certainty equivalent of \tilde{z} under v equals $e_v = 2928.5$, since

$$\frac{1}{2} \ln(4000) + \frac{1}{2} \ln(12000) = 8.8434 = \ln(6928.5).$$

Thus, e_v is smaller than e_u . Notice that the risk premium $\Pi_v = 1071.5$ under v is approximately twice the risk premium $\Pi_u = 535.9$. This was predicted by the Arrow-Pratt approximation, since A_v is equal to $2A_u$.

1.5 Decreasing absolute risk aversion and prudence

We have seen that risk aversion is driven by the fact that one's marginal utility is decreasing with wealth. In this section, we examine another question related to increasing wealth. Namely, we are interested in determining how the risk premium for a given zero-mean risk \tilde{z} is affected by a change in initial wealth w . Arrow argued that intuition implies that wealthier people are generally less willing to pay for the elimination of fixed risk. A lottery to gain or lose 100 with equal probability is potentially life-threatening for an agent with initial wealth $w = 101$, whereas it is essentially trivial for an agent with wealth $w = 1,000,000$. The former should be ready to pay more than the latter for the elimination of risk. We can check that this property holds for the square-root utility function, with $\Pi = 43.4$ when $w = 101$ and $\Pi = .0025$ when $w = 1,000,000$. If wealth is measured in euros, the individual would be willing to pay over 43 euros to avoid the risk when wealth is $w = 101$, whereas the same individual would not even pay one euro cent to get rid of this risk when wealth is one million euros! In the following, we characterize the set of utility functions that have this property.

The risk premium $\Pi = \pi(w)$ as a function of initial wealth w can be evaluated by solving the following equation

$$Eu(w + \tilde{z}) = u(w - \pi(w)) \tag{1.10}$$

for all w . Fully differentiating (1.10) with respect to w yields

$$Eu'(w + \tilde{z}) = (1 - \pi'(w))u'(w - \pi),$$

or, equivalently,

$$\pi'(w) = \frac{u'(w - \pi) - Eu'(w + \tilde{z})}{u'(w - \pi)}. \quad (1.11)$$

Thus, the risk premium is decreasing with wealth if and only if

$$Ev(w + \tilde{z}) \leq v(w - \pi(w)), \quad (1.12)$$

where function $v \equiv -u'$ is defined as minus the derivative of function u . Because the function v is increasing, we can also interpret it as *another* utility function. Condition (1.12) then just states that the risk premium of agent v is larger than the risk premium π of agent u . From Proposition 5, this is true if and only if v is more concave than u in the sense of Arrow-Pratt, that is, if $-u'$ is a concave transformation of u . For this utility v , the measure of absolute risk aversion is $A_v = A_{-u'} = -u'''/u''$. This measure has several uses, which will be made clearer later in this book. For this reason, without justifying the terminology at this stage, we will define $P(w) = -u'''(w)/u''(w)$ as the degree of absolute prudence of the agent with utility u . It follows from (1.12) that $-u'$ is more concave than u if and only if

$$P(w) \geq A(w)$$

for all w . We conclude that condition $P \geq A$ uniformly is necessary and sufficient to guarantee that an increase in wealth reduces risk premia. Because

$$A'(w) = A(w) [A(w) - P(w)],$$

condition $P \geq A$ is equivalent to the condition $A' \leq 0$. We obtain the following Proposition.

Proposition 6 *The risk premium associated to any risk \tilde{z} is decreasing in wealth if and only if absolute risk aversion is decreasing; or equivalently if and only if prudence is uniformly larger than absolute risk aversion.*

Observe that for the utility function $u(w) = \sqrt{w}$ satisfies this condition. Indeed, we have $A_u(w) = 0.5w^{-1}$, which is decreasing. This can alternatively be checked by observing that $v(w) = -0.5w^{-1/2}$ and $A_v(w) = P_u(w) = 1.5w^{-1}$, which is uniformly larger than $A_u(w)$. Notice that Decreasing Absolute Risk Aversion (DARA) requires that the third derivative of the utility function

be positive. Otherwise, prudence would be negative, which would imply that $P < A$, a condition that implies that absolute risk aversion would be increasing in wealth. Thus, DARA, a very intuitive condition, requires the necessary (but not sufficient) condition that u''' be positive, or that marginal utility be convex.

1.6 Relative risk aversion

Absolute risk aversion is the rate of decay for marginal utility. More particularly, absolute risk aversion measures the rate at which marginal utility decreases when wealth is increased *by one euro*.⁶ If the monetary unit would be the dollar, absolute risk aversion would be a different number. In other words, the index of absolute risk aversion is not unit free, as it is measured per euro (per dollar, or per yen).

Economists often prefer unit-free measurements of sensitivity. To this end, define the index of *relative* risk aversion R as the rate at which marginal utility decreases when wealth is increased *by one percent*. In terms of standard economic theory, this measure is simply the wealth-elasticity of marginal utility. It can be computed as

$$R(w) = -\frac{du'(w)/u'(w)}{dw/w} = \frac{-wu''(w)}{u'(w)} = wA(w). \quad (1.13)$$

Notice that the measure of relative risk aversion is simply the product of wealth and absolute risk aversion.

The (absolute) risk premium and the index of absolute risk aversion are linked by the Arrow-Pratt approximation and by Propositions 5 and 1.5. We can develop analogous kinds of results for relative risk aversion. Suppose that your initial wealth w is invested in a portfolio whose return \tilde{z} over the period is uncertain. Let us assume that $E\tilde{z} = 0$. Which share of your initial wealth are you ready to pay to get rid of this proportional risk? The solution to this problem is referred to as the relative risk premium $\hat{\Pi}$. This measure also is a unit-free measure, unlike the absolute risk premium, which is measured in

⁶In general, the *growth rate* for a function $f(x)$ is defined as $\frac{df(x)}{dx} \cdot \frac{1}{f(x)}$. Since marginal utility $u'(x)$ declines in wealth, its growth rate is negative. The absolute value of this negative growth rate, which is the measure of absolute risk aversion, is called the *decay rate*.

euros. It is defined implicitly via the following equation:

$$Eu(w(1 + \tilde{z})) = u(w(1 - \hat{\Pi})). \quad (1.14)$$

Obviously, the relative risk premium and the absolute risk premium are equal if we normalize initial wealth to unity. More generally, the relative risk premium for proportional risk \tilde{z} equals the absolute risk premium for absolute risk $w\tilde{z}$, divided by initial wealth w : $\hat{\Pi}(\tilde{z}) = \Pi(w\tilde{z})/w$. From this observation, we obtain that if agent v is more risk-averse than agent u with the same initial wealth, then agent v will be ready to pay a larger share of his wealth than agent u to insure against a given proportional risk \tilde{z} . Moreover, if σ^2 denotes the variance of \tilde{z} , then the variance of $w\tilde{z}$ equals $w^2\sigma^2$. Using the Arrow-Pratt approximation thus yields

$$\hat{\Pi}(\tilde{z}) = \frac{\Pi(w\tilde{z})}{w} \simeq \frac{\frac{1}{2}w^2\sigma^2 A(w)}{w} = \frac{1}{2}\sigma^2 R(w). \quad (1.15)$$

The relative risk premium is approximately equal to half of the variance of the proportional risk times the index of relative risk aversion. This can be used to establish a range for acceptable degrees of risk aversion. Suppose that one's wealth is subject to a risk of a gain or loss of 20% with equal probability. What is the range that one would find reasonable for the share of wealth π that one would be ready to pay to get rid of this zero-mean risk? From our various experiments in class, we found that most people would be ready to pay between 2% and 8% of their wealth. Because risk \tilde{z} in this experiment has a variance of $0.5(0.2)^2 + 0.5(-0.2)^2 = 0.04$, using approximation (1.15) yields a range for relative risk aversion between 1 and 4. This will be a useful information later in this book.

There is no definitive argument for or against decreasing relative risk aversion. Arrow originally conjectured that relative risk aversion is likely to be constant, or perhaps increasing, although he stated that the intuition was not as clear as was the intuition for decreasing absolute risk aversion. Since then, numerous empirical studies have offered conflicting results. We might also try to examine this question by introspection. If your wealth would increase, would you want to devote a larger or a smaller share of your wealth to get rid of a given zero-mean proportional risk? For example, what you pay to avoid the risk of gaining or losing 20% of your wealth, each with an equal probability? If the share is decreasing with wealth, you have decreasing relative risk aversion. There are two contradictory effects here

that need to be considered. On the one hand, under the intuitive DARA assumption, becoming wealthier also means becoming less risk-averse. This effect tends to reduce π . But, on the other hand, becoming wealthier also means facing a larger absolute risk $w\tilde{z}$. This effect tends to raise π . There is no clear intuition as to whether the first effect or the second effect will dominate. For example, many of the classic models in macroeconomics are based on relative risk aversion being constant over all wealth levels, which is implicitly assuming that our two effects exactly cancel each other out. Of course, there also is no a-priori reason to believe that the dominant effect will not change over various wealth levels. For instance, some recent empirical evidence indicates a possible "U-shape" for relative risk aversion, with R decreasing at low wealth levels, then leveling off somewhat before increasing at higher wealth levels.

1.7 Some classical utility functions

As already noticed above, expected utility (EU) theory has many proponents and many detractors. In chapter 12, we examine some generalizations of the expected utility criterion that satisfy those who find expected-utility too restrictive. But researchers in both economics and finance have long considered – and most of them still do – EU theory as an acceptable paradigm for decision making under uncertainty. Indeed, EU theory has a long and prominent place in the development of decision-making under uncertainty. Even detractors of the theory use EU as a standard by which to compare alternative theories. Moreover, many of the models in which EU theory has been applied can be modified, often yielding better results.

Whereas the current trend is to generalize the EU model, researchers often *restrict* EU criterion by considering a specific subset of utility functions. This is done to obtain tractable solutions to many problems. It is important to note the implications that derive from the choice of a particular utility function. Some results in the literature may be robust enough to apply for all risk-averse preferences, while others might be restricted to applying only for a narrow class of preferences. In this section, we examine several particular types of utility functions that are often encountered in the economics and the finance literature. Remember that utility is unique only up to a linear transformation.

Historically, much of the theory of finance has been developed during the

sixties by considering the subset of utility functions that are quadratic of the form

$$u(w) = aw - \frac{1}{2}w^2 \quad \text{for } w \leq a.$$

Note that the domain of wealth on which u is defined comes from the necessary requirement that u be nondecreasing, which is true only if w is smaller than a . This set of functions is useful because the expected utility generated by any distribution of final wealth is a function of only the first two moments of this distribution:

$$Eu(\tilde{w}) = aE\tilde{w} - 0.5E\tilde{w}^2.$$

Therefore, in this case, the expected utility theory simplifies to a mean-variance approach to decision-making under uncertainty. However, as already discussed, it is very hard to believe that preferences among different lotteries be determined only by the mean and variance of these lotteries.

Above $w = a$, marginal utility becomes negative. Since quadratic utility is decreasing in wealth for $w > a$, many people might feel this is not appropriate as a utility function. However, it is important to remember that we are trying to model human behavior with mathematical models. For example, if the quadratic utility function models your behavior quite well with $a = 100$ million euros, is it really a problem that this function declines for higher wealth levels? The point is that the quadratic utility might work well for more realistic wealth levels, and if it does, we should not be overly concerned about its properties at unrealistically high wealth levels. However, the quadratic utility function has another property that is more problematic. Namely, the quadratic utility functions exhibit increasing absolute risk aversion:

$$A(w) = \frac{1}{a-w} \Rightarrow A'(w) = \frac{1}{(a-w)^2} > 0.$$

For this reason, quadratic utility functions are not as much in fashion anymore.

A second set of classical utility functions is the set of so called constant-absolute risk-aversion (CARA) utility functions, which are exponential functions characterized by

$$u(w) = -\frac{\exp(-aw)}{a},$$

where a is some positive scalar. The domain of these functions is the real line. The distinguishing feature of these utility functions is that they exhibit constant absolute risk aversion, with $A(w) = a$ for all w . It can be shown that the Arrow-Pratt approximation is exact when u is exponential and \tilde{w} is normally distributed with mean μ and variance σ^2 . Indeed, we can take expectations to see that

$$\begin{aligned} Eu(\tilde{w}) &= \frac{-1}{\sigma a \sqrt{2\pi}} \int \exp(-aw) \exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right) dw \\ &= -\frac{1}{a} \exp(-a(\mu - 0.5a\sigma^2)) \left[\frac{1}{\sigma \sqrt{2\pi}} \int \exp\left(-\frac{(w-(\mu-0.5a\sigma^2))^2}{2\sigma^2}\right) dw \right] \\ &= -\frac{1}{a} \exp(-a(\mu - 0.5a\sigma^2)) = u(\mu - 0.5a\sigma^2). \end{aligned} \tag{1.16}$$

The third equality comes from the fact that the bracketed term is the integral of the density of the normal distribution $N(\mu - 0.5a\sigma^2, \sigma)$, which must be equal to unity. Thus, the risk premium is indeed equal to $0.5\sigma^2 A(w)$. In this very specific case, we obtain that the Arrow-Pratt approximation is exact. The fact that risk aversion is constant is often useful in analyzing choices among several alternatives. As we will later see, this assumption eliminates the income effect when dealing with decisions to be made about a risk whose size is invariant to changes in wealth. However this is often also the main criticism of CARA utility, since absolute risk aversion is constant rather than decreasing.

Finally, one set of preferences that has been by far the most used in the literature is the set of power utility functions. Researchers in finance and in macroeconomics are so accustomed to this restriction that many of them don't even mention it anymore when they present their results. Suppose that

$$u(w) = \frac{w^{1-\gamma}}{1-\gamma} \quad \text{for } w > 0.$$

The scalar γ is chosen so that $\gamma > 0$, $\gamma \neq 1$. It is easy to show that γ equals the degree of relative risk aversion, since $A(w) = \gamma/w$ and $R(w) = \gamma$ for all w . Thus, this set exhibits decreasing absolute risk aversion and constant relative risk aversion, which are two reasonable assumptions. For this reason, these utility functions are called the constant-relative-risk-aversion (CRRA) class of preferences. Notice that our definition does not allow for $\gamma = 1$. However, it is straightforward to show that function $u(w) = \ln(w)$ satisfies the property that $R(w) = 1$ for all w . Thus, the set of all CRRA utility

functions is completely defined by⁷

$$u(w) = \begin{cases} \frac{w^{1-\gamma}}{1-\gamma} & \text{for } \gamma \geq 0, \gamma \neq 1 \\ \ln(w) & \text{for } \gamma = 1. \end{cases} \quad (1.17)$$

As we will later see in this book, this class of utility functions eliminates any income effects when making decisions about risks whose size is proportional to one's level of wealth. For example, the relative risk premium $\widehat{\Pi}$ defined by equation (1.14) is independent of wealth w in this case. Assuming that relative risk aversion is constant enormously simplifies many of the problems often encountered in macroeconomics and finance.

1.8 Bibliographical references and extensions

The contribution by Pratt (1964) basically opened and closed the field covered in this chapter. It is however fair to mention that the measure of absolute risk aversion has been discovered independently by Arrow (1963) and de Finetti (1952). The paper by de Finetti was written in Italian and even today is not given the attention it deserves. The paper by Pratt is by far the most advanced in defining the notions of an increase in risk aversion and of decreasing absolute risk aversion. The orders of risk aversion are introduced by Segal and Spivak, (1990).

Ross (1981) challenged the idea that $A = -u''/u'$ is a good measure of the degree of risk aversion of an agent. Kihlstrom, Romer and Williams (1981) and Nachman (1982) showed that if initial wealth is uncertain, it is not true that an agent v who is more risk-averse than another agent u in the sense of Arrow-Pratt will be ready to pay more to get rid of another risk. Ross (1981) characterized the conditions on u and v that implies that $\Pi_v \geq \Pi_u$ even when initial wealth is uncertain and potentially correlated with the risk under scrutiny. This condition is of course stronger than $A_v \geq A_u$.

⁷We can also show that $u(w) = \ln(w)$ as a limiting case of the power utility function. To this end, rewrite the power utility function, using a linear transformation, as $u(w) = \frac{1}{1-\gamma}(w^{1-\gamma} - 1)$. Taking the limit as $\gamma \rightarrow 1$ and applying l'Hospital's rule, we obtain

$$\lim_{\gamma \rightarrow 1} u(w) = \lim_{\gamma \rightarrow 1} \frac{-(w^{1-\gamma}) \ln(w)}{-1} = \ln(w).$$

There is much contradictory empirical evidence on the shape of relative risk aversion as a function of wealth. Many authors have empirically estimated R assuming that we have CRRA. Fewer authors have examined whether R might be increasing or decreasing in wealth. A good summary of many of these results appears in Ait-Sahalia and Lo (2000).

References

- Ait-Sahalia, Y. and A. W. Lo, (2000), Nonparametric risk management and implied risk aversion, *Journal of Econometrics*, 94, 9-51.
- Arrow, K.J., (1963), Liquidity preference, Lecture VI in " *Lecture Notes for Economics 285, The Economics of Uncertainty*", pp 33-53, undated, Stanford University.
- Arrow, K.J., (1965), *Yrjo Jahnsson Lecture Notes*, Helsinki. Reprinted in Arrow (1971).
- Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing*. Chicago: Markham Publishing Co.
- Bernoulli, D., (1954), Exposition of a new theory on the measurement of risk, translated in english by Louise Sommer, *Econometrica*, vol. 22, pp. 23-36.
- Bernstein, P. L., (1998), *Against the Gods*, John Wiley and Sons.
- de Finetti, B., (1952), Sulla preferibilita, *Giornale Degli Economisti E Annali Di Economia*, 11, 685-709.
- Kihlstrom, R., D. Romer and S. Williams, (1981), Risk aversion with random initial wealth, *Econometrica*, 49, 911-920.
- Nachman, D.C., (1982), Preservation of 'More Risk Averse' Under Expectations, *Journal of Economic Theory*, 28, 361-368.
- Pratt, J., (1964), Risk aversion in the small and in the large, *Econometrica*, 32, 122-136.
- Ross, S.A., (1981), Some stronger measures of risk aversion in the small and in the large with applications, *Econometrica*, 3, 621-638.
- Segal, U. and A. Spivak, (1990), First order versus second order risk aversion, *Journal of Economic Theory*, 51, 111-125.

Yaari, M.E., (1987), The dual theory of choice under risk, *Econometrica*, 55, 95-115.

Chapter 2

The measures of risk

In Chapter 1, we defined the concept of risk aversion by considering the effect of the introduction of a zero-mean risk on welfare. That is, we assumed that the initial environment of the consumer was risk free. Our only conclusion from risk aversion was that the individual preferred no risk to a zero-mean risk. But what about choices among different zero-mean risks? For example, recall from the previous chapter that we might interpret Sempronius' situation as facing the risk $(-4000, 1/2; 4000, 1/2)$ with initial wealth $w = 8000$. The alternative, using two separate ships, can be thought of as facing the risk $(-4000, 1/4; 0, 1/2; +4000, 1/4)$ from the same initial wealth. We argued that the second alternative seemed more valuable, in some sense. In this Chapter, we will examine this question more closely and consider the comparison of such competing risks.

If one would know the utility function of the agent, ranking lotteries would be easy. For example, let us compare two different wealth prospects (i.e. two distributions of final wealth) \tilde{w}_1 and \tilde{w}_2 . The first is preferred to the second by an agent with utility function u if $Eu(\tilde{w}_1) \geq Eu(\tilde{w}_2)$. This preference order that is specific to a single utility function u is complete in the sense that for any pair $(\tilde{w}_1, \tilde{w}_2)$, either \tilde{w}_1 is preferred to \tilde{w}_2 , \tilde{w}_2 is preferred to \tilde{w}_1 , or we are indifferent to both wealth prospects. In this chapter, we consider several relatively weak restrictions on preferences, as for example that "agents are risk-averse" or that "agents are prudent". We want to find restrictions on the change in risk from \tilde{w}_1 to \tilde{w}_2 that is unanimously disliked by the group of agents under scrutiny. In that case, we say that \tilde{w}_1 dominates \tilde{w}_2 for this class of utility functions. As soon as the class is not limited to a single utility function, these preference orders are incomplete in the sense that it is not

true that for any pair of lotteries, one must necessarily dominate the other. Some people in the group may prefer the first, whereas other members in the group may prefer the second. Imposing unanimity in the group is a very strong constraint on the change in risk. Considering a larger group makes the constraint of unanimity stronger. On the other hand, if we do find unanimity among our group that \tilde{w}_1 dominates \tilde{w}_2 , it allows to choose \tilde{w}_1 over \tilde{w}_2 when these are our only two options.

The theory of stochastic dominance looks at certain statistical properties of the distributions of \tilde{w}_1 and \tilde{w}_2 , which allow us to infer unanimous agreement for certain classes of preferences. This is important not only for our understanding of individual behavior, but also for decision making designed to benefit a group, such as a corporate manager making decisions on behalf of the company's shareholders. In this book, we will consider basically three stochastic orders. In section 1, we first consider the most natural set of utility functions from what we have seen in chapter 1. We consider in that section the set of all risk-averse agents. It generates the concept of an increase in risk that was first examined by Rothschild and Stiglitz (1970). In section 2, we focus on the set of prudent agents, which yields the concept of an increase in downside risk introduced by Menezes, Geiss and Tressler (1980). Finally, in section 3, we assume only that agents have an increasing utility function. The corresponding stochastic order is called "first-order stochastic dominance."

2.1 Increases in risk

In this section, we characterize the changes in risk that make all risk-averse agents worse off. We focus the analysis on changes in risk which preserve the expected outcome, i.e., mean-preserving changes in risk. These changes are called "increases in risk". There are at least three equivalent ways to define them.

2.1.1 Adding noise

Consider the binary lottery faced by Sempronius using a single ship. This may be written as $\tilde{w}_1 \sim (4000, 1/2; 12000, 1/2)$. In the low wealth state, the ship is lost, whereas in the high wealth state, the ship succeeds in bringing the spices safely to the harbor. Let us assume that the ship contains 8000 pounds

of spices which will be sold at a unit price of one ducat. This environment generates a distribution \tilde{w}_1 for Sempronius' final wealth.

Suppose alternatively that the price at which the spices will be sold at the harbor is unknown at the time when the ship leaves the East Indies. More precisely, let us suppose that the unit price will be either 0.5 ducat or 1.5 ducat with equal probabilities. In this alternative environment, the final wealth is still 4000 in case of the ship being sunk. But conditional on the ship arriving safely in Europe, Sempronius' final wealth will be either 8000 or 16000 with equal probabilities, or $12000 + \tilde{\varepsilon}$, with $\tilde{\varepsilon} \sim (-4000, 1/2, 4000, 1/2)$. Because $E\tilde{\varepsilon} = 0$, the price uncertainty adds a zero-mean noise to Sempronius' final wealth conditional upon the no-loss state. Ex ante, the agent faces an uncertain wealth distributed as $\tilde{w}_2 \sim (4000, 1/2; 12000 + \tilde{\varepsilon}, 1/2)$. This situation describes what is called a compound lottery, i.e., a lottery for which some of the outcomes are themselves lotteries.

The intuition suggests that Sempronius should dislike this additional uncertainty. The reader can check that this is indeed the case e.g. if Sempronius' utility function is a square root. We have that

$$Eu(\tilde{w}_1) = \frac{1}{2}\sqrt{4000} + \frac{1}{2}\sqrt{12000} = 86.395,$$

whereas

$$Eu(\tilde{w}_2) = \frac{1}{2}\sqrt{4000} + \frac{1}{2}\left[\frac{1}{2}\sqrt{8000} + \frac{1}{2}\sqrt{16000}\right] = 85.606 < 86.395.$$

The same qualitative result would hold if Sempronius would have another concave utility function. In fact, adding a zero-mean noise conditional upon some specific state always reduces the expected utility of risk-averse agents, as we now show.

To keep the presentation relatively simple, suppose that \tilde{w}_1 can take n different possible values $\omega_1, \omega_2, \dots, \omega_n$. Let p_s denote the probability that \tilde{w}_1 takes value ω_s . Suppose that the alternative wealth distribution \tilde{w}_2 be obtained by compounding \tilde{w}_1 with zero-mean noises $\tilde{\varepsilon}_s$ for the different outcomes ω_s of \tilde{w}_1 . This means that each outcome ω_s of \tilde{w}_1 is replaced by $\omega_s + \tilde{\varepsilon}_s$ with $E\tilde{\varepsilon}_s = 0$:

$$\tilde{w}_2 = \tilde{w}_1 + \tilde{\varepsilon} \quad \text{with} \quad E[\tilde{\varepsilon} \mid \tilde{w}_1 = \omega_s] = E[\tilde{\varepsilon}_s] = 0.$$

The price uncertainty presented above is an example of this technique of adding noises to each possible outcome of the initial distribution of wealth.

With this notation, it is easy to show that any such alternative lottery \tilde{w}_2 makes all risk-averse agents worse off. Because $E\tilde{\varepsilon}_s$ is zero, risk aversion implies that $Eu(\omega_s + \tilde{\varepsilon}_s) \leq u(\omega_s)$. It follows that

$$Eu(\tilde{w}_2) = \sum_{s=1}^n p_s Eu(\omega_s + \tilde{\varepsilon}_s) \leq \sum_{s=1}^n p_s u(\omega_s) = Eu(\tilde{w}_1).$$

All risk-averse agents dislike adding zero-mean noises to the possible outcomes of their wealth.

2.1.2 Mean-preserving spreads in probability

The existence of price uncertainty in the situation faced by Sempronius can alternatively be seen as transferring probability masses. In Figure 2.1(a), we represent the probability distribution in the absence of price uncertainty. Figure 2.1(b) describes the probability distribution when the price uncertainty is taken into account. We see that adding noise $\tilde{\varepsilon} \sim (-4000, 1/2, 4000, 1/2)$ is equivalent to transfer half of the $1/2$ probability mass at 12000 to 8000, and the remaining of the probability mass at 12000 to 16000. By doing this, we do not modify the center of gravity of the probability distribution, i.e., we preserve the mean. In short, we construct what is called a "mean-preserving spread" of the probability distribution.

INSERT FIGURE 2.1 ABOUT HERE

Let $f_i(w)$ denote the probability mass of \tilde{w}_i at w . In the case of a continuous distribution, $f_i(\cdot)$ is the probability density of \tilde{w}_i . The following definition formalizes the concept of a mean-preserving spread, which is an operation consisting of the partial removal of probability mass from some interval I in order to transfer it outside this interval.

Definition 7 \tilde{w}_2 is a mean-preserving spread (MPS) of \tilde{w}_1 if

1. $E\tilde{w}_2 = E\tilde{w}_1$, and
2. there exists an interval I such that $f_2(w) \leq f_1(w)$ for all w in I , and $f_2(w) \geq f_1(w)$ for all w outside I .

Adding noise or constructing a sequence of MPS's are obviously two equivalent ways to increase risk. In some circumstances, it is easier to use one representation than the other. For example, let us compare distribution

$\tilde{w}_1 \sim (4000, 1/2; 12000, 1/2)$ to distribution $\tilde{w}'_2 \sim (2000, 1/2; 14000, 1/2)$. Obviously, the two distributions have the same mean, and the second is obtained by transferring some probability mass from interval $I = [4000, 12000]$ outside I . Thus, \tilde{w}_2 is an increase in risk of \tilde{w}_1 . It must be the case that \tilde{w}_2 is obtained from \tilde{w}_1 by adding some noise $\tilde{\varepsilon}_1$ to outcome 4000 and another noise $\tilde{\varepsilon}_2$ to outcome 12000 of \tilde{w}_1 . The reader may easily verify that indeed defining

$$\tilde{\varepsilon}_1 \sim (-2000, 5/6; 10000, 1/6) \quad \text{and} \quad \tilde{\varepsilon}_2 \sim (2000, 5/6; -10000, 1/6)$$

does the job.

It is often useful to translate the definition of a mean-preserving spread into a condition on the *cumulative* distribution functions of \tilde{w}_1 and \tilde{w}_2 . Let $F_i(w)$ denote the probability that \tilde{w}_i be no greater than w . That is, define $F_i(w) = \sum_{s|\omega_s \leq w} f_i(s)$ in the discrete case, and $F_i(w) = \int^w f_i(s)ds$ in the continuous case. In this latter case, the density function f_i is simply the derivative of F_i . To keep the level of technicality at a minimum, let us assume that all possible final wealth levels are in the interval $[a, b]$. Suppose that \tilde{w}_2 is an MPS of \tilde{w}_1 . Integrating by parts, preservation of the mean implies that

$$\int_a^b [F_2(s) - F_1(s)]ds = - \int_a^b s[f_2(s) - f_1(s)]ds = E\tilde{w}_2 - E\tilde{w}_1 = 0.$$

The fact that the expectation is preserved means that the area between F_1 and F_2 (counted "+" when F_2 is above F_1 and "-" otherwise) must sum up to zero. Also, by definition of an MPS, the derivative of F_2 is smaller (resp. larger) than the derivative of F_1 within the interval I (resp. outside I). Thus, F_2 must be larger than F_1 to the left of some threshold \hat{w} and F_2 must be smaller than F_1 to its right. We illustrate this property in Figure 2.2 in the continuous case, and in Figure 2.3 in the discrete case considered in the previous paragraph.

INSERT FIGURES 2.2 AND 2.3 ABOUT HERE

This so-called "single-crossing" property of MPS implies in particular that

$$S(w) = \int_a^w [F_2(s) - F_1(s)]ds \geq 0 \tag{2.1}$$

for all w , with an equality when w equals b . This integral condition is examined in more details in the next section.

2.1.3 The integral condition and risk-averse preferences

We now examine the problem of characterizing changes in risk that reduce the expected utility of all risk-averse agents. By integrating by parts, we obtain

$$Eu(\tilde{w}_i) = \int_a^b u(\omega) f_i(\omega) d\omega = u(\omega) F_i(\omega) \Big|_{\omega=a}^{\omega=b} - \int_a^b u'(\omega) F_i(\omega) d\omega,$$

or, equivalently, that

$$Eu(\tilde{w}_i) = u(b) - \int_a^b u'(\omega) F_i(\omega) d\omega.$$

It follows that

$$Eu(\tilde{w}_2) - Eu(\tilde{w}_1) = \int_a^b u'(\omega) [F_1(\omega) - F_2(\omega)] d\omega. \quad (2.2)$$

The difference in expected utility in transforming \tilde{w}_1 into \tilde{w}_2 is equal to the areas between F_1 and F_2 ("+" if F_1 is above F_2 , and "-" otherwise) weighted by the marginal value of wealth. Integrating by parts once again yields

$$Eu(\tilde{w}_2) - Eu(\tilde{w}_1) = -u'(\omega) S(\omega) \Big|_{\omega=a}^{\omega=b} + \int_a^b u''(\omega) S(\omega) d\omega,$$

where function S is defined by equation (2.1) and is such that $S'(w) = F_2(w) - F_1(w)$. Because we focussed the analysis on changes in risk that preserve the mean, we have that $S(a) = S(b) = 0$. The above equation thus simplifies to

$$Eu(\tilde{w}_2) - Eu(\tilde{w}_1) = \int_a^b u''(\omega) S(\omega) d\omega. \quad (2.3)$$

Equation (2.3) implies that all risk-averse agents dislike mean-preserving increases in risk, that is changes in risk for which the condition $S(w) \geq 0$ is satisfied for all w . This condition would indeed imply that the integrand in (2.3) is uniformly negative. Its integral in $[a, b]$ should therefore be negative. Condition $S(w) \geq 0$ for all w is also necessary to guarantee that every risk averter would unanimously prefer \tilde{w}_1 into \tilde{w}_2 . Indeed, suppose by contradiction that S is positive in some interval $J \subseteq [a, b]$. Let us consider the concave

utility function u that is linear outside J , and which is strictly concave in J . Then, from equation (2.3), agent u increases her expected utility from transforming \tilde{w}_1 into \tilde{w}_2 . The integrand $u''S$ is zero for w outside J and is positive for w in J .

To sum up, the condition that $S(w) = \int^w (F_2(s) - F_1(s)) ds$ be nonnegative is both a necessary and a sufficient condition for mean-preserving changes in risk to reduce the expected utility of all risk-averse agents. It was examined by Rothschild and Stiglitz (1970). Notice from equation (2.2) that the condition $S(w) \geq 0$ implies that agents with the concave utility functions $u_w(x) = \min(x, w) \forall w \in [a, b]$ all prefer risk \tilde{w}_1 to risk \tilde{w}_2 . We hope that this observation makes this integral condition less artificial.

There is a clear link between the integral condition $S \geq 0$ and the notion of a mean-preserving spread. It has been partly derived at the end of the previous section, where we have shown that a mean-preserving spread *implies* that S is nonnegative. Rothschild and Stiglitz (1970) have shown that the integral condition is *equivalent* to a sequence of mean-preserving spreads. In fact, they have proved the following Proposition, showing how several interpretations of a mean-preserving increase in risk are all the same.

Proposition 8 *Consider two random variables \tilde{w}_1 and \tilde{w}_2 with the same mean. The following four conditions are equivalent:*

- (a) *All risk-averse agents prefer \tilde{w}_1 to \tilde{w}_2 : $Eu(\tilde{w}_2) \leq Eu(\tilde{w}_1)$ for all concave functions u .*
- (b) *\tilde{w}_2 is obtained from \tilde{w}_1 by adding zero-mean noise terms to the possible outcomes of \tilde{w}_1 : $\tilde{w}_2 \stackrel{d}{=} \tilde{w}_1 + \tilde{\varepsilon}$, with $E[\tilde{\varepsilon} | \tilde{w}_1 = \omega] = 0$ for all ω , where " $\stackrel{d}{=}$ " means "equal in distribution."*
- (c) *\tilde{w}_2 is obtained from \tilde{w}_1 by a sequence of mean-preserving spreads.*
- (d) *$S(w) \equiv \int^w (F_2(w) - F_1(w)) dw \geq 0$ for all w .*

Any one of these four equivalent conditions define what we call an increase in risk from \tilde{w}_1 to \tilde{w}_2 . Correspondingly, a change from \tilde{w}_2 to \tilde{w}_1 is labelled a reduction in risk.

2.1.4 Preference for diversification

In Chapter 1, we have shown that Sempronius prefers to transfer the spices from the colonies by two ships rather than by only one. By doing so, he diversifies the risk. Let us now formalize this example by defining the random variable \tilde{x}_i which takes value 0 if ship i is sunk (probability $1/2$), and which takes value 1 otherwise. In short, \tilde{x}_i is distributed as $(0, 1/2; 1, 1/2)$. By assumption, the risks faced by the two ships are independent. If Sempronius puts his 8000 pounds of spice in ship 1, his final wealth would equal

$$\tilde{w}_2 = w + 8000\tilde{x}_1.$$

We assume here that the price of spice is risk free and normalized to unity. If he splits the goods in two equal parts to be brought to London in ships 1 and 2, his final wealth equals

$$\tilde{w}_1 = w + 8000 \left(\frac{\tilde{x}_1 + \tilde{x}_2}{2} \right) = w + 8000\tilde{y},$$

where $\tilde{y} = 0.5(\tilde{x}_1 + \tilde{x}_2)$ can be interpreted as the rate of success. The rate of success is distributed as $(0, 1/4; 1/2, 1/2; 1, 1/4)$. It is easy to check that \tilde{x}_1 can be obtained from \tilde{y} by adding the noise $\tilde{\varepsilon} \sim (-1/2, 1/2; +1/2, 1/2)$ conditional upon $\tilde{y} = 1/2$, as seen in Figure 2.4. It then follows from Proposition 8 that, independent of the utility function of Sempronius, he must prefer two ships to one ship as soon as this function is concave. Diversifying the transfer of spice to two ships is a way to diversify the risk faced by Sempronius. Not "putting all the eggs in the same basket" is a rational behavior for risk-averse agents.

INSERT FIGURE 2.4 ABOUT HERE

More generally, one can verify that, if \tilde{x}_1 and \tilde{x}_2 are two independent and identically distributed (i.i.d.) random variables, then $\tilde{y} = 0.5(\tilde{x}_1 + \tilde{x}_2)$ is a reduction of risk with respect to \tilde{x}_1 . We have that

$$\tilde{x}_1 = \tilde{y} + \tilde{\varepsilon} \quad \text{with} \quad \tilde{\varepsilon} = \frac{\tilde{x}_1 - \tilde{x}_2}{2}$$

and

$$E[\tilde{\varepsilon} \mid \tilde{y} = y] = E\left[\frac{\tilde{x}_1 - \tilde{x}_2}{2} \mid \frac{\tilde{x}_1 + \tilde{x}_2}{2} = y\right] = E\left[\tilde{x}_1 \mid \frac{\tilde{x}_1 - \tilde{x}_2}{2}\right] - E\left[\tilde{x}_2 \mid \frac{\tilde{x}_1 - \tilde{x}_2}{2}\right]$$

which must be equal to zero by symmetry. Thus, \tilde{x}_1 is riskier than \tilde{y} . In other words, diversification is a risk reduction device in the sense of Rothschild Stiglitz. All risk-averse agents should diversify their risks when possible. This guideline does not rely on any preference restrictions other than risk aversion.

2.1.5 And the variance?

The risk premium is the amount of money that the agent is ready to pay to eliminate the (zero-mean) risk. Facing the risk \tilde{w}_i or receiving its certainty equivalent $E\tilde{w}_i - \Pi_i$ generates the same expected utility. Consider two risky wealth prospects \tilde{w}_1 and \tilde{w}_2 with equal means. It is clear that \tilde{w}_1 is preferred to \tilde{w}_2 if and only if Π_2 is larger than Π_1 . Proposition 8 states the conditions on \tilde{w}_1 and \tilde{w}_2 that guarantee that Π_2 is larger than Π_1 . Notice that for small risks, we can use the Arrow-Pratt approximation

$$\Pi_i \simeq 0.5\sigma_i^2 A$$

to claim that \tilde{w}_1 is preferred to \tilde{w}_2 if and only if the variance of \tilde{w}_2 is larger than the variance of \tilde{w}_1 : $\sigma_2^2 \geq \sigma_1^2$. Should not it also be the case that this holds for larger risks as well? That is to say, could we not add another equivalent statement (e) to Proposition 8 that would be written as follows:

(e) *the variance of \tilde{w}_2 is larger than the variance of \tilde{w}_1 ?*

The answer is definitely no! In general, statistical moments of orders higher than 2 will matter to compare two random variables. By limiting the development of Taylor of the utility function to the second order, the Arrow-Pratt approximation is of no value for larger risks. The only exception is when the utility function is quadratic. A correct statement would then be that all risk-averse agents *with a quadratic utility function* prefer \tilde{w}_1 to \tilde{w}_2 if and only if the variance of the second is larger than the variance of the first.¹

However, it is easy to check that an increase in variance is a necessary, but not sufficient, condition for an increase in risk. Indeed, it is a necessary condition for those agents with a quadratic concave utility function to dislike this change in risk. Thus, it is a necessary for *all* agents with a concave function to dislike it.

¹Another strategy is to limit the set of random variables to those that can be parametrized by their mean and variance only, as the set of normal distributions.

It is noteworthy that the necessary condition that the variance is increased can be written as

$$\sigma_2^2 - \sigma_1^2 = \int_a^b w^2(f_2(w) - f_1(w))dw = 2 \int_a^b S(w)dw \geq 0.$$

This equation is a direct consequence of equation (2.3) applied for $u(w) = w^2$. We see that the increase in variance just means that the integral of function S must be nonnegative. The Rothschild-Stiglitz increase in risk is a much stronger requirement that S be uniformly nonnegative.

2.2 Aversion to downside risk

In this short section, we explore another set of changes in risk which are referred to as increases in downside risk. These changes have the property that they preserve both the mean and the variance of final wealth. To illustrate, consider again the distribution of final wealth $\tilde{w}_2 \sim (4000, 1/2; 12000 + \tilde{\varepsilon}, 1/2)$, with $\tilde{\varepsilon} \sim (-4000, 1/2, 4000, 1/2)$. Sempronius faces the price risk $\tilde{\varepsilon}$ in addition to the risk of losing his single ship. Observe that this is a situation where the additional zero-mean risk $\tilde{\varepsilon}$ is borne in the good state, i.e., when the ship arrives safely at the harbor of London. Consider an alternative situation where this zero-mean risk would be borne in the bad state. It would thus yield a final wealth distributed as $\tilde{w}_3 \sim (4000 + \tilde{\varepsilon}, 1/2; 12000, 1/2)$. Which of these two distributions of final wealth do you think Sempronius would prefer? To assist you in your choice, we represent these two distribution in Figure 2.5.

INSERT FIGURE 2.5 ABOUT HERE

Observe that the means are the same: $E\tilde{w}_2 = E\tilde{w}_3 = 8000$. The variance is also unchanged by the change in distribution: $\sigma_1^2 = \sigma_2^2 = 24 \times 10^6$. In fact, the reader can check that function S alternates in sign in the interval of final wealth levels $[0, 16000]$. Thus, this cannot be an increase in risk. Hence, by Proposition 8, some risk-averse agents will like this change, whereas others will dislike it. However, experiments have shown that most people in the real world prefer \tilde{w}_2 to \tilde{w}_3 . That is to say, they prefer to bear a zero-mean risk in the wealthier state. In other words, they dislike transferring a zero-mean risk from a richer to a poorer state. In this case, we say that they are averse to downside risk.

We are interested in determining a condition on the utility function that guarantees that an agent is averse to downside risk. Suppose that the agent is initially facing a risk that is characterized by $\tilde{w} \sim (z_1, 1/n; z_2, 1/n; \dots; z_n, 1/n)$. We assume for simplicity that these n states have the same probability of occurrence. Consider an additional risk $\tilde{\varepsilon}$ with a zero mean. The expected utility of final wealth depends upon the state i to which this additional risk is imposed. We denote

$$V_i = \frac{1}{n}Eu(z_i + \tilde{\varepsilon}) + \sum_{j \neq i} \frac{1}{n}u(z_j)$$

for the expected utility when $\tilde{\varepsilon}$ is borne in state i . Observe that

$$n(V_i - V_j) = [Eu(z_i + \tilde{\varepsilon}) - u(z_i)] - [Eu(z_j + \tilde{\varepsilon}) - u(z_j)] = \int_{z_j}^{z_i} [Eu'(\omega + \tilde{\varepsilon}) - u'(\omega)] d\omega.$$

Although intuition suggests the empirical observation that $V_i > V_j$ when $z_i > z_j$, we see from the above equation, that this is true if and only if

$$Eu'(\omega + \tilde{\varepsilon}) \geq u'(\omega)$$

for all ω . Because $\tilde{\varepsilon}$ is constrained only to have a zero mean, this condition is satisfied if and only if $u'(\cdot)$ is itself convex, i.e., if the agent is prudent. This is a direct application of Jensen's inequality. We thus obtain the following result:

Proposition 9 *An agent dislikes any increase in downside risk if and only if he is prudent.*

Prudence and aversion to downside risk are two equivalent concepts.

2.3 First-degree stochastic dominance

Up to now, we focused the analysis to changes in risk that preserve the mean. This is a strong requirement. For example, two portfolios with different shares invested in stocks typically have different expected returns. Or, purchasing more insurance typically induces a reduction in expected wealth, since the insurance premium likely contains a loading. More generally, most

decision making under uncertainty yields a trade-off between risk and (expected) return. In this section, we explore an important stochastic order named "First-degree Stochastic Dominance" (FSD) in which changes in mean are required.

There is often a discrepancy in the common use of the wording of "an increase in risk" between economists and the rest of the world. In common language, one often says that the risk is increased when the probability of an accident is increased. However, taken to the extreme this would imply that someone who always has an accident with the highest possible loss is the most "risky," whereas in a technical sense, there is no risk at all involved here: the lowest wealth value is realized with certainty! Of course, if the probability of an accident is increased, the expected final wealth of the risk bearer is reduced, which implies that this change in risk cannot be an increase in risk in the sense of Rothschild and Stiglitz. Economists say that the risk undergoes a dominated shift in the sense of FSD. More generally, any change in risk that is generated by a transfer of probability mass from high wealth states to low wealth states are said to be FSD-deteriorating. Such transfers of probability obviously raise $F(w)$, the probability that final wealth be no greater than w , for all w .

Definition 10 \tilde{w}_2 is dominated by \tilde{w}_1 in the sense of the first-degree stochastic dominance order if $F_2(w) \geq F_1(w)$ for all w .

It is obvious that all consumers in the real world dislike FSD-dominated shifts in the distribution of final wealth. Rewriting condition (2.2)

$$Eu(\tilde{w}_2) - Eu(\tilde{w}_1) = - \int_a^b u'(\omega)[F_2(\omega) - F_1(\omega)]d\omega, \quad (2.4)$$

we see that $Eu(\tilde{w}_2)$ is smaller than $Eu(\tilde{w}_1)$ if \tilde{w}_2 is dominated by \tilde{w}_1 in the sense of FSD and if u' is positive. These two conditions indeed imply that the integrand of the above equation is always positive. Suppose that the only restriction that we impose on the utility function is that it be nondecreasing: more wealth is preferred to less. This means that we allow for both risk aversion and risk-loving behavior. Then, equation (2.4) tells us that $F_2 - F_1$ nonnegative is a necessary and sufficient condition for $Eu(\tilde{w}_2)$ to be smaller than $Eu(\tilde{w}_1)$. To prove this suppose by contradiction that $F_2 - F_1$ is negative in the neighborhood of some ω_0 . Then, consider the nondecreasing utility function that is flat everywhere except in this neighborhood of ω_0 . For this

specific utility function, the integrand of (2.4) is zero everywhere except in the neighborhood of ω_0 where it is negative. Thus, the integral is negative, and this agent prefers \tilde{w}_2 to \tilde{w}_1 , a contradiction. We have thus just proven the equivalence of (a) and (b) in the following Proposition. The equivalence of (c) has been shown by several authors.

Proposition 11 *The following conditions are equivalent:*

- (a) *All agents with a nondecreasing utility function prefer \tilde{w}_1 to \tilde{w}_2 : $Eu(\tilde{w}_2) \leq Eu(\tilde{w}_1)$ for all nondecreasing functions u .*
- (b) *\tilde{w}_2 is dominated by \tilde{w}_1 in the sense of FSD: \tilde{w}_2 is obtained from \tilde{w}_1 by a transfer of probability mass from the high wealth states to lower wealth states, or $F_2(\omega) \geq F_1(\omega)$ for all ω .*
- (c) *\tilde{w}_1 is obtained from \tilde{w}_2 by adding nonnegative noise terms to the possible outcomes of \tilde{w}_2 : $\tilde{w}_1 \stackrel{d}{=} \tilde{w}_2 + \tilde{\varepsilon}$, where $\tilde{\varepsilon} \geq 0$ with probability one.*

Without surprise, this type of change in risk, where the probability of lower wealth states is increased, is disliked by a very wide set of agents.

Of course, we can combine various changes in risk. For example, combining any FSD-dominated shift in distribution with any increase in risk yields what is called a Second-degree Stochastically Dominated (SSD) shift in distribution. Obviously, SSD shifts are disliked by the set of agents with a nondecreasing and concave utility function. Combining any SSD shift with any increase in downside risk yields a Third-degree Stochastically dominated (TSD) shift in distribution. They are disliked by all prudent agents with a nondecreasing and concave utility function.

