Risk

These notes explore the conceptual tools that help investors in assessing the "riskiness" of their portfolios. We ask: Given two assets or portfolios A and B can we say which one is "riskier"? This turns out to be closely linked to the more general question of "portfolio choice". We ask: If an investor has the choice between two mutually exclusive portfolios, A and B, which of these two portfolios does he prefer? Can we specify certain general patterns of investment choice, whereby all "sensible" investors can be expected to prefer A over B if A has certain well-defined properties relative to B. For historical reasons, this whole area of investigation is called *utility theory* — we ask, what portfolio choices will give the highest "utility" to an investor if all investor's have similar preferences.

As a special application of this kind of query, we can then assess relative riskiness. Given a choice between two mutually exclusive portfolios A and B, both having the same expected payoffs (average payoffs), which of these portfolios will a "risk-averse" investor choose? We will explore three different notions of risk aversion that help answering such questions: (i) preference for certainty; (ii) variance aversion; (iii) stochastic dominance.

Keywords: Roy's Safety First Rule, Value at Risk, Domar-Musgrave Risk Measure, Conditional Value at Risk, Variance and Semi-Variance, Lower Partial Moments, Coherent risk Measures, Risk Aversion, Risk Premia, 1st and 2nd order stochastic dominance.

Reading: Chapter 18 from Hull as a supplement to these notes.

1 Background

This introductory section positions the topic of "Portfolio Choice" within the wider context of the course and sorts out a few preliminary issues.

Partial vs Total Payoffs: Investors buy various individual *assets* and combine these assets into overall *portfolios*. An investor cannot form a portfolio unless he is willing to purchase the individual assets out of which it is composed, but these individual assets are relevant to the investor only insofar as they contribute to the payoff profile of the overall portfolio of which they are a part. Ultimately, the investor ends up with one particular portfolio rather than any other portfolio, and his final income will then be given by the payoffs of this particular portfolio.

If we wish to stress that we are talking about the investor's ultimate overall portfolio, we talk of the "total payoffs" of the investor's overall portfolio (as opposed to the partial payoffs coming from individual assets), and we speak of "mutually exclusive investments" that are open to the investor (each of which offering an alternative "total payoff" to the investor).

Using assets to hedge risk: It is important to make the distinction between the riskiness of an asset and its value. The value of an individual asset depends on the relationship between the payoffs of that asset and the payoffs of the other assets in the investor's overall portfolio. Thus it is possible for an individual asset to be very risky and yet to be very desirable; an asset is desirable if its risky payoffs are *anti-correlated* with the payoffs of the other assets in the investor's overall portfolio. Such anti-correlated assets provide the opportunity to *hedge* risk and allow the investor to reduce the risk of his/her overall portfolio. Investors may not wish to hold such hedging assets in isolation, but they may wish to hold them as one component of their overall portfolios.

Hedging is a very important part of understanding how derivative securities and financial instruments are priced. We will see that the use of hedging assets can offset risk completely and allow the price of derivatives to be determined by relatively simple formulas.

Overall Income Risk: Stochastic Dominance: The present handout is concerned with another question. Here we no longer ask the question "How well does the riskiness of this particular individual asset combine with the riskiness of the other assets in the investor's portfolio?". Instead, we present the investor with a range of various complete or overall portfolios, neither of which may be altered, and ask the investor to choose one of them, reject all others, and subsequently live with the payoffs from the one portfolio that he has chosen. Thus we ask: "How does the riskiness of one particular overall portfolio (to which no further assets may be added) compare with the riskiness of some other, alternative such portfolio?". We ask: when given the choice between two mutually exclusive overall portfolio payoffs, which one does the investor regard as riskier (and hence less desirable) than the other?

How desirable is one overall portfolio compared with another? If investors are "risk-averse", then they will prefer "less risky" over "more risky" portfolios with the same average payoffs. What does it mean that a portfolio is "less risky" than another? There are a number of possible answers to this question and we will consider some of the alternatives that have been proposed. The concept of (second-order) *Stochastic Dominance* defined below is however the most important as it gives a theoretically rigorous answer. If portfolio A stochastically dominates portfolio B, then all "risk-averse" investors are well advised to prefer A over B; they should choose portfolio A and reject portfolio B. This however, does depend on what is meant by "risk-averse" and this we will consider later.

2 Preliminaries