From these three stochastic dominance orders, FSD is the most demanding one. For many applications however, it is considered as too broad to yield unambiguous comparative static results. In the principal-agent literature, for example, one usually uses the stricter concept associated to the Monotone Likelihood Ratio (MLR) order. We say that \tilde{w}_2 is dominated by \tilde{w}_1 in the sense of MLR if $f_2(\omega)/f_1(\omega)$ is nonincreasing in ω . One can check that MLR is a special case of FSD.

2.4 Bibliographical references and extensions

The origin of the concepts developed in the literature on stochastic dominance can be found in a old book by famous mathematicians Hardy, Littlewood and

Polya (1934). Its revival in the late sixties is due to Hadar and Russell (1969), Hanoch and Levy (1969) for the concepts of first-degree and second-degree stochastic dominance, while Rothschild and Stiglitz (1970) discussed mean-preserving increases in risk, a special case of SSD. Whitmore (1971) and Menezes, Geiss and Tressler (1980) were interested in third-degree stochastic dominance. The proof that diversification is liked by all risk-averse agents can be found in Samuelson (1967) and Rothschild and Stiglitz (1971).

References

- Geiss, C. , C. Menezes and J. Tressler, (1980), Increasing Downside Risk, *American Economic Review*, 70, 5, 921-931.
- Hadar, J., and W.R. Russell (1969), Rules for ordering uncertain prospects, *American Economic Review*, 59, 25-34.
- Hanoch, G., and H. Levy, (1969), Efficiency analysis of choices involving risk, *Review of Economic Studies*, 36, 335-346.
- Hardy, Littlewood and Polya, (1934), *Inequalities*, reprinted in 1997 by Cambridge University Press.
- Rothschild, M. and J. Stiglitz, (1970), Increasing risk: I. A definition, *Journal of Economic Theory*, 2, 225-243.
- Rothschild, M. and J. Stiglitz, (1971), Increasing risk: II Its economic consequences, *Journal of Economic Theory* 3, 66-84.
- Samuelson, P.A., (1967), General proof that diversification pays, *Journal of Financial and Quantitative Analysis*.
- Whitmore, G.A., (1970), Third-degree stochastic dominance, *American Economic Review*, 60, 457-459.

Part II

Risk management

Chapter 3

Insurance decisions

Insurance occurs when one party agrees to pay an indemnity to another party in case of the occurrence of a prespecified random event generating a loss for the initial risk-bearer. The most common example is an insurance policy, where the insurer is compensated by being paid a fixed premium by the policyholder. But many other contracts involve some form of insurance. For example, in share-cropping contracts, a landlord agrees to reduce the rent for his land in case of a low crop yield. In cost-plus contracts, a buyer agrees to pay a higher price if the producer incurs an unexpected increase in cost. In the case of income taxes, the state partially insures the losses of taxpayers by reducing the tax payment when incomes are low.

The shifting of risk is of considerable importance for the functioning of our modern economies.¹ Insurance allows for disentangling investment decisions from risk-taking decisions. Without it, we would certainly not have experienced the historical economic growth of the last century. Ford, Solvay, Rockefeller and the others would not have taken the investment risks that they actually took without the possibility to share the risk with shareholders and insurers. Similarly, many consumers might not purchase new expensive cars or houses if they did not have a possibility to insure them. Without an acceptable social net, young people would not engage in profitable but very risky investments in their human capital or in risky professional activities where their talents would mostlikely be recognized.

By pooling the risks of many policholders, the insurer can take advantage

¹Apparently, it was important in ancient economies as well. For example, ancient Chinese, Babylonian, Greek and Roman cultures already had various types of insurance and risk-sharing arrangements. See Outreville (1997).

of the Law of Large Numbers. So long as there is not much correlation between the insured risks of different policyholders, the insurer can diversify away its risk. For this reason, it is often convenient to think of the insurer as risk neutral – only the level of expected profits is what matters to the insurer. Indeed, the insurer might be thought of as a type of intermediary who collects and disperses funds amongst the policholders. So in some sense, it is essentially the policyholders who are insuring one another. This concept is often referred to as the mutuality principle.

Insurance is a particular example of a type of risk-transfer strategy known as hedging. Hedging strategies typically involve entering into contracts whose payoffs are negatively related to one's overall wealth or to one component of that wealth. Thus, for example, if wealth falls, the value of the contract rises, partially offsetting the loss in wealth. For instance, one might enter into contracts in the futures market to hedge against exchange-rate risk, when part of one's income is in a foreign currency. Or one might use an option contract on the S&P 500 Index to protect a pension fund against a percipitous fall in the value of stocks. Such options and futures contracts are typically based on financial-market data. Moreover, they contain various standardized attributes which make them fairly "liquid" assets, i.e., which allow them to be readily bought and sold in the market place. However, these hedging instruments typically entail another type of risk called a basis risk, which is a risk that the payoff does not offset losses exactly. For example, the value of one's pension fund is not likely to be perfectly correlated with the S&P 500 Index, and hence Index options will be an imperfect hedge.

Unlike these contracts, insurance is based on the level of one's own individual loss rather than some index. Since there is no financial market for this unique loss, insurance contracts are not easily tradeable in secondary markets, and transaction costs are high. Even if a policyholder needed more insurance for her home, it would not help her to buy your homeowners-insurance policy, since your policy will only pay when you have a loss, rather than when the policyholder has a loss. Thus, there is no secondary market for insurance contracts. In other words, compared to options and futures contracts, insurance is a rather "illiquid" asset. At the same time, insurance is a perfect hedge – the insurance indemnity is based on the occurrence of a prespecified loss. Insurance contracts do not contain the basis risk which is prevalent in options and futures contacts.²

²At least there is no basis risk in theory. In reality, the exact value of a loss is often not

There is an added value to the policyholder from insurance because policyholders are risk-averse, that is they dislike risk on their wealth. Consider an individual facing a random loss \tilde{x} to her wealth, where $\tilde{x} \geq 0$. An insurance contract stipulates a premium to be paid by the policyholder, P , and an indemnity schedule, $I(x)$, which indicates the amount to be paid by the insurer for a loss of size x . There is full coverage if the insurer reimburses the policyholder for the full value of any loss, so that $I(\cdot)$ is the identity function, $I(x) = x$. The actuarial value of the contract is the expected indemnity $EI(\tilde{x})$, which is the expected gross payoff from the insurance contract. The insurance premium is said to be actuarially fair (or often just "fair") if it is equal to the actuarial value of the contract, i.e., $P = EI(\tilde{x})$. When the premium is fair, the expected net payoff on the insurance contract is zero. The purchase of a full insurance contract at an actuarially-fair premium has the effect of replacing a random loss \tilde{x} by its expectation $P = E\tilde{x}$. The private value of such a contract is equal to the value of the Arrow-Pratt risk premium attached to the risk \tilde{x} by the policyholder. Indeed, if we let Π denote this Arrow-Pratt risk premium, then the maximum premium the individual would be willing to pay for a full-coverage insurance policy is $P = E\tilde{x} + \Pi$. This maximum premium is increasing with the policyholder's degree of risk aversion and with the riskiness of the loss. In other words, buying full insurance at a fair price would provide the policy owner with a surplus value of Π , when compared to the case of having no insurance.

When insurance prices are actuarially fair, the insurance decision is simple for risk-averse agents: full insurance is optimal, as we show below. But insurance contracts typically entail transaction costs in the real world. In many lines of casualty insurance, for example, transaction costs may be as large as 30% of the premium. When we add these costs into the picture, the optimal insurance decision is less obvious, since risk-averse policyholders must compare the marginal cost of more insurance to its marginal benefit coming from the risk reduction of the contract. In other words, there is a tradeoff between risk and expected final wealth. This chapter is mainly devoted to the analysis of this trade-off.

perfectly observable, or else subject to some debate. Likewise, insurance companies might not have sufficient funds to pay all of their liabilities. These possibilities would introduce a type of basis risk, but are ignored here in modeling a theory of insurance contracts.

3.1 Optimal insurance: An illustration

In this section, we examine the insurance problem faced by Sempronius when he has only one ship. He has an initial wealth of 4000 ducats that would be increased by 8000 ducats only if his ship arrives safely. Both the insurers and Sempronius evaluate the probability of this event to be equal to $1/2$. The insurers, who can diversify Sempronius's risk among a large set of shareholders, are assumed to be risk neutral. They offer a menu of insurance policies. A specific policy is fully described first by the indemnity I that is paid to Sempronius if his ship is sunk, and second, by the insurance premium P . In expectation, insurers will have to pay $I/2$ to Sempronius as an indemnity. $I/2$ is called the actuarial value of the policy. In addition to the indemnity, insurers bear various costs that have been evaluated to represent 10% of the actuarial value of the policy. Because insurance markets are competitive, an equilibrium condition is that expected profits be zero. This yields the following insurance tariff:

$$P(I) = (I/2) + 0.1(I/2) = 0.55 I.$$

It implies that Sempronius' expected wealth is decreasing with his insurance coverage:

$$\text{Expected wealth} = 8000 - 0.05I.$$

Sempronius must decide which insurance policy to purchase. Suppose that Sempronius has a square root utility function $u(z) = \sqrt{z}$. In Table 3.1 and in Figure 3.1, we compute Sempronius' expected utility as a function of the indemnity I . It equals

$$EU = 0.5\sqrt{4000 + I - P(I)} + 0.5\sqrt{12000 - P(I)}.$$

I	P	EU
0	0	86.395
1000	550	86.856
2000	1100	87.202
3000	1650	87.439
4000	2200	87.576
5000	2750	87.617
6000	3300	87.564
7000	3850	87.418
8000	4400	87.178

Table 3.1: Sempronius' expected utility as a function of his insurance coverage I .

INSERT FIGURE 3.1 HERE

We see that purchasing some insurance increases Sempronius' expected utility. The positive effect of reducing risk dominates the negative effect of reducing expected wealth. However, a closer look at these data shows that some insurance coverages are better than others. When I is small, a marginal increase in the insurance coverage has a net positive effect on Sempronius' expected utility. But when I is large, the opposite happens. The expected utility is concave in I . Its representation in Figure 3.1 is hump-shaped, with a maximum at $I^* = 4929.29$ ducats.

It is important to understand why expected utility is not a monotone function of the insurance coverage. We stressed in Chapter 1 that the cost of risk is approximately proportional to the square of the size of risk. If Sempronius purchases an insurance policy $(I, P(I))$, the variance of final wealth equals $(4000 - I/2)^2$. Using the Arrow-Pratt approximation, the cost of the uninsured risk is proportional to this variance. It implies that the marginal benefit of insurance, which is to marginally reduce the size of the uninsured risk, decreases linearly as insurance coverage increases. The marginal benefit is in fact approximately proportional to the derivative of the variance with respect to I , i.e., it is approximately proportional to $4000 - I/2$. When Sempronius is almost fully insured, i.e., when I is close to 8000, raising the coverage to full insurance has no benefit. Risk aversion is a second-order effect.

The cost of insurance equals the sure reduction of wealth corresponding to the deadweight transaction cost. It equals $0.5\lambda I = 0.05I$. The marginal cost of insurance is therefore a constant independent of the level of coverage I . Combining these observations explains why expected utility is hump-shaped with respect to the level of coverage, as shown in Figure 3.1.

3.2 Optimal coinsurance

Consider a risk-averse agent with initial wealth w_0 bearing a risk of loss \tilde{x} . Contrary to the illustration presented in the previous section, \tilde{x} needs not to be a binary random variable. Suppose that for each euro of indemnity paid by the insurance policy, the insurer incurs a cost λ of deadweight transaction

costs, including implicit costs. Obviously, more complex cost structures are also possible. Adding these transaction costs to the expected costs of the indemnity itself, the premium for insurance indemnity schedule $I(\cdot)$ must be equal to $P = (1 + \lambda)EI(\tilde{x})$. The level λ is often referred to as the loading factor, or the loading for profit and expenses. Thus, if expenses amount to 10 cents for each euro of indemnity, the competitive insurer will add 10 percent to the actuarial value of an insurance policy, in order to cover these expenses.

When the loss variable \tilde{x} is not binary, covering the risk can take many different forms characterized by function $I(\cdot)$. One that applies in many cases and is the easiest to work with from a modeling standpoint is a so-called coinsurance policy. That is, suppose that for a fixed premium the insurer agrees to reimburse the individual for a fixed fraction, β , of the loss. Thus $I(x) = \beta x$ for all x . The level β is called the coinsurance rate, while $1 - \beta$ is called the retention rate, since the policyholder retains this fraction of the loss. Quite often, β may be restricted, such as requiring $0 \leq \beta \leq 1$, though this need not be the case.

The coinsurance rate β is chosen a priori by the policyholder, given the following insurance pricing rule:

$$P(\beta) = (1 + \lambda)EI(\tilde{x}) = \beta P_0 \quad (3.1)$$

where $P_0 = (1 + \lambda)E\tilde{x}$ is the full insurance premium. The random final wealth of the policyholder with loss \tilde{x} and coinsurance rate β equals

$$\tilde{y} \equiv y(\tilde{x}, \beta) \equiv w_0 - \beta P_0 - (1 - \beta)\tilde{x}. \quad (3.2)$$

Note that when $\beta = 1$, we have full insurance coverage, and final wealth is non-random, $y(\tilde{x}, \beta) = w_0 - P_0$. On the other hand, when $\beta = 0$ we have the case where no coverage is purchased. The decision problem of the policyholder is to select an optimal coinsurance rate β^* .

Given this decision problem, the policyholder selects the coinsurance rate which maximizes the expected utility of her final wealth:

$$\max_{\beta} H(\beta) \equiv Eu(\tilde{y}) = Eu(w_0 - \beta P_0 - (1 - \beta)\tilde{x}). \quad (3.3)$$

Differentiating the objective function H twice with respect to β yields

$$H'(\beta) = \frac{\partial Eu(\tilde{y})}{\partial \beta} = E[(\tilde{x} - P_0)u'(\tilde{y})] \quad (3.4)$$

and

$$H''(\beta) = \frac{\partial^2 Eu(\tilde{y})}{\partial \beta^2} = E[(\tilde{x} - P_0)^2 u''(\tilde{y})]. \quad (3.5)$$

Observe that H'' is the expectation of the product of $(\tilde{x} - P_0)^2$, which is always positive, and $u''(\tilde{y})$, which is always negative, by risk aversion. Thus, H'' is the expectation of something which is negative with probability 1. It is therefore negative. This means that the expected utility of the policyholder is a concave function of the coinsurance rate. As for Sempronius' insurance problem, expected utility is hump-shaped with respect to the level of insurance that is measured here by β . It follows that the first-order condition

$$H'(\beta^*) = E[(\tilde{x} - P_0)u'(\tilde{y})] = 0 \quad (3.6)$$

is both necessary and sufficient for the maximization program (3.3). The fact that (3.5) is negative for all β and not just for β^* turns out to be important in our comparative static analyses below.

We can obtain some important insights by examining the sign of $\frac{\partial Eu(\tilde{y})}{\partial \beta}$ evaluated at $\beta = 1$,

$$H'(1) = E[(\tilde{x} - P_0)u'(w_0 - P_0)] = -\lambda u'(w_0 - P_0)E\tilde{x}. \quad (3.7)$$

This implies two important results. First, suppose that there are no transaction costs in the insurance process: $\lambda = 0$. In such a situation, $H'(1) = 0$ and the first-order condition (3.6) is satisfied with $\beta^* = 1$. In other words, when there is no insurance loading, the optimal contract is full insurance. This result is hardly surprising. When $\lambda = 0$, (3.1) and (3.2) together imply that $E\tilde{y} = Ey(\tilde{x}, \beta)$ is constant for all values of β . In other words, expected final wealth is not affected by our insurance choice. Therefore, a risk averter will prefer this expected wealth level with no risk at all, which is achievable by purchasing full insurance.

Suppose alternatively that there are nonzero transaction costs to the insurance process, $\lambda > 0$. This implies, from (3.7), that $H'(1)$ is negative. This means that reducing the rate of coverage from 100% to a marginally smaller rate raises the policyholder's expected utility. Since H'' is negative, so that $H(\beta)$ is strictly concave, it follows that the β^* at which H' vanishes must be

less than one. In other words, $\beta^* < 1$ and it is optimal for the policyholder to retain some of the risk. These results are summarized in the following Proposition, which is sometimes known as Mossin's Theorem.

Proposition 12 (*Mossin's Theorem*): *Full insurance ($\beta^* = 1$) is optimal at an actuarially fair price, $\lambda = 0$, while partial coverage ($\beta^* < 1$) is optimal if the premium includes a positive loading, $\lambda > 0$.*

When λ is positive, one might believe that full insurance could still be optimal if the degree of risk aversion of the policyholder is sufficiently high. However, this intuition is not correct, as shown in the above Proposition. The reason is that risk aversion is a second-order phenomenon, as reflected in the Arrow-Pratt approximation of the risk premium. For a very small level of risk, individual behavior towards risk approaches risk neutrality. By retaining some of the risk, $(1 - \beta) > 0$, the policyholder will save on transaction costs, thereby increasing her expected final wealth, which of course is beneficial if someone is risk neutral. When β tends to unity, this first-order effect must dominate the second-order effect of increasing the retained risk. Thus, full insurance will not be optimal.³

At the other extreme, when $\lambda > 0$ it may turn out to be the case that β^* is very small. If we restrict $\beta \geq 0$, then the first-order condition (3.6) might not hold for any value of β in our restricted range. In this case, we might find $H'(0) \leq 0$. Since $H(\beta)$ is strictly concave, we thus obtain a corner solution of no insurance, $\beta^* = 0$.

To illustrate this more clearly, let us examine the sign of $\frac{\partial Eu(\tilde{y})}{\partial \beta}$ evaluated at $\beta = 0$,

$$H'(0) = E[(\tilde{x} - P_0)u'(w_0 - \tilde{x})] = -\lambda E\tilde{x}Eu'(w_0 - \tilde{x}) + cov(\tilde{x}, u'(w_0 - \tilde{x})). \quad (3.8)$$

The covariance term is clearly positive, since u is concave due to risk aversion. The term $E\tilde{x}Eu'(w_0 - \tilde{x})$ is also positive. Since the loading factor λ only appears once, in a multiplicative fashion, in the right-hand side of (3.8), we see that $H'(0) > 0$ for $\lambda = 0$. Also, $H'(0)$ is linearly decreasing in λ

³This behavior results not just in expected-utility models with differentiable utility functions, but in any model exhibiting what is known as risk aversion of order 2, as defined by Segal and Spivak (1990).

and will be negative for all $\lambda > \lambda^*$, where $\lambda^* = \frac{\text{cov}(\tilde{x}, u'(w_0 - \tilde{x}))}{E\tilde{x}Eu'(w_0 - \tilde{x})}$. Thus, we see that whenever insurance is too expensive, in particular whenever $\lambda \geq \lambda^*$, no coverage will be optimal.

If we did not make the restriction that insurance be nonnegative, $\beta \geq 0$, then we would have $\beta^* < 0$ whenever $\lambda > \lambda^*$. Such contracts typically are not available in the insurance market, but are interesting to consider nonetheless. When the premium loading λ is excessively high, $\lambda > \lambda^*$, the policyowner would rather take a bet that her loss will occur than purchase insurance. She is willing to take on added risk (rather than insuring) in order to increase her expected final wealth. In a sense, the individual has what is known in the finance jargon as a "short position" in her insurance policy.⁴ Of course, if we do impose the restriction that $\beta \geq 0$, then the optimal coverage level will be $\beta = 0$, which follows from the concavity of $H(\beta)$ and the condition that $H'(0) < 0$.

In some cases, one can link transaction costs to undiversifiable risks. Obviously, many natural, environmental or technological risks are in the class of large risks that are difficult to eliminate by using the mutuality principle. We may question the insurer's risk-neutrality for these risks. Insurance companies will not provide fair insurance premiums for them. Indeed, shareholders will not be able to diversify the risk associated with the dividends paid by the insurance companies that cover these large risks. Shareholders will ask for a risk premium, which will increase the cost of funds for these companies. This cost will be passed on to policyholders through a larger premium rate for the component of individual risk that is systematic. This larger premium in turn will provide an incentive for the policyholders to retain a larger part of their individual risks. In short, the fact that the risk is systematic induces insurance premiums to contain a positive loading that has an effect equivalent to a transaction cost. In many instances, this effect can be so large as to preclude a market for insurance, or at least hinder it. For example, after the events of September 11, 2001, many businesses had insurance policies cancelled or else found premium increases too high to afford insurance. Also, insurance for many natural disasters requires a high level of coinsurance to

⁴In the market for life insurance, fairly new products known as viaticals and life settlements are essentially short positions in insurance policies. These contracts, which typically are designed for the terminally ill, pay the policyholder a lump sum of money now (akin to receiving an insurance premium), in return for a promise by the policyholder to pay a fixed sum of money upon his or her death. Of course, this payment is financed via the death-benefit proceeds of an existing life insurance policy.

obtain any coverage at all.

3.3 Comparative statics in the coinsurance problem

We examine the effect of a change in various parameters of the problem on the optimal coinsurance rate β^* . One natural question is about the effect of an increase in risk aversion on insurance demand. Intuition suggests that if agent 1 is more risk-averse than agent 2, agent 1 should have a larger insurance demand than agent 2. This intuition is correct, as stated in the following Proposition.

Proposition 13 *Consider two utility functions u_1 and u_2 that are increasing and concave, and suppose that u_1 is more risk averse than u_2 in the sense of Arrow-Pratt. Then, the optimal coinsurance rate β^* is higher for u_1 than for u_2 : $\beta_1^* \geq \beta_2^*$.*

Proof: If $\lambda = 0$, then $\beta_1^* = \beta_2^* = 1$ by Proposition 1, and hence we are done. Suppose that $\lambda > 0$ and let β_1^* be optimal for u_1 . Define $y_0 = w - P_0$ and note that $y_0 = w - (1 - \beta_1^*)x - \beta_1^*P_0$ evaluated at $x = P_0$. Without loss of generality, suppose that $u_1'(y_0) = u_2'(y_0)$. Because u_1 is more concave than u_2 in the sense of Arrow-Pratt, it must be the case that $u_1'(y_0) \geq u_2'(y_0)$ for all y smaller than y_0 (i.e. for all $x > P_0$), and $u_1'(y_0) \leq u_2'(y_0)$ for all y larger than y_0 (i.e. for all $x < P_0$). This implies that

$$(x - P_0)u_2'(w_0 - (1 - \beta_1^*)x - \beta_1^*P_0) \leq (x - P_0)u_1'(w_0 - (1 - \beta_1^*)x - \beta_1^*P_0) \text{ for all } x.$$

It follows that

$$\begin{aligned} H_2'(\beta_1^*) &= E[(\tilde{x} - P_0)u_2'(w_0 - (1 - \beta_1^*)\tilde{x} - \beta_1^*P_0)] \\ &\leq E[(\tilde{x} - P_0)u_1'(w_0 - (1 - \beta_1^*)\tilde{x} - \beta_1^*P_0)] = 0. \end{aligned}$$

The last equality follows from the first-order condition for the optimality of β_1^* . Hence, since H is concave, it follows that $\beta_2^* \leq \beta_1^*$. This concludes the proof. ■

If an increase in risk aversion raises the demand for insurance, it is natural that an increase in initial wealth w_0 reduces the demand for insurance if absolute risk aversion is a decreasing (DARA). This is shown in the following result.

Proposition 14 *An increase in initial wealth will*

1. *decreases the optimal rate of coinsurance β^* if u exhibits decreasing absolute risk aversion.*
2. *increase the optimal rate of coinsurance β^* if u exhibits increasing absolute risk aversion.*
3. *cause no change in the optimal rate of coinsurance β^* if u exhibits constant absolute risk aversion.*

Proof: We show here the proof for the first case, decreasing absolute risk aversion. The proofs for the other two cases are similar. Let β^* be optimal for $w = w_0$. Now consider

$$\frac{\partial H'(\beta)}{\partial w} = E[(\tilde{x} - P_0)u''(\tilde{y})]. \quad (3.9)$$

Since H is strictly concave in β , we only need to show that (3.9) is negative when evaluated at β^* . This will imply that the coinsurance level β^* is too high as wealth increases from w_0 . To see why this is indeed the case, recall that DARA implies that $-u'$ has the properties of a risk-averse utility function and is more risk-averse than u . Now, note that

$$\frac{\partial E[-u'(\tilde{y})]}{\partial \beta} = -E[(\tilde{x} - P_0)u''(\tilde{y})],$$

which equals the right-hand side of (3.9) in absolute value, but is opposite in sign. By the previous Proposition, we know that the level of coinsurance for $-u'$ is higher than the level for u . This implies that $-E[(\tilde{x} - P_0)u''(\tilde{y})] > 0$ when evaluated at $\beta = \beta^*$. In turn, it follows that (3.9) is negative at $\beta = \beta^*$. This concludes the proof. ■

We need to state a couple of caveats here. First, we need to point out that utility need not satisfy any of the conditions (i) – (iii) above. That is, preferences might exhibit decreasing risk aversion at some wealth levels and increasing risk aversion at others. Also, since decreasing absolute risk aversion is a fairly common assumption, it is often said that insurance is an "inferior good," since demand decreases with wealth. However, it is important here to stress the ceteris paribus assumption that the increase in initial

wealth does not modify the size of the risk borne by the policyholder. In the real world, wealthier consumers typically purchase bigger houses and more valuable cars. This tends to increase their optimal insurance budget in spite of the fact that their demand for insurance per unit of risk is decreasing. This can explain why the insurance sector has largely benefitted from the economic growth since the industrial revolution. This is not contradictory to the above Proposition.

Having examined the effect of changes in wealth on the demand for insurance, we can now easily look at how an increase in the price of insurance coverage affects demand. In particular, the following proposition examines the effects of an increase in the loading factor.

Proposition 15 *An increase in the premium loading factor $\lambda \geq 0$ will cause the optimal rate of coinsurance β^* to*

1. *increase if u exhibits constant or increasing absolute risk aversion.*
2. *possibly increase and possibly decrease if u exhibits decreasing absolute risk aversion*

Proof: Differentiate $H'(\beta)$ w.r.t λ to obtain

$$\frac{\partial H'(\beta)}{\partial \lambda} = -E\tilde{x}Eu'(\tilde{y}) - \beta E\tilde{x}E[(\tilde{x} - P_0)u''(\tilde{y})]. \quad (3.10)$$

The first term on the right-hand side of (3.10) is a negative substitution effect. When insurance becomes more expensive, we substitute away from insurance and towards other goods. However, a higher loading factor also induces a wealth effect, which is captured by the second term. Indeed, note that we may use the previous Proposition to re-write this term as

$$-\beta E\tilde{x}E[(\tilde{x} - P_0)u''(\tilde{y})] = -\beta E\tilde{x}\frac{\partial H'(\beta)}{\partial w}.$$

Clearly (3.10) will be negative whenever the second term is negative or zero, which occurs when u exhibits increasing and constant absolute risk aversion respectively. However, if u exhibits decreasing absolute risk aversion, then the wealth effect is positive: because the increase in λ makes the consumer "poorer," the individual behaves in a more risk aversion fashion and demands more insurance. Thus, the total effect on insurance demand depends on the relative magnitudes of the income and substitution effects. ■

Note that, under the fairly common assumption of DARA, insurance might well be a Giffen good, that is a good whose demand increases when price rises.⁵

Since an increase in risk aversion leads to an increased demand for insurance, it might seem natural to assume that an increase in the riskiness of the loss also should cause policyholders to increase their demand for insurance. However, this statement is not true in general. We know from Rothschild and Stiglitz that an increase in the riskiness of the loss \tilde{x} would cause expected utility to fall, so that the policyowner will be worse off. But this does not mean that more insurance will necessarily be purchased. Notice first that a mean-preserving risk increase of the loss \tilde{x} has no effect on the insurance premium βP_0 , since it is by assumption based on the expected loss $E\tilde{x}$.

From Rothschild and Stiglitz (1970), we know that a mean-preserving increase in the riskiness of \tilde{x} will always lead to a decrease in $E\phi(\tilde{x})$ if and only if $\phi(x)$ is a concave function. Define

$$\phi(x) \equiv (x - P_0)u'(w_0 - \beta P_0 - (1 - \beta)x).$$

From (3.6) it follows that an increase in the riskiness of \tilde{x} will always lead to a decrease in β^* if and only if $\phi(x)$ as defined above is concave. However, it is straightforward to show that this need not be the case. The fact that an increase in risk does not always lead to a higher demand for insurance was first discussed by Rothschild and Stiglitz (1971). The surprising result is that an increase in the risk of loss makes all risk-averse policyholders worse off, but some of them may well reduce their demand for insurance as a reaction to this change. Several authors have restricted the set of acceptable increases in risk or the set of acceptable utility function in order to get an unambiguous effect.⁶

3.4 The optimality of deductible insurance

There are various ways for policyholders to retain a share of the risk. One of the most common is to accept a straight deductible, in which the indemnity is either zero if the loss is less than a prespecified deductible level, or the loss minus the deductible level otherwise. Alternatively, the insurance contract

⁵Conditions that are both necessary and sufficient for insurance not to be a Giffen good are given by Briys, Dionne and Eeckhoudt (1989).

⁶A review of much of this literature can be found in Eeckhoudt and Gollier (2000).

can contain a coinsurance rule in which case the indemnity is a prespecified percentage of the loss, as we examined in the previous section. Other contractual forms of the indemnity also can be considered, such as upper limits on indemnities and so-called disappearing deductibles. For example, in the area of liability insurance, losses might not be bounded in size. Thus, we find upper-limit policies – policies in which the insurer pays full insurance but with a cap on the maximum indemnity – to be quite common in this line of insurance. Moreover, there is no reason to assume that the above mentioned types of contractual forms are mutually exclusive. For instance, in the area of health insurance, we might find contracts that contain a deductible (or “co-pay”) for each claim, a level of coinsurance above the deductible, and an upper limit that puts a cap on the aggregate indemnity.

Under a reasonable set of conditions, the optimal insurance contract always takes the form of a straight deductible. Any non-deductible insurance contract is dominated by a straight deductible contract with the same actuarial value. Deductibles provide the best compromise between the willingness to cover the risk and the limitation of the insurance deadweight cost. This result is due to Arrow (1971).

To understand more exactly why deductible policies are preferred, consider a model where the risk of loss \tilde{x} may take a finite number of possible values $\{x_1, \dots, x_n\}$. The uncertainty is represented by a vector of probabilities (p_1, \dots, p_n) where $p_i = \text{Prob}[\tilde{x} = x_i] > 0$, and $\sum_i p_i = 1$. Without losing generality, assume that $x_1 < x_2 < \dots < x_n$. A contract is characterized by a premium P and indemnity schedule $I(\cdot)$. This means that for each loss x_i , the insurance contract stipulates the indemnity $I(x_i)$ to be paid by the insurer in such a circumstance. As before, we assume that the insurance premium is to be paid ex-ante by the policyholder and is proportional to the actuarial value of the policy: $P = (1 + \lambda)EI(\tilde{x})$. The final wealth y of the policyholder, after purchasing policy (P, I) is

$$y(x) = w_0 - P - x + I(x), \quad (3.11)$$

if loss x occurs.

Finally, one generally assumes that insurance markets are constrained to provide policies with nondecreasing and nonnegative indemnity schedules, $I(x) \geq 0$ for all x . In other words, ex-post contributions from the policyholder are prohibited. There is a technical justification for imposing this

constraint. Indeed, the condition $\lambda > 0$ is not realistic when the indemnity is negative. In this case, the ex-post contribution of the policyholder would *reduce* transaction costs!

Proposition 16 *Suppose a risk-averse policyholder selects an insurance contract $(P, I(\cdot))$ with $P = (1 + \lambda)EI(\tilde{x})$ and with $I(x)$ nondecreasing and $I(x) \geq 0$ for all x . Then the optimal contract contains a straight deductible D ; that is $I(x) = \max(0, x - D)$.*

Proof: Suppose we have a deductible policy $(P, I(\cdot))$, so that $I(x_i) = 0$ for $x_i \leq D$ and $I(x_i) = x_i - D$ for $x_i > D$. We will show that any other indemnity schedule must cause wealth to be riskier in the sense of Rothschild and Stiglitz (1970), and hence must be less preferred by the policyholder. Consider an alternative insurance contract $(P, \hat{I}(\cdot))$ with the same premium P . By our pricing assumption, we know that $E\hat{I}(\tilde{x}) = P/(1 + \lambda)$, which is constant for a fixed P . Thus, if we increase the indemnity for one loss level, we must decrease it for others in order to preserve the mean indemnity.

Consider an increase in the indemnity for some loss level x_j by some amount $\varepsilon_j > 0$, so that $\hat{I}(x_j) = I(x_j) + \varepsilon_j$. First note that, since the indemnity must be nonnegative, $\hat{I}(x_i)$ cannot be reduced for any loss $x_i \leq D$. We thus must decrease the indemnity for one or more loss levels $x_i > D$ by amounts ε_i where $p_j\varepsilon_j = \sum p_i\varepsilon_i$. This leads to the following changes in final wealth.

At x_j : $\hat{y}(x_j) = y(x_j) + \varepsilon_j$, with $y(x_j) = w_0 - P - z_j$ and $z_j = \min(x_j, D)$.

At each x_i : $\hat{y}(x_i) = y(x_i) - \varepsilon_i$ with $y(x_i) = w_0 - P - D \leq y(x_j)$.

The new indemnity schedule \hat{I} yields a reduction in wealth in states i where wealth is small, and it yields an increase in wealth where it is large. Thus, it amounts to a mean-preserving increase in risk. It follows that any indemnity schedule $\hat{I}(\cdot)$ will be dominated by the deductible schedule $I(\cdot)$ with the same premium. Since we can make this argument for any level of the insurance premium P , this concludes the proof. ■

The idea of the proof is intuitive. One way to modify a deductible policy without changing the insurance budget of the policyholder is to reduce the indemnity when the loss exceeds the deductible in order to increase it when

the loss is smaller than the deductible. This change reduces final wealth at low wealth levels, and it raises it at larger wealth levels. This yields a mean-preserving spread in final wealth. Notice that this result is proven on the basis that policyholders dislike mean-preserving spreads of final wealth. Therefore, this Proposition does not rely on expected utility *per se*. Any model in which risk-averse agents dislike mean-preserving spreads will lead to the same conclusion, that deductibles are optimal, even if we do not invoke the expected-utility hypothesis as a basis for decision making.

To illustrate this result and the way it was proven, let us consider the case of a loss \tilde{x} which takes on the values 0, 50 and 100, each with equal probabilities. Assuming $\lambda = 0.2$, a contract with a pure coinsurance rate of 50%, i.e. with $I(x) = x/2$, can be purchased for a premium $P = 30$. Alternatively a contract with a straight deductible $D = 37.5$ also would have a premium equal to $P = 30$. Final wealth under the deductible policy is given by the lottery $\tilde{y}_D \equiv (w_0 - 30, 1/3; w_0 - 67.5, 2/3)$. Now define $\tilde{\theta} = (+12.5, 1/2; -12.5, 1/2)$. Note that $E\tilde{\theta} = 0$. Final wealth under the coinsurance policy can be written as

$$\tilde{y}_C \equiv (w_0 - 30, 1/3; w_0 - 67.5 + \tilde{\theta}, 2/3) = (w_0 - 30, 1/3; w_0 - 55, 1/3; w_0 - 80).$$

The final wealth distribution is illustrated in Figure 3.3. Observe how the wealth distribution under coinsurance in Figure 4.3 shows how \tilde{y}_C can be obtained by adding the noise term $\tilde{\theta}$ to the lowest realization of random wealth for the deductible policy, $y_D \equiv w_0 - 67.5$. Since this noise has a zero mean, the wealth distribution under coinsurance in Figure 3.2 is more risky in the sense of Rothschild and Stiglitz. We conclude that a contract with a 50% coinsurance rate will never be purchased, as it is dominated by a contract with a straight deductible for the same premium.

[INSERT FIG 3.2 ABOUT HERE]

The intuition of the Proposition also follows directly from its proof. A straight deductible insurance policy efficiently concentrates the effort of indemnification on only the largest losses. Any other insurance contract will compensate more for lower loss levels, while necessarily reducing the indemnity for some of the larger losses. The optimality of a straight deductible underscores the relevance of insurance for large risks. Small risks, i.e. risks whose largest potential loss is less than the optimal deductible should not be insured. The same principle applies to buying insurance for differing types of losses when they are covered under different insurance policies. A parent

is very willing to purchase life insurance against the important risk of a premature death. A homeowner is willing to purchase insurance for his or her house, which might be their most valuable physical asset. The owner of a new car is likely desirous of insurance to protect against damages, whereas the owner of an old car with much wear and tear might decide that insurance is simply not worth the cost.

3.5 Bibliographical references and extensions

The coinsurance problem was first examined by Mossin (1968). There have been many papers about the effect of a change in the distribution of a risk on its optimal exposure. Rothschild and Stiglitz (1971) were the first to observe that an increase in risk does not necessarily reduce the optimal demand for it by risk-averse agents. Gollier (1995) derived a necessary and sufficient condition on the change in the distribution of a risk to guarantee that all risk-averse consumers reduce their exposure to this risk. Milgrom (1981), Landsberger and Meilijson (1990) and Ormiston and Schlee (1993) showed that all shifts in distribution that satisfy the Monotone Likelihood Ratio (MLR) order have this property. Eeckhoudt and Gollier (1995) and Athey (1997) extended this result to the Monotone Probability Ratio (MPR) order. A change in distribution satisfies the MLR (resp. MPR) order if the ratio of the densities (resp. cumulative probabilities) is monotone with the realisation of the random variable. Meyer and Ormiston (1985) found that all strong increases in risk (SIR) also generate this result. A SIR is a mean-preserving spread in which all of the probability mass that is moved is transferred *outside* the initial support of the distribution.

The important result on the optimality of deductibles by Arrow (1971) has been followed by many others. Raviv (1979), Huberman, Mayers and Smith. (1983) and Spaeter and Roger (1997) for example provided alternative proofs to the result and extended it to include more general insurance pricing. Zilcha and Chew (1990), Karni (1992), Machina (1995) and Gollier and Schlesinger (1996) provided various proofs that do not use any specific decision criterion. They just rely on the assumption that policyholders dislike mean-preserving spreads in the distribution of final wealth. Gollier (1987) relaxed the constraint that the indemnity must be nonnegative in all states of nature.

In this chapter, we assumed that the distribution of the loss is exogenous

and common knowledge, and that losses are observable by the two parties. In many instances, the risk is affected by some preventive actions of the policyholder. If the insurer cannot observe these actions, it faces the moral hazard problem. When the distribution of the loss is not common knowledge, the insurer faces the adverse selection problem. When the loss incurred by the policyholder is observable by the insurer only at a cost, it faces the problem of insurance fraud (ex-post moral hazard) and audit. The economics of moral hazard, adverse selection and audit are now central elements of the economic theory, but they have initially been examined in the 70s in the insurance paradigm. A primer to this literature can be found in Salanié (1997). An introduction to it will be provided in the last part of this book.

References

- Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing*. Chicago: Markham Publishing Co.
- Athey, S., (1997), Comparative Statics under Uncertainty: Single Crossing Properties and Log-Supermodularity, Discussion paper, MIT.
- Briys, E., G. Dionne and L. Eeckhoudt (1989), More on Insurance as a Giffen Good, *Journal of Risk and Uncertainty*, 2, 415-420.
- Eeckhoudt, L., and C. Gollier, (1995), Demand for risky assets and the monotone probability ratio order, *Journal of Risk and Uncertainty*, 11, 113-122.
- Eeckhoudt, L., and C. Gollier, (2000), The Effects of Changes in Risk on Risk Taking: A Survey, in Dionne, G. (ed), *Contributions to Insurance Economics*, Boston: Kluwer Academic Press.
- Gollier, C. (1987). "The Design of Optimal Insurance without the Nonnegativity Constraint on Claims", *Journal of Risk and Insurance* 54, 312-324.
- Gollier, C., (1995), The Comparative Statics of Changes in Risk Revisited, *Journal of Economic Theory* , 66, 522-536.
- Gollier, C. and H. Schlesinger, (1995), Second-Best Insurance Contract Design in an Incomplete Market, *Scandinavian Journal of Economics*, 97, 123-135.
- Gollier, C. and H. Schlesinger, (1996), Arrow's Theorem on the Optimality of Deductibles: A Stochastic Dominance Approach, *Economic Theory*, 7, 359-363.
- Huberman, G., D. Mayers and C.W. Smith. (1983). Optimal Insurance Policy Indemnity Schedules, *The Bell Journal of Economics* 14, 415-426.
- Karni, E., (1992), Optimal Insurance: A Nonexpected Utility Analysis, in Dionne, G. (ed), *Contributions to Insurance Economics*, Boston: Kluwer Academic Press.

- Landsberger, M, and I. Meilijson, (1990), Demand for risky financial assets: A portfolio analysis, *Journal of Economic Theory*, 50, 204-13.
- Machina, M., (1995), Non-Expected Utility and the Robustness of the Classical Insurance Paradigm, in Gollier C. and M. Machina (ed.), *Non-expected Utility and Risk Management*, Boston: Kluwer Academic Publishers, reprinted from *The Geneva Papers on Risk and Insurance Theory*, 20, 9-50.
- Meyer, J. and M. Ormiston, (1985), Strong increases in risk and their comparative statics, *International Economic Review*, 26, 425-437.
- Milgrom, P., (1981), Good news and bad news: Representation theorems and applications, *Bell Journal of Economics*, 12, 380-91.
- Mossin, J. (1968). Aspects of Rational Insurance Purchasing, *Journal of Political Economy* 76, 533-568.
- Outreville, J. F. (1998), *Theory and Practice of Insurance*, Boston: Kluwer Academic Publishers.
- Ormiston, M. and E. Schlee, (1993), Comparative Statics Under Uncertainty for a Class of Economic Agents, *Journal of Economic Theory*, 61, 412-422.
- Raviv, A. (1979). "The Design of an Optimal Insurance Policy", *American Economic Review* 69, 84-96.
- Rothschild, M. and J. Stiglitz, (1971), Increasing risk: II Its economic consequences, *Journal of Economic Theory* 3, 66-84.
- Salanié, B., (1997), *The Economics of contracts: A primer*, MIT Press, Boston.
- Segal, U. and A. Spivak, (1990), First order versus second order risk aversion, *Journal of Economic Theory*, 51, 111-125.
- Spaeter, S. and P. Roger, (1997), The Design of Optimal Insurance Contracts: A Topological Approach, *The Geneva Papers on Risk and Insurance Theory*, 22, 5-20.
- Zilcha, I., and S.H. Chew, (1990), Invariance of the Efficient Sets When the Expected Utility Hypothesis is Relaxed, *Journal of*

Economic Behaviour and Organizations, 13, 125-131.

Chapter 4

Static portfolio choices

Financial markets are central to the functioning of our decentralized economies. Participants in these markets are risk-averse agents who are willing to take risk only if they receive appropriate rewards for this. Owning risky assets is compensated by higher expected returns on one's portfolio. Risk-averse households must determine their best trade-off between risk and expected return. A simple version of this problem is examined in the first section of this chapter. Organizing the economy in order to induce risk-averse people to accept to purchase risky assets is a vital condition for growth. Indeed, industrial investments, which must eventually be borne by the population, are risky. Without these investments, there would be no growth. Financial markets can be viewed as an institution that transfers entrepreneurial risks to consumers.

The decision problem of investors is in fact much more complex than just determining the best compromise between risk and performance. Investors face a myriad of possible investment opportunities. Selecting the composition of their portfolio requires comparing risks that are potentially correlated. Diversification is a key word in this environment, as we will see in the second section.

In this chapter, we focus on investors who consume their entire wealth at the end of the current period. By doing so, we isolate the investment problem from another of its essential characteristics in real life, namely time. Portfolio management has an intrinsic dynamic nature that will be examined in another chapter.

4.1 The one-risky-one-riskfree-asset model

4.1.1 Description of the model

Consider an agent who has a sure wealth w_0 that he can invest in one risk-free asset and in one risky asset. For ease of exposition, we can refer to the risk-free asset as a government bond, whereas the risky asset is a stock, or a portfolio of stocks. The risk-free return of the bond over the period is r . The return of the stock over the period is a random variable \tilde{x} . The problem of the agent is to determine the optimal composition $(w_0 - \alpha, \alpha)$ of his portfolio, where $w_0 - \alpha$ is invested in bonds and α is invested in stocks. The value of the portfolio at the end of the period may be written as

$$(w_0 - \alpha)(1 + r) + \alpha(1 + \tilde{x}) = w_0(1 + r) + \alpha(\tilde{x} - r) = w + \alpha\tilde{y}, \quad (4.1)$$

where $w = w_0(1 + r)$ is future wealth obtained with the risk-free strategy and $\tilde{y} = \tilde{x} - r$ is the so-called "excess return" on the risky asset. We assume in this chapter that the agent consumes all his wealth $w + \alpha\tilde{y}$ at the end of the period. The utility function u , which is assumed to be differentiable, increasing and concave, links the level of consumption at the end of the period to the utility attained by the consumer. We do not consider here the existence of any short-sale constraints, i.e., we allow α to be larger than w or less than zero. The problem of the investor is thus to choose α in order to maximize expected utility :

$$\alpha^* \in \arg \max_{\alpha} Eu(w + \alpha\tilde{y}). \quad (4.2)$$

This problem is formally equivalent to the program (3.3) describing the coinsurance problem of the previous chapter. To see this, define $w \equiv w' - P_0$, $\alpha \equiv (1 - \beta)P_0$ and $\tilde{y} \equiv (P_0 - \tilde{x})/P_0$ in (4.2), where P_0 is the premium for full coverage, β the coinsurance level and \tilde{x} here denotes the loss. Consequently,

$$Eu(w + \alpha\tilde{y}) = Eu[(w' - P_0) + (1 - \beta)P_0((P_0 - \tilde{x})/P_0)] = Eu(w' - \beta P_0 - (1 - \beta)\tilde{x})$$

Thus, we can interpret $\alpha = 0$ as starting at full insurance coverage, which is equivalent to having a 100 percent risk-free portfolio, i.e., to having all of our wealth invested in bonds in the portfolio problem. By increasing α (i.e.,

decreasing the coinsurance level β) the consumer accepts some of the risk in exchange for a higher expected final wealth. Here $\tilde{y} \equiv (P_0 - \tilde{x})/P_0$ can be interpreted as the return on coinsuring. In other words, retaining some share of a risk of loss is similar to purchasing a risky asset. In both cases, the problem is to determine the optimal, exposure to an exogenous risk. In both cases, risk-averse agents are willing to accept a positive exposure to the risk because of the positive expected net payoff of doing so.

By way of this formal link between the portfolio problem and the insurance problem, we directly obtain the following results from the previous chapter.

Proposition 17 *Consider problem (4.2) where \tilde{y} is the excess return of the risky asset over the risk-free rate, and α^* is the optimal dollar investment in the risky asset. The optimal investment in the risky asset is positive if and only if the expected excess return is positive: $\alpha^* = 0$ if $E\tilde{y} = 0$ and $\alpha^* E\tilde{y} > 0$ otherwise. Moreover, when the expected excess return is positive,*

- a) α^* is reduced when the risk aversion of the investor is increased in the sense of Arrow-Pratt;
- b) α^* is increasing in wealth if absolute risk aversion is decreasing.

Because risk aversion is second order in the expected utility model, the demand for the risky asset is positive as soon as the expected excess return, also known as the "equity premium", is positive. Thus, this model does not on its own explain why a large proportion of the population does not hold any stock. This "participation puzzle" has got various explanations. For example, investing in real-world markets involves some degree of knowledge about how these markets work. Consumers with a low optimal α^* might consider the cost of obtaining such knowledge to be too high.

The two comparative static properties of this standard portfolio problem are very intuitive. More-risk-averse people hold less-risky portfolios, and wealthier people have a larger demand for stocks, under decreasing absolute risk aversion. All existing empirical studies on households' portfolios obtain this positive relationship between stock holdings and wealth, thereby offering an additional argument in favor of DARA.

A special case that plays an important role in the theory of finance is when the utility function exhibits constant relative risk aversion. Let us assume accordingly that $u(c) = \frac{1}{1-\gamma} c^{1-\gamma}$ for all c , where γ is the degree of

relative risk aversion. Under this specification, the first-order condition to program (4.2) can be written as

$$E[\tilde{y}u'(w + \alpha^*\tilde{y})] = E[\tilde{y}(w + \alpha^*\tilde{y})^{-\gamma}] = 0. \quad (4.3)$$

Obviously, the solution to this equation is such that $\alpha^* = kw$, where k is a positive constant such that $E\tilde{y}(1 + k\tilde{y})^{-\gamma} = 0$. We conclude that under constant relative risk aversion, the optimal dollar amount invested in the risky asset is proportional to wealth. Or, in other words, it is optimal for CRRA investors to invest a fixed share of their wealth in stocks.

Proposition 18 *Under constant relative risk aversion, the demand for stocks is proportional to wealth: $\alpha^*(w) = kw$.*

For more general utility functions, determining the optimal demand may be more difficult. It can be useful to derive an approximate solution to this problem. Using a first-order Taylor approximation to $u'(w + \alpha^*y)$ around w , we can approximate the first-order condition $E\tilde{y}u'(w + \alpha^*\tilde{y}) = 0$ as

$$E\tilde{y}[u'(w) + \alpha^*\tilde{y}u''(w)] \simeq 0.$$

We thus obtain an approximation for the proportion of wealth invested in stocks:

$$\frac{\alpha^*}{w} \simeq \frac{\mu_{\tilde{y}}}{\sigma_{\tilde{y}}^2} \frac{1}{R(w)}, \quad (4.4)$$

where $R(w) = -wu''(w)/u'(w)$ is the degree of relative risk aversion evaluated at w , and $\mu_{\tilde{y}}$ and $\sigma_{\tilde{y}}^2$ are respectively the mean and the variance of the excess stocks return. Approximation (4.4) best fits the exact solution when $\sigma_{\tilde{y}}^2$ is small with respect to $\mu_{\tilde{y}}$. But it can be proven that it is exact when absolute risk aversion is constant and returns are normally distributed. To sum up, the optimal share of wealth invested in stocks is roughly proportional to the equity premium $\mu_{\tilde{y}}$, and inversely proportional to the variance of stock returns and to relative risk aversion.

4.1.2 The equity premium and the demand for stocks

We can use approximation (4.4) to get an idea on how much of the investor's wealth should be invested in stocks. Historical data on asset returns are

available several sources. Shiller (1989) and Kocherlakota (1996) provide statistics on asset returns for the U.S. over the period from 1889 to 1978. The average real return to Standard and Poor 500, a representative portfolio of U.S. stocks, has been 7% per year over this period, whereas the average short-term real risk free rate has been $r = 1\%$. The observed equity premium has thus been equal to $\mu_{\tilde{y}} = 6\%$ over the century. The standard deviation of the excess return was approximately equal to $\sigma_{\tilde{y}} = 16\%$.

Using reasonable degrees of relative risk aversion, we obtain unrealistically high shares of total wealth invested in stocks. For example, if R equals 2, approximation (4.4) yields a share equaling 117%. This means that this investor should borrow 17% of his wealth at the risk free rate to invest this loan together with his entire wealth in the stock market! An investor with an unrealistically high relative risk aversion of 10 should still invest 23% of his wealth in stocks. This surprising result is related to the so-called "equity premium puzzle" that will be discussed in more details later.

4.2 The effect of background risk

One way to explain the surprising large demand for stocks of the theoretical model is to recognize that there are other sources of risk on final wealth than the riskiness of assets returns. Consider for example labor income. For obvious reasons, wages are usually not fully insurable. To capture the effects of these types of risks, we can introduce a zero-mean background risk $\tilde{\varepsilon}$ to initial wealth w . This yields the following modified portfolio decision problem:

$$\alpha^{**} \in \arg \max_{\alpha} Eu(w + \tilde{\varepsilon} + \alpha\tilde{y}). \quad (4.5)$$

We want to compare α^{**} to α^* , the demand for the risky asset when there is no background risk. For the sake of simplicity, we assume that the risk on labor income is independent of the portfolio risk. Obviously any correlations between the risks is important, but we aim to show that, even in the case of statistical independence, there is often a predictable effect on decision making. Intuition might suggest that α^{**} should be smaller than α^* : independent risks should be substitutes. Since a risk averter is afraid of any bad luck with respect to the outcome of $\tilde{\varepsilon}$, he or she might try to compensate for the extra risk by behaving in a more cautious manor towards the level of

endogenous risk \tilde{y} . Because $\tilde{\varepsilon}$ and \tilde{y} are independent, the above problem can be rewritten as

$$\alpha^{**} \in \arg \max_{\alpha} Ev(w + \alpha\tilde{y}), \quad (4.6)$$

where the value function v is defined by $v(z) = Eu(z + \tilde{\varepsilon})$ for all z . This trick is very useful because we know the condition under which α^{**} (defined by (4.6)) is smaller than α^* (defined by (4.2)). Indeed, by Proposition 17, we just have to check whether v is more concave than u . In other words, the question becomes simply whether a zero-mean risk makes people more averse towards other independent risks. This is true if

$$-\frac{v''(z)}{v'(z)} = -\frac{Eu''(z + \tilde{\varepsilon})}{Eu'(z + \tilde{\varepsilon})} \geq -\frac{u''(z)}{u'(z)}, \quad (4.7)$$

for all $\tilde{\varepsilon}$ such that $E\tilde{\varepsilon} = 0$. This is equivalent to requiring that $Eh(z, \tilde{\varepsilon}) \leq 0$, where $h(z, \epsilon) = u''(z + \epsilon)u'(z) - u''(z)u'(z + \epsilon)$. This inequality holds if and only if h is concave in ϵ for all z . A necessary condition is that $h_{22}(z, 0)$ be negative, or, in the case where $u''' > 0$, that

$$-\frac{u''(z)}{u'(z)} \leq -\frac{u''''(z)}{u'''(z)} \quad (4.8)$$

for all z . This shows that this a priori simple and intuitive idea that independent risks must be substitutes requires a strong necessary condition on the fourth derivative of the utility function.

It can easily be shown that this condition is necessary but not sufficient to guarantee that any background risk makes investors more averse to other independent risks. We hereafter prove the following Proposition which provides a simple sufficient condition.

Proposition 19 *Consider the following three statements:*

1. *Any zero-mean background risk reduces the demand for other independent risks;*
2. *For all z , $-u''''(z)/u'''(z) \geq -u''(z)/u'(z)$;*
3. *Absolute risk aversion is decreasing and convex.*

Condition 2 is necessary for condition 1, under the assumption that u''' is positive. Condition 3 is sufficient for conditions 1 and 2.

Proof: It just remains to be proven that condition 3 is sufficient for condition 1. If $A(\cdot)$ denotes absolute risk aversion, we can write that

$$-Eu''(z + \tilde{\varepsilon}) = E[A(z + \tilde{\varepsilon})u'(z + \tilde{\varepsilon})].$$

Under decreasing absolute risk aversion, we have that the right-hand side of this equality is larger than $EA(z + \tilde{\varepsilon})Eu'(z + \tilde{\varepsilon})$. Moreover, because A is convex, $EA(z + \tilde{\varepsilon})$ is larger than $A(z)$. Combining these three observations implies condition (4.7), which is necessary and sufficient for property 1 to hold. ■

Notice that once we accept that absolute risk aversion is decreasing, it is natural to accept that it also is convex. In particular, it could not be concave everywhere, since a function cannot be positive, decreasing and everywhere concave. But this argument does not exclude the case where A would be locally concave. Observe also that concave power utility functions all have an absolute risk aversion that is decreasing and convex, which implies that independent risks are substitutes for investors having these preferences. But it is easy to find utility functions for which this is not the case.

4.3 Portfolios of risky assets

4.3.1 Diversification in the expected utility model

Suppose now that risk-averse investors can invest their wealth in two assets that are risky. To keep the argument simple, let us assume that these two assets have the same distribution of returns \tilde{x}_1 and \tilde{x}_2 that are independent and identically distributed.¹ What should be the optimal structure of their portfolios? To answer this question, we must solve the following program:

$$\max_{\alpha} Eu(\alpha\tilde{x}_1 + (w - \alpha)\tilde{x}_2), \quad (4.9)$$

where α is the amount invested in the first risky asset. Under risk aversion, the objective function is concave in the decision variable. The first-order condition is

$$E(\tilde{x}_1 - \tilde{x}_2)u'(\alpha^*\tilde{x}_1 + (w - \alpha^*)\tilde{x}_2) = 0.$$

¹The argument is easily extended to dependent but symmetric random variables.

It is obvious that the unique root to this equation is $\alpha^* = w/2$, since

$$E\tilde{x}_1 u' \left(\frac{w}{2}(\tilde{x}_1 + \tilde{x}_2) \right) = E\tilde{x}_2 u' \left(\frac{w}{2}(\tilde{x}_1 + \tilde{x}_2) \right),$$

because \tilde{x}_1 and \tilde{x}_2 can be interchanged because they are i.i.d.. Thus it is optimal for all risk-averse investor to perfectly balance their portfolio in this case.

The mechanism behind this result is risk diversification. In fact, all other portfolios are second-order stochastically dominated by the balanced one. They are thus all rejected by risk-averse investors. This is easily seen by observing that $\alpha\tilde{x}_1 + (w - \alpha)\tilde{x}_2$ is distributed as

$$\frac{\tilde{x}_1 + \tilde{x}_2}{2} + \tilde{\varepsilon},$$

where

$$\tilde{\varepsilon} \equiv \left(\alpha - \frac{1}{2} \right) (\tilde{x}_1 - \tilde{x}_2).$$

Thus, the return of any portfolio α is distributed as the return of the balanced portfolio plus a pure noise $\tilde{\varepsilon}$ satisfying

$$E \left[\tilde{\varepsilon} \mid \frac{\tilde{x}_1 + \tilde{x}_2}{2} \right] = 0.$$

This implies that accepting an unbalanced portfolio is equivalent to accepting zero-mean lotteries. To illustrate, let us consider two i.i.d. returns \tilde{x}_1 and \tilde{x}_2 that can take value 0% or 20% with equal probabilities. A perfectly balanced portfolio would yield a return \tilde{y} of either 0% with probability 1/4, 10% with probability 1/2, or 20% with probability 1/4. One can check that the single asset portfolio has a distribution of returns that can be duplicated by taking the balanced portfolio plus a zero-mean lottery $(-10\%, 1/2; +10\%, 1/2)$ conditional to obtaining 10% on this portfolio.

In chapter 1, we introduced the notion of risk aversion by showing that Sempronius is willing to diversify his assets, i.e., to share his wealth in two independent ships rather than in just one, if he has a concave utility function. This section formalized the link between the preference for diversification and the more standard definition of risk aversion, which is to dislike zero-mean risks.

Most real-world investors do diversify their portfolios. However, the theory predicts that individual portfolios should be diversified internationally, which is often not the case. This is the so-called "international diversification puzzle," which stems in part from a penchant towards buying more local stocks, a phenomenon known as "home bias."

4.3.2 Diversification in the mean-variance model

One simple measure of the beneficial effect of diversification is the reduction in the variance of the portfolio return that it generates. In the case of two i.i.d. assets, the variance of the portfolio return is minimized by selecting the perfectly balanced portfolio, which has a variance equaling the asset variance divided by a factor 2. Moreover, the expected return of the portfolio is independent of its composition. The 50-50 portfolio is optimal in our case where the expected return is the same for the two assets. When the expected returns vary, the investor must trade off the risk against the expected performance of the portfolio. By adding more of an asset with a lower expected return, we have the obvious detrimental effect of lowering the expected return on the portfolio. However, this detrimental effect may be countered by the beneficial effects of diversification. Additional difficulties may arise when asset returns are correlated. In this section, we limit the analysis to a special case of expected utility, namely the mean-variance approach. Obviously, there are many limitations of this approach, as we discussed in Chapter 1. On the other hand, this approach works perfectly well in the case where investors have constant absolute risk aversion and asset returns are normally distributed. Let us also remind the reader that, in spite of several limitations, the mean-variance approach is still a cornerstone in the modern theory of finance.

Assume there are n risky assets, indexed by $i = 1, \dots, n$. The return of asset i is denoted \tilde{x}_i , whose expectation is μ_i . The covariance between returns of assets i and j is denoted $\sigma_{ij} = E(\tilde{x}_i - \mu_i)(\tilde{x}_j - \mu_j)$. We assume that the variance-covariance matrix Σ can be inverted. There also exists a risk free asset whose return is $r \equiv x_0$. If we normalize the initial wealth to unity, final wealth equals

$$\begin{aligned} \tilde{z} &= 1 + x_0 \left(1 - \sum_{i=1}^n a_i\right) + \sum_{i=1}^n a_i \tilde{x}_i \\ &= 1 + x_0 + \sum_{i=1}^n a_i (\tilde{x}_i - x_0), \end{aligned} \quad (4.10)$$

where a_i is the share of wealth invested in asset i . The evaluation of this risk

follows a mean-variance approach, where the investor maximizes the certainty equivalent of final wealth which is approximated by $E\tilde{z} - 0.5AVar[\tilde{z}]$. A is the index of absolute risk aversion of the investor. By definition (??), the mean and the variance of final wealth are respectively written as

$$E\tilde{z} = 1 + x_0 + \sum_{i=1}^n a_i (\mu_i - x_0),$$

$$Var[\tilde{z}] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij}.$$

Differentiating the certainty equivalent wealth with respect to the share a_i invested in asset i , and setting it equal to zero, yields

$$\mu_i - x_0 - A \sum_{j=1}^n a_j^* \sigma_{ij} = 0,$$

or, in matrix format,

$$\mu - x_0 = Aa^*\Sigma, \quad (4.11)$$

where $\mu - x_0$ is the vector of excess expected returns, and a^* is the vector of optimal shares invested in the risky assets. The solution of this system is

$$a^* = \frac{1}{A} \Sigma^{-1} (\mu - x_0). \quad (4.12)$$

The investment in the risk free asset corresponds to the remaining wealth

$$a_0^* = 1 - \sum_{j=1}^n a_j^*.$$

When returns are independently distributed, which implies that Σ and Σ^{-1} are diagonal matrices, this equation can be written as

$$a_i^* = \frac{1}{A} \frac{\mu_i - x_0}{\sigma_{ii}},$$

which extends equation (4.4) to the case of more than one risky asset. It should be noticed that, in the independent case, the demand for a risky

asset is independent of the opportunity to purchase other risky assets. Independence of returns implies independence of demands, which is one of the counter-intuitive properties of the mean-variance model.

This solution has an important characteristic that much simplifies the operational advice that the modern theory of finance provides for investors. Namely, all investors, whatever their attitude to risk, should purchase the same portfolio of risky assets. To see this define $\alpha^* = \Sigma^{-1}(\mu - x_0)$ as the optimal portfolio of risky assets of the investor with a degree of risk aversion A equaling unity. Then, solution (4.12) states that the investor with a degree of risk aversion $A = 0.5$ should purchase a portfolio of risky assets with twice the quantity of each risky assets contained in portfolio α^* . This is done by investing less in the risk free asset. Thus, the structures of the two portfolios of risky assets are exactly the same. In short, all agents should purchase the same fund of risky assets. The only role of risk aversion is to affect the best balance between this fund and the risk free asset. The common fund of risky assets is often referred to as a "mutual fund" and the fact the the choice can be narrowed down to investing in this mutual fund and the risk-free asset leads the result to be referred to as the "mutual fund theorem" or the "two fund separation theorem." Indeed, the result can also be proven in more general contexts than the mean-variance context we show here.

This result is extremely powerful. It suggests that portfolio management is a simple problem on which people should not spend too much time and energy. This, again, is counter factual. However, it relies on two assumptions. First, it is assumed that financial markets are informationally efficient. This means that all information about future economic performances are already included in asset prices. Or, in other words, that all investors share the same information about risks. The second assumption is that investors have mean-variance preferences. Later in this book, we will come back to these two aspects of the "two-fund theorem".

4.4 Bibliographical references and extensions

The two-asset model has first been examined by Arrow (1963, 1965, 1971). The effect of an increase in risk aversion on the demand for the risky asset is discussed by Pratt (1964). There has been extensive research on the effect of a change in the distribution of the return of the risky asset that culminated in the paper by Gollier (1995) who obtained the necessary and sufficient

condition. The effect of an independent background risk on the demand for risky assets has been examined by Kimball (1993), Gollier and Pratt (1996) and Eeckhoudt, Gollier and Schlesinger (1996). The benefit of diversification for risk-averse agents is best explained in Rothschild and Stiglitz (1971). The international diversification puzzle is stated in French and Poterba (1991) and Baxter and Jermann (1997). The standard reference for the equity premium puzzle is Mehra and Prescott (1985). It has been the source of a large literature in which the simple two-asset model has been extended in several directions. Kocherlakota (1996) provides a survey of these lines of research.

Markowitz (1952) is the first to solve the static multiple assets problem in the mean-variance framework.

References

- Arrow, K.J., (1963), Liquidity preference, Lecture VI in " *Lecture Notes for Economics 285, The Economics of Uncertainty*", 33-53, undated, Stanford University.
- Arrow, K.J., (1965), *Yrjo Jahnsson Lecture Notes*, Helsinki. Reprinted in Arrow (1971).
- Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing*. Chicago: Markham Publishing Co.
- Baxter, M., and U.J. Jermann, (1997) The international diversification puzzle is worse than you think, *American Economic Review*, 87, 170-80.
- Eeckhoudt, L., C. Gollier and H. Schlesinger, (1996), Changes in background risk and risk-taking behavior, *Econometrica*, 64, 683-690.
- French, K., and J. Poterba, (1991), International diversification and international equity markets, *American Economic Review*, 81, 222-26.
- Gollier, C., (1995), The Comparative Statics of Changes in Risk Revisited, *Journal of Economic Theory*, 66, 522-536.
- Gollier, C. and J.W. Pratt, (1996), Risk vulnerability and the tempering effect of background risk, *Econometrica*, 64, 1109-1124.
- Kimball, M.S., (1993), Standard risk aversion, *Econometrica*, 61, 589-611.
- Kocherlakota, N.R., (1996), The Equity Premium: It's Still a Puzzle, *Journal of Economic Literature*, 34, 42-71.
- Markowitz, H., (1952), Portfolio Selection, *Journal of Finance*, 7, 77-91.
- Mehra, R. and E. Prescott, (1985), The Equity Premium: A Puzzle, *Journal of Monetary Economics*, 10, 335-339.
- Pratt, J., (1964), Risk aversion in the small and in the large, *Econometrica*, 32, 122-136.

Rothschild, M. and J. Stiglitz, (1971), Increasing risk: II Its economic consequences, *Journal of Economic Theory* 3, 66-84.

Chapter 5

Static portfolio choices in an Arrow-Debreu economy

In the previous chapter, we examined an economy in which the investment opportunity set was quite limited. Funds could be invested in portfolios containing only bonds and stocks. Suppose that an investor would like to take a bet on the event that the return on the Dow Jones this year be larger than 10%. More precisely, suppose that he wants to sign a contract with a counterpart on financial markets that would give him 1 dollar only if the return on the Dow Jones this year is larger than 10%, against the payment of a lump sum fee *ex ante*. He could be also willing to bet on other events, as whether some specific asset will have a return in between two prespecified values, or as whether the average temperature in Chicago will be larger than its level attained last year. Most of these types of investment opportunities were not available in the portfolio choice model presented in the previous chapter.

Both the strength and the weakness of the model presented in the previous section come from the linear relationship between final wealth and the return of each individual assets. It is a strength because this assumption yields simple operational advice for practitioners. It is a weakness because it artificially constrains the choice of investors. In particular, this model does not allow investors to purchase a wide variety of new financial instruments that have been developed by financial intermediaries worldwide over the last three decades. For example, assets called "options" have been developed that offer to their owners payoffs that are highly nonlinear in the return of the underlying assets. Many individual investors also favor the purchase of

”portfolio insurance”, a system in which a minimum return is guaranteed to the portfolio owners. In the following, we will assume that investors are allowed to take any such risk exposure. We will assume that investors can bet on any possible event. By enlarging the risk opportunity set, we can in fact ”complete” financial markets. The portfolio choice problem in this economy has first been examined by Arrow (1953) and Debreu (1959). We hereafter present the core of an Arrow-Debreu economy.

5.1 Arrow-Debreu securities and arbitrage pricing

Let us assume that there are S possible states of nature at the end of the period that are indexed by $s = 0, \dots, S - 1$. For the sake of a simple notation, we assume that there is a finite number of possible states. The probability of state s is denoted p_s . When the risk is limited to the randomness of assets returns, a state of nature is characterized by the vector of these realized returns. The assumption that markets are complete means that for each possible state s , there is an asset which provides a unit payoff to its owner if and only if state s occurs. These assets are called ”Arrow-Debreu securities.” Purchasing an Arrow-Debreu security associated to state s is equivalent to bet on that state.

In fact, what is important is not that Arrow-Debreu securities actually exist, but that they can be replicated with existing assets in the economy. Consider a simple example with two states of nature ($S = 2$), and a risky asset whose initial price is normalized to unity. Its final value in state s is P_s , $s = 0, 1$. In other words, the net asset return is $P_0 - 1$ in state 0, and $P_1 - 1$ in state 1. If, in addition, we assume that there exists a risk-free asset with return r in both states, then this is an example of an economy in which financial markets are complete. To see this, let us assume that $P_0 < 1 + r < P_1$; in other words the risky asset sometimes pays more and sometimes pays less than the risk-free asset.¹ Now, we can replicate the Arrow-Debreu security associated with state $s = 1$ by purchasing α units of

¹This is an equilibrium condition. If we suppose alternatively that $1 + r < P_0$ and $1 + r < P_1$, then purchasing an infinite number of the risky asset would be optimal, independent of the degree of risk aversion.

the risky asset and by borrowing B at the risk-free rate, in such a way that

$$\begin{cases} 0 = \alpha P_0 - (1+r)B \\ 1 = \alpha P_1 - (1+r)B \end{cases}$$

The first constraint states that this portfolio provides to its owner no revenue in state 0, whereas the second constraint means that the revenue in state 1 is one. The solution of this system is $\alpha = (P_1 - P_0)^{-1}$ and $B = P_0/(1+r)(P_1 - P_0)$. We can think of one share of a mutual fund that invests in our portfolio of α units of the risky asset together with B borrowed at the risk-free rate. Since the payout of this mutual fund is exactly the same in each state as the Arrow-Debreu security associated with state 1, the principle of no arbitrage tells us that the total out-of-pocket expenditure for one share of this mutual fund must equal the price of the Arrow-Debreu security associated with state 1. Thus, investors can "bet" on state 1 occurring by purchasing this mutual fund.

A similar exercise can be performed for the Arrow-Debreu security associated with state 0. In this example, there are two states, which implies that there are enough of two independent assets to "span" the entire set of all possible risk exposures. More generally, markets will be complete if there are at least as many assets whose vectors of state-contingent payoffs are linearly independent as there are number of states. Any set of assets with this property is said to "span the market." In any such market, we lose no generality and gain much simplicity by assuming that the Arrow-Debreu securities exist. Thus, investors can structure any set of state-contingent claims by investing in the appropriate portfolio of Arrow-Debreu securities².

In this chapter, devoted to portfolio choice, the prices of contingent claims are taken as given. Let Π_s denote the price of the Arrow-Debreu security associated with state s . This is the price to be paid to obtain 1 monetary unit if and only if state s occurs. These prices can be inferred from the price of real assets from which contingent claims can be duplicated. To illustrate, $\Pi_1 = \alpha - B = (1 - P_0/(1+r))/(P_1 - P_0)$ is the price of the Arrow-Debreu security associated with state $s = 1$ in our example above. Indeed, duplicating this asset requires purchasing α units of the risky asset whose price is 1, from which one must subtract a loan of B . By a simple arbitrage

²Readers with a course in linear algebra will recognize that we only require a set of securities whose vector payoffs form a "basis" for the vector space of state-contingent payouts.

argument, Π_1 must be equal to the cost of building this portfolio, since it provides exactly the same state-contingent profile of revenue. By the way, this is the general idea followed in the important literature on arbitrage pricing, as initiated by Black and Scholes (1973). If we assume that $P_0 < P_1 - 1$, then the Arrow-Debreu security associated with state $s = 1$ in our example is a call option with strike price $P_1 - 1$.³ In this example, we showed how to derive the price of this call option from the characteristics of the risk of its underlying asset, as did Black and Scholes (1973) in a much more complex environment.

Of course the implication works both ways: once we know the prices for our Arrow-Debreu securities, we can calculate the price of any financial asset by simply adding up the prices of Arrow-Debreu securities with the identical state payoffs. One particularly useful example is a risk-free bond with a payoff of 1 in each state of nature. The price of such a bond must be the discounted value of 1 unit of wealth, received with certainty at the end of the period:

$$P_B = (1 + r)^{-1}.$$

Moreover, holding a portfolio consisting of one unit of each and every Arrow-Debreu security also would have a risk-free payout of 1. Thus, using no-arbitrage arguments, we must have

$$\sum_{s=0}^{S-1} \Pi_s = P_B = (1 + r)^{-1}. \quad (5.1)$$

Define $\hat{p}_s \equiv \Pi_s(1 + r)$. If we consider the vector $(\hat{p}_0, \dots, \hat{p}_{S-1})$ of S positive scalars, it follows from (5.1) that the elements of this set sum to one. As a result, we can view the elements of this set as probabilities in their own right. Indeed, \hat{p}_s is often referred to as the risk-neutral probability for state s . By construction, in this manner, we see that the price of any asset in our complete market is simply its expected value, discounted at the risk-free rate, but with the expectation taken with respect to the risk-neutral probabilities $(\hat{p}_0, \dots, \hat{p}_{S-1})$, rather than the true probabilities (p_0, \dots, p_{S-1}) .

³A call option on a specific underlying asset provides to its owner the right to purchase this asset at a prespecified price (the "strike price") over a given period. Note that the option defined here would provide a gross payout of $P_1 - (P_1 - 1) = 1$ in state 1 and zero in state 0, since the owner of the option would not use it in state 0.

An asset with state-contingent payoffs $(y_0, y_1, \dots, y_{S-1})$ can be duplicated by a portfolio containing y_s units of the Arrow-Debreu security associated to state s , $s = 1, \dots, S$. By the standard arbitrage argument, its price must thus equal

$$P = \sum_{s=0}^{S-1} \Pi_s y_s = \frac{\sum_{s=0}^{S-1} \hat{p}_s y_s}{1+r} = \frac{\hat{E} \tilde{y}}{1+r}, \quad (5.2)$$

where \hat{E} denotes the expectation operator with respect to probabilities $(\hat{p}_1, \dots, \hat{p}_S)$. In a complete market economy, there exists a "risk-neutral" probability distribution such that the price of any asset can be expressed as the discounted value of the risk-neutral expectation of its future payoffs. The search for this risk-neutral distribution is developed in another chapter.

5.2 Optimal portfolios of Arrow-Debreu securities

The problem of the investor is to determine the demand for each of the Arrow-Debreu securities. Let c_s be the investment in the Arrow-Debreu security associated with state s . By construction, it is also the final wealth of the agent in that state. The program of the investor can thus be written as

$$\max_{c_0, \dots, c_{S-1}} \sum_{s=0}^{S-1} p_s u(c_s) \quad \text{subject to} \quad \sum_{s=0}^{S-1} \Pi_s c_s = w, \quad (5.3)$$

where w is the investor's initial wealth. In the following sections, we examine the properties of the optimal portfolio of Arrow-Debreu securities. Before proceeding, it is noteworthy that this decision problem is in general much more flexible for the investor than the one that we considered in the previous chapter. In the one-risky-one-riskfree-asset model, the state of nature is characterized by the realized excess return $s = y$, and the final wealth is constrained to be linear in it: $c_s = w + \alpha s$. In model (5.3), there is no such constraint, and the optimal risk exposure $\{c_s\}_{s=0, \dots, S-1}$ is constrained only by the budget constraint $\sum \Pi_s c_s = w$. It should also be observed that program (5.3) is a special case of the decision problem faced by consumers under certainty, as described in standard microeconomics textbooks. In an economy with S different goods, where Π_s denotes the price of good s , each consumer

selects the bundle (c_0, \dots, c_{S-1}) that maximizes his utility $U(c_0, \dots, c_{S-1}) = \sum p_s u(c_s)$ under the standard budget constraint. The only specific attribute of the above program compared to the standard textbook problem is the additive nature of the objective function U .

Because the objective function in (5.3) is a sum of concave functions of the decision variables, the following first-order conditions are both necessary and sufficient for optimality:

$$u'(c_s^*) = \xi \frac{\Pi_s}{p_s} \quad \text{for all } s = 0, \dots, S-1, \quad (5.4)$$

where ξ is the Lagrange multiplier associated with (5.3), equal to the marginal utility of additional wealth. We see that the optimal consumption depends upon the state only through the ratio $\pi_s \equiv \Pi_s/p_s$ of the state price per unit of probability.⁴ In other words, if there are two states with the same price per unit of probability for the associated contingent claims, it is optimal for the agent to purchase the same quantity of these claims. It is easy to check that not doing so would yield an increase in risk of final wealth. The intuition of this result is simple. Suppose, for example, that all π_s are the same. Since $\pi_s = (1+r)^{-1}(\widehat{p}_s/p_s)$, and since the $\{\widehat{p}_s\}$ and the $\{p_s\}$ each sum to one, it follows that $\widehat{p}_s = p_s \quad \forall s$. Hence, contingent claims are actuarially priced in the sense that the price of any asset in such an economy equals the true expected value of its contingent payoff, discounted at the risk-free rate. It is then optimal to fully insure the risk, i.e., to purchase a risk-free portfolio: $c_s^* = w(1+r)$ for all s . Purchasing a risky portfolio would here be equivalent to not insuring risk when insurance contracts are actuarially priced.

When some of the π_s are different, then \widehat{p}_s and p_s are not identical for some s and there is a scope for an optimal risk exposure. We might expect that the investor will reduce the demand for contingent claims whose prices are large relative to the probability of the corresponding state. This means taking risk. As stated above, equation (5.4) implies that we have a different level of consumption for each value of π_s and thus may write $c_s^* \equiv C(\pi_s)$ for all s . Since u' is decreasing due to risk aversion, the function C is well-defined. It is characterized by $C(\pi) = u'^{-1}(\xi\pi)$ for all π . It is a non-increasing function: one consumes less in more expensive states. The absolute value of

⁴Note that we can rewrite (5.2) using true expectations, rather than risk-neutral expectations, as $P = \sum_{s=0}^{S-1} p_s \pi_s y_s = E\tilde{\pi}\tilde{y}$. The set $\{\pi_s\}$ is often referred to in the finance literature as the pricing kernel.

the slope of the C function is a local measure of the exposure to risk. At the limit, if C is a constant function, the agent does not take any risk. By fully differentiating the first-order condition $u'(C) = \xi\pi$ with respect to π , we obtain that

$$C'(\pi) = \frac{\xi}{u''(C)} = -\frac{T(C(\pi))}{\pi}, \quad (5.5)$$

where $T(C) = -u'(C)/u''(C)$ is the inverse of the measure of Arrow-Pratt risk aversion, often referred to as the local measure of absolute risk tolerance. Thus, more risk-tolerant people (i.e., less risk-averse people) will take more risk. In Figure 5.1, we compare the optimal portfolio C_1 of agent 1 to the optimal portfolio C_2 of agent 2 who is more risk-tolerant in the sense of Arrow-Pratt. Note that we cannot have either consumption curve totally above or below the other, since both must satisfy the same budget constraint in (5.3). Observe that equation (5.5) implies that at any crossing point, such as π_0 , the slope of C_2 is larger than the one of C_1 in absolute value. This further implies that we must get only a single crossing as illustrated in the figure.

[INSERT FIGURE 5.1 ABOUT HERE]

When the utility function exhibits constant relative risk aversion, one can solve the problem analytically. Suppose that $u'(C) = C^{-\gamma}$. Then the first-order condition yields $C(\pi) = \lambda\pi^{-1/\gamma}$, where λ equals $\xi^{-1/\gamma}$. The budget constraint, which can be rewritten as $E[\tilde{\pi}C(\tilde{\pi})] = w$, yields in turn

$$E[\tilde{\pi}C(\tilde{\pi})] = \lambda E\tilde{\pi}^{\frac{\gamma-1}{\gamma}}.$$

Thus, the optimal portfolio of an agent with constant relative risk aversion is such that

$$C(\pi) = \left[\frac{\pi^{-\frac{1}{\gamma}}}{E\tilde{\pi}^{\frac{\gamma-1}{\gamma}}} \right] w. \quad (5.6)$$

Observe that, as in the one-risky-one-riskfree-asset model, the demand for assets is proportional to wealth when relative risk aversion is constant.

5.3 A simple graphical illustration

We return to the simple case of two states of nature, $S = 2$. In order to focus on the risk aspects of the model, we assume that the risk-free rate is

zero, $r = 0$. The set of all possible contingent claims is represented by the positive orthant in Figure 5.2. The 45° line represents the locus of claims with equal consumption in both states of the world, which is referred to as the certainty line. The set of claims $\{(c_0, c_1)\}$ for which $p_0c_0 + p_1c_1 = w$ for some positive scalar w represent an iso-expected-value locus, i.e., the set of claims with mean wealth k . To analyze preferences, set expected utility equal to a constant k , $p_0u(c_0) + p_1u(c_1) = k$. The set of claims $\{(c_0, c_1)\}$ for which expected utility is constant defines an indifference curve in state-claims space: $c_1 = f(c_0; k)$

[INSERT FIGURE 5.2 ABOUT HERE]

Using implicit differentiation, we can find the slope of the indifference curve through any point in state-claims space,

$$\left. \frac{dc_1}{dc_0} \right|_{EU} = \frac{\partial f(c_0; k)}{\partial c_0} = -\frac{p_0 u'(c_0)}{p_1 u'(c_1)}.$$

The absolute value of dc_1/dc_0 is the marginal rate of substitution (*MRS*) between states. It is the rate of trade-off, at the margin, between one unit of consumption in state 0 and consumption in state 1, for which the consumer is just indifferent. If we evaluate the *MRS* at claims for which we have certainty, $c_0 = c_1$, it follows that the *MRS* $= p_0/p_1$ is independent of the preferences of the investor. This is another expression of the fact that individuals are neutral to the introduction of small risk, i.e., that there is second-order risk aversion under expected utility.

$$\begin{aligned} \frac{\partial^2 f(c_0; k)}{\partial c_0^2} &= -\frac{p_0}{p_1} \left[\frac{u''(c_0)}{u'(c_1)} - \frac{u'(c_0)u''(c_1)}{(u'(c_1))^2} \frac{\partial f(c_0; k)}{\partial c_0} \right] \\ &= \frac{p_0 u'(c_0)}{p_1 u'(c_1)} \left[A(c_0) + A(c_1) \frac{p_0 u'(c_0)}{p_1 u'(c_1)} \right], \end{aligned}$$

where $A(c) = -u''(c)/u'(c)$ is absolute risk aversion. It implies that the indifference curve is everywhere convex.

To operationalize the above information, consider the portfolio problem with state prices equal to the probabilities for each state, $\Pi_0 = p_0$ and $\Pi_1 = p_1$. Since $r = 0$, this implies that the line $p_0c_0 + p_1c_1 = w$ in Figure 5.2 represents the budget constraint. It follows from the above graphical information that $c_0^* = c_1^* = w$ is the optimal set of contingent claims on consumption. That is, the consumer should buy w units of each of the two Arrow-Debreu securities.

[INSERT FIGURE 5.3 ABOUT HERE]

Now let us suppose that $\Pi_0 > p_0$, which in turn implies here that $\Pi_1 < p_1$. The budget line thus becomes $\Pi_0 c_0 + \Pi_1 c_1 = w$. As illustrated in Figure 5.3, this new budget line will be steeper. Moreover, at $c_0 = c_1 = w$, we now have $MRS = p_0/p_1 < \Pi_0/\Pi_1$. Thus, transferring contingent wealth from state 0 to state 1 at market prices will increase the agent's expected utility. Indeed, the optimal level of contingent consumption will be one for which $c_0^* < w < c_1^*$, such as illustrated in Figure 5.3, where $MRS = \Pi_0/\Pi_1$. If we consider the contingent claim (w, w) as the individual's initial endowment, the optimal set of trades is to sell $w - c_0^*$ units of the Arrow-Debreu security associated with state 0, and to use the proceeds to purchase $c_1^* - w$ units of the Arrow-Debreu security associated with state 1. Since $\pi_1 \equiv \Pi_1/p_1 > 1 > \pi_0 \equiv \Pi_0/p_0$, we see that $C(\pi)$ is decreasing in π , as our theory predicted.

5.4 Bibliographical references and extensions

Arrow (1953) applied the general equilibrium theory developed by himself and Debreu (1959) to the case of financial markets under uncertainty. It is not the place here to present the developments that followed the discovery of the complete market framework. They span most of the modern theory of finance. The book by LeRoy and Werner (2001) provides a recent presentation of this theory.

References

- Arrow, K.J., (1953), Le rôle des valeurs boursières pour la répartition la meilleure des risques. *Econométrie*, Paris: CNRS [Translated as: Arrow, K.J., (1964), The role of securities in the optimal allocation of risk-bearing, *Review of Economic Studies*, 31, 91-96.]
- Debreu, G., (1959), *Theory of value*, Wiley, New York.
- LeRoy, S., and J. Werner, (2001), *Principles of financial economics*, Cambridge University Press, Cambridge, UK.

Chapter 6

Consumption and saving

Up to now, we have assumed that the decision maker lives for one period. This obscures the important intertemporal dimension of risk. In real life, agents can often postpone risk to the future. Indeed, this ability to postpone risky choices adds value of its own, usually referred to as a "real-option value." Decision makers also can choose to disseminate gains and losses from their current risk exposure over several periods, which is a type of time diversification of the effects of risk on their ultimate consumption. Alternatively, they might hope to recoup some of their current losses by taking on more risk in the future. In the extreme, this may lead to a type of "go-for-broke" strategy, such as a Las-Vegas gambler who bets all of his small amount of remaining wealth on one last gamble in the hopes of recouping his large losses. Finally, agents can alter their planned levels of consumption and saving in the expectation of dealing with uncertainty in the future, such as saving a bit more in earlier periods as type of insurance against future risks. In the next two chapters, we examine the relationship between risk and time. We first focus on the impact of risk on the optimal timing of consumption and saving.

6.1 Consumption and saving under certainty

As a benchmark case, we start with the characterization of optimal consumption under certainty. Assume that an agent lives for a known number of periods. We denote time by the n dates, $t = 0, \dots, n - 1$. The agent is endowed with a flow of sure incomes y_t , where y_t denotes the revenue that

is received with certainty at date t . We assume that there exists an efficient credit market with a constant risk-free interest rate r for both borrowing and lending. In each period, the agent decides how much to consume, which implicitly defines how much is saved or borrowed. If c_t denotes the consumption at date t , the dynamic budget constraint may be written as

$$z_{t+1} = (1 + r)[z_t + y_t - c_t], \quad t = 0, \dots, n - 1 \quad (6.1)$$

where z_t is the cash transferred from date $t - 1$ to date t . This can also be interpreted as the aggregate saving at date t . We assume that the initial cash z_0 is zero. Because lenders will not agree to provide credit to agents that will not be able to repay their debt later on, there is an ultimate constraint $z_n \geq 0$: the agent cannot die with a negative net debt position. Recall that we are assuming certainty with regards to both wealth and income, so that this constraint may be imposed on debt financing. Using the dynamic budget constraint (6.1) recursively, condition $z_n \geq 0$ may be rewritten as

$$\sum_{t=0}^{n-1} \frac{y_t - c_t}{(1 + r)^t} \geq 0. \quad (6.2)$$

This is the lifetime budget constraint of the agent. It states that the net present value of the flow of savings $y_t - c_t$ must be nonnegative; or equivalently that the present value of lifetime consumption cannot exceed the present value of lifetime income. More directly, we can write

$$\sum_{t=0}^{n-1} \Pi_t c_t \leq w_0, \quad (6.3)$$

where $\Pi_t \equiv (1 + r)^{-t}$ and $w_0 \equiv \sum \Pi_t y_t$ is lifetime wealth, i.e., the net present value of the flow of revenues. Observe that Π_t is the price of a zero-coupon bond with maturity at date t , i.e. an asset that generates a unit cash flow only at date t . For a given value of w_0 , the agent's objective is to choose an optimal consumption path (which in turn defines an optimal savings plan) across time.

We now describe the preferences of a young consumer over the set of all feasible lifetime consumption flows. Let $U(c_0, c_1, \dots, c_{n-1})$ denote his lifetime utility if he selects consumption plan $c = (c_0, c_1, \dots, c_{n-1})$. We assume that function U is increasing and concave. The optimal consumption is obtained

by solving the following program:

$$\max_c U(c_0, c_1, \dots, c_{n-1}) \text{ subject to budget constraint (6.3).} \quad (6.4)$$

This decision problem is described in Figure 6.1 in the case of two periods. the optimal consumption plan is characterized by point A , where the indifference curve is tangent to the budget line AB .

[INSERT FIGURE 6.1 ABOUT HERE]

Before discussing the solution to this problem in a more analytical way, suppose that the agent can choose a revenue profile (y_0, \dots, y_{n-1}) in a given opportunity set. This situation is typical of an investment problem in which an investor has several investment choices, each yielding a different set of cash-flows. The structure of problem (??) obviously implies that the optimal revenue profile is the one that maximizes its net present value (NPV) $w_0 = \sum \Pi_t y_t$, independent of the temporal consumption preferences of the decision maker. This result is known as Fisher's Separation Theorem. It sustains the NPV rule, which is one of the most important rules in economics: every investor should choose the investment which maximizes the net present value $\sum y_t(1+r)^t$ of its cash flow.

Problem (??) is not much different from the static decision problem of an agent consuming n different physical goods in the classical theory of demand. The general properties of these demand functions are well-known. We hereafter make an additional assumption on temporal preferences that enriches the model. Namely, we introduce an independence axiom stating that the preference order over the consumption pair (c_0, c_1) does not depend upon the consumption path over the remaining $n-2$ periods.¹ This precludes phenomena such as the formation of consumption habits. This independence assumption implies that the utility function U must be separable: $U(c) \equiv \sum u_t(c_t)$, where u_t is the (intra-period) utility of consumption at time t . To distinguish it from the intertemporal utility function U , we label u_t as the "felicity function" of consumption at date t . A common assumption is that felicity functions are proportional to each other: $u_t(\cdot) = p_t u(\cdot)$, for some increasing and concave function u and for some scalar $p_t > 0$. This set of assumptions has been adopted by most researchers over the last fifty years. Without loss of generality, we normalize p_0 to unity. Thus p_t can be interpreted as the discount factor for felicity $u(c_t)$ occurring at date t . If p_t is

¹More precisely, if $(a, b, c_2, \dots, c_{n-1})$ is preferred to $(d, e, c_2, \dots, c_{n-1})$, then $(a, b, x_2, \dots, x_{n-1})$ is preferred to $(d, e, x_2, \dots, x_{n-1})$ for all (x_2, \dots, x_{n-1}) .

less than unity, it can be interpreted as a proportional loss of utility due to postponing consumption, i.e. it indicates a preference for consuming sooner rather than later. It is important to dissociate p_t , a psychological parameter, from r , a financial variable. Parameter p_t serves as a discount factor on felicity, whereas the interest rate r is only useful in discounting monetary flows, as we have seen above.

Using these restrictions, we can rewrite consumption problem (??) as

$$\max_c \sum_{t=0}^{n-1} p_t u(c_t) \text{ subject to } \sum_{t=0}^{n-1} \Pi_t c_t = w_0. \quad (6.5)$$

The first-order conditions for this problem may be written as

$$p_t u'(c_t) = \xi \Pi_t, \text{ for } t = 0, \dots, n-1 \quad (6.6)$$

together with the budget constraint, where ξ is the Lagrange multiplier associated with problem (6.5). The Lagrange multiplier ξ is simply the marginal lifetime utility of an increase in the present value of wealth

This problem is formally equivalent to the static Arrow-Debreu portfolio problem (5.3) of the previous chapter. We simply replace "states of nature" with dates, probabilities with discount factors, and Arrow-Debreu securities with zero-coupon bonds. This equivalence is striking between figures 5.2 and 6.1. This equivalence has several important consequences for the remainder of this book. The most obvious ones are summarized below.

6.1.1 Aversion to consumption fluctuations over time

First, we can interpret the concavity of the felicity function u in the context of the consumption-saving problem under certainty as an aversion to consumption fluctuation from period to period. The fact that marginal utility is decreasing with respect to consumption provides an incentive for the decision maker to smooth consumption over time. To see this, consider the special case with $p_t = 1$ for all t , and with $r = 0$, which implies that $\Pi_t = 1$ for all t . It follows from the first-order conditions (6.6) that $u'(c_t) = \xi$ in each period, so that the optimal consumption path does not exhibit any fluctuation in consumption from period to period: $c_t = w_0/n$ for all t . This is a situation where the optimal consumption plan A is on the 45 degree line in Figure 6.1. If incomes fluctuate over the life cycle of the consumer, the optimal saving

strategy is to lend any extra income above w_0/n , or to borrow the extra money in case of the period income being smaller than w_0/n . The concavity of u implies that second order conditions are satisfied, so that lifetime utility is maximized via consuming an equal amount in each period.² We conclude that when there is no impatience ($p_t = 1$), and a zero interest rate ($r = 0$), it is optimal to smooth consumption over time if the felicity function is concave. Thus, the assumption $u'' < 0$ expresses an aversion to consumption fluctuation over time. This result is a complete analogy with the aversion to consumption fluctuations across states of nature that we make in the static Arrow-Debreu portfolio problem. In the latter model, this implies that full insurance is optimal when asset prices are actuarially fair, i.e., when state prices equal probabilities in every state.

One can measure the intensity of the desire to smooth consumption over time by considering a situation without any credit market, so that $c_t = y_t$. Suppose that the income y_0 at date 0 is strictly less than the income y_1 at date 1. Since the marginal utility of consumption is larger at date 0 than at date 1, $u'(y_0) \geq u'(y_1)$, we know that the agent would not be willing to exchange one unit of consumption today for one unit of consumption tomorrow. Accepting such a deal would increase the discrepancy between date-0 and date-1 consumption, and would reduce his lifetime utility. If asked to sacrifice one unit of consumption today, the agent will demand more than one unit of consumption tomorrow as compensation. One way to measure the intensity of this resistance to trade today's consumption for consumption tomorrow is to define an additional reward $k > 0$ that must be given to the agent at date 1 to compensate for the loss in consumption at date 0. Assuming changes in the consumption level that are sufficiently small, k is defined by the following condition:

$$u'(y_0) = (1 + k)u'(y_1).$$

The left-hand side of this equality is the marginal cost (in utils) of reducing consumption today, whereas the right-hand side is the marginal benefit (also in utils) of raising future consumption by a factor $1 + k$. In other words, k is defined so that the marginal utility loss by giving up 1 unit of consumption at date 0 must equal the marginal increase in utility by adding $1 + k$ units of consumption at date 1. If y_1 is close to y_0 , we can use a first-order Taylor expansion of $u'(y_0)$ around y_1 to obtain

²If u were convex, this solution would yield a minimum lifetime utility, and it is easy to show that maximal utility is achieved by consuming only in one period.

$$k \simeq \frac{y_1 - y_0}{y_1} \left[\frac{-y_1 u''(y_1)}{u'(y_1)} \right]. \quad (6.7)$$

The resistance to intertemporal substitution is approximately proportional to the growth rate of consumption. The multiplicative factor, $\gamma(y) \equiv -yu''(y)/u'(y)$, is hereafter called the measure of relative fluctuation aversion, or the relative degree of resistance to intertemporal substitution of consumption. Obviously, $\gamma(y)$ is an analogue to the Arrow-Pratt measure of relative risk aversion. Both measure the percentage decline in marginal utility relative to a small percent increase in wealth (consumption). In our current setting, this is a local measure of the consumer's aversion to moving consumption from a date with lower consumption to a date with slightly higher consumption. Note also that there is an equivalence between the measures "in the small" and "in the large." That is, $\gamma(y)$ is related to the approximation of k (6.7) in exactly the same way that the Arrow-Pratt approximation for the risk premium is related to the measure of risk aversion. There has been much attention given to empirical estimates of $\gamma(y)$, which is widely believed to be somewhere between 1 and 5 for most consumers.

6.1.2 Optimal consumption growth under certainty

In general, the real interest rate is not zero, and agents are impatient. Let us assume here that consumers use exponential discounting: $p_t = \beta^t$, for some scalar β less than unity. This yields a rate of pure preference for the present $\delta \equiv (1 - \beta)/\beta$ that is positive. In other words, $\beta = (1 + \delta)^{-1}$ and multiplying the felicity $u(c_t)$ by β^t is equivalent to discounting felicity at a constant rate δ per period. Using a constant rate to discount future utils is important for the time consistency of consumer decisions, as we will see in the last section of this chapter.

The presence of impatience and a positive return on savings presents two countervailing reasons not to smooth consumption completely over time. A higher level of impatience, i.e. a higher δ , induces agents to prefer consumption earlier in life. In other words, impatience tends to bias preferences in favor of consumption paths that decrease over time. On the other hand, a higher interest rate makes savings more attractive. It biases consumption choices in favor of consumption paths that are increasing over time. These

two contradictory effects must be combined with the aversion to consumption fluctuations to characterize the optimal consumption growth under certainty.

As an illustration, we solve this problem analytically in the special case where the felicity function exhibits a constant relative degree of aversion to consumption fluctuations. Suppose that $u(c) = c^{1-\gamma}/(1-\gamma)$, where γ is the constant degree of fluctuation aversion. Using the analogy with the Arrow-Debreu problem together with condition (5.6), or solving first-order conditions (6.6) immediately yields the solution

$$c_t = c_0 a^t, \quad (6.8)$$

where $a = ((1+r)/(1+\delta))^{1/\gamma}$, and c_0 is some initial consumption that is selected to satisfy the lifetime budget constraint. Thus, when the psychological discount is exponential and relative aversion is constant, it is optimal for consumers to let their consumption grow at a rate g , where

$$g \equiv \left(\frac{1+r}{1+\delta} \right)^{1/\gamma} - 1 \simeq \frac{r-\delta}{\gamma}. \quad (6.9)$$

The quality of the approximation is better when δ does not differ much from r . The optimal growth rate of consumption is positive when r is larger than δ . It is easy to check that this property holds independent of the specification of the felicity function. This is a case where the speculative motive for savings dominates the effect of impatience. We also see the intuitive effect of the aversion to fluctuations: an increase in γ reduces the optimal growth rate of consumption over time. But it is optimal to smooth consumption completely over the lifetime only when $r = \delta$, such as was the case in previous section where we had $r = \delta = 0$.

In the real world, consumption growth is subject to business cycles, which forces consumption to fluctuate over time. This has been the topic of much research over the last thirty years. This cycle around the secular trend has a negative impact on consumer welfare. Suppressing it, i.e. smoothing out business cycles, would be beneficial to consumers who dislike consumption fluctuations around the optimal growth rate g . However, Lucas (1987) showed that the importance of the effect of the business cycle on welfare has been largely overestimated by the profession. One can measure the cost of business cycle by the reduction in the growth rate of consumption that the representative agent would accept in exchange for the complete elimination of business cycles. Using data on consumption fluctuations in the U.S., Lucas showed that

business cycles "cost" a reduction of much less than one-tenth of one percent in the annual growth rate of the U.S. economy. This is totally insignificant! The reason is simple: exactly as risk aversion is a second-order effect in the additive EU model, the aversion to consumption fluctuations is a second-order effect in the time-additive lifecycle model. In other words, consumers have an extremely low aversion to small fluctuations in consumption. Lucas concluded that economists should concern themselves with the determinants of long-term growth rather than with the reduction of volatility.

6.2 Uncertainty and precautionary savings

Assuming that consumers have a sure income flow is clearly an unrealistic assumption. In this section, we introduce uncertainty into the picture. We consider a simple two-date model with a sure income y_0 in period 0, but an uncertain income \tilde{y}_1 in the second period. We assume that this risk is exogenous. For example, the consumer might plan for the future knowing that his future labor income is subject to changes that may be higher or lower than anticipated. Consumers select how much to save at date 0 in order to maximize their expected lifetime utility:

$$\max_s V(s) = u_0(y_0 - s) + Eu_1((1+r)s + \tilde{y}_1). \quad (6.10)$$

Observe that we don't need to assume at this stage that $u_1 = \beta u_0$, as we did in the previous section. Denote the optimal saving under uncertainty by s^* . The first-order condition for s^* is written as

$$u'_0(y_0 - s^*) = (1+r)Eu'_1((1+r)s^* + \tilde{y}_1). \quad (6.11)$$

It is important to observe that the willingness to save is determined by the expected marginal utility of future consumption.

The uncertainty affecting future incomes introduces a new motive for saving. The intuition is that it induces consumers to raise their wealth accumulation in order to forearm themselves to face future risk. This is the so-called precautionary motive for saving, and it relies on a prudent behavior. Its theoretical foundation can be derived by comparing s^* to the optimal saving \hat{s} when the uncertain future income \tilde{y}_1 is replaced by its expectation:

$$\max_s \widehat{V}(s) = u_0(y_0 - s) + u_1((1+r)s + E\tilde{y}_1).$$

Let \hat{s} denote the solution to this maximization program. We want to determine whether the optimal saving under uncertainty is larger than when the uncertainty is removed: $s^* > \hat{s}$. Because \hat{V} is concave in s , which is easily verified, this is the case if and only if $\hat{V}'(s^*)$ is negative. This condition means that reducing saving marginally from s^* raises lifetime utility under certainty. In other words, there will be a precautionary demand for savings if and only if

$$\begin{aligned}\hat{V}'(s^*) &= -u'_0(y_0 - s^*) + (1+r)u'_1((1+r)s^* + E\tilde{y}_1) \\ &= (1+r)[u'_1((1+r)s^* + E\tilde{y}_1) - Eu'_1((1+r)s^* + \tilde{y}_1)] \leq 0,\end{aligned}$$

where the second equality is obtained by using condition (6.11). Therefore, the level of precautionary saving is positive if and only if

$$Eu'_1((1+r)s^* + \tilde{y}_1) \geq u'_1((1+r)s^* + E\tilde{y}_1). \quad (6.12)$$

By Jensen's inequality, this follows whenever u'_1 is convex, or equivalently whenever u''_1 is positive. This condition is referred to as "prudence", a concept that has already been introduced in the first chapter. Thus, prudence is necessary if we require that precautionary saving be positive for all possible distributions of the future risk. A consumer who has a concave marginal utility function, to the contrary, would reduce savings because of the future risk. This individual would exhibit what is called "imprudent behavior." Thus, prudence corresponds to the positivity of the third derivative of the utility function, exactly as risk aversion relies on the negativity of its second derivative. An agent can exhibit a risk-averse and imprudent behavior, for example by insuring risk at an unfair premium and by reducing its saving in the face of a non-insurable future risk. Or, following the definitions, a prudent person can be a risk-lover.

There is a link however between (decreasing) risk aversion and prudence. Recall from Chapter 1 that

$$A'(w) = A(w)[A(w) - P(w)]$$

where $A(w)$ is the Arrow-Pratt measure of absolute risk aversion and $P(w) = -u'''(w)/u''(w)$ is the measure of absolute prudence, which under risk aversion is positive only if $u''' > 0$. Thus, absolute risk aversion is decreasing if and only if $P(w) > A(w)$ for all w . Because we took decreasing absolute risk aversion (DARA) as a natural assumption, so should we take prudence.

One also can measure the intensity of the precautionary saving motive. This can be done by answering the following question: what would be the sure reduction in future income that would have the same effect on savings as the introduction of the future risk? Let ψ be this "precautionary premium". It is defined implicitly by the following equality:

$$Eu'_1(w + \tilde{y}_1) = u'_1(w + E\tilde{y}_1 - \psi), \quad (6.13)$$

where w is the accumulated wealth before the second date. This condition states that the willingness to save, which is measured by the expected marginal utility of future consumption, is not affected by the replacement of the risk by its expectation diminished by ψ . The precautionary premium is seen to be positive whenever the agent is prudent, i.e. whenever $u''_1 > 0$. It is useful to observe at this stage that the precautionary premium is equivalent to the risk premium defined before, but where the utility function u_1 would be replaced by the marginal utility function $-u'_1$. The precautionary premium and the risk premium are the sure reductions in wealth that have the same effects as adding the risks to the expected marginal utility and to the expected utility respectively. This implies that all results that we obtained previously for risk aversion and the risk premium can be transferred to prudence and the precautionary premium, by simply replacing u_1 with $-u'_1$.

For example, one can use the Arrow-Pratt approximation for the risk premium to obtain an equivalent one for the precautionary premium:

$$\psi \simeq \frac{1}{2}P(w + E\tilde{y}_1)\sigma_{\tilde{y}_1}^2, \quad (6.14)$$

where P is called the degree of absolute prudence. Recall from Chapter 1 that P is equivalently the index of absolute risk aversion for the utility function $v(w) = -u'_1(w)$, where v is risk averse whenever u_1 is prudent. This equivalence also follows since ψ is simply Pratt's risk premium for utility v . Similarly, the precautionary premium ψ is decreasing in wealth if and only if absolute prudence is decreasing in wealth.

To see how the precautionary premium affects savings more directly, consider the simple case where the risk-free rate of savings equals the discount rate for time preference, and set both equal to zero, i.e. $r = \delta = 0$. In particular, lifetime utility is assumed to be $U(c_0, c_1) = u(c_0) + u(c_1)$. We also assume that $E\tilde{y}_1 = y_0$, so that the individual has the same expected income

at dates 0 and 1. Suppose first that \tilde{y}_1 is non-risky with $\tilde{y}_1 \equiv y_0$. In this setting, the first-order condition (6.11) implies that $u'(y_0 - s) = u'(y_0 + s)$. As already observed, it follows that the optimal savings is zero, $s^* = 0$, since u is strictly concave. In other words, the consumer simply consumes his current income in each period, $c_0^* = c_1^* = y_0$.

Now suppose that \tilde{y}_1 is risky so that the first-order condition is $u'(y_0 - s) = Eu'(E\tilde{y}_1 + s) = u'(y_0 + s - \psi)$. The second equality above derives from our definition of the precautionary premium. Solving for the optimal savings, we obtain $s^* = \psi/2$. Thus, if the consumer is prudent, there will be a precautionary demand for savings, $s^* > 0$. Moreover, an individual who is more prudent will have a higher value of the precautionary premium ψ , in the same way that an individual who is more risk averse has a higher risk premium. Consequently, a more prudent consumer will save more than his less prudent counterpart. It is also interesting to note that if the felicity function is quadratic, an assumption that is not uncommon in the finance literature, we have $\psi = 0$ and, hence, there is no precautionary savings motive.

6.3 Risky savings and precautionary demand

In the previous section we considered only a labor-income risk. The individual had a risk-free savings alternative but was unsure about how much income would be earned at date 1. We now look at a model in which labor income is known, but the rate of return on savings is risky. We abstract from the portfolio problem of Chapter 4 and assume that there exists only one fund for risky savings, paying a return of $1 + \tilde{r}$, where $E\tilde{r} \equiv r_0 > 0$. We consider a consumer with an investment horizon of two periods. Since lifetime income is known with certainty, we assume without loss of generality that all income is paid at date $t = 0$. Letting w_0 denote this wealth, the consumer's objective is

$$\max_s V(s) \equiv u(w_0 - s) + \beta Eu((1 + \tilde{r})s). \quad (6.15)$$

The first-order condition for this program is

$$u'(w_0 - s) = \beta E[(1 + \tilde{r})u'((1 + \tilde{r})s)]. \quad (6.16)$$

The second-order condition is easily shown to hold under risk aversion. In fact, the objective function $V(s)$ is easily seen to be concave in s .

Consider first the case with a risk-free savings rate of r_0 , as in the previous section. The first-order condition in this case becomes $u'(w_0 - s) = \beta(1 + r_0)u'((1 + r_0)s)$. In order to focus on the effects of risk, consider once again a simple case where the expected rate of return on savings equals the discount rate for time preference, i.e. $r = \delta$, so that $\beta = (1 + r_0)^{-1}$. Hence s^* satisfies $w_0 - s^* = (1 + r_0)s^*$. As expected, the optimal savings s^* is such that there is no fluctuation in consumption between dates, $c_0^* = c_1^*$. We next turn to the question of whether adding risk to the return on savings leads to a higher level of savings. It turns out in this setting that prudence alone is not sufficient to lead to an increase in the level of savings. In fact, there are two competing influences at work. On the one hand, the riskiness of returns makes savings less attractive than a risk-free rate with the same average return. But on the other hand, the date-1 risk will induce a precautionary motive to the prudent consumer. It turns out that we need a sufficiently high level of prudence in order to have the precautionary motive dominate, as we show next.

Since $V(s)$ is concave, it follows from (6.16) that uncertainty in the rate of return will cause the optimal level of savings to rise whenever

$$E[(1 + \tilde{r})u'((1 + \tilde{r})s)] > (1 + r_0)u'((1 + r_0)s).$$

This inequality will hold if the function $h(R) \equiv Ru'(Rs)$ is convex in R . Straightforward calculations show that $h''(R) = 2su''(Rs) + s^2Ru'''(Rs)$. Assume that $u'' < 0$ and that savings are not zero, since c_1 would also be zero in that case. It follows that $h'' > 0$ if the following inequality holds

$$\frac{-zu'''(z)}{u''(z)} > 2, \quad (6.17)$$

with $z = Rs$. The left-hand side of (6.17) is simply a measure of relative prudence, i.e. the absolute prudence measure $P(z)$ multiplied by the wealth level z . Thus, from (6.17) we obtain the following comparative static property of an increase in risk of the return on saving:

$$\text{Optimal savings } s^* \text{ will } \left\{ \begin{array}{l} \text{increase if relative prudence exceeds 2} \\ \text{remain the same if relative prudence equals 2} \\ \text{decrease if relative prudence is less than 2.} \end{array} \right\}$$

Of course, relative prudence need not satisfy any of the above conditions. However one case in which it does is the case where the felicity function is of

the CRRA type, namely $u(c) = c^{1-\gamma}/(1-\gamma)$, where γ is the constant degree of risk aversion.³ In this case, straightforward calculations show that relative prudence is equal to $\gamma+1$. Hence, in the case of CRRA preferences, we have⁴

Optimal savings s^* will $\left\{ \begin{array}{l} \text{increase if relative risk aversion exceeds 1} \\ \text{remain the same if relative risk aversion equals 1} \\ \text{decrease if relative risk aversion is less than 1.} \end{array} \right\}$

6.4 Time consistency

When the model contains only two consumption dates as in the previous two sections, every future action can be planned in advance at date 0, with no possibility to change one's mind. At the second date, the agent just consumes what he has on his saving account. When there are more than two dates, what has been planned at $t = 0$ can be revised at $t = 1$. If you decided at date $t = 0$ to purchase an expensive good that you planned to pay next period ($t = 1$), you may still decide at $t = 1$ to postpone the repayment in order to maintain your high consumption level. Thus, consumers may have a time consistency problem. To see this, let us reexamine the consumption-saving problem under certainty that is described by (6.5) with $\Pi_t = (1+r)^{-t}$ and $n \geq 3$. At $t = 0$, the consumer plans the consumption profile (c_0, \dots, c_{n-1}) for his remaining lifetime that maximizes his lifetime utility $\sum_{t=0}^{n-1} p_t u(c_t)$ subject to his lifetime budget constraint $\sum_{t=0}^{n-1} \Pi_t c_t = w_0$. Remember that p_t is the factor that is used to discount the felicity occurring t dates from the current date. Using the first-order condition (6.6) for $t = 1$ and 2, a condition for the consumption plan to be optimal as seen from $t = 0$ is written as

$$\text{planned choice: } u'(c_2) = \frac{p_1}{(1+r)p_2} u'(c_1). \quad (6.18)$$

This consumption rule must be satisfied in order to spend efficiently the money saved from date $t = 0$. Anticipating how he will spend this money saved, the agent determines his optimal initial consumption c_0 . Solving the

³We refer to "risk aversion" here rather than "fluctuation aversion" since our focus is on risk, and not on simply the timing of consumption in a risk-free setting. It is important to note that a problem with the approach we are using is its inability to fully distinguish between these two phenomena. This has been the focus of much research.

⁴The case for $\gamma = 1$ is found directly by assuming $u(c) = \ln c$.

system of equation (6.6) together with the budget constraint, the entire consumption profile $(c_0, c_1, c_2, \dots, c_{n-1})$ is selected.

Now, let us consider the situation that emerges at date $t = 1$. Wealth has been depleted by the initial consumption c_0 , but it has also been augmented by return r on savings. At date $t = 0$, the agent planned to consume c_1 at $t = 1$. However, he is ready to reconsider this choice. His welfare for his remaining lifetime can be written as

$$\sum_{t=1}^{n-1} p_{t-1} u(c_t).$$

The indexes of the p parameters and the c variables are important here. Observe in particular that the felicity $u(c_2)$ occurring one period from the current date $t = 1$ is discounted at p_1 , the discount factor for a one period horizon. Maximizing this objective function with the budget constraint $\sum_{t=1}^{n-1} \Pi_{t-1} c_t = (w_0 - c_0)(1 + r)$ yields the following first-order condition:

$$\text{actual choice: } u'(c_2) = \frac{p_0}{(1+r)p_1} u'(c_1). \quad (6.19)$$

Equations (6.18) and (6.19) are equivalent only if $p_1/p_2 = p_0/p_1$. This is equivalent to require that $p_t = a\beta^t$ for $t = 0, 1$ and 2 , or that discounting be exponential.⁵ Extending this condition for all t implies that the optimal consumption choice c_1 as actually selected at $t = 1$ is not different from the one that has been planned at $t = 0$. There is no time-consistency problem under exponential discounting.

The problem is more complex when the consumer does not use specification $p_t = a\beta^t$ for the discount factors. Suppose for example that p_2 is larger than p_1^2/p_0 . From (6.18) and (6.19), we derive that the consumption level c_1 that is actually selected at $t = 1$ is larger than the one that was planned at $t = 0$. There is a time-consistency problem in that case. When determining his initial consumption, the agent cannot trust himself to limit his consumption in the future. This is typical of an addictive behavior: a smoker finds it beneficial to smoke today conditional on his commitment to stop smoking tomorrow. But when tomorrow arrives, the smoker finds it again beneficial to smoke, thereby postponing his decision to stop to the next day, and so

⁵This terminology comes from the fact that the continuous-time equivalent of this discount function is $p(t) = e^{-\delta t}$.

on. We may suspect that such addictive behavior also arises for other goods, yielding a global consumption addiction problem. For people facing this problem, long-term saving plans with no possibility to withdraw may be welfare improving in spite of the lack of flexibility of these saving plans. The time consistency problem may explain why a large fraction of the population in developed countries accepts to finance short-term consumption with credit card loans at rates as large as 20%, and still saves money for the long run at 5%.

6.5 Bibliographical references and extensions

The understanding of consumption behavior is probably one of the most important challenges in modern macroeconomics. There has been a lot of developments in this area of research since the seminal papers of Modigliani and Brumberg (1954) and Friedman (1957). These developments refer to the theory of real business cycles, which will not be covered here. Estimations of the relative degree of resistance to consumption fluctuations can be found in many different papers. Hall (1988) found an estimation around 10, whereas Epstein and Zin (1991) found a value ranging from 1.25 to 5. An experiment on this has been performed by Barsky, Juster, Kimball and Shapiro (1997). The first formal analysis of precautionary savings is due to Leland (1968), Sandmo (1970) and Drèze and Modigliani (1972). Kimball (1990) coined the term of prudence, and examined the properties of the precautionary premium.

There is an important literature on the effect of liquidity constraints on optimal saving rates. If consumers cannot borrow money when there is a negative temporary shock on their incomes, they will be more willing to accumulate wealth *ex-ante*. This "buffer stock" leads to a new motive to save (Deaton (1991) and Carroll (1997)).

Strotz (1956) was the first to discuss the time consistency problem of consumers using a discount factor that does not decrease exponentially with time horizon. Pollack (1968) solved this time consistency problem using a game-theoretic approach where the different players are the different selves of the consumer living at the different periods. Laibson (1997) reexamined this question to explain various facts on credit markets. There is now a wide and lively literature on "hyperbolic discounting".

References

- Barsky, R.B., F.T. Juster, M.S. Kimball and M. Shapiro, (1997), Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study, *Quarterly Journal of Economics*, 537-79.
- Carroll, C.D., (1997), Buffer-stock saving and the life cycle/permanent income hypothesis, *Quarterly Journal of Economics*, 112, 1-55.
- Deaton, , A., (1991), Saving and liquidity constraints, *Econometrica*, 59, 1221-48.
- Drèze, J.H. and F. Modigliani, (1972), Consumption decisions under uncertainty, *Journal of Economic Theory*, 5, 308-335.
- Epstein, L.G., and S. Zin, (1991), Substitution, Risk aversion and the temporal behavior of consumption and asset returns: An empirical framework, *Journal of Political Economy*, 99, 263-286.
- Freidman, M., (1957), *A theory of the consumption function*, Princeton: Princeton University Press.
- Hall, R.E., (1988), Intertemporal substitution of consumption, *Journal of Political Economy*, 96, 221-273.
- Kimball, M.S., (1990), Precautionary savings in the small and in the large, *Econometrica*, 58, 53-73.
- Laibson, D., (1997), Golden eggs and hyperbolic discounting, *Quarterly Journal of Economics*, 62.
- Leland, H.E., (1968), Dissertation, Stanford University.
- Modigliani, A. and R. Brumberg, (1954), Utility analysis and the consumption function: An interpretation of the cross-section data, in (ed. K. Kurihara) *Post-Keynesian Economics*, New Brunswick, NJ: Rutgers University Press.
- Pollack, R.A., (1968), Consistent Planning, *Review of Economic Studies*, 35.

Sandmo, A., (1970), The effect of uncertainty on saving decisions,
Review of Economic Studies, 37, 353-360.

Strotz R., (1956), 'Myopia and Inconsistency in Dynamic Utility
Maximization', *Review of Economic Studies*, 23.

Chapter 7

Dynamic portfolio management

Investors most often view their financial investment over a large period of time. In many instances, earlier investment decisions are not irreversible. This implies that investment management has an obvious dynamic nature. An important question is therefore whether the investment advices that can be deduced from a static model, as those developed in chapters 4 and 5, can be used to determine the optimal dynamic portfolio strategy. In other words, the problem is to determine how do future investment opportunities affect the short-term investment choice. Similarly, one can be interested in determining the effect of one's investment horizon on the riskiness of his portfolio? Popular treatments suggest that short horizons often lead to excessively conservative strategies. Thus, the decisions of corporate managers, graded on their quarterly earnings, are said to focus too much on safe, short-term strategies, with underinvestment say in risky R & D projects. Privately-held firms, it is widely believed, secure substantial benefit from their ability to focus on longer-term projects. Mutual fund managers, who get graded regularly, are also alleged to focus on strategies that will assure a satisfactory short-term return, with long-term expectations sacrificed. In the formal literature, the horizon-riskiness issue has received the greatest attention addressing portfolios appropriate to age.

Samuelson (1989) and several others have asked: "As you grow older and your investment horizon shortens, should you cut down your exposure to lucrative but risky equities?" Conventional wisdom answers affirmatively, stating that long-horizon investors can tolerate more risk because they have more time to recoup transient losses. This dictum has not received the backing of scientific theory, however. As Samuelson (1963, 1989) in particular

points out, this “time-diversification” argument relies on a fallacious interpretation of the Law of Large Numbers: repeating an investment pattern over many periods does not cause risk to wash out in the long run. This fallacy is illustrated by the following question raised by Samuelson (1963):

I offered some lunch colleagues to bet each \$200 to \$100 that the side of a coin they specified would not appear at the first toss. One distinguished scholar (...) gave the following answer: “I won’t bet because I would feel the \$100 loss more than the \$200 gain. But I’ll take you on if you promise to let me make 100 such bets”.

This story suggests that independent risks are complementary. However, Samuelson went ahead and asked why it would be optimal to accept 100 separately undesirable bets. The scholar answered:

”One toss is not enough to make it reasonably sure that the law of averages will turn out in my favor. But in a hundred tosses of a coin, the law of large numbers will make it a darn good bet.”

Obviously, this scholar misinterprets the Law of Large Numbers! It is not by accepting a second independent lottery that one reduces the risk associated with the first one. If $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ are independent and identically distributed random wealth variables, $\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_n$ has a variance n times as large as the variance of each of these risks. What is stated by the Law of Large Numbers is that $\frac{1}{n} \sum_{i=1}^n \tilde{x}_i$ — not $\sum_{i=1}^n \tilde{x}_i$ — tends to $E\tilde{x}_1$ almost surely as n tends to infinity. It is by subdividing — not adding — risks that they are washed away by diversification.

7.1 Backward induction

Solving dynamic decision problems requires understanding the method generally known as “backward induction”. Suppose that you have to make a sequence of two decisions α_0 in period 0, and α_1 in period 1. Decision α_0 is about some risk exposure whose payoff $z(\alpha_0, x)$ depends upon the realization x of a random variable \tilde{x} . It is important to notice that x is observed after

selecting α_0 , but before decision α_1 is taken. Your objective ex ante is to maximize the expectation of a function U of $(\alpha_0, \alpha_1, \tilde{x})$:

$$\max_{\alpha_0, \alpha_1} EU(z(\alpha_0, \tilde{x}), \alpha_1) \quad (7.1)$$

Backward induction consists in first solving the second period problem for each possible outcome that could prevail at the beginning of that period. This set of outcomes is entirely summarized by the payoff z obtained in the first period. The optimal strategy α_1^* in the second period will in general depend upon z , which is hereafter called the state variable of the dynamic program. This second period problem contingent to "state z " is written as

$$v(z) = \max_{\alpha_1} U(z, \alpha_1). \quad (7.2)$$

The optimal value of the objective given z is denoted $v(z)$. Function v is called the value function, or the Bellman function. One then solves the first period problem by selecting the risk exposure α_0 that maximizes the expectation of value function $Ev(z(\alpha_0, \tilde{x}))$. By doing so, the decision maker internalizes the effect of his future contingent strategy on his welfare U , given the definition of v . He is what we call "dynamically consistent." This technique transforms any dynamic problem into a sequence of static problems through the value function.

7.2 The dynamic investment problem

In this section, we examine the effect of the opportunity to take risk in the future on the willingness to take risk in the short run. In other words, will an investor with a longer planning horizon be willing to invest a higher proportion of wealth in risky stocks as opposed to safer bonds? We assume that the investor has the objective of maximizing the expected utility of his accumulated wealth at a specific date. This is the case, for example, when the investment is targeted for retirement. This money is not used for intermediary consumption. In the standard terminology, this is called an investment problem. We will introduce intermediary consumption later on in this chapter. We also assume here that risks are independent over time, a condition that will be relaxed in the last section of this chapter.

One can illustrate the problem examined in this section as follows. Building on Samuelson's question, suppose that you are offered to bet on whether

a fair coin will land Head or Tail. You get 3 times your stake if it lands H, and you lose it otherwise. Suppose that, given your risk aversion, you want to bet α on this single gamble. Now, suppose that you are said that you will be allowed to bet sequentially on two independent draws of the coin. How does this affect your bet α' on the initially draw of the coin? This question is equivalent to the effect of time horizon on the optimal investor's portfolio composition.

We consider the following more general question. An investor who is endowed with wealth w_0 lives for two periods. At the beginning of each period, he has the opportunity to take some risk whose realization will be observed at the end of the corresponding period. It is important to notice that the investor will observe his loss or gain on the risk that he took in the first period before deciding how much risk to take in the second period. This makes the problem intrinsically dynamic, and it introduces flexibility, an essential element of dynamic risk management. To illustrate, under DARA, investors will take less risk in the second period if they suffered heavy losses on their portfolio in the first period.

To be more specific, we suppose that the second period problem is an Arrow-Debreu portfolio decision. There are S possible states of nature $s = 0, \dots, S-1$. The uncertainty prevailing over this second period is characterized by the vector of probabilities (p_0, \dots, p_{S-1}) . Π_s is the unit price of the Arrow-Debreu security associated to state s . We assume that the risk-free rate is zero. This implies that a claim paying one euro in every state of nature must itself cost one euro; $\sum_s \Pi_s = 1$. In other words, if the investor does not take any risk in the second period, he will end up with the same final wealth as in the first period. Given the wealth z accumulated at the end of the first period, the investor selects a portfolio (c_0, \dots, c_{S-1}) which maximizes the expected utility of his wealth at the end of the period subject to his budget constraint:

$$v(z) = \max_{c_0, \dots, c_{S-1}} \sum_{s=0}^{S-1} p_s u(c_s) \text{ subject to } \sum_{s=0}^{S-1} \Pi_s c_s = z. \quad (7.3)$$

This is equivalent to problem (7.2) with $\alpha_1 = (c_1, \dots, c_{S-1})$ and

$$U(z, \alpha_1) = p_0 u \left(\frac{z - \sum_{s=1}^{S-1} \Pi_s c_s}{\Pi_0} \right) + \sum_{s=1}^{S-1} p_s u(c_s).$$

In period zero, the investor must take a risky decision α_0 that yields a payoff $z(\alpha_0, x)$ which depends upon the realization x of some random variable \tilde{x} . In particular this can be another portfolio choice problem. The optimal exposure to the risk in period 0 is obtained by solving the following program:

$$\alpha_0^* \in \arg \max_{\alpha_0} Ev(z(\alpha_0, \tilde{x})) \quad (7.4)$$

We want to determine the impact of the opportunity to take risk in the second period on the optimal exposure to risk in the first period. To do this, we compare the solution α_0^* obtained from the dynamic program (7.4) to the optimal exposure to risk in the first period when there is no such an option to take further risk in the second period. The short-lived investor, as well as the myopic, investor would select the level $\hat{\alpha}_0$ that would maximize the expected utility of $z(\alpha_0, \tilde{x})$:

$$\hat{\alpha}_0 \in \arg \max_{\alpha_0} Eu(z(\alpha_0, \tilde{x})). \quad (7.5)$$

We see that the only difference between programs (7.5) and (7.4) is that the utility function u in the first is replaced by the value function v in the second. It is another way to say that the effect of the future is entirely captured by the characteristics of the value function. In our context, the opportunity to take risk in the future raises the willingness to take risk today if v is less concave than u in the sense of Arrow-Pratt. This is a consequence of Proposition 17 in the case of a one-risky-one-riskfree portfolio problem in the first period.

The optimal exposure to risk in the first period is larger than the myopic one if the value function v defined by program (7.3) is less concave than the original utility function u , i.e. v is more risk tolerant than u . The degree of absolute risk tolerance of v is characterized in the following Proposition.

Proposition 20 *The value function for the Arrow-Debreu portfolio problem (7.3) has a degree of absolute risk tolerance given by*

$$T_v(z) = -\frac{v'(z)}{v''(z)} = \sum_{s=0}^{S-1} \Pi_s T(c_s^*), \quad (7.6)$$

where c^* is the optimal solution to problem (7.3) and $T(\cdot) = -u'(\cdot)/u''(\cdot)$ is the absolute risk tolerance for final consumption.

Proof: The optimal solution to program (7.3) is hereafter denoted $c^*(z)$. It satisfies the following first-order condition:

$$u'(c_s^*(z)) = \xi(z)\pi_s \quad s = 0, \dots, S-1, \quad (7.7)$$

where $\pi_s = \Pi_s/p_s$ is the state price per unit of probability as developed in Chapter 5. Fully differentiating condition (7.7) with respect to z and eliminating π_s yields

$$c_s^{*'}(z) = -\frac{\xi'(z)}{\xi(z)}T(c_s^*(z)). \quad (7.8)$$

Fully differentiating the budget constraint yields in turn

$$\sum_{s=0}^{S-1} \Pi_s c_s^{*'}(z) = 1. \quad (7.9)$$

Replacing $c_s^{*'}(z)$ by its expression in (7.8) implies that

$$-\frac{\xi'(z)}{\xi(z)} = \left[\sum_{s=0}^{S-1} \Pi_s T(c_s^*(z)) \right]^{-1}. \quad (7.10)$$

Finally, fully differentiating $v(z)$, which by definition equals $\sum_s p_s u(c_s^*(z))$, implies that

$$v'(z) = \sum_{s=0}^{S-1} p_s u'(c_s^*(z)) c_s^{*'}(z) = \xi(z) \sum_{s=0}^{S-1} \Pi_s c_s^{*'}(z) = \xi(z).$$

The second equality above follows from the first-order condition whereas the third equality is due to (7.9). This confirms the classical result that the Lagrange multiplier associated to the budget constraint of the consumer equals the shadow price of wealth. From this result, we see that $v''(z) = \xi'(z)$ and $T_v(z) = -\xi(z)/\xi'(z)$. The Proposition then follows immediately from equation (7.10). ■

Now, remember that we assumed that the risk-free rate in the second period is zero, in turn implying that $\sum_s \Pi_s = 1$. This assumption eliminated a potential wealth effect for those who are allowed to invest in the second period. Then, property (7.6) states that the absolute risk tolerance of the value function is a weighted average of the degree of risk tolerance of final

consumption.¹ This property allows us to compare the degrees of concavity of u and v . Suppose for example that u is HARA, i.e., that T is linear in c . It implies that $T_v(z) = \Sigma_s \Pi_s T(c_s^*) = T(\Sigma_s \Pi_s c_s^*) = T(z)$. Thus, when u is HARA, the value function v has the same degree of concavity as u : $v(\cdot) = Ku(\cdot)$. This implies that the two programs (7.4) and (7.5) have exactly the same solution. In other words, under HARA preferences the option to take risk in the future has no effect on the optimal exposure to risk today: myopia is optimal. Ceteris paribus, young and old investors should select the same portfolio composition.

Suppose alternatively that the utility function u exhibits a convex absolute risk tolerance. Applying Jensen's inequality, it follows that $T_v(z) = \Sigma_s \Pi_s T(c_s^*) > T(\Sigma_s \Pi_s c_s^*) = T(z)$: the opportunity to take risk in the future raises the tolerance to current risks. The assumption that T'' is nonnegative is compatible with the intuition that a longer time horizon should induce more risk-taking.² On the other hand, if T'' is nonpositive, a longer time horizon for investment should imply a more conservative investment in the short run.

Proposition 21 *Suppose that the risk free rate is zero. In the dynamic Arrow-Debreu portfolio problem with serially independent returns, a longer time horizon raises (resp. reduces) the optimal exposure to risk in the short term if the absolute risk tolerance $T(\cdot) = -u'(\cdot)/u''(\cdot)$ is convex (resp. concave). In the HARA case, the time horizon has no effect on the optimal portfolio.*

When the long term investment is targeted for consumption at a specific date, whether the investor should modify his risk exposure as the time horizon recedes is an empirical question which relies on the convexity, linearity, or concavity of absolute risk tolerance. Of course none of these conditions on absolute risk tolerance need hold for all wealth levels. It is possible to have a risk tolerance that is sometimes convex and sometimes concave. For such an individual, we will not be able to predict the effect of a longer planning horizon on investment strategy. Depending on the circumstances, this individual sometimes will invest more in stocks and other times will invest more in bonds than would be invested under myopia.

¹In the terminology of the theory of finance, it is a martingale.

²This assumption relies on the sign of the fourth derivative of the utility function.

One can think about the convexity/concavity of absolute risk tolerance by introspection. Remember that, by (4.4), the euro amount optimally invested in stocks is approximately proportional to T . Under DARA, it is increasing in wealth. The question is whether it is increasing at an increasing rate as wealth increases. If it is, this would be an argument for a convex T , and for a positive effect of time horizon length on risk taking. Most theoretical models in finance use HARA utility functions. In these models, myopia is optimal, which simplifies much the analysis. One can, however, suspect that this assumption is made for simplicity rather than for realism. Econometric tests for HARA preferences are extremely scarce in the literature.

7.3 Time diversification

In the investment problem, there is a single prespecified consumption date. This implies that all risks taken in life are borne on that date. In most instances however, investors will want to use their portfolios to finance consumption throughout their lifetimes.³ This has an important advantage, which is to allocate current risks on wealth into small risks on consumption over a long time horizon. This produces an important time-diversification effect, which makes people with a longer planning horizon willing to take more risk.

To explain this, let us consider a simple model where the agent has the opportunity to take a risk at date $t = -1$. More specifically, we assume that the payoff for the initial risk-taking game is $z(\alpha_0, \tilde{x})$, where α_0 is a decision variable and \tilde{x} is a random variable. The agent then consumes over the remaining n dates numbered $t = 0, \dots, n - 1$. We assume that the agent can save and borrow at a zero interest rate, and that he has no risk-taking opportunities from date $t = 0$ on. Moreover, at each period he earns a labor income y .

This problem exhibits the same dynamic structure as presented in section 7.1. To determine the optimal exposure to risk in the first period, one needs first to solve the consumption-saving problem occurring after the risky outcome is revealed. For a given wealth z accumulated prior to date $t = 0$,

³Only tax incentives can work against this possibility.

we can write

$$v(z) = \max_c \sum_{t=0}^{n-1} p_t u(c_t) \text{ subject to } \sum_{t=0}^{n-1} c_t = z + ny, \quad (7.11)$$

where p_t is the discount factor associated to date t , and $z + ny$ is the lifetime wealth. With this value function v , one can determine the level of optimal initial risk taking by solving $\max_{\alpha_0} Ev(z(\alpha_0, \tilde{x}))$. This level of risk is increasing in the degree of risk tolerance of the value function v .

As already observed in the previous Chapter, the structure of this consumption-saving problem is essentially the same as that of the static Arrow-Debreu portfolio problem in Proposition 20. The main difference is that we need to assume here that $\Pi_s = 1$ for all s , i.e. a euro's worth of consumption at time t costs one euro today (since we assume that the risk-free rate is zero). Consequently, it follows from Proposition 20 that

$$T_v(z) = \sum_{t=0}^{n-1} T(c_t^*), \quad (7.12)$$

where c_t^* is the optimal solution of problem (7.11). In the consumption-saving problem with certainty for dates $t = 0, n - 1$ and with a zero interest rate, the degree of tolerance to the risk on initial wealth equals the sum of the absolute tolerances to risk on consumption over the lifetime of the consumer.

We now want to examine the effect of an increase in n on $T_v(z)$. For simplicity, suppose that consumers are not impatient, so that $p_t = 1$ for all t . Then, it is optimal to smooth consumption completely: $c_t^* = y + (z/n)$ at every date t . In this setting, all gains and losses on the initial risk are allocated equally over the remaining n periods of consumption. Property (7.12) can thus be written as

$$T_v(z) = nT(y + (z/n)).$$

For a small initial risk (z small), the absolute tolerance to risk on wealth is proportional to the lifetime of the gambler. Thus, an agent who expects to live twice as long as another agent with the same yearly income would invest approximately twice as much in stocks as his shorter-lived counterpart at date $t - 1$. This is the real meaning that should be given to the notion of "time diversification".

Of course, we assumed here that there is only a single point in time where consumers may take risk. In the real world, consumers can own stocks and take risk at any time. This more realistic assumption would not change our result in the HARA case. Indeed, using backward induction together with Proposition 21, adding the opportunity to take risk in the future would not change the concavity of the value function at any specific date when HARA is assumed. Agents are myopic to future risks in that case, and property (7.12) would still hold.

Another issue is related to the existence to liquidity constraints. Time diversification works well only if consumers are allowed to borrow money at an acceptable loan rate when they face an adverse shock on their incomes, and that their cash-on-hand is depleted. This is an unrealistic assumption. Agents with no liquidity reserve cannot smooth a negative income shock by borrowing money from their bank. They cannot time-diversify. It implies that they should be much more averse to income risks. This is an additional argument in favor of decreasing absolute risk aversion.

7.4 Portfolio management with predictable returns

In section 7.2, we examined a portfolio-decision problem in which the opportunity set for investment was invariant over time. In the real world, it is often the case that this opportunity set is stochastic and that there is some predictability in its changes. Predictability can come for example from the existence of serial correlation in stock returns. The existence of mean-reversion in stocks returns has recently been recognized: a high return of the risky portfolio today generally implies a lower expected portfolio return tomorrow. Receiving some good news today often means bad news for the future opportunity set. Alternatively, predictability might come from some type of learning process as investors try to estimate the return distribution based on observed data. If there is some parameter uncertainty, for example about the size of the equity premium, the observation of a large stock return in one year might yield an upward Bayesian updating of the distribution of future returns.

In this section, we consider the effect of such predictability on the optimal dynamic portfolio. Obviously, investors will follow a flexible strategy where

the optimal risk exposure is made conditional on the opportunity set. But investors also will try to anticipate any shocks to this set. More specifically, investors can consider the possibility hedging against any bad news about their future opportunity set. Of course, this is easier to accomplish if shifts are statistically related to current returns. The demand for stocks which is due to this anticipation is called the "hedging demand" for stocks. Because stocks are thought to be safer in the long run than in the short run, intuition suggests that an investor with a longer planning horizon will take more risk early in life.

For the sake of simplicity, we limit the analysis here to the case of constant relative risk aversion γ with a two-period time horizon. Constant relative risk aversion implies myopia with respect to the time horizon in the absence of predictability. We assume that the economy has one risk-free asset with a zero return and one risky asset whose return in period t is denoted by \tilde{x}_t , $t = 0, 1$. The opportunity set in the second period is thus completely described by \tilde{x}_1 . Predictability comes from an assumption that the distribution of \tilde{x}_1 is correlated to \tilde{x}_0 . We assume that $E\tilde{x}_0 > 0$ and $E[\tilde{x}_1 | x_0] > 0$ for any x_0 . Investors invest only for their retirement at the end of the second period so that there is no intermediary consumption.

In order to determine the optimal demand for the risky asset in the first period, and in particular its hedging component, it is necessary to follow the method presented in section 7.1. We begin with solving the problem faced by investors in the second period for each possible situation. What is new here is that a situation is described not only by the wealth z accumulated at that time, but also by the realized return x_0 of the risky asset in the first period. More specifically, the value function v is defined by

$$v(z, x_0) = \max_{\alpha} E \left[\frac{(z + \alpha \tilde{x}_1)^{1-\gamma}}{1-\gamma} \mid x_0 \right]. \quad (7.13)$$

From Proposition 18, we know that the optimal solution of this program is a separable function $\alpha_1(z, x_0) = a(x_0)z$. This implies in turn that the value function is separable, with $v(z, x_0) = h(x_0)z^{1-\gamma}/(1-\gamma)$, where

$$h(x_0) = E \left[\frac{(1 + a(x_0)\tilde{x}_1)^{1-\gamma}}{1-\gamma} \mid x_0 \right].$$

We can now turn to the first-period decision problem. This can be written

as

$$\alpha_0^* = \arg \max_{\alpha} H(\alpha) = E \left[h(\tilde{x}_0) \frac{(w_0 + \alpha \tilde{x}_0)^{1-\gamma}}{1-\gamma} \right]. \quad (7.14)$$

In order to determine the hedging component of the demand for the risky asset, we compare α_0^* to the demand for the risky asset when there is no predictability, i.e., when \tilde{x}_1 is independent of \tilde{x}_0 . In this case, we know that myopia is optimal. Thus, without predictability, investors solve

$$\alpha_0^m = \arg \max_{\alpha} E \left[\frac{(w_0 + \alpha \tilde{x}_0)^{1-\gamma}}{1-\gamma} \right].$$

The hedging demand is defined as $\alpha_0^* - \alpha_0^m$. This hedging demand will be positive if the derivative of H evaluated at α_0^m is positive. In other words, the hedging demand for the risky asset is positive if

$$H'(\alpha_0^m) = E [\tilde{x}_0 h(\tilde{x}_0) (w_0 + \alpha_0^m \tilde{x}_0)^{-\gamma}] \geq 0$$

whenever $E [\tilde{x}_0 (w_0 + \alpha_0^m \tilde{x}_0)^{-\gamma}] = 0$.

To consider a specific type of predictability, we examine the case where an increase in x_0 deteriorates the distribution of \tilde{x}_1 in the sense of second-order stochastic dominance (SSD). A special case is when the stochastic process $(\tilde{x}_0, \tilde{x}_1)$ exhibits mean-reversion. For example, suppose that the conditional distribution of \tilde{x}_1 can be written as $\tilde{x}_1 | x_0 = -kx_0 + \tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is assumed to be independent of \tilde{x}_0 . Because any SSD shift in \tilde{x}_1 reduces the expected utility of final wealth, this assumption implies that $\partial v / \partial x_0$ is negative. Since $v(z, x_0) = h(x_0) z^{1-\gamma} / (1-\gamma)$, it follows that h' must be negative when $\gamma < 1$, and h' must be positive when $\gamma > 1$.

Suppose that relative risk aversion γ is larger than unity. Because h' must be positive in this case, it follows that for all x_0 ,

$$x_0 h(x_0) (w_0 + \alpha_0^m x_0)^{-\gamma} \geq x_0 h(0) (w_0 + \alpha_0^m x_0)^{-\gamma}.$$

Taking the expectation, it implies in turn that

$$H'(\alpha_0^m) \geq h(0) E [\tilde{x}_0 (w_0 + \alpha_0^m \tilde{x}_0)^{-\gamma}] = 0.$$

Thus, the hedging demand is positive when relative risk aversion is larger than unity. If instead we have relative risk aversion less than unity, $\gamma < 1$, then h' is negative, and the above inequality is reversed. This result is summarized in the following proposition.

Proposition 22 *Suppose that an increase in the first-period return deteriorates the distribution of the second-period return in the sense of second-order stochastic dominance. Then, the hedging demand for the risky asset is positive (resp. negative) if constant relative risk aversion is larger (resp. smaller) than unity.*

Another way to interpret this result is that when relative risk aversion is constant and larger than unity, a longer time horizon should induce investors to take more risk. The contrary is true if constant relative risk aversion is less than unity. Notice that when investors have a logarithmic utility function ($\gamma = 1$), myopia is still optimal in the presence of predictability.

The choice of the initial portfolio risk is driven by the slope of the marginal value of wealth at the end of the initial period. This marginal value of wealth depends upon the future opportunity set. If predictability reduces the marginal value of wealth in states where it is large, and if it raises it in states where it is small, then predictability has the same effect as a reduction in risk aversion: it raises the optimal level of risk in the portfolio. Consequently, we see that the central step of the analysis is to determine the effect that our SSD-deteriorating shift in the return of the risky asset will have on the marginal value of wealth. In the special case of mean-reversion, we can see two different effects of an increase in x_0 . The first effect is a wealth effect: because the expected return in the second period becomes smaller, so is the expected final wealth. This raises the marginal value of wealth, since v is concave in z . The second effect is a precautionary effect: investors will invest less in the risky asset, thereby reducing the exposure to the risk. Under prudence, this reduces the marginal value of wealth. The global effect of an increase in x_0 on the marginal value of wealth is thus ambiguous. When relative risk aversion is constant and larger than unity,⁴ the wealth effect always dominates the precautionary effect, and the hedging demand is positive. When relative risk aversion is less than unity, the wealth effect is dominated by the precautionary effect.

⁴Notice that constant relative risk aversion is larger than unity if and only if absolute prudence is smaller than twice the absolute risk aversion. This explains why this condition implies that the precautionary effect is dominated by the wealth effect.

7.5 Bibliographical references and extensions

Merton (1969) and Samuelson (1969) were the first to solve the dynamic portfolio problem in a continuous-time economy with HARA utility functions. Mossin (1968) proved that HARA functions are the only ones for which myopia is optimal when there is no serial correlation in returns. Deaton (1991) and Carroll (1997) examine the effect of liquidity constraints on the optimal saving behavior.

The book by Campbell, Lo and MacKinlay (1997) provides an extensive analysis of stock returns. Barberis (2000) estimates significant predictability of US stocks returns. The implied standard deviation of ten-year returns is 23.7 percent, much smaller than the 45.2 percent value implied by the standard deviation of monthly returns. Kim and Omberg (1996) showed that this is indeed the case if constant relative risk aversion is larger than unity. Campbell and Viciera (1999) and Barberis (2000) estimated this hedging demand numerically. The effect of the return predictability on the optimal structure of the initial portfolio is surprisingly large. For an agent with a relative risk aversion equaling 10 and a ten-year time horizon, the optimal investment in stocks is about 40% of current wealth without predictability. It goes up to 100% when mean-reversion is taken into account. Kandel and Stambaugh (1996) solved a model in which there is mean-reversion in stock returns, but with some estimation risk on the parameters of the mean-reversion.

References

- Barberis, N., (2000), Investing for the long run when returns are predictable, *Journal of Finance*, 55, 225-64.
- Campbell, J.Y, A.W. Lo and A. C. MacKinlay, (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton.
- Campbell, J., and L. Viciera, (1999), Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics*, 114, 433-95.
- Carroll, C.D., (1997), Buffer-stock saving and the life cycle/permanent income hypothesis, *Quarterly Journal of Economics*, 112, 1-55.

- Deaton, A., (1991), Saving and liquidity constraints, *Econometrica*, 59, 1221-48.
- Kandel, S., and R. Stambaugh, (1996), On the predictability of stock returns: An asset allocation perspective, *Journal of Finance*, 51, 385-424.
- Kim, T.S., and E. Omberg, (1996), Dynamic nonmyopic portfolio behavior, *Review of Financial Studies*, 9, 141-61.
- Merton, R.C., (1969), Lifetime portfolio selection under uncertainty: The continuous- time case, *Review of Economics and Statistics*, 51, 247-257.
- Mossin, J., (1968), Optimal multiperiod portfolio policies, *Journal of Business*, 215-229.
- Samuelson, P.A., (1969), Lifetime portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics*, 51, 239-246.

Chapter 8

Risk and information

The nature of risks is to be sensitive to the arrival of new information. My beliefs about my lifetime expectancy may be much affected if I got to be tested positive for some bad genes. I may change my beliefs about the frequency of floods in my area because of new scientific information relative to global warming. Financial news emerge every day that affect my perception of the riskiness of my investments.

Information is useful because of the Bayesian updating of the risk that it generates. Information is useful because it allows for Bayesian updating of probability distributions, so that better decisions can be made than if no information would be available. For example, if you hear the weather forecast in the morning predicting that it will likely rain, you might carry your umbrella with you to work or to school. But if your office has a window and it sunny at lunchtime, you might decide not to carry your umbrella to lunch. Compare this to your lunch companion, who took her umbrella to lunch only because she had no window in his office and did not have the same information to update her forecast. Young people might observe signals about their future productivity, which allow them to better select their education. I can treat the disease for which I have been tested positive. Entrepreneurs can obtain information linked to the probability of success for their investment project. Policy-makers may wait for better scientific knowledge about a risk, such as global warming, before committing a large amount of funding to a potential solution. Technological innovations provide information to economic forecasters about the prospect of future growth of the economy. A common feature of these examples is that informative signals might be observed before being required to make a final decision about the

exposure to some risk.

Expecting better information should affect both the welfare of the decision maker and the optimal management of risk. This chapter first examines the effect of information on welfare, and then addresses the problem of risk management with information.

8.1 The value of information

8.1.1 An example

Information is valuable because it allows for a better management of risk. To illustrate this point, consider a simple static insurance problem. Sempronius has an initial wealth of 4000 ducats. If his ship crosses the ocean safely, his wealth will be increased by 8000 ducats. Sempronius, who is risk-averse, is contemplating the possibility to purchase insurance. To simplify the problem, we assume that the insurance company offers a single contract. The contract stipulates that if Sempronius' ship is sunk, an indemnity of 8000 ducats will be paid. The premium associated to this full insurance contract is 4400 ducats. If the probability of damage is $1/2$, this premium corresponds to a loading factor of 10%, as in the illustration presented in section 3.1.

Sempronius' utility function is $u(z) = \sqrt{z}$. If he purchases the full insurance contract, his expected utility equals $\sqrt{12000 - 4400} = 87.178$. Because of full insurance, this is independent of Sempronius' beliefs about the likelihood of an accident. Let $p = p^0$ denote Sempronius' subjective probability of success in the absence of any information. At this stage, it is useful to leave p^0 unspecified. If he decides not to purchase any insurance, his expected utility equals

$$p^0 \sqrt{12000} + (1 - p^0) \sqrt{4000} = 46.299p^0 + 63.246.$$

His decision problem can thus be written as

$$V^0 = V(p^0) = \max \{87.178, 46.299p^0 + 63.246\} \quad (8.1)$$

where V^0 is Sempronius' maximum expected utility in the absence of information. It is a function $V(p^0)$ of his subjective probability of success. This problem can be solved by looking at Figure 8.1. As long as the subjective probability of success p is less than a threshold $p_c = 0.517$, it is optimal

to fully insure the risk because $87.178 > 46.299p + 63.246$. Otherwise, no insurance is optimal. Remember that in this simple example, there is no possibility to partially insure the risk. This implies that the optimal expected utility $V(p)$ as a function of the probability of success p is piecewise linear. However, the essential property of the maximum expected utility is that it is convex in the probability of success. If the subjective probability of success equals $p^0 = 0.5$ when Sempronius has no information, his optimal strategy is to leave his shipment uninsured.

[INSERT FIGURE 8.1 ABOUT HERE]

Suppose now that Sempronius can obtain information for free before determining to insure or not. For example, his fellows can give him information about pirates on the route, or about the weather. Suppose also that the insurer is unaware of the existence of such information available to its customer. Before getting information, Sempronius expects either a bad signal or a good signal about the success of his enterprise, with respectively probability q and $1 - q$. Using Bayes rule, he computes that the posterior probability of success is either $p^g > p^0$ if he receives the good signal, or $p^b < p^0$ if he receives the bad signal. Of course, we have that $qp^g + (1 - q)p^b = p^0$. It means that prior to observing the signal, the probability of success is as before. We will hereafter suppose that $p^0 = 0.5$, $q = 0.5$, $p^g = 0.75$ and $p^b = 0.25$. We want to determine whether expecting this information makes the decision maker better off ex ante.

To answer this question, we need to apply backward induction, as presented in section 7.1. We first solve the decision problem for each possible signal received by the agent. We then use these contingent solutions to compute the expected utility prior to the observation of the signal. Contingent on a given signal, this problem is solved exactly as in (8.1), where we replace p^0 by either p^g or p^b . This solution has been described earlier using the dashed curve in Figure 8.1. When the signal is bad, the probability of success is only $p^b = 0.25$, which implies that it is optimal to fully insure the risk. It yields a sure final wealth equaling $12000 - 4400 = 7600$, and a final utility equaling $V(p^b) = 87.178$ with certainty. On the contrary, if a good signal is perceived, the probability of success $p^g = 0.75$ is large enough to induce Sempronius not to insure the risk. In fact, the implicit loading of the premium becomes so large with such beliefs that insurance becomes undesirable. It implies an expected utility equaling $V(p^g) = 46.299 * 0.75 + 63.246 = 97.970$. This is illustrated in Figure 8.2. The unconditional expected utility, i.e., the expected utility before observing the signal, is then simply $V^i = qV(p^g) + (1 - q)V(p^b) =$

92.574. This is depicted as V^i in Figure 8.2.

[INSERT FIGURE 8.2 ABOUT HERE]

Information is equivalent to introducing a mean-preserving spread in the probability of success. Because, as seen in Figure 8.1, the maximum expected utility V is a convex function of the probability of success p , information raises welfare : $V_i = qV(p^g) + (1 - q)V(p^b) > V_0$. Figure 8.2 illustrates why. The investor has the option to not insure, if he receives the good signal. In this case, it is optimal to exercise this option, allowing him to raise his expected utility. Information combined with flexibility is valuable to the decision maker. In our numerical example, information has a monetary value equaling $\kappa = 970$ ducats, since

$$V^i = u(12000 - 4400 + \kappa),$$

i.e, Sempronius is indifferent between getting the information or receiving κ ducats.

It is important to see that information is valuable only because of the sensitiveness of the ex-post decision to the signal. Suppose that, contrary to what we have illustrated in Figure 8.2, both p^g and p^b would be both smaller than p_c . In this case, $p^t u(12000) + (1 - p^t)u(4000)$ is smaller than $u(7600)$ both for $t = g$ and for $t = b$. Hence, because of the linearity of V in interval $[p^b, p^g]$,

$$\begin{aligned} V^i &= qV(p^g) + (1 - q)V(p^b) \\ &= V(qp^g + (1 - q)p^b) = V^0, \end{aligned}$$

so that it would be optimal to not exercise this option. That is, Sempronius should insure the risk regardless of which signal is received. If this occurs, the information would be useless and the agent would not be willing to pay for such information. Information is valuable only if observing some of the possible signal would reverse your decision. To illustrate, the value of a genetic test would be zero if no treatment is available to cure the illness, or even to stabilize it.

8.1.2 A general model

The property that the value of information is nonnegative is a general result that does not depend upon the particular decision problem or upon the information structure. Consider any decision problem where the final utility

$U(s, \alpha)$ is a function of the state of the world s and of a decision variable α . Suppose that there are S possible states of nature. The uncertainty can thus be described by a vector of probabilities $P = (p_1, \dots, p_s, \dots, p_S)$, where $\sum_s p_s = 1$. Now we can define the indirect utility function as

$$V(P) = \max_{\alpha} \sum_{s=1}^S p_s U(s, \alpha). \quad (8.2)$$

One can describe the decision problem without information as determining $V^0 \equiv V(P^0)$, the maximum expected utility given the distribution P^0 of the states of nature.

We now wish to compare this environment without information with an environment where the decision maker can observe a signal before choosing α . Suppose that there can be M possible messages $m = 1, \dots, M$. The probability of receiving message m is denoted by q^m , with $\sum_m q^m = 1$. The posterior probability distribution of the states of nature given message m is denoted $P^m = (p_1^m, \dots, p_s^m)$. Notice that the unconditional probability of state s is $\sum_m q^m p_s^m$. We assume that this is equal to the probability p_s^0 of state s in the environment without information. Using a vector notation, this means that

$$P^0 = \sum_{m=1}^M q^m P^m. \quad (8.3)$$

This states that the underlying risk is the same in the two environments. The maximum expected utility of the decision maker, who must decide upon α before observing the informative message, can be written as

$$V^i = \sum_{m=1}^M q^m \max_{\alpha} \sum_{s=1}^S p_s^m U(s, \alpha) = \sum_{m=1}^M q^m V(P^m). \quad (8.4)$$

The value of information will be nonnegative whenever V^i is at least as great as V^0 , i.e.

$$\sum_{m=1}^M q^m V(P^m) \geq V \left(\sum_{m=1}^M q^m P^m \right). \quad (8.5)$$

Obviously, this will be true for all possible information structures if and only if the function V , which is defined by equation (8.2), is convex in P . Observe that, by definition, V is the upper envelope of linear functions $\sum_s p_s U(s, \alpha)$ of (p_1, \dots, p_S) . This must be convex. We thus obtain the following Proposition.

Proposition 23 *In the expected utility model, the value of information is always nonnegative: $V^i \geq V^0$.*

Proof: Let α^0 and α^m denote the optimal decision respectively without information and conditional to signal m . Observe that $V(P^m)$ is larger than

$$\sum_{s=1}^S p_s^m U(s, \alpha^0).$$

Otherwise α^0 would be a better decision than α^m when signal m is received. Because this must be true for all $m = 1, \dots, M$, it implies that

$$\begin{aligned} V^i &= \sum_{m=1}^M q^m V(P^m) \\ &\geq \sum_{m=1}^M q^m \sum_{s=1}^S p_s^m U(s, \alpha^0) \\ &= \sum_{s=1}^S \left(\sum_{m=1}^M q^m p_s^m \right) U(s, \alpha^0) \\ &= \sum_{s=1}^S p_s^0 U(s, \alpha^0) = V^0. \end{aligned}$$

This concludes the proof. It means that V is convex. ■

This proof provides a simple intuition to this result. An informed decision maker always can do at least as well as an uninformed decision maker by deciding to ignore the information. This shows once again that the value of information comes from the ability of the informed decision maker to adapt the decision in a more efficient way to the circumstances. It is worth noting that the linearity of expected utility with respect to probabilities is an essential ingredient for this result to hold. Interestingly, risk aversion does not play any role here. The information always has a nonnegative value, even to a risk-loving decision maker.

Up to now, we compared a situation with some information to a situation with no information. There is an extensive literature on comparing two information structures, where the question is to determine whether one is unambiguously better than the other. We will not examine this topic here, except to note that Figure 8.2 is very helpful to think about this in the

framework of our example. It is apparent from this figure that if the posterior probabilities p^b and p^g are "spread" further away from the mean p^0 , then the value of information will increase. More generally, because V is convex, any mean-preserving spread of the information structure in the space of posterior probabilities will make all expected-utility maximizers better off.

8.1.3 Value of information and risk aversion

Because information allows for a better management of the risk, it would seem intuitive that more risk-averse decision-makers should value information more. However, this is not true in general. We can illustrate this by going back to the two-state example presented in section 8.1.1. Let w^b and w^g denote the certainty-equivalent wealth levels of the decision maker when he receives the signals p^b and p^g respectively. Also let w^0 and w^i denote the certainty equivalent respectively in the absence and in the presence of information. The monetary value κ of information can implicitly be defined by

$$u(w^0 + \kappa) = qu(w^g) + (1 - q)u(w^b) = u(w^i). \quad (8.6)$$

In other words, the agent is indifferent between being informed, or not being informed and getting a compensating premium κ . It means that $\kappa = w^i - w^0$. We examine the effect of an increase of the concavity of u on κ .

Suppose that p^0 is smaller than p_c , the critical probability of success below which it is optimal to fully insure the risk of failure. Without information, the agent does not take any risk and $w^0 = 7600$ ducats. Of course, if the bad signal is received, the willingness to insure is reinforced, and $w^b = w^0$. The only interesting case is when p^g is larger than p_c , in which case the arrival of a good signal induces the agent to self-insure. In this case, we can expand (8.6) to obtain

$$u(7600 + \kappa) = q[p^g u(12000) + (1 - p^g)u(4000)] + (1 - q)u(7600).$$

It follows that κ can be interpreted as the certainty equivalent of a lottery whose payoffs are 4400, -3600 and 0 respectively with probabilities qp^g , $q(1 - p^g)$ and $1 - q$. As we know from Chapter 1, any increase in risk aversion reduces the certainty equivalent κ of this risk. Thus, this is an example where an increase in risk aversion *reduces* the value of information. The intuition is simple: we have a case where information induces the agent to take a risk

that he would not take in the absence of information. In this case, it is clear that a more risk-averse agent would value the information less, because the benefit of the risk exposure is smaller.

[INSERT FIGURE 8.3 ABOUT HERE]

In Figure 8.3, we draw the value of information κ as a function of the degree of relative risk aversion γ by assuming the Sempronius' utility function is $u(z) = z^{(1-\gamma)}/(1-\gamma)$. As before, we assume that $p^0 = 0.5$, $q = 0.5$, $p^g = 0.75$ and $p^b = 0.25$. In the absence of information, it is optimal to insure the risk if relative risk aversion is larger than $\gamma^* = 0.375$. When a bad signal is observed with $p^b = 0.25$, the subjective loading factor perceived by Sempronius is negative, which implies that it is always optimal to fully insure the risk. When a good signal is observed ($p = p^g = 0.75$), the subjective loading factor is 120%. It can be computed that full insurance is optimal in that case only if relative risk aversion is larger than $\gamma = 2.390$, far outside the range of γ depicted in Figure 8.3.

As explained above, the value of information is decreasing with γ in this range $\gamma \geq \gamma^*$ of relative risk aversion. Suppose alternatively that $\gamma < \gamma^*$. For such low degrees of risk aversion, it is optimal to self-insure without information because of the positive loading factor. In this range of smaller risk aversion, the risk exposure ex-ante is larger without information than when information can be obtained, since it is still optimal to self-insure when a bad signal is observed. Symmetrically to the previous case, a bad signal induces the agent to insure a risk that he would have left uninsured in the absence of information. Here, information reduces the optimal risk exposure ex ante. It implies that an increase in risk aversion has a negative effect on w^0 that is larger than on w^i . It implies that $\kappa = w^i - w^0$ is positively affected by an increase in risk aversion. Globally, the value of information is hump-shaped in Figure 8.3.

8.2 Comparative statics analysis

Until now, we have examined the effect of information on welfare. We next examine the effect of information on behavior: how do these informative signals affect optimal actions that are to be taken ex ante, i.e. taken prior to the arrival of the signal? To answer this question, we consider the following

two-period model:

$$\max_{\alpha_0} u(\alpha_0) + \sum_{m=1}^M q^m \max_{\alpha \in B(\alpha_0)} \sum_{s=1}^S p_s^m U(s, \alpha, \alpha_0) \quad (8.7)$$

At date 0, the decision maker selects α_0 which yields felicity $u(\alpha_0)$ for this first period. At the beginning of the next period, the agent observes a signal m which affects his beliefs about the distribution of the states of nature \tilde{s} . He then chooses α to maximize his expected utility, conditional on having received signal m . The dynamic nature of this problem comes from the fact that the decision in the first period affects the second period in two ways. First, the choice set for α in the second period may be constrained by the initial choice of α_0 . This is stipulated in (8.7) by the constraint $\alpha \in B(\alpha_0)$. Second, the initial choice of α_0 at date 0 might directly affect utility in the second period. This is why we assume in program (8.7) that U depends upon α_0 .

The decision maker is in a situation of probabilistic uncertainty in the first period. He does not know which of the M probability distributions P^m , $m = 1, \dots, M$, is the true one. However, he knows the probability distribution (q^1, \dots, q^M) over the set $\{P_m\}$. In short, he faces what is sometimes known as parameter risk: uncertainty about the parameters of the distribution for the risk that he must manage. He must make an initial decision prior to resolving this uncertainty. In real life, we are surrounded by these kinds of difficult decision problems. The timing of the decision is crucial when the uncertainty is evolving over time, i.e., when there is some resolution of the uncertainty over time.

For example, one can reinterpret the dynamic portfolio decision problem with predictability, as in section 7.4, using our signaling terminology. Alternatively, a firm must decide to invest in new production capacities before knowing the future demand for its product. The firm knows that delaying the decision will allow it to get additional information, but it is costly to wait because of the lost revenues. Or consider a situation with scientific uncertainty such as the problem of the reduction of greenhouse gas emissions and the risk of global warming. We must decide whether to act immediately, without knowing the exact size of the risk, or to wait for scientific progress that might –or might not – confirm the level of danger. Many scientific advisers and environmentalists recommend that we use the precautionary principle in such a situation; in other words we should "play it safe."

We want to determine the effect of information on the choice of α_0 . To do this, we compare the optimal α_0 of program 8.7 with the optimal initial decision when there is no early resolution of the uncertainty. We illustrate this general model with specific examples in the following two subsections.

8.2.1 Real-option value and irreversibility

Early economic models of information considered the case where the selection of α_0 modified the opportunity set $B(\alpha_0)$ in the future, but had no direct effect on future utility ($U_{\alpha_0} \equiv 0$). Examples of this phenomenon abound in the real world. In many instances, investments are irreversible: a firm that invests in new factory cannot easily decide to disinvest. Greenhouse gases emitted in the past can hardly be removed from the atmosphere by human intervention. Many genetically modified organisms cannot be eliminated once they have been introduced in the environment. Once a hydroelectric dam is built in a beautiful valley, one cannot go back and restore the valley to its initial state. Of course, this irreversibility might also preclude using newer and better technology: burying radioactive waste in reasonably secure containers today might not be considered if we had information that a new scientific method was on the horizon, one which would completely neutralize the radioactivity.

There is a clear link between information and irreversibility. Because new information may cause the decision maker to regret his initial decision, irreversibility is a problem. Stated differently, when beliefs are expected to evolve over time, it is often important to preserve some flexibility in decision making. In fact, flexibility has emerged recently as an essential aspect of risk management. To see why, we consider a very simple and still classical example.

A risk-neutral firm must determine if and when to invest in a risky project. The project generates a net cash-flow of x_0 in period 0 and \tilde{x}_1 in period 1. We assume that $E\tilde{x}_1$ is positive. The investment is irreversible in the sense that if the investment is made in period 0 ($\alpha_0 = 1$), the firm cannot disinvest in period 1.¹ If α_t denotes the production capacity in period t , where $t = 0, 1$, this irreversibility means that $B(\alpha_0 = 0) = \{0, 1\}$, whereas $B(\alpha_0 = 1) = \{1\}$. This means that selecting not to invest in the project at

¹An extension of the irreversibility problem is the abandonment problem, in which the firm can reverse its decision, but only at some high cost.

date 0 ($\alpha_0 = 0$), allows for more flexibility in the future. "Not investing" is a reversible decision, whereas "investing" is an irreversible decision. The firm is assumed to discount future cash-flows at rate r . We compare two cases. In the first case, the firm does not expect to receive any information about the distribution of \tilde{x}_1 before the end of period 1. In the second case, there is a complete early resolution of uncertainty at the end of the first period: x_1 is revealed with certainty.

Without any early resolution of uncertainty, it is optimal to invest immediately if and only if x_0 is positive. Indeed, the firm must compare the discounted expected cash flow $x_0 + (1+r)^{-1}E\tilde{x}_1$ if it invests immediately to the one that it will obtain if it delays the decision. Waiting to invest would yield a smaller net present value (NPV) of $(1+r)^{-1}E\tilde{x}_1$.

We now turn to the optimal strategy when an early resolution of uncertainty is expected. In this case, it is optimal to invest immediately only if

$$x_0 + \frac{E\tilde{x}_1}{1+r} \geq \frac{E \max(0, \tilde{x}_1)}{1+r}. \quad (8.8)$$

The right-hand side of inequality (8.8) is the NPV when the firm decides not to invest in the first period, and then follows a strategy of investing in the second period only when it is optimal to do so conditional to the information, i.e., only when x_1 is positive. Thus, it might be optimal in this setting for the firm not to ever invest in the project. Inequality (8.8) may be rewritten as

$$x_0 \geq \frac{E \max(-\tilde{x}_1, 0)}{1+r}. \quad (8.9)$$

The right-hand side, which is positive, measures the benefit of waiting, which is the ability to avoid the loss $-x_1$ when x_1 is negative. Of course, the cost of this delay is the opportunity cost of receiving x_0 .

The minimum value of x_0 that will cause the firm to invest immediately is larger when the uncertainty evolves over time. In other words, taking into account the resolution of the uncertainty causes the decision maker to value flexibility for the future. This is a very general result that holds for all decision problems (8.7) where $U_{\alpha_0} \equiv 0$. For example, in the real world, this value of flexibility tends to favor delaying the use of genetically-modified foods, taking stronger early actions against global warming, and preserving environmental assets.

A firm that had to decide at date 0 whether or not to invest, would usually decide to invest if and only if the NPV $x_0 + (1+r)^{-1}E\tilde{x}_1$ is positive. Indeed in much of the finance literature, we often see this so-called "NPV rule" espoused. However, if delaying the decision is a possibility, and if some information will be revealed in the future, then the correct cost-benefit analysis is to use rule (8.8), which accounts for a premium $(1+r)^{-1}[E \max(0, \tilde{x}_1)]$ derived from the more flexible strategy. This premium is called the "real option value" of the ability to delay the decision. There is a large literature that provides various methods to compute these option values in more realistic setting.

8.2.2 Savings and the early resolution of uncertainty

Irreversibility is not the only element that affects the optimal early decision in the presence of an evolving uncertainty. Here we consider an example where the set B itself does not depend upon α_0 . A consumer lives for 3 periods, $t = 0, 1, 2$ and has an initial wealth of w_0 . There is a single asset whose return is risk-free and is normalized to zero. We also assume that the agent is patient ($\beta = 1$). At the last date of consumption, the agent earns an uncertain income \tilde{x} . If there is no early resolution of uncertainty, i.e., when the realization of \tilde{x} is observed only at the beginning of the third period, this decision problem can be written as follows:

$$\max_{\alpha_0} u(w_0 - \alpha_0) + \max_{\alpha} [u(\alpha_0 - \alpha) + Eu(\alpha + \tilde{x})], \quad (8.10)$$

where α_0 and α are the savings respectively at the end of period 0 and period 1. Because the agent is averse to consumption fluctuations ($u'' < 0$), we know that it is optimal to smooth consumption over the first two periods: $w_0 - \alpha_0 = \alpha_0 - \alpha$. This implies that $\alpha = 2\alpha_0 - w_0$. Thus, the above problem can be rewritten as

$$\max_{\alpha_0} H(\alpha_0) = 2u(w_0 - \alpha_0) + Eu(2\alpha_0 - w_0 + \tilde{x}). \quad (8.11)$$

We want to determine the effect of informative signals on the optimal savings decision, which must be made before this information is received. There is a simple intuition for why an earlier resolution of uncertainty might induce a smaller level of savings. Better information allows time for better diversifying the future risk. This implicit reduction in risk provides an incentive for prudent agents to reduce their precautionary savings.

Suppose that the uncertainty is fully resolved at the end of the first period. It then is optimal for the agent to smooth his consumption perfectly over the remaining two periods. The problem at date 0 then becomes

$$\max_{\alpha_0} u(w_0 - \alpha_0) + 2Eu\left(\frac{\alpha_0 + \tilde{x}}{2}\right). \quad (8.12)$$

Its first-order condition for this problem is

$$u'(w_0 - \alpha_0^i) = Eu'\left(\frac{\alpha_0^i + \tilde{x}}{2}\right) \quad (8.13)$$

where α_0^i is the optimal saving of the informed agent, ex ante.

Since the function H is strictly concave in α_0 , the optimal level of saving with a late resolution of uncertainty is larger than α_0^i if and only if $H'(\alpha_0^i)$ is positive. This condition can be rewritten as

$$Eu'(2\alpha_0^i - w_0 + \tilde{x}) \geq u'(w_0 - \alpha_0^i). \quad (8.14)$$

Now, let us make the following two changes in variables: let $z = w_0 - \alpha_0^i$ and let $\tilde{y} = 1.5\alpha_0^i - w_0 + 0.5\tilde{x}$. Conditions (8.13) and (8.14) then can be rewritten respectively as

$$Eu'(z + \tilde{y}) = u'(z) \quad \text{and} \quad Eu'(z + 2\tilde{y}) \geq u'(z). \quad (8.15)$$

Let the function g be defined as $g(k) = Eu'(z + ky) - u'(z)$. We know from (8.15) that $g(0) = g(1) = 0$. Moreover, g will be convex whenever u is prudent. Consequently, $g(2)$ must be positive, as we wished to demonstrate. Thus, a complete early resolution of the income uncertainty at the end of the first period reduces the level of savings by all prudent agents. The result would be the opposite if the agents would be imprudent.

8.3 The Hirshleifer effect

In the first section of this chapter, we explained that information always has a nonnegative value for the decision maker. In making this observation, it was assumed that the information does not affect the other parameters of the environment for the decision maker. In a sense, the information is private, not public. When the information is public, the story is quite different, as we see now in an illustrative example.

Suppose that, as in section 3.2, all risk-averse agents face the idiosyncratic risk of a damage \tilde{x} . Suppose also that there is a competitive insurance market with no transaction cost ($\lambda = 0$) and no asymmetric information. We know that in this case that it is an equilibrium for agents to fully insure their risk at the actuarially-fair premium $E\tilde{x}$. This is a first-best outcome, since all risk-averse agent are fully covered, and the aggregate risk is diversified away by the Law of Large Numbers.

Now suppose that a new technology is introduced that allows both parties to obtain information on \tilde{x} at zero cost. To keep the presentation simple, let us assume that this technology provides perfect information. It will tell everyone – policyholders and insurers alike – who will suffer damage, together with the size of the damage. If insurance markets open only after this information is made available, there is nothing to insure anymore. From the point of view of the insurers, one cannot insure a realized risk. Viewed *ex ante*, this makes everyone worse off, because the possibility to insure at fair price has been eliminated. The value of information is here negative. The cost of information equals the risk premium associated to \tilde{x} , since agents bear risk \tilde{x} rather than its mean $E\tilde{x}$. This is the so-called Hirshleifer effect.

To see the problem here, let us suppose that exactly one half of the population will suffer a damage of size D , and the other half will suffer no damage. With no information, everyone can buy insurance based on a probability of damage of $p = 1/2$. However, this perfect signal tells us that the "truth" is that exactly one-half of the population have $p = 0$ while the other half of the population has $p = 1$. In a sense, the no-information case allows one to buy insurance against being an unfavorable loss type (with $p = 1$). This possibility disappears once we know which individuals will suffer the loss and which will not.

What can we do against this? One possibility would be to organize insurance markets before information is available. The long-term insurance contract would then also insure against bad news. However, it may be difficult to guarantee that no one has the information at the time when the contract is signed. Otherwise, insurers would face the adverse selection problem that we explain later in this book. Also, it may be difficult to implement a system where policyholders obtaining good news will not be able to cancel their contracts. A second and more drastic solution would be to ban this new technology. This would be hard to organize in a globalized world. Also, this is highly likely to be counterproductive, as information has a positive value for organizing prevention more efficiently, for example. One can alter-

natively prohibit insurers the access to the information, but this also create an adverse selection problem – only individuals who knows they will have an accident demand insurance. Finally, one can socialize the risk through some form of compulsory insurance, in the fashion of a social security system or nationalized health care.

Let us debate the issue at a more prosaic level. Consider health insurance and the recent developments in biogenetics. It is expected that in the near future, one will be able to predict the evolution of health for an individual. Even if not 100 percent perfect, the information that a particular individual has a probability of say $p = 0.99$ of developing cancer in the next five years makes insurance a virtual impossibility *ex post*. Of course, suppressing these tests and hence this knowledge might deny the individual access to medical treatments, or to at least planning more optimally for his shorter expected life span. Being able to diagnose future diseases early is undoubtedly a noble project for medical research.. But this might entail the undesired consequence of destroying the basis for health insurance, a crucial source of welfare in our modern societies. This is one of the reasons that genetic testing is such a controversial topic. Other examples abound, such as earthquake insurance, or life insurance. All long-term insurance contracts, where individual risks evolve over time in a Markovian way, can be examined in light of the Hirshleifer argument.

8.4 Bibliographical references and extensions

Hirshleifer and Riley (1992) provide an extensive analysis of the value of information. In this chapter, we limited the analysis to the comparison of two situations, one of the two without any information. Blackwell (1951) was interested in the comparison of any pair of information structures. He raised the following question: under which condition does the second information structure make all decision makers better off than with the first structure, independent of both their attitude towards risk and the decision problem under scrutiny? Blackwell (1951) characterized such a restrictive notion of refined information structure, which is seldom satisfied. Cremer (1982) and Kihlstrom (1984) provided alternative proofs of Blackwell's characterization. Therefore, an increase of risk aversion reduces the value of information, which is the certainty equivalent of the posterior certainty equivalents of the optimal signal-dependent portfolios.

Treich (1997) and Persico (1998) examined the relationship between risk aversion and the value of information particularly for the portfolio decision problem. Persico obtains a positive relation with CARA preferences and normal returns. Treich obtains the same result without assuming CARA, but only in the case of small portfolio risks.

Epstein (1980) provides a general method for the comparative statics analysis of an early resolution of uncertainty. Arrow and Fischer (1974) and Henry (1974) were the first to stress the importance of irreversibility in cost-benefit analysis. The general result that more information induces the selection of a more flexible early action has been widely recognized in the modern theory of real investment. McDonald and Siegel (1984), Pindyck (1991) and Dixit and Pindyck (1994) show that it may be optimal to delay an investment with a positive marginal net present value (NPV) if more information is expected to come about the distribution of future cash flows. Eeckhoudt, Gollier and Treich (2003) derive the conditions under which information reduces precautionary savings.

Hirshleifer (1971) examined the effect of information in competitive markets. Schlee (2001) provides conditions on preferences that guarantee that all agents are made worse off by information in an Arrow-Debreu exchange economy.

References

- Arrow, K.J. and A.C. Fischer, (1974), Environmental preservation, uncertainty and irreversibility, *Quarterly Journal of Economics*, 88, 312-319.
- Blackwell, D., (1951), Comparison of Experiments, in J. Neyman (ed.) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 93-102.
- Cremer, J., (1982), A simple proof of Blackwell's "comparison of experiments" theorem, *Journal of Economic Theory*, 27, 439-443.
- Dixit, A.K., and R.S. Pindyck, (1994), *Investment under uncertainty*, Princeton University Press, Princeton.

- Eeckhoudt, L., C. Gollier and N. Treich, (2003), Optimal consumption and the timing of the resolution of uncertainty, *European Economic Review*, forthcoming.
- Epstein, L.S., (1980) , Decision-making and the temporal resolution of uncertainty, *International Economic Review*, 21, 269-284.
- Henry, C., (1974) , Investment decisions under uncertainty: the irreversibility effect, *American Economic Review*, 64, 1006-1012.
- Hirshleifer, J., (1971), The private and social value of information and the reward to inventive activity, *American Economic Review*, 61, 561-574.
- Hirshleifer, J. and J.G. Riley, (1992) , *The analytics of uncertainty and information*, Cambridge University Press.
- Kihlstrom, R.E., (1984), A Bayesian exposition of Blackwell's theorem on the comparison of experiments, in M. Boyer and R.E. Kihlstrom (eds.) *Bayesian Models of Economic Theory*, Elsevier.
- McDonald, R. and D. Siegel, (1984), The value of waiting to invest, *Quarterly Journal of Economics*, 101, 707-728.
- Persico, N., (1998), Information acquisition in affiliated decision problems, mimeo, UCLA.
- Pindyck, R., (1991) , Irreversibility, uncertainty and investment, *Journal of Economic Literature*, 29, 1110-1148.
- Schlee, E.E., (2001), The value of information in efficient risk sharing arrangements, *American Economic Review* , 91, 509-524.
- Treich, N., (1997), Risk tolerance and value of information in the standard portfolio model, *Economics Letters*, 55, 361-363.

Chapter 9

Optimal prevention

Up to now, our analysis has considered only financial decisions based upon risk. In many circumstances, it may be possible to alter the risk itself. Indeed one reads in the newspaper most every week about some new discovery with regards to genetic engineering. By manipulating DNA codes, scientists hope to one day alter the probabilities of diseases. Of course, we can think of more mundane examples of engineering designed to mitigate risk. Consider a firm, for example, that in addition to fire insurance can use fireproof materials in construction and/or invest in a fire sprinkler system. This type of investment alters the risk distribution itself as opposed to insurance, which simply alters the financing of a risk's consequences. These types of risk-reducing activities in general are often referred to as loss control. Exactly how the distribution is altered might be rather complex. Plus, the engineering devices themselves might entail risks of their own. Perhaps our fireproof building material contains some toxic substance and will suffer the fate of asbestos, leading to long-term liability losses. Or perhaps our sprinkler system will erroneously be triggered even though there is no fire, leading to losses stemming from water damage. In this chapter we look at one type of loss control.

Loss prevention - or "self protection" - is an effort undertaken to reduce the probability of an untoward event. What is the level of effort that would maximize the expected utility u of the representative agent in the economy? In many instances, the cost-benefit analysis of prevention is examined under the assumption of risk neutrality. However, this implies that only reductions in the average size of the loss matter: there is no desire to reduce the variability of losses if this entails a slightly higher average loss. In reality, and most especially for most catastrophic risks like global warming, earthquakes

or epidemic diseases, the assumption of risk neutrality is not reasonable. We should take into account risk aversion in the cost-benefit analysis of preventive actions. For example, potential severe losses due to possible future lawsuits might lead a pharmaceutical firm to conclude that a new drug is not worth bringing to the marketplace at all. The most extreme case of loss prevention is to avoid a risk completely.

9.1 Prevention under risk neutrality

As a base case, let us first consider a risk-neutral agent who faces the risk of losing an amount L with probability p . The agent can invest in preventive measures that reduces the probability of damage. If e is the amount of money invested in prevention, the probability of damage L is $p(e)$. We assume that p is twice differentiable, decreasing and convex function : $p' < 0$ and $p'' \geq 0$. The convexity condition means that the preventive activity exhibits decreasing marginal productivity. The decision problem of the risk-neutral agent is to select e to minimize the the net expected cost of the risk, taking into account of the cost of prevention. This objective may be written as

$$e^n \in \arg \min_{e \geq 0} C(e) \equiv e + p(e)L. \quad (9.1)$$

Because C is convex in e , the first-order condition $C'(e^n) = 0$ is both necessary and sufficient for a minimum. Finally assume that $C'(0) < 0$, so that constraint $e \geq 0$ is not binding.

The optimal preventive investment e^n for the risk-neutral agent, assuming an interior solution, is defined by

$$-p'(e^n)L = 1. \quad (9.2)$$

The left-hand side of the equality is the marginal benefit of prevention. This is simply the reduction in the expected loss generated by one more euro invested in prevention. Thus, equation (9.2) is the classical optimality condition stating that the marginal cost must equal the marginal benefit. Typically, this yields a positive probability of damage ($p(e^n) > 0$), because the full elimination of the risk is usually extremely costly. This is the case, for example, when $p'(e) < 0$ for all e and $\lim_{e \rightarrow \infty} p'(e) = 0$. This is a situation where zero-risk (i.e. avoidance), even if technically attainable, is not economically feasible.

9.2 Risk aversion and optimal prevention

The risk neutrality hypothesis is a good approximation only when the risk is small, or when it can be diversified away by the market. In this section, we consider the more general framework of risk aversion. Consider an expected-utility maximizer who is endowed with wealth w_0 and who faces the risk of losing the amount L with probability $p(e)$. The decision problem now can be written as

$$e^* \in \arg \max_{e \geq 0} V(e) = p(e)u(w_0 - e - L) + (1 - p(e))u(w_0 - e). \quad (9.3)$$

With probability $p(e)$, final wealth is $w_0 - e - L$, otherwise it is $w_0 - e$. Notice that, when u is linear, program (9.1) and (9.3) are equivalent, since $V(e) = a - bC(e)$ for pair $(a, b > 0)$. We want to compare the optimal prevention e^* of the EU-maximizer to e^n , the optimal prevention of the risk-neutral decision maker. It might seem that risk-averse agents should invest more in risk prevention, and risk-loving ones should invest less. We hereafter show that this is not necessarily the case.

Before doing so, however, we need to say a few words about the second order condition, which is not necessarily satisfied even under risk aversion. We have

$$\begin{aligned} V''(e) &= -p''[u(w_0 - e) - u(w_0 - e - L)] \\ &\quad + 2p'[u'(w_0 - e) - u'(w_0 - e - L)] + Eu'', \end{aligned}$$

where $Eu'' = pu''(w_0 - e - L) + (1 - p)u''(w_0 - e)$. The first term is negative, whereas under risk aversion, the second and third terms are respectively positive and negative. It turns out that one cannot guarantee that V is concave without placing additional restrictions on u and p . We hereafter assume that V is concave without examining these restrictions explicitly.

Because we assume that V is concave, e^* will be larger than e^n if and only if $V'(e^n)$ is positive. Evaluating this derivative yields

$$V'(e^n) = -p'(e^n)[u(w_0 - e) - u(w_0 - e - L)] - Eu',$$

where $Eu' = pu'(w_0 - e - L) + (1 - p)u'(w_0 - e)$. Using condition (9.2), we see that $V'(e^n)$ is positive if and only if

$$\frac{u(z) - u(z - L)}{L} \geq p(e^n)u'(z - L) + (1 - p(e^n))u'(z), \quad (9.4)$$

where $z = w_0 - e^n$. This condition can be rewritten as

$$-p(e^n)[u'(z) - u'(z - L)] \leq \frac{u(z) - u(z - L)}{L} - u'(z). \quad (9.5)$$

It is easy to check that the right-hand side of this inequality is positive under risk aversion. Because $[u'(z) - u'(z - L)]$ is negative under the same assumption, we obtain that risk aversion raises the optimal investment in prevention if and only if the probability of loss that is optimal for the risk-neutral agent is smaller than some critical threshold \bar{p} , where

$$\bar{p} \equiv \frac{\frac{1}{L}[u(z) - u(z - L)] - u'(z)}{[u'(z - L) - u'(z)]}.$$

There is in fact a simple intuition for why risk aversion does not necessarily raise the optimal investment in prevention. Introducing any form of risk aversion would always raise prevention only if more prevention yielded a second-order dominant shift in the distribution of final wealth. However, this can never be the case, since more prevention yields a reduction in wealth in the worst case (when damage L occurs). Indeed, prevention reduces wealth in both states of the world. The benefit of prevention comes from making the better of the two states more likely. Sufficiently risk-averse agents will find lowering wealth in the worst state to be extremely painful (in terms of utility loss). Among them is the infinitely risk-averse agent who wants to maximize the minimum final wealth. This agent will *never* invest in loss prevention.

It is easy to check that the critical threshold \bar{p} equals $1/2$ if the utility function is quadratic, i.e., when the degree of prudence is zero. This is also easy to understand. The quadratic agent measures risk by its variance $\sigma^2 = p^n(1 - p^n)L^2$, which has a maximum at $p^n = 1/2$. If p^n is less than $1/2$, an increase in loss prevention reduces both p and σ^2 , which is desirable for risk-averse quadratic agents. Thus, when $p^n < 1/2$, quadratic risk aversion comes to reinforce the willingness to spend effort for prevention. But if p^n is larger than $1/2$, an increase in loss prevention reduces p , but it raises σ^2 . In this case, quadratic risk aversion tends to reduce risk prevention. In the limit case where the risk-neutral agent selects $p^n = 1/2$, the effect on the variance is nil for small changes in prevention. It follows that all quadratic agents will also select $p^* = 1/2$ in this case. But other risk-averse agents without a quadratic utility function may behave differently. The difficulty is that the critical threshold \bar{p} is utility-dependent, and thus varies from agent to agent. In the next section, we examine conditions on preferences that makes \bar{p} smaller or larger than $1/2$.

9.3 Prudence and optimal prevention

Common wisdom would seem to suggest that prudent people should invest more in prevention. This is quite the opposite in fact, as we show now. To isolate the effect of prudence ($u''' > 0$) or imprudence ($u''' < 0$), we limit the analysis to the case where the risk-neutral agents select $p^n = 1/2$. As we have just seen, this implies that risk aversion does not affect the optimal investment in prevention for quadratic utility, since quadratic preferences exhibit zero prudence, $u''' = 0$. In such a situation, would a prudent agent invest more in prevention than the risk neutral agent? Using the analysis from the previous section, more prevention is optimal if condition (9.5) with $p^n = 1/2$ is satisfied, i.e., if

$$\frac{u'(z-L) + u'(z)}{2} \leq \frac{u(z) - u(z-L)}{L}. \quad (9.6)$$

We know that for the agent with quadratic utility, (9.6) is satisfied via an equality. Suppose that the agent is imprudent ($u''' < 0$). Then, using Jensen's inequality for each possible value of the integrand below, we obtain that

$$u(z) - u(z-L) = \int_{z-L}^z u'(x)dx \geq \int_{z-L}^z \left[\frac{z-x}{L}u'(z-L) + \frac{x-z+L}{L}u'(z) \right] dx.$$

Solving the integral yields

$$u(z) - u(z-L) \geq 0.5L [u'(z-L) + u'(z)],$$

which is equivalent to condition (9.6) for e^* to be larger than e^n . In the case of prudence ($u''' > 0$), we obtain the opposite result.

Proposition 24 *Suppose that the risk-neutral agent optimally selects an effort e^n such that the probability of loss p^n is equal to $1/2$. Then, all prudent agents select a level of effort smaller than e^n , whereas all imprudent agents select a level of effort larger than e^n .*

The intuition is that prudence raises the marginal value of wealth, thereby reducing the willingness to consume wealth in order to finance prevention. In other words, the prudent agent has a higher value for precautionary savings, and thus prefers to save more as a protection against loss (and consequently invest less in loss prevention) than an imprudent agent.

9.4 Bibliographical references and extensions

The optimality of loss-prevention activities was first examined by Ehrlich and Becker (1972), who termed such activity "self protection." They also showed that insurance and prevention could be either complements or substitutes. The effects of the potential risks in loss-prevention activities themselves was examined by Briys, Schlesinger and Schulenburg (1991). Dionne and Eeckhoudt (1985) showed that an increase in risk aversion has an ambiguous effect on the optimal level of effort. Briys and Schlesinger (1990) showed that this effect is ambiguous because more prevention does not generate a reduction of risk in the sense of Rothschild-Stiglitz. Jullien, Salanié and Salanié (1999) showed that an increase in risk aversion raises the optimal level of effort if and only if the initially optimal probability of loss is less than a utility-dependent threshold. Chiu (2000) is the first to show that the third derivative of the utility function plays a role in the determination of this threshold. Jewitt (1989) and Athey (2002) examine the effect of risk aversion on the optimal choice for more general decision problems, that include prevention as a special case.

References

- Athey, S., (2002), Monotone comparative statics under uncertainty, *Quarterly Journal of Economics*, 117, 187-223.
- Briys, E., and H. Schlesinger, (1990), Risk aversion and the propensities for self-insurance and self-protection, *Southern Economic Journal*, 57, 458-467.
- Briys, E., and H. Schlesinger and J.-M. Schulenburg, (1990), Reliability of risk management: Market insurance, self-insurance and self-protection reconsidered, *Geneva Papers on Risk and Insurance Theory* 16, 45-59.
- Chiu, W.H., (2000), On the propensity to self-protect, *Journal of Risk and Insurance*, 67, 555-578.
- Dionne, G. and L. Eeckhoudt, (1985), Self-insurance, self-protection and increased risk aversion, *Economics Letters*, 17, 39-42.
- Ehrlich, I., and G. Becker, (1972), Market insurance, self insurance and self protection, *Journal of Political Economy*, 80, 623-648.

- Jewitt, I., (1989), Choosing between risky prospects: the characterization of comparative statics results, and location independent risk, *Management Science*, 35, 60-70.
- Jullien, B., B. Salanié and F. Salanié, (1999), Should more risk-averse agents exert more effort?, *Geneva Papers on Risk and Insurance Theory*, 24, 19-28.

Part III

Risk sharing

Chapter 10

Efficient allocations of risks

This chapter examines the efficient allocation of risk within an economy. Up to now, we have mostly considered individual decisions related to risk and time. In order to focus on the properties of the risk sharing itself, we will assume that there are no transaction costs associated with the transfer of risk. One can easily characterize efficient risk sharing in two particular exchange economies. In the first economy, an infinite number of risk-averse individuals bear an independent and identically distributed risk. Suppose that these individuals create a mutual agreement whereby everyone would transfer his own risk to the pool in exchange for the mean outcome. By the Law of Large Numbers, this arrangement is technically feasible. Furthermore, this outcome is obviously a Pareto-efficient allocation of risk: everyone exchanges a risky wealth for its mean, i.e. all risk-averse individuals are fully insured. Alternatively, consider an economy in which all risk-averse agents have the same attitude toward risk and where they bear the same perfectly correlated risk. In this economy there are no possible gains from risk sharing. Thus autarky is Pareto-efficient, and it is a competitive allocation.

More realistically, agents are not alike and individual risks are hopefully not perfectly correlated, but they are not completely idiosyncratic either. People are not completely alike, as they don't have the same wealth, and it is likely that there is some heterogeneity in their degree of risk aversion. In such an heterogeneous economy, it is essential that agents be allowed to trade their risk. In this chapter, we describe the set of allocations that are Pareto-efficient. We do not address the question of whether decentralized economies are able to attain such desirable allocations of risks. Whether such a market mechanism exists and how it works are topics left for the next chapter.

10.1 Risk sharing: An illustration

Let us go back to the Bernoullian story. Suppose that Sempronius and Jacobus both have a shipment that is expected to come back to London latter this month. Sempronius' boat comes from the East Indies, whereas Jacobus' boat brings spices back from Cuba. There is a common belief that each boat has a probability to be sunk equaling $1/2$, and that these two possible events would be independent. Each owner is expecting a benefit of 8000 ducats in case of success of their own enterprise. Let \tilde{x}_S and \tilde{x}_J be the random benefit of respectively Sempronius and Jacobus. They are independent and distributed as $(0, 1/2; 8000, 1/2)$.

Suppose that Sempronius and Jacobus decide to create a joint venture. The contract stipulates that the two entrepreneurs would bring the spices from their own shipment (conditional to its success) in the warehouse of the new company. The revenue generated by the selling of the spices contained in the warehouse would be spilt equally between the two shareholders. Would this sharing of the individual risk acceptable for bothy parties? The answer to this question is already in chapter 1, where we showed that Sempronius is better off by diversifying his investmpent in two ships rather than in one. This result is independent of the shape of Sempronius' utility function, as long as it is concave. The joint venture has the effect to replace the distribution of Sempronius' benefit from \tilde{x}_S to $(\tilde{x}_S + \tilde{x}_J)/2$. This yields a mean-preserving reduction in risk in the sense of Rothschild and Stiglitz, as shown in section 2.1.4. The symmetric argument can be made for Jacobus. This risk-sharing arrangement, which consists in a risk pooling, is Pareto-improving.

Suppose that Sempronius and Jacobus have mean-variance preferences with

$$U_S = E\tilde{c}_S - 0.5A_S Var(\tilde{c}_S)$$

and

$$U_J = E\tilde{c}_J - 0.5A_J Var(\tilde{c}_J)$$

where \tilde{c}_i is the final benefit of agent i , and A_i is his absolute risk aversion. These preferences can be interpreted as an Arrow-Pratt approximation of the true preferences of the two investors. Under these specifications, the social surplus of the risk-sharing contract presented above equals

$$\frac{4000^2}{4} [A_S + A_J].$$

We now show that the contract presented above is not efficient if the two investors have different degrees of risk aversion. Suppose for example that Sempronius is more risk-averse than Jacobus: $A_S > A_J$. The initial situation is such that the two agents have 50% of the shares of the joint venture. Jacobus can propose to Sempronius to purchase an additional 1% of the firm against a lump sum payment P . What is the minimum price P_{\min} that Sempronius is going to ask for this transfer of property rights? It is given by

$$4000 - (4 \times 10^6 A_S) = 3920 + P_{\min} - (3.84 \times 10^6 A_S),$$

or

$$P_{\min} = 80 - 0.16 \times 10^6 A_S$$

Similarly, the maximum price that Jacobus is ready to pay to get this additional 1% of shares is such that

$$4000 - (4 \times 10^6 A_J) = 4080 - P_{\max} - (4.16 \times 10^6 A_J),$$

or

$$P_{\max} = 80 - (0.16 \times 10^6 A_J).$$

Because A_J is smaller than A_S , we conclude that P_{\max} is larger than P_{\min} . This means that any exchange of this 1% share for a price between P_{\min} and P_{\max} is Pareto-improving. From an initial fifty-fifty sharing of the aggregate risk, it is socially efficient to transfer some of the aggregate risk to the less risk-averse agent.

It is interesting to determine the efficient sharing of the aggregate risk between Sempronius and Jacobus. Consider a contract in which member i would get a share α_i of the total benefit of the products brought back from the colonies by those of the two boats that would come back safely to London. Of course, $\alpha_S + \alpha_J = 1$. Moreover, the contract stipulates that agent i gets a lump sum P_i from the company in addition to his share on the benefit. Because the company has no free cash, it must be that $P_S + P_J = 0$. We determine the feasible contract $(\alpha_S, \alpha_J, P_S, P_J)$ that maximizes the sum of the two agents' ex ante utility $U_S + U_J$. Using conditions $\alpha_J = 1 - \alpha_S$ and $P_J = -P_S$, this sum equals

$$8000 - 16 \times 10^6 [\alpha_S^2 A_S + (1 - \alpha_S)^2 A_J].$$

It is maximized with

$$\alpha_S^* = \frac{T_S}{T_S + T_J},$$

where $T_i = 1/A_i$ is the agent i 's degree of absolute risk tolerance. For example, if Jacobus' risk tolerance is twice as large as Sempronius' risk tolerance, Jacobus should bear two-third of the joint venture risk. The risk pooling and the bias of the sharing of the aggregate risk towards those who are more risk-tolerant are the two characteristics of an efficient risk sharing, as shown in the remaining of this chapter.

10.2 Description of the economy and definition

We consider a simple static exchange economy with n risk-averse agents. Individual i has a twice differentiable increasing and concave utility function u_i , $i = 1, \dots, n$. The description of the uncertainty in the economy is as described in Chapter 5. At the beginning of the period, nobody knows which state of nature will prevail at the end of the period. Agent i is endowed with $\omega_i(s)$ units of the single consumption good in state s . Thus, agent i faces risk if there exists at least one pair of states of nature (s, s') such that $\omega_i(s) \neq \omega_i(s')$. Because he is risk-averse, the agent seeks insurance at an actuarially reasonable price. But he is also willing to accept risk from other consumers if he gets a good expected return on this activity. We assume that there is an agreed-upon probability distribution over the states.

To examine which risk allocations and which risk transfers are socially efficient, we take the broadest view possible and allow for any kind of contingent transfers, a long as they are socially feasible. The set of possible risk contract is complete. An allocation is characterized by n functions $c_1(\cdot), \dots, c_n(\cdot)$, where $c_i(s)$ is the consumption of agent i in state s . An allocation of risk is socially feasible if, in each possible state of nature, the aggregate consumption equals the total wealth $z(s) = \sum_i \omega_i(s)$ available in that state:

$$\sum_{i=1}^n c_i(s) = z(s), \quad \text{for all } s \text{ in the support of } \tilde{s}. \quad (10.1)$$

In Figure 10.1, we describe an economy with two agents and two states of nature by using the Edgeworth box. The edges of the box measure the

total quantity of the consumption good in the two states. Observe that there is some aggregate uncertainty, since there is more to consume in state $s = 0$ than in state $s = 1$. The set of allocations yielding full insurance for agent i is the 45° line starting from the origin of agent i . Because these two lines do not cross, there is no possibility to fully insure the two agents against the macroeconomic risk. We also draw the indifference curve of the two agents that contains their initial endowment. They describe their expected utility in autarchy. In a sense, Figure 10.1 combines two figures 5.2, with the one corresponding to agent $i = 2$ being upside down. Any point in the Edgeworth box describes a feasible allocation of resources.

[INSERT FIGURE 10.1 ABOUT HERE]

An allocation of risk is Pareto efficient if it is feasible and if there is no other feasible allocation that raises the expected utility of one consumer without reducing the expected utility of at least one of the others. In Figure 10.2, we describe the set of all Pareto-efficient allocations of risk in a two-agent two-state economy.

[INSERT FIGURE 10.2 ABOUT HERE]

As is well-known, a Pareto-efficient allocation can be obtained by a maximization procedure. Let $(\lambda_1, \dots, \lambda_n)$ be a vector of positive scalars. Then, the solution to the following program is a Pareto-efficient allocation:

$$\max_{c_1(\cdot), \dots, c_n(\cdot)} \sum_{i=1}^n \lambda_i Eu_i(c_i(\tilde{s})) \text{ s.t. constraints (10.1),} \quad (10.2)$$

where \tilde{s} describes the uncertainty about the state of nature. The proof of this result is easily seen by contradiction. Suppose that the solution to (10.2) is not Pareto-efficient. Then, by definition, there would be another allocation that satisfies constraints (10.1) that would increase $Eu_i(c_i(\tilde{s}))$ for some i without reducing the $Eu_j(c_j(\tilde{s}))$ of the other $j \neq i$. But this would contradict the assumption that the initial allocation maximizes $\sum_i \lambda_i Eu_i$.

It also can be shown that any Pareto-efficient allocation of risk can be expressed as the solution of program (10.2) for some vector $(\lambda_1, \dots, \lambda_n)$. In the following, we take this vector as given. This means that we consider a specific efficient allocation, i.e., a specific point along the contract curve AB in Figure 10.2. But the properties that we examine are common to all efficient allocations.

10.3 Characterization of efficient allocations of risk

An important insight into this problem comes from switching the sum and the expectation in program (10.2) so that it can be rewritten as

$$\max_{c_1(\cdot), \dots, c_n(\cdot)} E \left[\sum_{i=1}^n \lambda_i u_i(c_i(\tilde{s})) \right] \text{ s.t. } \sum_{i=1}^n c_i(s) = z(s), \text{ for all } s. \quad (10.3)$$

This is the consequence of the expected-utility hypothesis, which makes the objective additive with respect to states. Obviously, the optimal solution to this problem can be obtained by solving it state by state, i.e., by solving the following sequence of much simpler programs

$$\max_{c_1(s), \dots, c_n(s)} \sum_{i=1}^n \lambda_i u_i(c_i(s)) \text{ s.t. } \sum_{i=1}^n c_i(s) = z(s), \quad (10.4)$$

for each s . Program (10.4) can be interpreted as a decision problem under certainty in which the group must share a cake of size $z(s)$ in order to maximize a weighted sum of the members' utility. The proof that the optimal solution $c_1^*(\cdot), \dots, c_n^*(\cdot)$ generated by the sequence of cake-sharing programs (10.4) is the optimal solution of program (10.3) can be obtained by contradiction. Suppose that this is not the case: there is a feasible allocation $\hat{c}_1(\cdot), \dots, \hat{c}_n(\cdot)$ other than the one generated by the sequence (??) that yields a larger value for $E[\sum_i \lambda_i u_i(c_i(\tilde{s}))]$. That is obviously not possible, since it would mean that for at least one state of nature s , this alternative allocation would yield $\sum_i \lambda_i u_i(\hat{c}_i(s))$ that is larger than what can optimally be attained in state s with solution $c_1^*(\cdot), \dots, c_n^*(\cdot)$. This is a contradiction. It also is noteworthy that program (10.4) has an objective function that is concave in the decision variables. Therefore, its solution is unique and its first-order conditions together with the feasibility constraint are both necessary and sufficient for efficiency. Observe that the probability distribution of \tilde{s} completely disappeared from the picture to describe the set of efficient allocations of risk.

The first-order conditions for program (10.4) can thus be written as

$$\lambda_i u'_i(c_i(s)) = \mu(s), \quad (10.5)$$

for all $i = 1, \dots, n$, and for all possible s . The Lagrange multiplier μ associated to the feasibility constraint is a function of the parameter s of the

corresponding program. Condition (10.5) can be rewritten as

$$\frac{u'_i(c_i(s))}{u'_i(c_i(s'))} = \frac{u'_j(c_j(s))}{u'_j(c_j(s'))} \quad (10.6)$$

for all pairs (i, j) and (s, s') . The two sides of the above equality correspond to the marginal rates of substitution of consumption in states s and s' , respectively for agents i and j . Condition (10.6) is the classical efficiency condition stating that marginal rates of substitution must be equalized across agents. Graphically, it means that indifference curves in Figure 10.2 must be tangent at an efficient allocation.

The simplification allowed by program (10.4) allows us to more easily characterize efficient risk-sharing allocations. In particular, we will examine in sequence the two main components of efficiency. The first one is the mutuality principle which states that all diversifiable risks must be "washed out" by a mutual arrangement. This principle leads to full insurance for the case of independent and identically distributed risks. The second principle deals with sharing the socially undiversifiable risk. So long as $z(s)$ is not a constant, full insurance for everyone is not feasible.

10.3.1 The mutuality principle

The mutuality principle is easy to obtain from the sequence of state-contingent cake-sharing programs (10.4). What is it that differentiates the program associated with state s from the one associated with state s' ? The only difference comes from the possibility that the aggregate wealth levels $z(s)$ and $z(s')$ are different. Thus, if two states have the same aggregate wealth $z(s) = z(s')$, then the solutions of the programs for states s and s' must be identical: $c_i(s) = c_i(s')$, for all $i = 1, \dots, n$. In words, if there are two states with the same aggregate wealth, then for any consumer, the same amount of the consumption good is optimal in each these two states. The following Proposition states the mutuality principle.

Proposition 25 *A necessary condition for an allocation of risk to be Pareto efficient is that whenever two states of nature s and s' have the same level of aggregate wealth $z(s) = z(s')$, then for each agent i consumption in state s must be the same as in state s' : $c_i(s) = c_i(s')$ for $i = 1, \dots, n$.*

The simplest illustration of this property is when there is no aggregate uncertainty in the economy. When $z(s) = z$ for all s , all agents should enjoy a state-independent consumption plan, i.e. they should be fully insured. In other words, when individual risk can be completely diversified away, it is socially efficient to do so. Quite an obvious recommendation indeed. Figure 10.3 illustrates such a situation. The Edgeworth box is a square and the two 45° lines coincide when there is no macroeconomic risk. The set of all Pareto-efficient allocations corresponds to this line.¹ In autarchy, the risk borne by agent 1 is perfectly negatively correlated with the risk borne by agent 2. This is as if the two agents with an initially sure endowment would gamble against each other on the occurrence of a spot on the sun this year, or on whether a coin will land head or tail. The mutuality principle states that it is Pareto optimal to withdraw the gamble, or equivalently, that an insurer fully covers the risk borne by the two players.

[INSERT FIGURE 10.3 ABOUT HERE]

The mutuality principle however says in fact much more than that. It states that Pareto-efficient consumption of any agent i in state s does not depend directly upon the individual's wealth $\omega_i(s)$ in that state. Rather, it only depends upon the aggregate wealth $z(s) = \sum_i \omega_i(s)$ in that state. This means that if individual endowments were socialized and the n members of the pool gave their endowment to the pool, the planner could achieve an efficient allocation by reallocating the collected wealth according to a rule that does not depend upon who gave how much to the pool. In short, a Pareto-efficient c_i depends upon s only through $z(s)$. Of course, market considerations such as those in the next chapter rely upon more than just Pareto efficiency. For example, if participation in any reallocation scheme is voluntary, we would also need to add a constraint that the agent is at least as well off by participating as he would be by not participating. This additional constraint would restrict the set of Pareto-efficient allocations to the core.

The mutuality principle tells us that risks that can be fully diversified must be completely washed out. In the real world, mutualizing risks does not typically diversify the risk completely. A social risk or macroeconomic risk remains that comes from the fact that the aggregate wealth z is in general not independent of the state of nature. The economy faces random

¹The traditional tangency condition for efficiency is easy to check in this case. Indeed, remember from section 5.3 that along the 45° line, the marginal rate of substitution between $c_i(0)$ and $c_i(1)$ equals p_0/p_1 for the two agents $i = 1$ and 2.

events that generate phases of recession or expansion. The mutuality principle leaves unanswered the question of how to share the undiversifiable risk. The next section deals with the problem of how to share this inescapable type of macroeconomic risk.

10.3.2 The sharing of the macroeconomic risk

Before examining the efficient allocation of the undiversifiable risk, it is noteworthy that program (10.4) shares the same structure as programs (5.3) and (6.5). It should not be surprising then that we will encounter results in this chapter that are quite similar to results that were obtained in previous chapters. Only the context and the terminology will change.

We normalize states in such a way that $z(s) = s$ for all s . From the mutuality principle, this is without loss of generality. Indeed, if we have two states with the same aggregate wealth, the social planner should allocate it in exactly the same way. The first-order conditions for program (10.4) can thus be rewritten as

$$\lambda_i u'_i(c_i(z)) = \mu(z), \quad (10.7)$$

for all $i = 1, \dots, n$, and for all possible z . What is of interest here is the relationship between c_i and z . If c_i is independent of z , this would mean ex ante that agent i would be fully covered against fluctuations of the aggregate wealth z . On the other hand, if $c'_i(z) = 1$, this means that agent i bears all of the macroeconomic risk, and that he insures all other consumers. Thus, the derivative of c_i with respect to z measures the share of the macroeconomic risk born by agent i . Fully differentiating the first-order condition (10.7) yields

$$\lambda_i u''_i(c_i) c'_i(z) = \mu'(z).$$

Eliminating λ_i in this equation using condition (10.7) implies that

$$c'_i(z) = T_i(c_i(z)) \frac{-\mu'(z)}{\mu(z)},$$

where $T_i(c) = -u'_i(c)/u''_i(c)$ is the absolute risk tolerance of agent i . Now, observe that fully differentiating the feasibility constraint $\sum_j c_j(z) = z$ implies

that $\sum_j c'_j(z) = 1$. Combining this with the above equation implies that

$$\frac{-\mu'(z)}{\mu(z)} \sum_{j=1}^N T_j(c_j(z)) = 1. \quad (10.8)$$

Finally, these last two equations together imply that

$$c'_i(z) = \frac{T_i(c_i(z))}{\sum_{j=1}^N T_j(c_j(z))}. \quad (10.9)$$

This is an important result. It is very intuitive. To understand it, let us first digress as to what risk tolerance for an individual describes.

Recall that risk aversion represents the rate of decline in marginal utility that would occur for a one euro increase in wealth. Thus risk tolerance is just the inverse: risk tolerance equals one hundred times the number of euros it would take to yield a 1% decline in marginal utility. The aggregate risk tolerance is the sum of individual risk tolerances, $\sum_{j=1}^N T_j(c_j(z))$. Thus, aggregate risk tolerance is one hundred times the amount of aggregate wealth it would take to make all consumers experience a one-percent drop in marginal utility. Equation (10.9) thus states that the share of the undiversifiable risk borne by each agent equals the share of his own risk tolerance in the aggregate risk tolerance. The larger the risk tolerance of agent i relative to the aggregate risk tolerance of the group, the larger the risk agent i should bear. Notice that condition (10.9), which is a differential equation, must hold for all $i = 1, \dots, n$. Thus, condition (10.9) is in fact a system of differential equations. Solving it requires n initial conditions, say $c_1(0), \dots, c_n(0)$. Once again there is an infinite number of efficient allocations that can be identified either by this set of initial conditions, or by the weighting vector $(\lambda_1, \dots, \lambda_n)$.

Note that (10.9) implies that $c'_i(z) > 0$ for all i if individuals are all risk-averse. This implies that each agent must have more wealth in state s than in state s' whenever $z(s) > z(s')$. This property of Pareto efficiency is known as comonotonicity. Graphically, it implies that the set of Pareto-efficient solutions are in between the two 45° lines in Figure 10.2: everyone should consume more in state 0 than in state 1, since $z(0) > z(1)$. It implies unanimous agreement among all agents as to which state is the best state, which state is the second-best state, and so on. Of course, this is not typically the case in the real world. The state in which you win a lottery is better for you than the state in which your neighbor wins the same lottery, for example.

10.3. CHARACTERIZATION OF EFFICIENT ALLOCATIONS OF RISK 173

However, lotteries and other types of market transactions are not based solely on efficient risk sharing. From a risk sharing perspective, it makes sense that everyone will agree on the ranking of the states.

Several particular cases are worth examining. Suppose for example that there exists one risk neutral agent, say $i = 1$, in the economy: $T_1(c) \rightarrow \infty$. Condition (10.9) directly implies that $c'_1(z) = 1$, and that $c'_i(z) = 0$ for all $i \neq 1$. All risk-averse agents are fully insured against individual risks by this single risk-neutral agent.

Another interesting example is when all agents have a constant degree of absolute risk tolerance (CARA): $T_i(c) = t_i$ for all c . The system of differential equations is here particularly simple to solve, since the right-hand side of (10.9) does not depend upon z . It follows that

$$c_i(z) = c_{i0} + \frac{t_i}{\sum_j t_j} z, \quad (10.10)$$

for some vector (c_{10}, \dots, c_{n0}) of lump-sum transfers, with $\sum_i c_{i0} = 0$.

More generally, the system can be solved analytically if individual risk tolerances are linear in consumption (i.e. HARA utility) with the same slope. Let $T_i(c) = t_i + \alpha c$. Indeed, the system can then be rewritten as

$$c'_i(z) = \frac{t_i + \alpha c_i(z)}{t + \alpha z} \quad (10.11)$$

where we use the feasibility constraint together with $t = \sum_j t_j$. The reader can verify that the solution to this differential equation is simply

$$c_i(z) = c_{i0} + \frac{t_i + \alpha c_{i0}}{\sum_j t_j} z. \quad (10.12)$$

The solution (10.12) is what is known as a linear sharing rule. In the special case where all consumers have the same constant relative risk aversion ($t_i = 0$ so that $T_i(c) = \alpha c$ for all i), condition (10.11) states that $c'_i(z) = z^{-1} c_i(z)$. It follows that each agent i should consume a constant share of the aggregate wealth $c_i(z) = k_i z$, where $\sum_j k_j = 1$. In the Edgeworth box of Figure 10.4, we draw the sets of efficient allocations of risk in these special cases.

[INSERT FIGURE 10.4 ABOUT HERE]

10.4 Aggregation of preferences

Suppose that a group of agents is in a position to implement an efficient sharing of risks among the group's members, characterized by a specific weighting vector $(\lambda_1, \dots, \lambda_n)$. This group faces a decision problem under uncertainty. In this section, we examine the attitude towards risk of this group. If the group is society as a whole, for example, we might need to decide upon some public policy such as how much to invest in global warming problems or how to structure a social security program. If the group is a firm, i.e., a set of shareholders, this might be a capital-budgeting decision as to which projects the firm should undertake. If the group is a household, the problem under scrutiny might be about job selection or the investment in education for the children. How should the fact that risks can be efficiently shared within the group affect the decision making of the group? In other words, can we assign a level of risk tolerance to the group?

As earlier in this chapter, the planner of the group's allocation of the risk has the objective to maximize the weighted sum of the members' expected utility. If this group can decide which risk it should select among different options, it should select the one that maximizes this weighted sum. Let us define function v as

$$v(z) = \sum_{i=1}^n \lambda_i u_i(c_i(z)), \quad (10.13)$$

where $c_1(\cdot), \dots, c_n(\cdot)$ is the efficient allocation associated to the weights $\lambda_1, \dots, \lambda_n$. With this notation, the problem of the planner is just to select the aggregate risk that maximizes $Ev(\tilde{z})$. Observe that v will be increasing and concave if all of the individual utilities u_i are. Hence, the problem for the planner of a group of expected-utility maximizers is itself an expected-utility maximization problem.² In a certain sense, we can think of the planner as individual representing the group, who makes decisions based upon the utility v . For this reason, we often refer to the planner as the "representative agent" for this group.

Observe also that the risk tolerance of the group, i.e. of the representative agent, can be measured by the function T_v , with $T_v(z) \equiv -v'(z)/v''(z)$. We can link the degree of risk tolerance of the representative agent to the

²This is not necessarily the case outside the world of Arrow and Debreu.

distribution of risk tolerances in the group by observing that

$$v'(z) = \sum_{i=1}^n \lambda_i u'_i(c_i(z)) c'_i(z) = \mu(z), \quad (10.14)$$

since $\lambda_i u'_i(c_i(z)) = \mu(z)$ and $\sum_i c'_i(z) = 1$. Thus, we obtain that

$$T_v(z) = -\frac{\mu(z)}{\mu'(z)} = \sum_{i=1}^N T_i(c_i(z)), \quad (10.15)$$

because of condition (10.8). The absolute risk tolerance of an efficient group equals the sum of individual risk tolerances in the corresponding state. Notice that in general, the choice of the weighting vector $(\lambda_1, \dots, \lambda_n)$ will affect the risk tolerance of the group through variations in the allocation of consumption.

Let us consider the special case of HARA utility for the individuals in the group, so that all individual degrees of absolute risk tolerance are linear with the same slope: $T_i(c) = t_i + \alpha c$. Equation (10.15) implies in this case that $T_v(z) = t + \alpha z$. This is an example where the choice of the allocation of risk among the set of all efficient ones, i.e., the choice of $(\lambda_1, \dots, \lambda_n)$ does not matter for determining the attitude towards risk of this group. The inequity in the wealth distribution does not affect the group's willingness to take risk. Under DARA, for example, the larger risk tolerance of wealthier agents just compensates the smaller risk tolerance of the poorer ones. If we assume more specifically that all agents have the same constant relative risk aversion γ , i.e., if $t_i = 0$ and $\alpha = \gamma^{-1}$, we obtain that the representative agent also has a constant relative risk aversion γ .

This is also an example where there is unanimity within the group about the group's attitude to risk. To see this, consider the attitude of agent i to the group's risk \tilde{z} given the transfer $c_i(\tilde{z})$ of this aggregate risk to agent i . This can be measured by the degree of concavity of the implicit utility function v_i , where $v_i(z) \equiv u_i(c_i(z))$ for all z . This is how agent i calculates his final utility in a given state z . Given that

$$v'_i(z) = u'_i(c_i(z)) c'_i(z) = u'_i \left(c_{i0} + \frac{t_i + \alpha c_{i0}}{\sum_j t_j} z \right) \frac{t_i + \alpha c_{i0}}{\sum_j t_j},$$

the degree of concavity of v_i equals

$$-\frac{v'_i(z)}{v''_i(z)} = -\frac{u'_i(c_i) \frac{t_i + \alpha c_{i0}}{\sum_j t_j}}{u''_i(c_i) \left[\frac{t_i + \alpha c_{i0}}{\sum_j t_j} \right]^2} = \frac{t_i + \alpha c_i}{\frac{t_i + \alpha c_{i0}}{\sum_j t_j}} = t + \alpha z = T_v(z).$$

Thus, the attitude towards the aggregate risk of each member of the pool is identical. Therefore, there is unanimity on the management of risk followed by the planner. It happens that these two important properties hold only in this special case of linear risk tolerances with the same slope. For groups where preferences do not satisfy this condition, it is always possible to find a risky choice problem for which some members of the pool disagree on the risk policy followed by the planner.

10.5 Bibliographical references and extensions

Borch (1962) was the first to examine the efficient allocations of risk in an economy of risk-averse agents. Wilson (1968), Rubinstein (1974) and Constantinides (1982) demonstrated the mutuality principle, and he showed how to share the undiversifiable risk in the economy. Leland (1980) and Varian (1985) were more specifically interested in the characterization of efficient risk sharing and collective preferences when consumers have heterogeneous beliefs. Arrow and Lind (1970) discussed whether society should be neutral towards risk. Finally, transaction costs have been introduced into the efficient risk-sharing model by Arrow (1971) and Raviv (1979), as discussed in the chapter on insurance decisions. Townsend (1984) tests the existence of efficient allocations of risk in small villages of developing countries.

References

- Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing*. Chicago: Markham Publishing Co.
- Arrow, K.J., and R.C. Lind, (1970), Uncertainty and the evaluation of public investment decision, *American Economic Review*, 60, 364-378.
- Borch, K. (1962). "Equilibrium in a Reinsurance Market", *Econometrica* 30, 424-444.
- Constantinides, G.M., (1982), Intertemporal asset pricing with heterogeneous consumers and without demand aggregation, *Journal of Business*, 55, 253-67.
- Leland, H.E., (1980), Who Should Buy Portfolio Insurance?, *Journal of Finance*, 35, 581-596.
- Raviv, A. (1979). "The Design of an Optimal Insurance Policy", *American Economic Review* 69, 84-96.
- Townsend, R.M., (1984), Risk and insurance in village India, *Econometrica*, 62, 539-592.
- Rubinstein, M., (1974), An aggregation theorem for securities markets, *Journal of Financial Economics*, 1, 225-244.
- Varian, H., (1985), Divergence of opinion in complete markets, *Journal of Finance*, 40, 309-317.
- Wilson, R. (1968). "The theory of syndicates", *Econometrica* 36, 113-132.

Chapter 11

Asset pricing

In the previous chapter, we assumed that there was a benevolent social planner in society who imposed an efficient allocation of risk. In this chapter, we turn to decentralized decision making by the individuals. Each individual has access to a complete set of competitive markets of Arrow-Debreu securities to exchange risk. Many questions can be raised about the functioning of this economy. Why do these individuals, all with their own self interest in mind, find it worthwhile to trade with each other? How will assets be priced in this market? Can such decentralized decision making generate a Pareto-efficient allocation of risk? Can this model explain the large equity premium that has been observed in developed countries over the last century? Is it compatible with the observed low return on bonds over the same period? What is the shape of the yield curve in this economy? In this chapter, we will attempt to provide some answers to these important questions.

11.1 Competitive markets for Arrow-Debreu securities

In this section, we reconsider the economy that was defined in the past chapter by introducing markets for risk exchanges. There is some uncertainty about the state \tilde{s} that will occur at the end of the period. We have n risk-averse agents, $i = 1, \dots, n$, and agent i is endowed with an initial state-dependent wealth claim $\omega_i(\tilde{s})$. As in chapter 5, we assume that agents can trade risk within a complete market. Let $\pi(s)$ denote the price per unit of probability of the Arrow-Debreu security associated with state s . If $p(s)$ de-

notes the probability of state s , this means that one must pay $p(s)\pi(s)$ to obtain one unit of the consumption good if and only if state s occurs.¹ The decision problem of agent i is to find a portfolio $c_i(\cdot)$ of Arrow-Debreu securities that maximizes his expected utility under a budget constraint. This can be written as

$$\max_{c_i(\cdot)} Eu_i(c_i(\tilde{s})) \quad (11.1)$$

subject to

$$E\pi(\tilde{s})c_i(\tilde{s}) = E\pi(\tilde{s})\omega_i(\tilde{s}). \quad (11.2)$$

Equation (11.2) is the budget constraint, which states that the value of the selected portfolio cannot exceed $E\pi(\tilde{s})\omega_i(\tilde{s})$, the market value of individual i 's initial endowment. As was examined in chapter 5, the first-order condition for this program can be written as

$$u'_i(c_i(s)) = \xi_i\pi(s), \quad (11.3)$$

for all s . Solving equations (11.2) and (11.3) yields demand functions $c_i = C_i(\pi)$ whose main properties have been described before.

To complete the description of the model, we need to add the market clearing condition $\sum_i [c_i(s) - \omega_i(s)] = 0$, or equivalently,

$$\sum_{i=1}^n c_i(s) = \sum_{i=1}^n \omega_i(s) = z(s), \quad (11.4)$$

for all s . Condition (11.4) simply states that the aggregate consumption in state s cannot exceed what is available in that state, which is denoted $z(s)$. This condition is, in fact, the feasibility condition.

It is fair to recognize the main weakness of the complete-markets model which is to assume that all individual risks can be traded on financial markets. More realistically, there are many risks that are not tradeable. Labor income risk typically cannot be traded. It is very difficult to transfer all of the risk associated with real estate, an important fraction of the households' wealth. More recently, we have found a lack of tradeability for much of the

¹Recall from section ?? that the price for one unit of wealth contingent upon the occurrence of state i was defined as $\Pi(s) \equiv p(s)\pi(s)$. The set of prices defined per unit of probability, $\{\pi(s)\}$, is referred to as the pricing kernel for this economy.

risk associated with terrorism. If all risks were indeed tradeable, this would imply that the total capitalization of financial markets, i.e., the market value of all assets traded, must equal the total wealth of the economy, which is very far from reality, even in the U.S. economy. Still, the assumption that markets are complete remains a cornerstone of modern asset-pricing theory.

11.2 The first theorem of welfare economics

All economists have somewhere in their background knowledge the essential result that, under certain conditions, a competitive allocation is Pareto-efficient. These conditions are the absence of externalities and of asymmetric information, the completeness of market, and some additional technical conditions on preferences that are fulfilled for risk-averse expected-utility maximizers. So, it is no surprise that this result also holds in the economy presented here. This can be seen directly by comparing conditions (10.7) and (11.3), as we explain below.

Let us start with the mutuality principle. We will take some liberty in notation and use "s" to denote both the state of nature and the Arrow-Debreu security associated with that state of nature. We have seen in chapter 5 that the individual demand for a specific Arrow-Debreu security s depends only upon its relative price $\pi(s)$. Moreover, this individual demand is decreasing in π . Thus, if there are two states s and s' with the same price $\pi(s) = \pi(s')$, every agent must have the same demand for asset s and asset s' , and thus the same consumption in these two states: $c_i(s) = c_i(s')$. Consequently, the aggregate consumption levels must be the same, which is possible only if $z(s) = z(s')$. In other words, if there are two states with the same aggregate wealth, the competitive state prices must be the same, and for any arbitrary i , agent i will consume the same amount of the consumption good in these two states. Thus, individual consumption at equilibrium and competitive state prices depend upon the state only through the aggregate wealth available in the corresponding state. This implies that the competitive allocation of risk satisfies the mutuality principle, and that all diversifiable risk is washed out in equilibrium. As in the previous chapter, we hereafter denote $z(s) = z$, for all s .

Let us rewrite condition (11.3) as

$$\xi_i^{-1} u'_i(c_i(z)) = \pi(z). \quad (11.5)$$

This is equivalent to the first-order condition (10.7) for an efficient allocation of risk by taking $\xi_i^{-1} = \lambda_i$ for all i , and $\mu(\cdot) \equiv \pi(\cdot)$. We conclude that competitive markets allocate the macroeconomic risk in a Pareto-efficient manner: decentralized decision making yields efficient sharing. In the Edgeworth box of Figure 11.1, we draw the competitive equilibrium of an economy with two agents and two states of nature. We see that the competitive equilibrium lies on the curve AB of Pareto-efficient allocations. Moreover, both consumers are better off at the competitive equilibrium than in autarchy.

Recall that there is a "representative agent" for each efficient allocation of risk. This implies, in particular, that the attitude towards risk of any economy implementing a Pareto-efficient allocation can be duplicated by an economy with a single expected-utility maximizing agent with concave utility function v . From the first theorem of welfare economics, we know that this must also be true in the special case of the competitive allocation.

The existence of a representative agent simplifies the analysis. From (10.14), the marginal utility of the representative agent in a given state equals the equilibrium state price itself: $v'(z) = \mu(z) \equiv \pi(z)$. From condition (10.15), recall that the risk tolerance of v evaluated at aggregate wealth z equals the sum of individual risk tolerances in that state:

$$T_v(z) = \sum_{i=1}^N T_i(c_i(z)). \quad (11.6)$$

We are now in a position to explore some characteristics of asset prices in such an economy.

11.3 The equity premium

In this section, we want to compare the prices and expected returns of two particular portfolios. The first portfolio is the risk-free portfolio, i.e., a portfolio that yields one unit of the consumption good in all states. We normalize its price to $E\pi(\tilde{z}) = 1$. This means that the risk-free rate in this economy is zero. Because our model is static at this stage, this is without loss of generality. Because we know that the price kernel $\pi(\cdot)$ is linked to the preferences of the representative agent through $\pi(\cdot) = v'(\cdot)$, this condition normalizes the representative utility function in such a way that $E v'(\tilde{z}) = 1$.

The second portfolio provides a constant share α of the aggregate wealth in the economy. This means that the portfolio yields αz units of the con-

sumption good in state z , for all z . Because the entire individual wealth is subject to trading, this portfolio can be interpreted as a fully diversified mutual fund of all assets of the economy. The price of this equity portfolio is equal to the sum of the values of the assets that it contains, i.e., it equals $E\alpha\tilde{\pi}(\tilde{z})$. Its expected payoff is $E\alpha\tilde{z}$. The expected return, which is often called the return on equity, equals

$$\phi = \frac{E\tilde{z}}{E\tilde{\pi}(\tilde{z})} - 1 = \frac{E\tilde{z}}{E\tilde{z}v'(\tilde{z})} - 1, \quad (11.7)$$

The difference between ϕ and the risk-free rate is called the equity premium. This is the amount of excess return (i.e. return in excess of the risk-free return) that is earned on average by holding the risky mutual fund. Because the risk-free rate here is assumed to be zero, ϕ is at the same time the return on equity, as well as the equity premium.

Because agents are risk-averse, they will not invest in the stock market at all unless the average equity return is larger than the risk-free rate in the economy. But for this exchange economy to be in equilibrium, it is necessary that the entire aggregate risk \tilde{z} be borne by individuals, through their assets holding. This is stated in the feasibility (or market-clearing) condition (11.4). Thus, the equity premium must be larger than the risk-free rate, large enough to induce voluntary assumption of the entire aggregate risk. But how large is large enough? Suppose that the aggregate consumption was z_0 in the previous period, and that the aggregate consumption at the end of the current period is a random variable \tilde{z} . Because each period is very short, \tilde{z} is a small risk around z_0 . Because the risk free rate is zero, it must be that $1 = Ev'(\tilde{z}) \simeq v'(z_0)$. Using a first-order Taylor approximation similar to the one used to obtain equation (4.4), we derive that

$$E\tilde{z}v'(\tilde{z}) \simeq z_0v'(z_0) + z_0^2\sigma^2v''(z_0) \simeq z_0 [1 - \gamma\sigma^2],$$

where $\sigma^2 = Var(\tilde{z}/z_0)$ is the variance of the growth rate of aggregate consumption, and $\gamma = -z_0v''(z_0)/v'(z_0)$ is the relative risk aversion of the representative agent. Suppose that all agents in the economy have the same constant relative risk aversion γ . We know from the previous chapter that this implies that the representative agent also will have relative risk aversion γ . Using these approximations, we can rewrite equation (11.7) as

$$\phi \simeq \frac{z_0}{z_0 [1 - \gamma\sigma^2]} - 1 \simeq \gamma\sigma^2. \quad (11.8)$$

The variance σ^2 of the yearly growth rate of GDP per capita has been around 0.0006, or 0.06% over the last century in the United States.² Thus, for empirically feasible levels of relative risk aversion between 1 and 4, the equity premium should be somewhere in between 0.06% and 0.24% per year.

Is this prediction about how risks should be priced in our economy compatible with the data? Let us remind some historical data on financial returns, as those presented by Kocherlakota (1996) and documented in Chapter 4. They provided statistics on asset returns for the U.S. over the period from 1889 to 1978. We concluded that the observed equity premium has been equal to 6% over that period. This important difference between the value predicted by the model (less than one-fourth of a percent per year) and the observed value of 6% per year is what economists refer to the equity premium puzzle.³

This makes a huge difference in the long term. For example, Ibbotson Associates showed that one dollar invested in the stock of large US companies in 1925 would have yielded a portfolio value of 2351 dollars in 1998, assuming that all earned dividends are reinvested during this period. Compare this to the same dollar invested in long term bonds over this period. The 1998 value of this bond would be only 44 dollars, including reinvested interest. True, holding stocks has been risky, in particular in the first half of the century, as shown in Figure 11.2. It has been shown that all investors with an investment horizon of 20 years always earned more from their portfolio by investing in stocks than by investing in bonds, independent of the time they entered the market during the century. Because of the small size of the sample, this does not prove first-order stochastic dominance, however. Indeed, it may be possible that we experienced luck during the period, with no big, long downturns of the economy, in spite of the two world wars. We see in Figure 11.3 that other developed countries experienced similar patterns in their financial markets.

[INSERT FIGURE 11.2 ABOUT HERE]

[INSERT FIGURE 11.3 ABOUT HERE]

The level of relative risk aversion consistent with U.S. data typically falls

²We consider the GDP per capita rather than the GDP to take into account of the important growth rate of the U.S. population during the period.

³Not all empirical evidence is in agreement however. For example, evidence by Jorion and Goetzmann (1999) shows that the equity premium might not be quite as high as is typically supposed. Thus, there may be an empirically smaller “puzzle” that actually needs to be rationalized by the theory.

at unrealistically high levels, such as 15 or 40. With relative risk aversion equal to 20, for example, a person with 100 dollars of initial consumption would have a certainty equivalent of less than three and one-half dollars for a lottery ticket paying a prize of 100 dollars with a probability of $p = 0.5$. There has been an enormous effort over the last 15 years to understand and to solve the equity premium puzzle. Several explanations have been proposed for this high equity premium: uninsurability of labor risks, the peso problem, habit formation, consumption externalities, violations of expected utility, liquidity constraints, transaction costs, participation costs, inequity in the distribution of wealth, persistence of pessimism, and many others. It appears that there is no single explanation of the puzzle, in particular if we want to solve it jointly with the other puzzle that we will see later in this chapter, namely, the risk-free rate puzzle.

11.4 The capital asset pricing model

In the previous section, we priced a very specific portfolio, i.e., the fully diversified portfolio consisting a fixed percentage claim on the aggregate equity available in the economy. We now examine the pricing of specific assets within this economy. Before going to the details, we want first to stress an important characteristic of asset pricing when markets are complete. Consider a risky asset, i.e., an asset whose value depends upon the state of nature that will prevail at the end of the period. Suppose first that this risk is statistically independent of the market risk \tilde{z} . Suppose for example that, whatever the GDP z , the firm generates a value q with probability p , and zero otherwise. What should its market value be ex ante? Because of the mutuality principle, we know that the state prices depend only upon z , not upon the success of the specific firm. The ex ante value of the firm is thus equal to $E[\pi(\tilde{z})(qp + 0(1 - p))] = qp$, since we normalized $E\pi(\tilde{z})$ to unity. This means that individual risks that are not correlated to the market risks are actuarially priced. This result is a direct consequence of the fact that risk aversion is a second order effect in the expected utility framework. Adding a small share ε of independent risk $(q, p; 0, 1 - p)$ to the market portfolio has the same effect on welfare than increasing wealth by $\varepsilon pq - 0.5\varepsilon^2 p(1 - p)q^2 A$, where A is absolute risk aversion. If ε is small, investors will value this investment at εpq , which is actuarially fair.

Observe that the actuarial fair pricing of diversifiable risks in the economy

implies that all agents will fully insure against diversifiable risks in equilibrium. This is compatible with Pareto efficiency.

Let us now consider an asset whose return may be correlated to the aggregate risk. A risky asset is fully characterized by the payoff that it provides to its owner in each possible state of nature. Consider an asset q whose payoff in state z is $q(z)$. From the mutuality principle, we only need to know the relationship between z and q . If factors other than z influence the firm's final value, $q(z)$ represents the expected payoff of the firm conditional on z . The market value of this firm ex ante equals

$$P(q) = Eq(\tilde{z})\pi(\tilde{z}) = Eq(\tilde{z})v'(\tilde{z}). \quad (11.9)$$

This is the asset pricing formula common to all models with complete markets. Using no-arbitrage arguments, we just need to know function $\pi \equiv v'$ to price any asset on the market. This is why the function π is often called the pricing kernel.

As in the previous section, one can obtain an approximation formula by assuming that \tilde{z} is a small risk around some z_0 : $\tilde{z} = z_0 + k\tilde{\varepsilon}$, where we now assume that $E\tilde{\varepsilon} = 0$. By normalizing the risk-free rate to zero, so that $v'(z_0) = 1$, it follows that

$$P(q) \simeq Eq(\tilde{z}) - \frac{\gamma}{z_0} cov(q(\tilde{z}), \tilde{z}), \quad (11.10)$$

where $\gamma = -z_0 v''(z_0)/v'(z_0)$ is the relative risk aversion of the representative agent measured at the expected level of aggregate wealth, and $cov(q(\tilde{z}), \tilde{z})$ is the covariance of $q(\tilde{z})$ and \tilde{z} . This is a central equation for understanding the Capital Asset Pricing Model, or "CAPM." It tells us that the current market value of an asset equals its future expected payoff minus a risk premium which takes the form of the product of risk aversion and a measure of the asset risk. This is reminiscent of the Arrow-Pratt formula. But the asset risk is here measured by the covariance of the future asset value and the aggregate wealth in the economy, not the variance of $q(\tilde{z})$. This is again the direct consequence of the mutuality principle. If the asset value is not correlated to the market value, the asset will be actuarially priced. Notice that the risk premium becomes negative for assets that are negatively correlated with aggregate wealth. The intuition is simple. These are assets whose integration in the market portfolio reduces the risk of the portfolio. This hedging property raises their attractiveness for risk-averse agents, and consequently increases

their market values to levels above their expected payoffs in equilibrium. In other words, these assets will have a rate of return less than the risk-free rate.

The statistical relationship between the asset return and the market return is often measured by the so-called beta of the asset, which is defined by

$$\beta(q) \equiv \frac{\text{cov}(q(\tilde{z}), \tilde{z})}{\sigma_{\tilde{z}}^2}. \quad (11.11)$$

If we consider the special case where $q(\tilde{z}) = \tilde{z}$, equation (11.10) yields the price of our market portfolio:

$$P_M \simeq E(\tilde{z}) - \frac{\gamma}{z_0} \sigma_{\tilde{z}}^2. \quad (11.12)$$

Solving (11.12) for γ/z_0 and using this together with (11.11) in equation (11.10) above yields

$$Eq(\tilde{z}) - P(q) = \beta(q)[E(\tilde{z}) - P_M]. \quad (11.13)$$

This equation is essentially a modified version of the so-called security-market line from the CAPM.⁴

The beta of the market portfolio, of course, is unity. Fully diversifiable assets have a zero beta. The following table provides the estimated beta of a few existing assets:

firm	β	firm	β
Moulinex	1.8	LVMH	0.8
Renault	1.6	British Telecom	0.7
HSBC	1.5	Glaxo Wellcome	0.6
Royal and Sun	1.3	Carrefour	0.5
Pechiney	1.2	L'Oréal	0.4

⁴If we concern ourselves with returns, rather than with prices, then we can assume that one share of each security has a price of 1, with the payoffs expressed now as the gross earnings per share. In this setting, equation (11.13) tells us that the return on asset q equals its beta times the expected return on the market portfolio. This is precisely the security-market-line equation whenever the risk-free rate is equal to zero.

11.5 Two fund separation theorem

One can reexamine the equilibrium allocation of risks in this framework. In equilibrium, people select portfolios whose final values are comonotonic, as they depend in an increasing way upon the realization of a single random variable \tilde{z} . This dependence can be quite complex, however. But in the special case of linear risk tolerances $T_i(c) = t_i + \alpha c$, we know that this relationship is linear. Linearity has an important consequence in a decentralized economy. It implies that it is sufficient to limit asset supply to two funds. There would be a fund offering a risk-free portfolio in zero net supply. In addition, there would be another fully diversified fund of stocks. It would be a fund whose unique aim is to duplicate the market performance z , state by state. By allowing people to sell all their risk to the second fund, and to purchase shares of the two funds, one can obtain any consumption plan that depends linearly upon z . Thus, one can duplicate the equilibrium allocation with just these two funds. We do not need to organize complete contingent markets. This is the two-fund separation property, which holds under linear absolute risk tolerances. It is also sometimes referred to as a "mutual-fund theorem," since it implies that all investors only require one mutual fund of risky stocks, together with a risk-free asset, for investment purposes. As said before, linear risk tolerance is in fact the only case within an expected-utility framework where the two-fund separation property holds.

[INSERT FIGURE 11.4 ABOUT HERE]

Many investors however do not want linear consumption plans, such as the investors $i = 1$ and 2 whose optimal portfolios are illustrated in Figure 11.4 for example. In this case, agent 1 purchases what is known as portfolio insurance. This means that he gets a minimum guaranteed income from his portfolio. This can be done by buying put options whose underlying asset value is the mutual fund. Intuition might suggest that the most risk-averse agents would purchase that kind of instrument at equilibrium for much the same reason that risk averters prefer deductible insurance policies. However, this is not true when the two-fund separation property holds, i.e., when risk tolerances are linear with the same slope. More risk-averse investors will rather purchase more of the risk free asset. To sum up, portfolio insurance seems attractive for risk-averse investors. However, it should be reminded that to each investor purchasing portfolio insurance, there must be someone on the other side of the market that would be ready to sell portfolio insurance, a very risky business. At equilibrium with HARA preferences, the price of

portfolio insurance would be too large for that anyone would want to purchase it.

Still, it is possible to understand why such financial instruments are often traded in real-world markets. First, they are useful for helping agents to get rid of their individual risks that cannot be diversified away by the market. Indeed differences in these individual risks alone can be enough to cause some investors to sell options, while others buy options. Second, agents may have risk tolerances that are not linear, or which are linear but not with the same slope. Finally, contrary to what we implicitly assume here, people often form different beliefs on the probability distribution of the states of nature. One analyst's "hot stock" recommendation may be another analyst's "dog".

11.6 Bond pricing

We normalized the risk-free rate to zero in our analysis above. In fact, because there is only one consumption date in the model, the notion of an interest rate has no meaning. We need to introduce saving in the model to discuss interest rates. It is easiest to extend the above model to two consumption dates. We do this in the next two sections. We then show how this framework easily extends to multiple periods.

11.6.1 The risk-free rate

Agent i is now endowed with a fixed wealth of ω_{i0} units of the consumption good at date 0, and with the state contingent claim for $\omega_i(s)$ units at the second date in state s , for all s . We denote aggregate wealth at date 0 as $\omega_0 = \sum_i \omega_{i0}$. As in chapter 6, it is assumed that the agent maximizes the discounted value of the flow of expected utility over his lifetime, i.e., he maximizes $u_i(c_{i0}) + \beta E u_i(c_i(\tilde{s}))$, where c_{i0} is consumption at date 0, $c_i(s)$ is consumption in date 1 in state s , and β is the discount factor.⁵

At date 0, consumers face both a saving decision and a portfolio decision. The numeraire is the consumption good. It costs $\pi(s)$ (per unity of probability) units of the consumption at date 0 to purchase a contract that guarantees the delivery of one unit of the good at date 1 if and only if state s occurs. Let $r = [E\pi(\tilde{s})]^{-1} - 1$ be the risk-free rate in the economy. Indeed, it

⁵The notation for β as used here is not to be confused with the beta from the CAPM.

costs $E\pi(\tilde{s}) = (1+r)^{-1}$ at date 0 to get one unit of the good with certainty at date 1. The decision problem of agent i can now be written as

$$\max_{c_{i0}, c_i(\cdot)} u_i(c_{i0}) + \beta E u_i(c_i(\tilde{s})) \quad (11.14)$$

subject to

$$c_{i0} + E\pi(\tilde{s})c_i(\tilde{s}) = \omega_{i0} + E\pi(\tilde{s})\omega_i(\tilde{s}). \quad (11.15)$$

The market clearing conditions are $\sum_i c_{i0} = \omega_0$ together with condition (11.4). We see that the additive nature of the problem is not affected by introducing an initial saving/consumption decision. Therefore, all properties that we examined previously, such as the mutuality principle and the sharing rule of the macro risk, still hold in this dynamic model. The only difference is due to the fact that we cannot normalize the risk-free rate to zero, since the numeraire has already been fixed to be the consumption good. The first-order conditions for this problem are given by $\beta u'_i(c_i(s)) = \xi_i \pi(s)$ together with $u'_i(c_{i0}) = \xi_i$ and the constraints. One can eliminate the Lagrangian multiplier ξ_i to write

$$\beta u'_i(c_i(s)) = u'_i(c_{i0})\pi(s) \quad \forall s. \quad (11.16)$$

Condition (11.16) states that the marginal gain in expected utility from having an extra date-1 unit of consumption in state s exactly equals the utility loss of financing this via current consumption. Since this is true for all states s , this condition implies that there is no possible gain for switching contingent consumption between the states at date 1, or for switching consumption between date 0 and date 1. This condition is often called the Euler equation.

In order to characterize the risk-free rate in this economy, it is necessary to use the representative agent. Because of the additive nature of the objective in (11.14), and because markets are complete, there must be a representative agent who behaves towards risk and time exactly as the economy as a whole at the competitive equilibrium. Let v be the utility function of this representative agent. Preferences for our representative agent are fully characterized by the agent's risk tolerance as described in equation (11.6). As in (11.16), the first-order condition for the decision problem of the representative agent is written as

$$\beta v'(c(s)) = v'(c_0)\pi(s) \quad \forall s, \quad (11.17)$$

where $(c_0, c(\cdot))$ denotes the optimal consumption plan of the representative agent. Obviously we cannot interpret this condition in the same way as we did for (11.16). The representative agent consumes all wealth available at every date and state. It is thus the prices that must be adjusted correctly in (11.17). In other words, in order to have an equilibrium, we need in addition that $c_0 = z_0$, and $c(s) = z(s)$ for every state s . Combining all this, we get the equilibrium conditions

$$\pi(s) = \frac{\beta v'(z(s))}{v'(z_0)} \quad \forall s. \quad (11.18)$$

By taking the expectation, we obtain that

$$E\pi(\tilde{s}) = \frac{\beta E v'(\tilde{z})}{v'(z_0)}.$$

It thus follows that

$$r = \frac{v'(z_0)}{\beta E v'(\tilde{z})} - 1 = \frac{v'(z_0)}{\beta E v'(z_0(1 + \tilde{x}))} - 1 \quad (11.19)$$

where \tilde{x} is the growth rate of the economy. This is a new important pricing formula which allows us to determine the current market value of any asset whose future payoff is certain. The equilibrium risk-free rate r depends upon the rate of impatience β , on the concavity of the representative utility function, on the current level of aggregate wealth z_0 , and on the distribution of future growth \tilde{x} . We examine this relationship in more detail in the next subsection.

11.6.2 Factors affecting the interest rate

First of all, consider the case where future growth is identically zero $\tilde{x} \equiv 0$. If $\delta = \beta^{-1} - 1$ denotes the rate of pure preference for the present, then the equilibrium risk-free rate of interest will equal δ , $r = \delta$. Let $B_1 = (1 + r)^{-1}$ denote the price of a pure-discount bond, also called a zero-coupon bond, paying one unit of consumption with certainty at date 1. Thus, with zero growth in the economy, $B_1 = \beta$. If the expected growth rate is positive but certain, $\tilde{x} = g > 0$, it follows from (11.19) that we must have $B_1 < \beta$, i.e. $r > \delta$. This is because agents are averse to consumption fluctuations over time. Therefore, in order to induce them to accept a current consumption

low enough to be compatible with current aggregate wealth, a larger interest rate is necessary. The larger the growth rate g , the larger the equilibrium risk-free rate. We call this the wealth effect. This explains why interest rates tend to go up during booms and down during recessions.

To illustrate, consider the case where individuals all have constant relative risk aversion γ . Then we already seen that this is also true for the representative agent. We can thus write the zero-coupon bond price as

$$B_1 = \frac{\beta E v'(z_0(1 + \tilde{x}))}{v'(z_0)} = \beta E(1 + \tilde{x})^{-\gamma}. \quad (11.20)$$

From (11.20) we can confirm the two results above, that $B_1 = \beta$ when $\tilde{x} \equiv 0$ and $B_1 = \beta(1 + g)^{-\gamma} < \beta$ when $\tilde{x} \equiv g > 0$. Moreover, note that B_1 is decreasing in the degree of risk aversion γ in the case where $\tilde{x} \equiv g$. A higher γ lowers the price of a zero-coupon bond, i.e. increases the equilibrium risk-free rate of interest. Since there is no risk at date 1 and since g is positive, we know that consumers consume more at date 1 than at date 0. An increase in the concavity of v would cause the ratio $v'(z_0(1 + g))/v'(z_0)$ to fall. Thus, consumers would find it advantageous to shift some of their consumption from date 1 to date 0; they would purchase fewer zero-coupon bonds. Of course this not possible in equilibrium. Hence, the price of these bonds would need to fall to maintain our equilibrium.

Suppose now that \tilde{x} is random with $E\tilde{x} \equiv 0$: there is zero growth on average but we may experience either expansion or recession with positive probabilities. Since constant relative risk aversion implies constant relative prudence $P^r = 1 + \gamma$, we know that $E v'(\tilde{z}) = v'(E\tilde{z} - \psi)$, where ψ denotes the precautionary premium for \tilde{z} . It follows that $B_1 = \beta(1 - \psi)^{-\gamma} > \beta$. When the future growth of the economy is uncertain, prudent people tend to save more due to their precautionary savings demand. At a fixed current aggregate wealth z_0 , this cannot be an equilibrium. Therefore, the interest rate must be reduced in order to lower the demand for savings. This explains why the interest rate tends to go down when more uncertainty accumulates. We call this the precautionary effect.

If we have a positive rate of growth on average, $E\tilde{x} \equiv g > 0$, but this rate of growth is risky, we see that the wealth effect (to shift consumption from date 1 to date 0) and the precautionary effect (to shift wealth from date 0 to date 1) work in opposite directions: $B_1 = \beta(1 + g - \psi)^{-\gamma}$. We cannot tell a priori which of these effects might dominate. Moreover, the

effect of an increase in γ is also ambiguous, since it also will also increase the precautionary premium ψ .

If we do not have constant relative risk aversion, of course the analysis is even more complex. However, if we assume that both the expected growth rate g and the level of risk are small, we can approximate the risk-free rate as ⁶

$$r \simeq \delta + R[g - 0.5\sigma^2 P^r], \quad (11.21)$$

where σ^2 is the variance of growth rate of aggregate consumption, R is relative risk aversion and P^r is relative prudence, both evaluated at the expected future wealth $\mu = z_0(1 + g)$. Here we see the various trade-offs between the average growth rate and its riskiness in determining the equilibrium risk-free rate of interest. Since any change in risk aversion necessitates a change in relative prudence, we should be cautious about predicting the effect about a change in risk aversion.

One can try to bring this theory to the data. The expected growth rate of the US economy has been around 2% during the 20th century. So, the consumption smoothing term gR alone generates a risk-free rate between 2 to 8 percent per year, assuming a measure of relative risk aversion somewhere between 1 and 4. It is even larger if we add the effect of impatience. This is a problem, however, since the mean real interest rate of the US economy during this period has been much lower, around 1%. This illustrates the risk-free rate puzzle. One could hope that the precautionary effect helps in explaining this phenomenon, but this is not the case. The variance of the growth rate of the economy has been relatively small, around 0.0006. This implies that one would need a very large degree of relative prudence to solve the puzzle. Observe that when relative risk aversion is a constant γ , relative prudence is a constant $\gamma + 1$. So, even with a large level of relative risk aversion $\gamma = 4$, the precautionary term $0.5\sigma_x^2 R P^r = 0.5\sigma_x^2 \gamma(1 + \gamma)$ equals $0.006 = 0.6\%$ per year, a very small impact with respect to the consumption smoothing term $gR = 8\%$. Hence, the empirical risk-free rate remains a puzzle from a theoretical point of view.

⁶Let $\mu = z_0(1 + g)$. Since the risk is small, $Ev'(z_0(1 + \tilde{x}))$ is approximately equal to $v'(\mu)[1 + 0.5z_0^2\sigma_x^2 A(\mu)P(\mu)]$, where $A(c) = -v''(c)/v'(c)$ and $P(c) = -v'''(c)/v''(c)$ are respectively absolute risk aversion and absolute prudence of the representative agent. Since the expected growth is small, we can approximate $v'(z_0)$ by $v'(\mu)[1 + z_0gA(\mu)]$. We also use the approximation $(1+y)^{-1} \simeq 1-y$ for small y . Finally, we approximate $R = z_0(1+g)A(\mu)$ by $z_0A(\mu)$. This approximation can be shown to be exact in a continuous time framework.

11.6.3 The yield curve

Adding a longer time horizon to our model is not difficult. Up to now, we were interested in determining the risk-free rate for cash flows occurring within the next 12 months. We can perform the same exercise for cash flows obtained in 24 months. There is no reason a priori to believe that the risk-free rate will be independent of the time horizon. In reality, they differ. Both the expected growth and the uncertainty attached to it depend upon the time horizon that we consider. The so-called yield curve describes the relationship between the interest rate and time horizon.⁷ Most of the time, the observed yield curve in the economy is increasing: one gets a better return when investing in long term bonds than when investing in bonds with a shorter maturity. But it may be possible in some circumstances to have an inverted yield curve.

The expected growth of the economy over the next two years is larger than the expected growth rate over the next year alone. Thus, the positive wealth effect is larger over two years than over one year. Similarly, the uncertainty on the growth of the economy is larger over 24 months than over 12 months. It implies that the precautionary effect will also be larger. Because the wealth effect and the precautionary effect work in opposite direction, it is not clear whether the interest rate is increasing or decreasing with the maturity of the asset.

Let \tilde{x}_{01} denote the random growth rate of the economy over the next 12 months, and \tilde{x}_{12} denote the growth rate between months 12 and 24. Thus, the GDP per capita 2 years from now is fully described via the random variable $z_0(1 + \tilde{x}_{01})(1 + \tilde{x}_{12})$. If we assume that the representative agent is patient ($\beta = 1$), the price for a zero-coupon bond maturing in 1 year, is as in equation (11.20):

$$B_1 = \frac{1}{1 + r_{01}} = \frac{Ev'(z_0(1 + \tilde{x}_{01}))}{v'(z_0)}.$$

Similarly, the price for a zero-coupon bond maturing in 2 years, B_2 , satisfies the following condition

$$B_2 = \frac{1}{(1 + r_{02})^2} = \frac{Ev'(z_0(1 + \tilde{x}_{01})(1 + \tilde{x}_{12}))}{v'(z_0)},$$

⁷In the finance jargon, this is also often called the zero curve, since it describes the interest rates used to price zero coupon bonds.

where r_{02} denotes the rate of return *per year* for this bond. Suppose that \tilde{x}_{01} and \tilde{x}_{12} are independent and identically distributed, and that the representative agent has a constant relative risk aversion of γ . Then, we obtain that

$$\begin{aligned} B_2 &= \frac{1}{(1+r_{02})^2} = \frac{Ez_0^{-\gamma}(1+\tilde{x}_{01})^{-\gamma}(1+\tilde{x}_{12})^{-\gamma}}{z_0^{-\gamma}} \\ &= [E(1+\tilde{x}_{01})^{-\gamma}]^2 \\ &= \left[\frac{E(z_0(1+\tilde{x}_{01}))^{-\gamma}}{z_0^{-\gamma}} \right]^2 = \left[\frac{Ev'(z_0(1+\tilde{x}_{01}))}{v'(z_0)} \right]^2 = \frac{1}{(1+r_{01})^2}. \end{aligned}$$

Thus, with constant relative risk aversion and no serial correlation in the growth rate of the economy, the yield curve is flat, $r_{01} = r_{02}$. There should be no reward for long-term investors. This is a case where the increased wealth effect and the increased precautionary effect counterbalance each other.

11.7 Bibliographical references and extensions

The building blocks of asset pricing theory have been shaped by many different authors, such as Markowitz, Merton, Samuelson, Black, Scholes and many others. The list is too long to be presented here. Lucas (1994) proposes a very simple model of an exchange economy to solve the asset pricing problem. Mehra and Prescott (1985) and Weil (1989) use Lucas' model to present the equity premium puzzle and the risk-free rate puzzle respectively. Cass and Stiglitz (1970) prove that the two-fund separation property, which is key requirement in the capital asset pricing model, holds under expected utility only when absolute risk tolerances are linear with the same slope. Leland (1980) shows who should purchase portfolio insurance at equilibrium when linear risk tolerance not hold. Franke, Stapleton and Subrahmanyam (1998) show how, even in an economy with linear risk tolerances, differences in background risks can be enough to cause some investors to sell options, while others buy options.

Cox, Ingersoll and Ross (1985) consider a very general economy with production to characterize the yield curve. Several models prior to 1985 proposed pricing formulas that relied on non-maximizing behaviors by investors.

References

- Cass, D. and J. Stiglitz, (1970), The structure of investor preferences and asset returns, and separability in portfolio allocation, *Journal of Economic Theory*, 2, 122-60.
- Cox, J., J. Ingersoll and S. Ross, (1985), A theory of the term structure of interest rates, *Econometrica*, 53, 385-403.
- Dimson, E., P. Marsh and M. Staunton, (2002), *Triumph of the Optimists: 101 Years of Global Investment Returns*, Princeton University Press, Princeton.
- Franke, G., R. Stapleton and M. Subrahmanyam, (1998), Who buys and who sells options: The role of options in an economy with background risk, *Journal of Economic Theory* 82, 89-109.
- Ibbotson Associates, (1999), Stocks, bonds, bills and inflation : 1999 yearbook, Ibbotson Associates, Chicago, www.ibbotson.com.
- Jorion, P., Goetzmann, W.N., (1999), Global stock markets in the twentieth century, *Journal of Finance*, 54, 953-980.
- Leland, H.E., (1980), Who Should Buy Portfolio Insurance?, *The Journal of Finance*, 35, 581-596.
- Lucas, D.J., (1994), Asset pricing with undiversifiable income risk and short sales constraints: Deepening the equity premium puzzle, *Journal of Monetary Economics*, 34, 325-341.
- Mehra, R. and E. Prescott, (1985), The Equity Premium: A Puzzle, *Journal of Monetary Economics*, 10, 335-339.
- Weil, P., (1989), The equity premium puzzle and the risk free rate puzzle, *Journal of Monetary Economics*, 24, 401-21.

Part IV
Extensions

Chapter 12

Asymmetric Information

Until now we have maintained an assumption of complete information by all agents in the market place. Of course in many situations, people might have different information on which to base their economic decisions. In this chapter we focus on the case where some market participants have private information, which is information that is known only to themselves. Sometimes this information is difficult, if not impossible, to fully reveal to others. For example, it would be easy for you to convince someone that you know how to play the piano, assuming that you do know how. Thus "playing the piano" would be information that can be easily verified. However, suppose a local store offered a free piano to you if you could prove to them that you do not know how to play the piano? How would you convince them that you are not a gifted pianist who is just pretending not to know how to play? "Not playing the piano" would be information that is quite difficult to verify.

An example of a market with private information is the market for used cars. The purchase of a used car, especially one that is no longer on a manufacturer's warranty, places much risk on the buyer. What if that beautiful car that appears to be in such excellent condition actually suffers from some severe problems, which will only manifest themselves after the car has been driven for a while? One problem is that low-quality cars, often called "lemons," are not always easy to distinguish from high-quality cars until after their purchase. As a result, we are likely to see prices in the used-car market in which the existence of "lemons" forces the owners of high-quality cars to accept a lower price than they would accept in a market where automobile quality was transparent to everyone. In the extreme, the owner of a high-quality used car might decide not to sell the car at all, since the low

market price is less than the owner's reservation price for selling the car.

On the other hand, it might be possible to provide some kind of credible signal that a car is not a lemon. For example, the seller of a high-quality car might offer a limited warranty to the buyer, whereas such a warranty would be too costly to offer for the low-quality seller. Or, the owner of the high quality car might be willing to pay for an inspection by a local independent mechanic. At this stage, it is useful to note that it is the high-quality car owner who must bear the cost of signalling that his car is not a "lemon."

Now consider a related problem in which the car has been inspected and revealed to be high-quality car. Suppose the car was sold for a high price to someone who lives far away. The original owner offers to drive the car to the buyer's home and turn over the keys in person. Since the sale has already been made, the original owner no longer has an economic incentive to drive the car in a careful manner. Thus, the car might be driven more harshly than if the inspection and sale had not yet taken place. Moreover, the original owner might discover some new problem during the delivery trip. If this new problem is not readily visible, the original owner does not have much incentive to reveal this new problem to the buyer. Or suppose that, instead of the original owner, the new owner hires some university student to deliver the car to his home, with an agreed-upon fee to be paid in advance? What incentive would the University student have to drive carefully?

In this section we analyze the effects of private information on market contracts. In cases like the market for "lemons," we will show how contracts may be designed to illicit truthful revelation of private information. In the case of the driver/deliverer of the new car, we will show how particular contracts might induce desirable behavior.

12.1 Adverse selection

If there was just one market price for all used cars, that price would be more attractive to the current owners of "lemons" and less attractive to the current owners of high-quality cars. As a result, there would be a disproportionate share of low-quality cars for sale. This is a classic example of a phenomenon known as "adverse selection." Indeed, the first real economic analysis of this problem propelled George Akerlof to a Nobel Prize.

Another classic example is an insurance market, as first examined by Rothschild and Stiglitz (1976). To make the example more concrete, let us

consider the auto insurance market. In this market, various laws allow the insurance companies to base their prices on several observable features of the insured, such as past driving history and the type of car she drives. Let us assume that we restrict ourselves to drivers who all fall within the same insurance classification and thus would face the same menu of prices from the insurer. Of course, all drivers within this classification are not identical. To focus on the adverse selection aspects, we assume that individuals are identical except for their probabilities of a loss. A loss, if it occurs, is of a fixed size L . There are two types of drivers within this risk classification, "good" drivers and "bad" drivers, with loss probabilities given by p_G and p_B respectively. We assume that $0 < p_G < p_B < 1$. Each driver is assumed to know his or her own loss probability, whereas the insurer cannot observe an individual's type and only knows the distribution of types within the population.

12.1.1 Full insurance

As a point of comparison, consider first a market in which p_G and p_B are public information and where the insurer is free to offer a different menu of contracts to each type. Because insurers can diversify the independent individual risks, they are assumed to be risk neutral and seek to maximize their expected profits. The market is assumed to be in a long-run competitive equilibrium so that the expected profit is zero for each of the insurance companies and for each supplied contract. This means in particular that good drivers pay a premium that is equal to p_G dollar per dollar of indemnity, and that bad drivers pay a larger premium rate p_B . Since this implies actuarially-fair pricing, we know from Mossin's Theorem (??) that full insurance would be optimal, if offered. To see that full insurance would indeed be offered in such a market, suppose to the contrary that some level of coinsurance $\alpha^* < 1$ was the highest available level of coverage. We know from Chapter 3 that a level of coinsurance $\alpha^* + \varepsilon$ at a fair price would increase the consumer's expected utility, for any $\varepsilon > 0$ such that $\alpha^* + \varepsilon \leq 1$. But in this case, it must be true that offering $\alpha^* + \varepsilon$ at a price slightly higher than an actuarially-fair price also would be preferred to α^* at a fair price. As a result, the insurer offering $\alpha^* + \varepsilon$ would earn an expected profit. In a long-run competitive equilibrium, we will assume that competition drives this profit out. The end result is that there must be full insurance at a fair price offered to each type of driver. The good drivers pay a premium of $P_G = p_G L$ while the bad drivers pay a

premium of $P_B = p_B L$.

Now let us suppose that driver type is private information. Each driver knows his or her own type, but this information is not observable nor verifiable by the insurer. If we continue to offer full coverage insurance policies at two separate premium rate, P_G and P_B , it is likely that all consumers would claim to be good drivers in order to pay price P_G , since $P_G < P_B$. This implies that insurers will all lose money and the market will not be able to survive with these prices. One possibility of the market surviving with full insurance policies is for the insurers to all charge a premium of P_B . In this case, the good risks will either decide that the price P_B is too high and that they are better off with no insurance; or the good risks will decide that full insurance with a premium of P_B is better than no insurance. If the premium P_B is too high, the good drivers will buy no insurance at all. In this case, we have an extreme example of the "lemons" phenomenon, since only bad drivers buy insurance.¹ This is illustrated in Figure 12.1.

[INSERT FIGURE 12.1 ABOUT HERE]

In Figure 12.1, we draw a state-claims diagram, as introduced in Chapter 5. Both types of drivers have an initial state claim of $(w, w - L)$. Both also have the same von Neumann-Morgenstern utility function; however their expected utilities differ due to their differing probabilities of a loss. The preferences of the two agents satisfy the important single-crossing property. It means that at any point in Figure 12.1, as at point A for example, the indifference curve of the low risk type is steeper than the indifference curve of the high risk agent. This is the graphical expression of the fact that high risk agents are willing to pay a larger premium in the good state to get more to consume in the bad state, which is more likely. In other words, high risk agents are more reluctant to buy a high deductible insurance. The so-called "fair-price lines" are also drawn for both the bad-risk and the good-risk types. These lines represent all of the possible state claims that arise from paying a fair premium of $p_t \alpha L$ in return for a net payment by the insurer of $(1 - p_t) \alpha L$ in the event of a loss, for $t = B, G$. The bad risk price line is flatter due to the higher probability of a loss p_B , which of course entails a higher premium. With full information, both types of drivers will be offered full insurance, with their final contingent wealth claims at the points labelled **B** and **G** for the bad and the good risks respectively.

¹The analogous used-car market situation would be a market in which the only used cars that are for sale are "lemons."

12.1.2 Pooling contracts

The above analysis assumes that only full-coverage insurance is available. Of course, unlike used cars, we can sell partial insurance contracts. To this end let us suppose that an insurance contract consists of a pair, specifying both a premium P and an indemnity level α where the indemnity itself is understood to be αL . A Rothschild-Stiglitz equilibrium in this market is defined as follows:

Definition 26 *A set of contracts is an equilibrium set of contracts if*

- (i) *All contract pairs that are offered earn an expected profit of zero, and*
- (ii) *There is no other contract that could be added to the equilibrium set of contracts that would earn a positive expected profit.*

Consider the full insurance contracts described above. For the good risks to demand any contract, their final wealth must be preferred to their wealth with no insurance. That is, we must have the contingent claim $(w - P, w - P) \succeq_G (w, w - L)$, where " \succeq_G " denotes the preference relation for good drivers and P denotes the premium offered for full insurance. This inequality represents the so-called "individual rationality constraint" for the good drivers. It is simply a constraint for participating in the market.

To analyze insurer profits, we assume that there is public information available on the proportion of bad drivers in the population. Let λ denote this proportion of bad drivers and assume that $0 < \lambda < 1$. Consider a full coverage insurance premium of $P_\lambda \equiv [\lambda p_B + (1 - \lambda)p_G]L$. Such a contract would break even, on average, if both types of drivers bought this contract. Of course the bad drivers will want to buy it, since the price is even lower than a fair price. But, this contract will, of course, lose money if only the bad drivers buy it. However, if the good drivers find this contract better than no insurance, the contract will earn a positive profit on the good drivers. Given the design of the premium P_λ above, this full coverage insurance contract would earn an expected profit of zero. Such a contract that is purchased by both types of individuals is called "pooling contract." If such a contract is an equilibrium contract, we refer to the equilibrium as a "pooling equilibrium."

The idea for this type of contract was carried to the extreme in the United States in the early 1990's, when then-First-Lady Hillary Clinton was attempting to set up such a pooling contract for health insurance. Indeed, the idea then was even more encompassing; it was to pool everyone, not just pool those individuals within a particular risk classification. However, there is

a problem with pooling those with observable classification differences. We can illustrate this by assuming that type is observable with our automobile-insurance pooling contract above, so that we have in essence two observable classifications. Suppose the government imposed such a pooling contract as the only available contract, with no choice allowed to obtain private insurance from elsewhere. Obviously, the good drivers would be the losers in this setting, since they will not be offered insurance at a price of $P_G = p_G L$. The problem then becomes one of voter disenfranchisement. The good risks are disenfranchised and will resist this government action. But what if we restrict ourselves to pooling within one particular risk classification, and where type is not observable, will we then be able to support such a pooling contract as an equilibrium contract?

[INSERT FIGURE 12.2 ABOUT HERE]

The situation in this case turns out to be somewhat similar. The good risks are once again disenfranchised. To see this, consider the situation depicted in Figure 12.2. The full coverage pooling contract leads to the contingent wealth claim C , which offers full insurance to both types at the same premium P_λ . At C , the insurers earn a zero profit. The good risks are seen to prefer C to zero coverage, since the contingent wealth claim C is on a higher indifference curve than $(w, w - L)$. However, full coverage pooling cannot be an equilibrium. Consider a partial insurance contract such as the one leading to contingent wealth D in Figure 12.2. This contract will be preferred by the good drivers to the full-coverage pooling contract, but it will not be preferred to the full-coverage contract by the bad drivers. Since D lies below the fair-price line for the good drivers, this new partial insurance contract will earn a positive expected profit. As a result, an equilibrium cannot contain the full coverage pooling contract.

The problem is essentially the same for any fair-priced pooling contract, not just for full insurance. Since the good-risk indifference curve will always be steeper than the bad-risk indifference curve everywhere along the fair-price line for pooling contracts, we can always find a new contract that will attract only the good drivers and make an expected profit. If your competing insurers only offer a pooling contract, this means that you can offer a new contract with a larger deductible and a smaller premium. Because high risk agents are more reluctant towards larger deductible, there exist a reduction in premium that is large enough to attract low risk agents, but it is not large enough to attract the high risk agents. Because the low risk agents initially paid a premium which is much larger than the actuarial value of their policy,

this new contract is profitable. We thus have the following conclusion.

Proposition 27 *In a Rothschild-Stiglitz equilibrium under adverse selection, there cannot exist a pooling equilibrium.*

So if there are no pooling contracts in equilibrium, what other type of equilibrium contracts might be possible?

12.1.3 Separating contracts

Since the expected utility of the good drivers and bad drivers is different, we can design a set of contracts in such a way that each type prefers a different contract in equilibrium. Such an equilibrium is called a "separating equilibrium." Since the insurers cannot directly observe or verify an individual's type, these contracts serve as a mechanism by which insurance can segregate the population into the good drivers and the bad drivers. Since each individual is free to select his or her own contract, the mechanism is often referred to as a "self-selection mechanism." It is also an example of what is called a "revelation mechanism," since by self selecting a contract, the individual reveals his or her type to the insurer.²

To see how this type of contract is designed, we need to introduce a new type of constraint called the "incentive-compatibility constraint." In the Rothschild-Stiglitz model, this constraint is that each type of driver likes its own insurance contract better than the alternative contract. Let (P_t, α_t) denote the contract purchased by type t , $t = B, G$, and let \succsim_t denote the preference ordering of type t . There are two incentive constraints in our insurance model, one for each risk type: (i) $(P_B, \alpha_B) \succsim_B (P_G, \alpha_G)$, and (ii) $(P_G, \alpha_G) \succsim_G (P_B, \alpha_B)$.

Since each type of contract offered must earn an expected profit of zero, we first note that the bad drivers must be offered full insurance at a fair price, $\alpha_B = 1$ and $P_B = p_B L$. If this were not the case, if the bad drivers were offered a different amount of insurance, new contract offerings could lead to a profit in the same manner as discussed in section 12.1.1. The insurers would not be concerned if good drivers also bought this contract, since any contract

²One might think that once the insurers know the individual's type, they can charge a premium based on that information. However, we assume that type is revealed only by the actual purchase of insurance. If the insurer could renegotiate all contracts, this would cause individuals to behave strategically, so as not to always reveal their true type.

that breaks even or earns a profit from the bad drivers will also earn a profit, on average, from any good drivers who purchase it. As a result, long-run competition ensures that the bad drivers will be offered full insurance at a fair price, with a contingent-wealth claim of B in Figure 12.1. With the bad-risk contract $(p_B L, 1)$ in place, the good-risk contract must be designed so as to not attract the bad drivers. Competition will force the insurers to offer as much insurance as possible to the good drivers at a fair price for the good drivers. Thus, the incentive compatibility constraint for the bad drivers, (i) above, must be binding. The good drivers are offered the contract that yields contingent-wealth claim G' in Figure 12.3. This contract leaves the bad drivers just indifferent to their full-coverage contract, and we will assume that they opt for the full coverage contract with contingent wealth B . Thus, this pair of contracts causes the two types of drivers to reveal their type via their contract choice.

[INSERT FIGURE 12.3 ABOUT HERE]

However, is this pair of separating contracts an equilibrium? As it turns out, there may be a problem. Certainly no other separating contracts could dominate the separating contracts discussed above, but what about a pooling contract? In particular, suppose the fraction of bad drivers λ is relatively small, so that the actuarially-fair pooling price line is represented by the line labelled "Pooling price 1" in Figure 12.3. In that case, consider the pooling contract that leads to the contingent-wealth claim C' in the figure. This contract is preferred by both types of individuals to their respective separating contracts. Hence both the bad drivers and good drivers would purchase it, if it were offered as an alternative to their separating contracts. Moreover, since contingent claim C' lies below "Pooling price 1," it would earn an expected profit if both types of drivers purchased it. Thus, part (ii) of Definition 26 is not satisfied by our pair of separating contracts. Of course, we have already seen in Proposition 27 that a pooling equilibrium cannot exist. Hence, in this situation, a Rothschild-Stiglitz equilibrium fails to exist.

If the proportion of bad drivers λ is relatively large, such as illustrated by the pooling price line labelled "Pooling price 2" in Figure 12.3, then no fair-priced pooling contract can attract the good drivers. Hence, the separating contracts defined above are indeed an equilibrium. Thus, we obtain the following result, which summarizes what we have discussed here:

Proposition 28 *If there are sufficiently many bad drivers in the population,*

then a Rothschild-Stiglitz equilibrium consists of separating contracts in which the bad drivers receive full insurance at a fair price, while the good drivers receive partial coverage at a fair price. In this context, the number of bad drivers is sufficiently large whenever no fairly price pooling contract attracts the good drivers.

If an equilibrium does exist, it is interesting to note that adverse selection does not affect the welfare of the bad drivers at all. They receive full insurance at a fair price, exactly the same as if there was full information. It is only the good drivers who are affected by accepting less-than-full coverage. In the same way that the owner of a high-quality automobile had to bear any cost of signalling that his car was not a "lemon" in the market for used cars, the good driver here must bear the cost of signalling that he is not a bad driver by purchasing partial insurance.

12.2 Moral hazard

Moral hazard deals with hidden actions or with the fact that effort is typically not observable. For example, an individual with insurance might not drive as carefully as she would if she had to pay for all of her own losses. Or consider the effect of airbags on driver caution. Knowing that there is state-of-the-art protection in the event of an accident might cause a driver to be less cautious. Consider the incentive effect of an airbag compared to that of a device that was mounted in the steering wheel and was triggered exactly the same as your airbag, except that this alternative device triggered a bomb that was attached beneath the car, so that any front-end impact caused an explosion and immediate death as opposed to a nice inflated pillow that offers protection from injury. In which car would you have the incentive to drive more carefully?

The fact that one's market choices might lead to incentives that alter one's behavior is the general problem of "moral hazard." In particular, we focus on the impact of contracts on one's behavior. Although the settings for moral hazard can be quite complex, much can be gleaned by examining the simplest case possible, namely a case where there are only two levels of effort. Again, to keep the story more concrete, we maintain the setting of an insurance market with only two possible loss states. Either no loss occurs or a loss occurs and is of size L . Rather than two types of individuals, we

will consider only one individual, but with two possible effort levels. With no effort the probability of an accident is p_N and with effort the probability is p_E , where we assume that $0 < p_E < p_N < 1$. The setting is not too unlike that of the previous section under adverse selection, except that individual can now choose his or her own type. Of course everyone would choose to take the effort if effort is costless. We thus assume that there is a cost for taking effort that is measured in utility terms for an individual. In particular, taking effort costs the individual e units of utility.

In order to study how effort is affected by market contracts, let us first take a look at how the decision to take effort is affected by an individual's distribution of random wealth. When there are only two states of nature, we are considering two competing probability distributions. Consider the indifference curve through a contingent-wealth claim on the certainty line, such as at E in Figure 12.4. Since wealth at E is the same in both states, the probabilities p_E and p_N play no role in calculating the expected utility of wealth. Let us suppose that the expected utility at E with no effort is equal to k . If effort were costless, utility would also be k with effort; but because effort has a utility cost of c , the individual's expected utility with effort is $k - c$. The single crossing property also holds here: indifference curves when effort is taken are steeper at every contingent wealth claim, since $p_E < p_N$. Since expected utility of wealth is lower at E when effort is taken, consider a higher indifference curve, one for which the expected utility of wealth with effort is also k .³ In Figure 12.4, this occurs on the indifference curve with effort, passing through contingent claims BDC .

[INSERT FIGURE 12.4 ABOUT HERE]

Now consider the individual's choice about whether or not to take effort. At claim B in Figure 12.4, the individual has k units of utility when effort is taken. On the other hand, B lies above the indifference curve for k units of utility when no effort is taken (the indifference curve drawn through points EAD). Consequently, if the individual's contingent-wealth claim is B , the individual will decide not to take effort, and achieve a utility higher than k . Similarly, we can consider the contingent-wealth claim denoted by A , for which k units of utility are obtained when no effort is taken. If the individual with contingent-wealth claim A decides to take effort, she will be on a lower

³Of course this will require $(1 - p_E)u(y_1) + p_E u(y_2) = k + e$, where y_1 and y_2 are wealth in the no-loss state and loss state respectively. We assume that such contingent claims exist and that y_1 and y_2 are both strictly positive.

indifference than the one through BDC . Indeed, we can see that, as drawn, taking effort with contingent claim A would lead to a utility somewhere between $k - c$ and k . Thus, if the individual's contingent-wealth claim was A , no effort would be taken.

Note that at claim D , the individual is indifferent as to whether or not effort is taken. When effort is endogenized, the reader can easily verify that the state-contingent claims E , A , D and C all have utility k , after optimizing with respect to the effort level. As a result, they can all be considered to lie on the same indifference curve, where we allow the level of effort to be chosen optimally. This indifference curve for k units of utility is depicted in Figure 12.5, together with some indifference curves for other levels of satisfaction. We can see how the "kinks" on each indifference curve divide that level of utility into state claims that induce effort and those that do not induce any effort. Each "kink" is itself a contingent-wealth claim for which the individual is indifferent between taking effort and not taking effort. Thus, by considering these claims for which effort is indifferent, we can naturally divide the space of state-contingent claims into two regions: one region for which effort is optimal, and second region for which no effort is optimal. For the sake of clarity, we will assume that the individual takes effort if she is indifferent between taking effort and not taking effort.

[INSERT FIGURE 12.5 ABOUT HERE]

In general, claims closest to the 45 degree certainty line will lead to no effort. For these claims, the difference in wealth between the loss state and the no-loss state is literally "not worth the effort." On the other hand, for claims such as C , where there is a fairly substantial difference in wealth between the no-loss state and the loss state, the individual finds the cost of effort to be worthwhile. That is to say that the expected monetary reward from decreasing the likelihood of the loss state will increase the individual's expected utility by more than c , the cost of effort. Thus, even though we cannot observe or verify whether the individual takes effort, we have information about what the individual's incentive is, based on her contingent wealth. This allows contracts to be written, based on this information.

Let us see how an insurer can use this information in offering insurance contracts. If we maintain the assumption of long run competition, any profits will be driven away. We thus look at only actuarially-fair contracts. Of course individuals would prefer to have full insurance at a fair price, but we know that with full insurance, the individual would take no effort. Consequently, the price for full insurance is the no-effort fair price, $P = p_N L$. Only

if the level of insurance is small enough will the individual have an incentive to take any effort. This is illustrated in Figure 12.6. If I denotes the initial contingent-wealth claim with no insurance, then the highest level of insurance that can be sold at the lower with-effort price is that level which takes the individual to wealth D . This level of insurance is labeled as α_D in the diagram. Any additional insurance coverage at this low price would cause the individual to stop taking any effort and, consequently, the insurer would expect to make a loss on the contract. If more coverage than α_D is desired by the individual, the insurer will charge the higher "no-effort" price. This leads to a non-linear set of insurance prices:

$$P(\alpha) = \left\{ \begin{array}{ll} \alpha p_E L & \text{for } \alpha \leq \alpha_D \\ \alpha p_N L & \text{for } \alpha > \alpha_D. \end{array} \right\}$$

[INSERT FIGURE 12.6 ABOUT HERE]

For any fixed level of effort and assuming that insurance is offered at a fair price, the individual's expected utility is monotonically increasing in the level of insurance coverage for all levels of coverage below full insurance. Thus α_D is the most preferred level of insurance that is offered at the lower "with-effort" price on the non-linear pricing schedule. Of course, at the higher "no-effort" price, full coverage would be most preferred, leading to the contingent-wealth claim labelled N in Figure 12.6. As a consequence, we only need to compare two contracts: the ones with $\alpha = 1$ and with $\alpha = \alpha_D$. Whichever of these two levels of insurance coverage leads to a higher expected utility is the one chosen by the individual. As drawn in Figure 12.6, D lies above the indifference curve through N . Thus, the individual would choose $\alpha = \alpha_D$, leading to the contingent-wealth claim D as well as a decision to take the effort to reduce the probability of a loss. Although not drawn in the diagram, it also would be possible to have preferences for which full coverage is preferred, together with a decision not to take any effort.

It is interesting to compare our moral-hazard solution presented here to the separating equilibrium in the adverse-selection model. In both cases, the consumer has (essentially) a choice between two contracts: one contract in which a limited amount of coverage is offered at a low price, and another contract in which full coverage is offered but at a higher price. In the adverse selection model, the limited-coverage contract at the low price is set in such a way as to segregate the good risks, i.e. those with a lower probability of a loss. In the moral-hazard model, the limited-coverage contract is set in such

a way as to segregate good behavior, i.e. behavior that lowers the probability of a loss.

12.3 The principal-agent problem

We now consider a variant of the moral hazard model that has wide array of applications. Consider an individual or a firm that has two possible final wealth levels, X_1 and X_2 where $X_1 > X_2$. The probability of state i is p_i for $i = 1, 2$. We will refer to this individual or firm as the "principal." In many circumstances, it is possible for the principal to hire someone to increase the likelihood of the good outcome in state 1. For example, consider someone who is being sued and who hires a lawyer to help win the case. Or consider a firm trying to win a construction contract from a city and suppose that the firm hires a consulting company to present its proposal to the city council. The lawyer and consulting company are examples of what is referred to as the "agent." The agent thus works on behalf of the principal to increase the likelihood of state 1 (or, equivalently, to decrease the likelihood of state 2). Suppose we pay the agent a fixed fee for its service. If we ignore long-term considerations, such as reputation effects, what incentive does the agent have to take any effort? Of course if effort is observable, we can write a clause into any contract denying payment in the event that no effort is taken. But in a situation where effort is both unobservable and unverifiable by the principal, we cannot base compensation on the level of effort taken. If the lawyer works extremely hard and increases our probability of winning the case, we might still lose due to plain bad luck. Or if the city council grants our construction contract, it might be that we were awarded the contract in spite of the fact that the consulting firm did not work very hard. In other words, the principal cannot be sure, even *ex post*, whether effort was taken by the agent or not.⁴

12.3.1 Binary effort with a risk-neutral principal

We will keep the model as simple as possible by assuming that the agent has only two possible choices: effort or no effort, similar to our insured in the

⁴In our two-state model, it is easy to define the effect of effort as increasing the probability of the good state. With more than two states, the beneficial effect of effort is a bit trickier. It usually involves a better set of outcomes via first-order stochastic dominance, with some additional restrictions.

moral hazard model above. Once again, effort is assumed to have a direct cost in terms of expected utility, with effort reducing utility by an amount e . Of course, if the agent takes no effort, which is often referred to as "shirking" in the principal-agent literature, the principal would prefer to pay the agent zero. That is, the principal does not want to pay the agent who shirks. At the same time, the agent needs to find it worthwhile to work for the principal.

We assume for now that the principal is risk neutral, with the agent being risk averse. Since a fixed-fee form of compensation to the agent will not induce the agent to take any effort and since effort itself is unobservable, we need to examine whether it might be possible to induce the agent to take effort by offering compensation in the form of a contingent claim. Let (a_1, a_2) denote the contingent claim that is paid by the principal to the agent, where $a_1, a_2 \geq 0$. The principal thus retains the claim $(X_1 - a_1, X_2 - a_2)$.⁵ Although the principal receives less wealth in every state of the world after paying the agent, the principal will have a higher probability of state 1 by hiring the agent, assuming that the agent does not shirk. Of course, if we do not have $a_1 > a_2$, there is not going to be any incentive for the agent to take effort. On the other hand if we make a_1 so much larger than a_2 that $X_1 - a_1 < X_2 - a_2$, the principal will have no desire to hire the agent.

The method by which we can induce effort is the same as we analyzed in the previous section. Reinterpreting no-loss-state wealth as a_1 and loss-state wealth as a_2 , Figure 12.5 shows the set of contingent-claims contracts for the agent that will induce effort. To operationalize the principal-agent model, we will assume that the agent is offered a contract that yields a level of utility that is identical to the level of utility obtainable from the agent's next best alternative. Let this level of utility be denoted by k . The level of effort is denoted as e and is assumed here to be either zero or one. The corresponding probability of state i is $p_i(e)$, with $p_2(e) = 1 - p_1(e)$. The cost of effort is denoted c . We assume that $p_1(1) > p_1(0)$.

Our objective then is to find the contract (a_1, a_2) to solve the following program

$$\max_{a_1, a_2, e} p_1(e)(X_1 - a_1) + p_2(e)(X_2 - a_2) \quad (12.1)$$

⁵We could add a fixed wealth to both the claim of the principal and that of the agent, but that does not affect the model and only adds notation. We will therefore conveniently assume that (a_1, a_2) and $(X_1 - a_1, X_2 - a_2)$ are the final contingent-wealth claims of the agent and of the principal respectively.

subject to

$$p_1(e)u(a_1) + p_2(e)u(a_2) - ce = k \quad (12.2)$$

$$p_1(1)u(a_1) + p_2(1)u(a_2) - c \geq p_1(0)u(a_1) + p_2(0)u(a_2). \quad (12.3)$$

Thus, the principal chooses the contingent pay to the agent (a_1, a_2) and, indirectly, the level of agent effort e that maximize its expected payoff. It might seem odd at first that we include a choice of e , since e is unobservable by the principal. In our model with just two levels of effort, this effectively amounts to the principal somehow choosing $e = 1$. How this occurs becomes clear once we examine the two constraints. The constraint (12.2) is the individual-rationality constraint of the agent. The agent needs to be at least as well off as it would be in the next-best alternative. In the principal-agent set-up, this constraint is also referred to as the "participation constraint" for the agent. Constraint (12.3) is the agent's incentive-compatibility constraint. Recall that in the adverse-selection model the incentive-compatibility constraint guaranteed that each type of risk liked its own contract at least as well as it liked the other type's contract. In the current principal-agent setting, the incentive-compatibility constraint guarantees instead that the agent likes working at least as much as shirking, in which case we assume that the agent takes the effort. In effect, the incentive-compatibility constraint (12.3) guarantees that $e = 1$. In other words, although the principal cannot observe effort e directly, it can effect a choice of $e = 1$ by an appropriate choice of a_1 and a_2 .

[INSERT FIGURE 12.7 ABOUT HERE]

Figure 12.7 illustrates the solution to this problem. Here we show the payment to the agent as a contingent claim. Just as in the moral-hazard case, we can partition the space into contingent payments for which effort is optimal for the agent and payments for which taking no effort is optimal. The incentive-compatibility constraint (12.3) forces our solution for the optimal (a_1, a_2) to lie somewhere within the "with-effort" set of contingent payoffs. The participation constraint (12.2) forces us to be on the agent's indifference curve for which $Eu = k$ in this "with-effort" set of contingent payments. The contracts that satisfy the two constraints lie on curve DE. Given these constraints, the contingent payment to the agent that maximizes

the expected payoff to the principal is precisely the one that minimizes the expected payment to the agent. This is illustrated by the principal's indifference curves as drawn in Figure 12.7. These indifference curves are parallel straight lines with a slope of $-p_1(1)/p_2(1)$. The principal, as well as the agent, would receive the same expected payoff everywhere along such a line so long as effort is taken by the agent. We know from Chapter 5 that risk aversion of the agent implies that the agent's marginal rate of substitution, $p_1(1)u'(a_1)/p_2(1)u'(a_2)$, is less than $p_1(1)/p_2(1)$ for all $a_1 > a_2$. This leads to a corner solution at D in Figure 12.7, with a contingent payment to the agent of (a_1^*, a_2^*) . In other words, we pay the agent barely enough to make the effort worthwhile.

Note that the "reserve utility" of the agent, k , might be viewed as arbitrary. That is, we might think about changing k in (12.2). In this sense, we can view the solution to (12.1) as being Pareto efficient: we maximize the principal's "expected utility" (which is the same as its expected payoff under risk neutrality) subject to a given level of expected utility k for the agent. Since we are only Pareto efficient with respect to a reduced set of possibilities, namely those satisfying both the constraints (12.2) and (12.3), the solution is often referred to as a "second-best" solution. That is, we must qualify the way in which our solution is "Pareto efficient."

To see what this difference entails, consider a world in which effort could be observed. Notice that the solution of paying the agent (a_1^*, a_2^*) is not Pareto efficient in this case. In this full information world, let us make the agent's contingent payment conditional upon the agent exerting effort. With no effort, the agent gets paid zero. In this case, we can find other contracts, such as at contingent payment R in Figure 12.7, for which both the principal and the agent are better off. The agent will be on a higher indifference curve while at the same time receiving a lower expected payment, meaning that the expected payoff to the principal is higher. In other words, at R , both the principal and the agent are better off. The contingent payment at D is thus Pareto dominated in the sense of efficient risk sharing. However, for the case where effort is fixed at $e = 1$, note that all of the contingent payments to the agent that Pareto dominate D (i.e. all contingent payments that offer more utility with a lower expected payment to the agent) involve contingent payments to the agent for which no effort, $e = 0$, would be optimal. In other words, a contract such as R is not incentive compatible. Thus R does not dominate D in our world with unobservable effort. In a very real sense, compared to the "first-best" case with full information and risk-sharing

efficiency, the principal needs to "overpay" the agent in the good state in order to induce the agent to take effort. This is precisely the reason most lawyers for the plaintiff in a liability case are paid on a so-called contingent-fee basis: the lawyer receives an exceptionally high payment if he wins the case, but receives little or nothing if the case is lost.

One might think that this inefficient risk-sharing property of our solution is due to the fact that we have a risk-neutral principal. In the next section we show that this is indeed not the case, while extending our model to a more realistic case with a continuum of effort levels by the agent.

12.3.2 Continuous effort with a risk-averse principal

Let the risk-averse utility of the principal be given by the function v . We assume that effort can be any level $e \geq 0$, with $p_1(e)$ increasing and concave, that is $p_1'(e) > 0$ and $p_1''(e) < 0$. In other words, there are diminishing marginal returns to the agent's effort. In order to avoid a corner solution, we further assume that $p_1(e) < 1 \forall e$, so that it is not possible to guarantee state 1 with certainty, regardless of how much effort is used. The cost of effort is assumed to be c units of utility per unit of effort. We can now write the principal's objective as follows.

$$\max_{a_1, a_2, e} p_1(e)v(X_1 - a_1) + p_2(e)v(X_2 - a_2) \quad (12.4)$$

subject to

$$p_1(e)u(a_1) + p_2(e)u(a_2) - ce \geq k \quad (12.5)$$

$$p_1'(e)[u(a_1) - u(a_2)] - c = 0. \quad (12.6)$$

The inequality (12.5) is the individually-rationality constraint, the same as before. Since preferences are assumed to be continuous, we will treat this constraint as effectively an equality constraint, fixing the agent's utility at its reservation level. Equation (12.6) is the incentive-compatibility constraint. Since we now allow for continuous effort levels, this equation identifies the optimum level of effort by the agent. Indeed, for a given level of contingent

payment to the agent of (a_1, a_2) , this equation is simply the agent's own first-order condition to maximize its level of effort w . Once again, although the principal cannot directly observe or verify the agent's level of effort, the principal will bring about the agent's optimal effort, as determined by (12.6), by choosing the contingent payment scheme. Obviously both the principal and the agent have no reason to contract with each other unless both $a_1 > a_2$, and $(X_1 - a_1) > (X_2 - a_2)$. Since we also have $p_1''(e) < 0$, (12.6) implies a unique level of effort e for any relevant (a_1, a_2) .

We can write the Lagrangean for (12.4) as

$$\begin{aligned} L(a_1, a_2, w, \lambda, \mu) = & p_1(e)v(X_1 - a_1) + p_2(e)v(X_2 - a_2) \\ & + \lambda[p_1(e)u(a_1) + p_2(e)u(a_2) - ce - k] \\ & + \mu\{p_1'(e)[u(a_1) - u(a_2)] - c\}. \end{aligned}$$

This leads to the following first-order conditions:

$$-p_1(e)v'(X_1 - a_1) + \lambda p_1(e)u'(a_1) + \mu p_1'(e)u'(a_1) = 0, \quad (12.7)$$

$$-p_2(e)v'(X_2 - a_2) + \lambda p_2(e)u'(a_2) + \mu p_2'(e)u'(a_2) = 0, \quad (12.8)$$

and

$$\begin{aligned} p_1'(e)[v(X_1 - a_1) - v(X_2 - a_2)] + \lambda\{p_1'(e)[u(a_1) - u(a_2)] - c\} \\ + \mu p_1''(e)[u(a_1) - u(a_2)] = 0. \end{aligned} \quad (12.9)$$

Our constraints, (12.5) and (12.6), are the final two first-order conditions.

To characterize the solution, consider first the condition (12.9). The first term is obviously positive, whereas the second term must be zero by the incentive compatibility constraint. Since $p_1''(e) < 0$ and $a_1 > a_2$, it follows that the optimal value of the Lagrange multiplier μ must be strictly positive. This turns out to be quite important. To see why, rewrite the first two first-order conditions above as follows:

$$v'(X_1 - a_1) = \left[\lambda + \mu \frac{p_1'(e)}{p_1(e)} \right] u'(a_1) \quad (12.10)$$

and

$$v'(X_2 - a_2) = \left[\lambda + \mu \frac{p_2'(e)}{p_2(e)} \right] u'(a_2). \quad (12.11)$$

Recall from Chapter 10 that efficient risk sharing (here in the sense of "first best") requires that there exists some constant, call it γ , such that $v'(X_i - a_i) = \gamma u'(a_i)$ for each state $i = 1, 2$. Clearly that is not the case here, since $p_1'(e) > 0$ while $p_2'(e) < 0$. Further note that, since $\mu > 0$, it follows from (12.11) that our other Lagrange multiplier λ must also be positive. Using (12.10) and (12.11), we obtain

$$MRS_v \equiv \frac{p_1(e) v'(X_1 - a_1)}{p_2(e) v'(X_2 - a_2)} > \frac{p_1(e) u'(a_1)}{p_2(e) u'(a_2)} \equiv MRS_u,$$

where MRS_v and MRS_u denote the marginal rates of substitution for the principal and for the agent respectively, both evaluated at the optimal payment for the agent, (a_1^*, a_2^*) . This inequality implies that both the principal and the agent would be better off, in terms of efficient risk sharing, if the agent traded some state-1 wealth for some state-2 wealth. However, just as in the case with only two effort levels, reducing the state-1 payment to the agent would reduce the agent's level of effort. Thus we see that, once again, our second-best solution involves in a certain sense "overpaying" the agent in state 1 in order to induce the proper amount of effort.

12.4 Bibliographical references and extensions

The adverse-selection problem was modeled by Akerlof (1970). The analysis of equilibrium in a competitive market is due to Rothschild and Stiglitz (1976). Since the Rothschild-Stiglitz model often has no equilibrium, several authors considered non-Nash types of solutions to extend the results. Wilson (1977), for example, shows how pooling contracts might be an equilibrium if we allow nonprofitable contracts to be removed from the market. Riley (1979) allows for strategic responses to new contracts. His equilibrium was applied explicitly to the Rothschild-Stiglitz insurance setting by Engers and Fernandez (1987). Miyazaki (1977) and Spence (1978) use Wilson's model of equilibrium, but require only that the insurer earn zero profit, allowing for profits from some contracts to subsidize losses from others.

The moral hazard model as presented here is adapted from Stiglitz (1983). The principal-agent problem was introduced by Ross (1973) and our model is adapted from the so called "first-order approach" of Holmström (1979). A potential problem with the optimization in Holmström's model, essentially that a local maximum might not yield a global maximum, was discovered by Grossman and Hart (1983). However, some reasonable restrictions avoid this potential pitfall, as was pointed out by Rogerson (1985).

References

- Akerlof, G. (1970). "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84, 488-500.
- Engers, M. and L. Fernandez (1987), "Market Equilibrium with Hidden Knowledge and Self-Selection," *Econometrica* 55, 425-439.
- Grossman, S., and O. Hart (1983). "An Analysis of the Principal-Agent Problem," *Econometrica* 51, 7-45.
- Holmström, B. (1979). "Moral Hazard and Observability," *Bell Journal of Economics* 10, 74-91.
- Miyazaki, H. (1977). "The Rat Race and Internal Labor Markets," *Bell Journal of Economics* 8, 394-418.
- Riley, J. (1979), "Informational Equilibria," *Econometrica* 47, 331-359.
- Rogerson, W. (1985). "The First-Order Approach to Principal-Agent Problems," *Econometrica* 53, 1357-1368.
- Rothschild, M., and J. Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90, 629-650.
- Spence, M. (1977). "Product Differentiation and Performance in Insurance Markets," *Journal of Public Economics* 10, 427-447.
- Stiglitz, J. (1983). "Risk, Incentives and Insurance: The Pure Theory of Moral Hazard," *Geneva Papers on Risk and Insurance* 8, 4-33.
- Wilson, C. (1977). "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory* 16, 167-207.

Chapter 13

Alternative decision criteria

The purpose of this book has been to characterize the properties of optimal decisions in the face of risk, both at the individual and collective levels. As modelled thus far, any decision under uncertainty can be described as essentially selecting a specific lottery L^i from a choice set $C = \{L^1, L^2, \dots\}$. If there is a finite number S of possible states of nature, together with an objective probability distribution, each lottery can be defined as a vector $L = (x_0, p_0; x_1, p_1; \dots; x_{S-1}, p_{S-1})$, where x_s is the outcome in state s , and p_s is the corresponding state probability. An essential component of modelling choice under uncertainty is to determine the way by which individuals make their choice in C . As in the general theory of consumer choice, it is quite innocuous to claim that, at the individual level, the individual chooses a strategy in order to maximize welfare. This statement leaves open the problem of how to evaluate welfare under uncertainty.

At the highest level of generality, welfare must be a function of the set of possible state-dependent outcomes and of the corresponding probabilities; that is, it must be a function of L . Let $V(L)$ denote the welfare of the decision maker who would select lottery L . Many possibilities exist for the form of V , each of them capturing a different type of economic behavior. For example, $V(L) = \sum_{s=0}^{S-1} p_s x_s$ would model risk-neutral preferences. Or consider the case of $V(L) = \min\{x_s\}$, which captures "infinite risk aversion." An individual with such preferences would be the extreme pessimist, who always assumes that the worst outcome will happen.

Throughout this book, we have followed the assumption first made by Daniel Bernoulli in 1738, that individual welfare can be measured by computing the expected utility of the outcome of the lottery. That is, we assumed

that each agent has a utility function u such that his welfare, conditional on selecting any lottery $L = (x_0, p_0; x_1, p_1; \dots; x_{S-1}, p_{S-1})$, takes the following specific form¹:

$$V(x_0, p_0; x_1, p_1; \dots; x_{S-1}, p_{S-1}) = \sum_{s=0}^{S-1} p_s u(x_s). \quad (13.1)$$

It is interesting to note that our example of risk-neutral preferences is a special case of expected utility, with $u(x) = x$. The preferences of the person who is infinitely risk averse ($V(L) = \min\{x_s\}$) can be modelled via expected utility using for example function $u(x) = -e^{-Ax}/A$, with A tending to infinity.

Although it is widely used, the Expected Utility (EU) preference functional (13.1) is very restrictive. In particular, it is additive with respect to the states of nature, and it is linear with respect to the probabilities. The property of additivity with respect to the states of nature implies that the marginal value of the outcome in state s is independent of the outcome in any other state $s' \neq s$:

$$\frac{\partial^2 V}{\partial x_s \partial x_{s'}} = 0.$$

This additivity/independence property is an essential aspect of the EU model, one that we analyze in more detail in the next section. This property greatly simplifies the analysis of choices compared to a more general consumption theory, although it does so at a cost. For example the EU model (13.1) makes the assumption that the marginal value of consumption in one state is independent of the contingent amount consumed in other states.

Notice that the case of uncertainty is not the only environment in which economists restrict the decision criterion to be additive. As we have seen in chapter 6, the classical consumption/saving model assumes that consumers select their lifetime consumption profile in order to maximize the discounted value of their flow of felicity over their lifetime. This welfare functional is also additive with respect to time. It implies, for example, that, seen from the age

¹We pointed out in Chapter 1 that the utility function u was unique only up to a so-called affine transformation. In this sense, the utility is cardinal. However, the preference functional V is equivalent to any increasing transformation of itself. In other words, welfare V is ordinal.

of 20, the marginal lifetime utility of consumption at age 25 is independent of the level of consumption at the age of 24.

An overwhelming majority of researchers still use additive choice functionals to model optimal behavior with respect to risk and time. However, this does not necessarily imply that the additive models are best in describing human behavior or in making policy recommendations. For example, consider a world in which we define $y_s = x_s - w$. If x_s and w describes respectively current and past wealth levels, y can be interpreted as the increase in wealth. Suppose we define the incremental-wealth lottery $\widehat{L} = (y_0, p_0; y_1, p_1; \dots; y_{S-1}, p_{S-1})$. We can define a welfare functional of the form $\widehat{V}(w; \widehat{L})$. In this setting, we can think of w as a "reference point". Unless the welfare functional $\widehat{V}(w; \widehat{L}) = \widehat{V}(0; L)$ for each initial wealth w , as is the case in the EU model, then welfare is obviously dependent on more than just the distribution of final wealth. This shows how preferences themselves might be subject to a particular arbitrary choice of the reference point w , which depends upon the framing of the choice problem – a point made well by Kahneman and Tversky (1979) in their Prospect Theory.

Or suppose that we know the lottery outcomes x_s , but we do not know the precise probability for each state? This situation is often referred to as "ambiguity" or as "parameter uncertainty." If a preference functional is linear in the probabilities, we can simply use a mean probability for each state of nature. However, some research has shown that such ambiguity is disliked by the decision maker and affects behavior, a concept known as "ambiguity aversion."

The number of alternative models is growing fast in the literature. Most of them generalize the EU model presented in this book. This suggests that the EU model will remain a cornerstone of the economic theory of risk. The EU theory is simple, parsimonious and quite successful in explaining a wide set of empirical facts. Moreover, many of the types of behavior that we are trying predict are the same in these alternative models as they are in the EU model. Our objective in this chapter is to provide an introduction to a few of these alternative models. Admittedly, our coverage is only a small sampling of the vast literature on this topic. We start with a description of a few of the main attacks on expected utility theory.²

A final remark is in order here. It is noteworthy that x_s describes the out-

²These attacks are quite numerous. A good survey of the classic theoretical arguments is in Machina (1987).

come relevant to the agent's well-being in state s . It can be multidimensional to include, in addition to consumption, parameters describing the health status, the environment, and so on. When those parameters are uncertain, this raises the question of the interaction between these different sources of risk. Moreover, the aversion to risk on consumption can be substantially different from the aversion to risk on health. The risk management problem becomes even more complex when a lower health usually goes with a smaller labour income stream. Finally, the outcome in some specific states can be a lottery. For example, a risky situation described by $\widehat{L} = (L, p; a, 1 - p)$ is a situation where you get lottery L with probability p , or you get a . We say that \widehat{L} is a compound lottery.

13.1 The independence axiom and the Allais' paradox

Consider a game consisting in choosing between two lotteries L^1 and L^2 . If you can choose between them, which one would you prefer? Suppose that you prefer L^1 . This means that $V(L^1)$ is larger than $V(L^2)$, otherwise you would have preferred L^2 . Let us now modify the rules of the game. We consider now lotteries \widehat{L}^1 and \widehat{L}^2 , where \widehat{L}^i is a compounded lottery in which you get lottery L^i with probability p and another outcome a with probability $1 - p$. If you are offered to choose between \widehat{L}^1 and \widehat{L}^2 , which of the two would you select? If you are an expected-utility maximizer, which implies that V is linear in p , it must be true that

$$V(\widehat{L}^i) = V(L^i, p; a, 1 - p) = pV(L^i) + (1 - p)u(a).$$

Therefore, if you prefer L^1 to L^2 , it also must be true that you prefer \widehat{L}^1 to \widehat{L}^2 . The choice between these two compounded lotteries is independent of the irrelevant common outcome a . This property is often referred to as the "independence axiom". This axiom, introduced by von Neumann and Morgenstern (1944), is a founding stone of expected utility theory. When combined with a few other technical axioms, it *implies* that preferences must be linear in probabilities. This is the expected utility theorem.

The independence axiom is appealing on normative grounds. It is also a quite natural assumption. To illustrate, suppose you are contemplating going to diner downtown either to restaurant A or to restaurant B. You decide that,

13.1. THE INDEPENDENCE AXIOM AND THE ALLAIS' PARADOX 225

all things considered, you prefer restaurant A. Now you learn that your car has developed a fault which means that, with some probability, it will break down after a few miles and prevent you from reaching either destination. Is your preference between restaurants A and B affected by the fact that your choice is no longer a definite one, but only a choice conditional on your not having to spend a few hours repairing the car? If you believe that there is no reason to switch your choice, your preferences satisfy the independence axiom.

When dealing with choices involving only wealth, the oldest and most famous challenge to the independence axiom has been proposed by Allais (1953). The paradox that he raised has generated thousands of papers. Allais proposed the following experiment. An urn contains 100 balls that are numbered from 0 to 99. They are four lotteries whose monetary outcomes depend in different ways on the number that is written on the ball that is taken out of the urn. The outcome expressed, say, in thousands of euros, are described in the following table.

Lottery	0	1-10	11-99
M^1	50	50	50
M^2	0	250	50
\widehat{M}^1	50	50	0
\widehat{M}^2	0	250	0

Outcome as a function of the number of the ball

As in our presentation of the independence axiom, decision makers are subjected to two games. In the first game, they are asked to choose between M^1 and M^2 , whereas in the second game, they must choose between \widehat{M}^1 and \widehat{M}^2 . Many people report that they prefer M^1 to M^2 in the first game, but they prefer \widehat{M}^2 to \widehat{M}^1 in the second. Notice that since M^1 and M^2 have the same outcome when the number of the ball is larger than 10, the independence axiom tells us that these people prefer L^1 with a sure outcome of 50 to L^2 which takes value 0 with probability 1/11 and value 250 with probability 10/11. The paradox is that the same argument can be used with the opposite result when considering the preference of \widehat{M}^2 over \widehat{M}^1 ! Thus, many of the people in Allais' experiment do not behave in a way compatible with the independence axiom.

There is another simple way to verify that the pair of choices $M^1 \succ M^2$ and $\widehat{M}^2 \succ \widehat{M}^1$ is incompatible with EU preferences. This can be done

by contradiction. Suppose that the decision maker is an expected utility maximizer with utility function u . The first choice implies that

$$u(50) > 0.01u(0) + 0.1u(250) + 0.89u(50),$$

or equivalently,

$$u(50) > \frac{1}{11}u(0) + \frac{10}{11}u(250). \quad (13.2)$$

That is, L^1 is preferred to L^2 . Similarly, the preference of \widehat{M}^2 over \widehat{M}^1 implies that

$$0.11u(50) + 0.89u(0) < 0.9u(0) + 0.1u(250),$$

or equivalently,

$$u(50) < \frac{1}{11}u(0) + \frac{10}{11}u(250). \quad (13.3)$$

That is, L^2 is preferred to L^1 . Obviously, there is no utility function u that can satisfy both conditions (13.2) and (13.3).

13.2 Rank-dependent expected utility

One generalization of the EU model that can solve Allais' paradox has been proposed by Quiggin (1982) who weakened the independence axiom. This generalized criterion is now called the Rank-Dependent Expected Utility (RDEU) model. Consider the lottery $L = (x_0, p_0; x_1, p_1; \dots; x_{s-1}, p_{s-1})$, and assume without loss of generality that $x_0 \leq x_1 \leq \dots \leq x_{s-1}$. The index i of the outcome x_i can be interpreted as its rank in the size of the outcomes.

Consider the earlier example of our infinitely risk-averse agent, with $V(L) = \min\{x_s\}$. Since V is ordinal, if $0 < p_0 \leq 1$ we can easily write:

$$V(\overline{x}_0, p_0; x_1, p_1; \dots; x_{s-1}, p_{s-1}) = u$$

potential outcome. For example the lottery $L^1 = (1.01, 0.5; 1.02, 0.5)$ would be preferred to the lottery $L^2 = (1000000, 0.99; 1, 0.01)$.

A less extreme decision maker would consider the possibility that, with probability $1 - p_0$, the outcome will be at least equal to x_1 . This can be done, for example, by considering the following welfare valuation:

$$V(x_0, p_0; x_1, p_1; \dots; x_{S-1}, p_{S-1}) = u(x_0) + f(1 - p_0) [u(x_1) - u(x_0)]. \quad (13.5)$$

In addition to the utility of the sure outcome, the decision maker takes into account the minimum additional utility $u(x_1) - u(x_0)$ that will be generated if state 0 does not occur. This case occurs with probability $1 - p_0$. It is natural to assume that the welfare generated by this prospect is increasing in this probability, i.e., that f is an increasing function. Of course, we also need to assume that $f(0) = 0$: when the worst outcome x_0 is certain, outcome x_1 does not matter for evaluating the welfare of the risk bearer. Similarly, if state 0 is a zero-probability event, outcome x_0 should not matter for welfare. This requires that $f(1) = 1$. For $0 < p < 1$, we might wish to assume that $f(p) < p$, which implies that the agent places a lower weight on the probability of the good news that x_0 did not occur, which is a characteristic of pessimism. In the limiting case where $f(p) = 0$ for all $p < 1$, specification (13.5) is equivalent to (13.4).

We can go one step further by also taking account of the other possible minimal increment to utility $u(x_2) - u(x_1)$ that would be obtained when states 0 and 1 do not occur. The probability of such event is $1 - p_0 - p_1$. This aspect is captured using the following specification:

$$V(x_0, p_0; x_1, p_1; \dots; x_{S-1}, p_{S-1}) = u(x_0) + f(1 - p_0) [u(x_1) - u(x_0)] + f(1 - p_0 - p_1) [u(x_2) - u(x_1)]. \quad (13.6)$$

Extending this idea to the $S - 2$ remaining possible increments in utility, we obtain

$$V(L) = u(x_0) + \sum_{s=1}^{S-1} f \left(\sum_{t=s}^{S-1} p_t \right) [u(x_s) - u(x_{s-1})]. \quad (13.7)$$

This welfare functional defines the rank-dependent expected utility model. Preferences are referred to as "rank dependent" because a change in the ranking of the outcomes (x_0, \dots, x_{S-1}) on the real line would affect the functional. The only case in which there is no effect of a change in the ranking is when

f is the identity function. In this particular case, equation (13.7) can be rewritten as

$$V(L) = \sum_{s=0}^{S-1} p_s u(x_s)$$

which of course is the Expected-Utility criterion. Thus, EU is the special case of RDEU with a linear transformation function of decumulative probabilities. We can now examine the effect of a non-linear probability transformation function f on choices made under uncertainty.

First, observe that the RDEU criterion does not satisfy the independence axiom. Therefore, it does not fall prey to the criticism of Allais' paradox. To see this, observe that a RDEU decision maker prefers M_1 over M_2 if and only if

$$u(50) > u(0) + f(0.99) [u(50) - u(0)] + f(0.1) [u(250) - u(50)],$$

or equivalently,

$$u(50) > (1 - \pi_1)u(0) + \pi_1 u(250) \quad (13.8)$$

with $\pi_1 = f(0.1)/(1 + f(0.1) - f(0.99))$. On the other hand, under RDEU, preferring \widehat{M}_2 over \widehat{M}_1 implies that

$$u(0) + f(0.11) [u(50) - u(0)] < u(0) + f(0.1) [u(250) - u(0)],$$

or, equivalently,

$$u(50) < (1 - \pi_2)u(0) + \pi_2 u(250) \quad (13.9)$$

with $\pi_2 = f(0.1)/f(0.11)$. Contrary to the case where f is the identity function (EU), conditions (13.8) and (13.9) are not incompatible.

These two conditions, (13.8) and (13.9), can be reinterpreted within an EU framework to be saying that an EU-maximizer with utility u has a certainty equivalent for lottery $N_1 \sim (0, 1 - \pi_1; 250, \pi_1)$ that is smaller than 50, whereas he has a certainty equivalent for lottery $N_2 \sim (0, 1 - \pi_2; 250, \pi_2)$ that is larger than 50. Of course, a necessary condition for this to be possible is that π_2 be larger than π_1 . Since $f(1) = 1$, it follows that

$$\frac{1}{2}f(0.1) + \frac{1}{2}f(1) > \frac{1}{2}f(0.11) + \frac{1}{2}f(0.99).$$

Because $(0.1, 1/2; 1, 1/2)$ is a mean-preserving spread of $(0.11, 1/2; 0.99, 1/2)$, this suggests that f must be convex.

We also can examine the conditions under which a RDEU-maximizer behaves in a risk-averse manner. We will do this first for a special case which is dual to EU theory, where u is the identity function. In this "dual theory" first examined by Yaari (1987), the agent is risk-neutral when f is also linear. Remember that an agent is risk-averse if the certainty equivalent of any risk is smaller than its mean. Consider the binary risk $(x_0, p_0; x_1, 1 - p_0)$ with $x_0 < x_1$. In this case, we can use condition (13.5) to say that the agent is risk averse if and only if the following condition holds:

$$x_0 + f(1 - p_0)[x_1 - x_0] \leq x_0 + (1 - p_0)[x_1 - x_0]. \quad (13.10)$$

Of course, this is true if and only if $f(1 - p_0) \leq 1 - p_0$. Since p_0 is arbitrary, we conclude that risk aversion holds in the dual theory if $f(p) \leq p$ for all $p \in [0, 1]$. It is easy to check that this is also the necessary and sufficient condition when we relax the condition that the lottery is binary. The intuition is that lowering the probability transformation function, one reduces the welfare effect of the potential increments in utility above $u(x_0)$. Notice that there is a clear link between the convexity of f and this condition of risk aversion. Indeed, $f(p) < p$ is necessary (but not sufficient) for the convexity of f , taking into account the conditions that $f(0) = 0$ and $f(1) = 1$.

Two important remarks must be made here. First, in the RDEU model, concavity of u does not imply that the agent dislikes all mean-preserving spreads of outcomes. As shown by Chew, Karni, and Safra (1987) in the general RDEU model, this more demanding concept of risk aversion also requires that f be convex. That is, in the general RDEU model, the agent is averse to any mean-preserving spread if and only if both u is concave and f is convex. Under this condition, any result that we presented in this book whose proof relied exclusively on aversion to mean-preserving spreads remain true in the RDEU framework. This is the case for example for the optimality of a straight deductible contract in insurance, as proven in Proposition 16. It can be shown that this is also the case for the efficiency of the mutuality principle in the allocation of risk. Indeed, the mutuality principle guarantees that there exists no reallocation of risk that generates a mean-preserving contraction of all individual consumption plans in the economy.

Second, it is important to note that whenever f is non-linear, risk aversion will yield a first-order effect. Consider an agent with initial wealth w_0 who

must bear a zero mean risk $(-k, 1/2; k, 1/2)$, where $k > 0$ can be interpreted as the size of the risk. In the dual theory, welfare can be measured as

$$w_0 - k + f(1/2)2k = w_0 - k(1 - 2f(1/2)). \quad (13.11)$$

This is the certainty equivalent of the risky wealth of the agent. Under the condition of risk aversion, $f(1/2) < 1/2$ and the certainty equivalent is smaller than the expected wealth w_0 . But the interesting point here is to observe that the risk premium $(1 - 2f(1/2))k$ is increasing proportionally with the size of the risk. This is typical of first-order risk aversion. This is quite different from the EU model where the risk premium is approximately proportional to the square of the size of the risk (second-order risk aversion). This has several implications. For example, risk-averse RDEU investors may prefer to stay 100% invested in bonds in spite of a positive equity premium. Or, a risk-averse RDEU consumer may prefer to purchase full insurance in spite of a positive loading factor on the insurance premium.

When there are only two states of nature with outcomes x_0 and x_1 , the RDEU welfare functional can be written as

$$V = \begin{cases} (1 - f(1 - p_0))u(x_0) + f(1 - p_0)u(x_1) & \text{if } x_0 < x_1 \\ f(p_0)u(x_0) + (1 - f(p_0))u(x_1) & \text{if } x_0 \geq x_1. \end{cases} \quad (13.12)$$

In Figure 13.1, we depict the indifference curves associated to these preferences. The main difference with EU curves is that there is a kink at the 45° line that illustrates first-order risk aversion. As shown in this figure, this kink allows for the optimality of full insurance even when the insurance premium is actuarially unfair.

INSERT FIGURE 13.1 ABOUT HERE

13.3 Ambiguity aversion

The linearity of the EU functional with respect to probabilities implies that EU-maximizers are neutral with respect to any uncertainty on the state probabilities. To illustrate this feature, let us reconsider the framework presented in section 8.1.2. To simplify the analysis, since we are only concerned with probability changes, we fix the set of possible outcomes (x_0, \dots, x_{S-1}) , so that a (simple) lottery can be represented by the vector of probabilities $P = (p_0, \dots, p_{S-1})$ associated to these outcomes. We now introduce some uncertainty that affects this vector of probabilities. Suppose that there is some

hidden information that would affect our beliefs about the distribution of outcomes. If this information would become known, the agent could revise his beliefs of the distribution of outcomes. This can be formalized by assuming that there are M possible signals $m = 1, \dots, M$. If signal m is observed, the revised probability distribution of outcomes would be $P^m = (p_0^m, p_1^m, \dots, p_{S-1}^m)$. If q^m denotes the probability of receiving signal m , a (compound) lottery would be described by the vector $(P^1, q^1; \dots; P^M, q^M)$. This is a situation where there is some uncertainty about the true distribution of outcomes, which we describe as a type of "ambiguity" about the risk.

The agent must choose between different compounded lotteries. Contrary to what was assumed in section 8.1, we assume here that the agent must select his preferred lottery *before* observing the signal. To illustrate, this is the kind of decision problem that we face when determining whether to reduce greenhouse gas emissions in spite of our limited knowledge about the effect of these gas. The case of genetically modified food and new drugs are similar. An EU-maximizer would evaluate his welfare as follows:

$$\begin{aligned} V(P^1, q^1; \dots; P^M, q^M) &= \sum_{m=1}^M q^m \sum_{s=0}^{S-1} p_s^m u(x_s) \\ &= \sum_{s=0}^{S-1} p_s^0 u(x_s), \end{aligned}$$

where $p_s^0 = \sum_m q^m p_s^m$ is the prior probability of state s . This means that agents should make their choices as if they would face no uncertainty about the distribution $P^0 = (p_0^0, p_1^0, \dots, p_{S-1}^0)$ of outcomes. This implies that EU agents should be indifferent between two compounded lotteries with the same prior distribution. Because of the linearity of the EU functional with respect to probabilities, the uncertainty on probabilities has no effect on welfare.

This can be illustrated by the following example adapted from Ellsberg (1961). Consider an urn containing 90 balls. Thirty of them are red balls. The remaining sixty balls can be either black or white. The proportion of black and white balls is not announced. It is common knowledge that the number $m \in \{0, \dots, 60\}$ of black balls is taken at random from a uniform distribution ($q^m = 1/61$). A ball is taken out from the urn, and the player gets a prize that depends upon the color of that ball, as described in the following table. The player is confronted with a choice between the lotteries L^a and L^b prior to getting any information about the composition m of the

urn. Ellsberg observed in laboratory experiments that many players prefer L^a over L^b . This is not compatible with the EU model since the two lotteries L^a and L^b have the same prior probability distribution $(0, 2/3; 50, 1/3)$. It suggests that the player penalizes lottery L^b because of the ambiguity on the probability of success.

Lottery	Red	Black	White
L^a	50	0	0
L^b	0	50	0
M^a	50	0	50
M^b	0	50	50

Prize as a function of the color of the ball

This can be confirmed by considering the choice between lotteries M^a and M^b using the same urn. We observe that many players who preferred L^a over L^b also prefer M^b over M^a . This is in spite of the fact that M^a and M^b have the same prior distribution $(0, 1/3; 50, 2/3)$. This also suggests some form of aversion to ambiguous probabilities, since the choice of M^a yields an ambiguous probability of success, whereas M^b has unambiguous probabilities. Only their expected probabilities are the same.

Under Expected Utility Theory, the first choice can be explained only by assuming that the probability of a black ball with L^b be less than $1/3$. Thus, a proponent of the EU-model could claim that players prefer L^a over L^b because they do not trust the experimenter about the composition of the urn. This absence of confidence could induce players to believe that the experimenter biased the probability of the winning black balls below $1/3$. However, such beliefs are incompatible with the preference $M^b \succ M^a$, which requires that the probability of a black ball be larger than $1/3$.

Gilboa and Schmeidler (1989) proposed a decision criterion that can explain this paradox. The decision maker computes his expected utility for each possible posterior probability distribution P^m . His welfare is measured as the minimum of these various expected utility valuations:

$$V(P^1, q^1; \dots; P^M, q^M) = \min_{m=1}^M \sum_{s=0}^{S-1} p_s^m u(x_s). \quad (13.13)$$

In a sense, players believe that they play against nature which always reacts to their choices by selecting the worst possible probability distribution. This

is an extreme form of pessimism. It does have the advantage that the players do not need to know the distribution (q^1, \dots, q^M) of the hidden information. When there are only two states with outcomes x_0 and x_1 , the welfare function can be written as

$$V = \begin{cases} p_0^{\max}u(x_0) + (1 - p_0^{\max})u(x_1) & \text{if } x_0 < x_1 \\ p_0^{\min}u(x_0) + (1 - p_0^{\min})u(x_1) & \text{if } x_0 \geq x_1, \end{cases} \quad (13.14)$$

where $p_0^{\max} = \max_m p_0^m$ and $p_0^{\min} = \min_m p_0^m$. Observe that this function is equivalent to the RDEU function (13.12) if we set $1 - f(1 - p_0) = p_0^{\max}$ and $f(p_0) = p_0^{\min}$. It follows that the behavioral properties of ambiguity aversion are much the same as those generated by rank-dependent expected utility. For example, ambiguity aversion implies that full insurance may be optimal even if the insurance premium is actuarially unfair based on the prior distribution of loss.

Ambiguity aversion can explain choices $L^a \succ L^b$ and $M^b \succ M^a$. Indeed, from (13.13), the welfare generated by lottery L^b equals $u(0)$ whereas the welfare generated by lottery L^a equals $(1/3)u(50) + (2/3)u(0) > u(0)$. Similarly, M^b is preferred to M^a because

$$\frac{2}{3}u(50) + \frac{1}{3}u(0) > \frac{1}{3}u(50) + \frac{2}{3}u(0).$$

This illustrates how the decision criterion of Gilboa and Schmeidler (13.13) exhibits aversion to ambiguity.

13.4 Prospect theory and loss aversion

Some authors, such as Kahneman and Tversky (1979), have suggested that there is a discontinuity of the marginal utility at some wealth level W . This is the case for example with the following utility function:

$$u(w) = \begin{cases} aw & \text{if } w \leq W \\ aW + b(w - W) & \text{if } w > W. \end{cases} \quad (13.15)$$

If $a > b$ as in Figure 13.2, this function is globally concave in wealth. Using laboratory experiments, Kahneman and Tversky claimed that a is approximately twice as large as b . This means that a one-euro loss has approximately the same effect on welfare as a gain of two euros, in absolute value. We say that the agent is loss-averse in this case.

INSERT FIGURE 13.2 ABOUT HERE

Notice that the Arrow-Pratt approximation cannot be used here since it requires that the utility function be differentiable. In fact, loss aversion is a case of first-order risk aversion at W . This can be checked by evaluating the risk premium for an agent with utility function (13.15) and initial wealth $w_0 = W$ who faces risk $\tilde{x} \sim (-k, 1/2; +k, 1/2)$. When $a > b$, it can be checked that

$$Eu(w_0 + \tilde{x}) = u\left(w_0 - \frac{k}{2}\left(1 - \frac{b}{a}\right)\right),$$

which implies a positive risk premium $\Pi = 0.5(1 - (b/a))k$. This is linear with the size of risk k . When initial wealth is at a point of non-differentiability of a piecewise utility function, the risk premium is proportional to the size of risk. Small risks matter.

Absolute risk aversion is zero everywhere except at W where it is not defined. However, utility function (13.15) can be approximated by a smooth function that exhibits zero risk aversion everywhere except in a small neighborhood of W where absolute risk aversion tends to infinity. In this sense, this function exhibits increasing absolute risk aversion below W , and decreasing absolute risk aversion above it. In their Prospect Theory, Kahneman and Tversky (1979) proposed a S-shaped utility function with a kink at perceived initial wealth at W , a convex branch below W and a concave branch above W . This is illustrated in Figure 13.3. The kink at the current wealth level, yielding first-order risk aversion, also means that the agent has a prospective behavior: what matters for welfare is not how much wealth he has but rather how wealth changed compared to his "reference wealth" W . The convexity of u in the loss region means that an uncertain loss is preferred to a sure loss. The concavity of u in the gain region means that the agent prefers a sure gain to an uncertain gain.

INSERT FIGURE 13.3 ABOUT HERE

The key departure of note in Prospect Theory is that utility u as defined above is conditional on the reference wealth level W . Their hypothesis is that decision makers frame each decision with reference to some status quo or some initial wealth level W . Thus, the framing of the problem at hand becomes crucial. Another economic decision, or even the same problem with a different framing will lead to a different objective and, hence, a different optimal solution. For example, suppose you just found out that you won 10,000 Euros, which was on its way to you as cash in the mail, but with a

50 percent chance that it will never arrive. Whether you frame the problem as having an initial wealth of 10,000 with a 50 percent chance of losing it all, or as having an initial wealth of zero with a 50 percent chance of gaining 10,000, will make a difference in your decision choices under prospect theory.

In this static framework with a fixed reference point, the theory of loss aversion is nothing else than a particular case of the EU model. Prospect theory is enriched in a dynamic framework by the idea that the reference point W changes with time. For example, W can be the wealth level of the previous period. Under this specification, the hypothesis is that consumers and investors extract direct utility from changes in wealth over time. We can also enrich the model by combining this assumption with rank-dependent transformations of cumulative probabilities, yielding the so-called Cumulative Prospect Theory (Tversky and Kahneman (1992)).

13.5 Some Concluding Thoughts

Although the expected-utility model has received much criticism in the literature, it remains as a cornerstone for modern research. Part of the reason for this might be that competing theories to date each suffer some flaws of their own. A current trend in economics and in finance is to take more of a "behavioral" view of decision making. However, it seems that modeling human behavior with any precision will remain an impossible task.

Consider, for example, that in May 2000 seven states in the USA jointly sponsored the "Big Game Lottery," which had the largest payout ever for a lottery at \$363 million. The lottery brought in a total revenue of \$565 million. Thus, the mean expected payout on every dollar invested in this lottery was less than 65 cents. Theory thus implies that millions of people, everyone who purchased lottery tickets, must have been risk loving. Now consider that, for this particular lottery, the main prize was split equally between two winners, who were two of the many participants who exhibited this risk-loving behavior. Suppose that these two winners were offered a chance to flip a fair coin one time, to award the entire \$363 million to either one or the other winner. Would either of these winners prefer the coin toss to simply taking their one-half share of \$181.5 million?

Of course buying lottery tickets provides some satisfaction in and of itself. In particular, as stressed by Caplin and Leahy (2001), buyers may savor the anticipatory feeling of winning the big prize, before the name of the winner

is announced. Who would not like to daydream about winning \$363 million? We doubt that purchasing a lottery ticket by itself is an indication of risk-loving behavior, although this example shows how any model, with EU theory being no exception, must be used with caution.³ One approach to this dilemma is to model decision methodology as unique to the individual situation at hand, which unfortunately does not allow one to make any positive predictions or normative assessments.

In presenting expected-utility modeling throughout this text, the authors make no claim that it gives an accurate description in every situation. However does the theory give us any usable information as to how market decisions are made and how financial assets are priced? We believe it does. Moreover, economic and financial decisions are typically modelled in isolation, whereas many complicated decisions in the real-world are often intertwined. Using EU theory as a starting point and adapting the theory by including such modifications as acknowledging background risks or applying hyperbolic discounting can hopefully improve the explanatory power of decision models.

13.6 Bibliographical references and extensions

This chapter presents a perspective on the development of non-expected utility models. Much of the relevant literature already has been reviewed in the text. An excellent and easily readable critique on the EU model can be found in the paper by Machina (1987). In a recent counting, the number of decision-criteria alternatives to the EU model is larger than 40. Our presentation here obviously covers only a small part of this large and growing literature. However, the RDEU model is often considered as the most promising alternative to EU theory, together with variations on Prospect Theory as developed by Kahneman and Tversky (1979). These two models have been combined by Tversky and Kahneman (1992) in a model known as Cumulative Prospect Theory. An abundance of the behavioral arguments against EU theory, as well as against many other decision models are presented in Kahneman (2003). A very lucid set of arguments as to why these behavioral

³Rabin (2000), for example, makes a strong case against using expected-utility for small gambles. Based only upon the concavity of utility u , he show how even mild risk aversion for small slightly unfavorable gambles leads to implausible rejections of large highly favorable gambles.

objections really only call for the modification, rather than the destruction, of existing theories, is presented in Glaeser (2003).

References

- Allais, M., (1953), Le comportement de l'homme rationnel devant le risque, Critique des postulats et axiomes de l'école américaine, *Econometrica*, 21, 503-46.
- Caplin, A.J., and J. Leahy, (2001), Psychological expected utility theory and anticipatory feelings, *Quarterly Journal of Economics*, 106, 55-80.
- Chew, S., E. Karni and Z. Safra, (1987), Risk aversion in the theory of expected utility with rank dependent preferences, *Journal of Economic Theory*, 42, 370-381.
- Ellsberg, D., (1961), Risk, ambiguity, and the Savage axioms, *Quarterly Journal of Economics*, 75, 643-69.
- Gilboa, I. and D. Schmeidler, (1989), Maximin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18, 141-153.
- Glaeser, E., (2003), Psychology and the market, *NBER Working Paper No. 10203*.
- Kahneman, D., (2003), Maps of bounded rationality: Psychology for behavioral economists, *American Economic Review*, 93, 1449-1475.
- Kahneman, D., Tversky, A., (1979), Prospect Theory: An analysis of decision under risk, *Econometrica*, 47, 263-291.
- Machina, M., (1987), Choice under uncertainty: Problems solved and unsolved, *Journal of Economic Perspectives*, 1, 121-54.
- Rabin, M., (2000), Risk aversion and expected-utility theory: A calibration theorem, *Econometrica*, 68, 1281-1292.
- Tversky, A., and D. Kahneman, (1992), Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 5, 297-323.
- Quiggin, J., (1982), A theory of anticipated utility, *Journal of Economic Behavior and Organization*, 3, 323-343.

- von Neumann, J. and O. Morgenstern, (1944), *Theory of Games and Economic Behavior*, Princeton: Princeton University Press. 2nd Ed., 1947.
- Yaari, M.E., (1987), The dual theory of choice under risk, *Econometrica*, 55, 95-115.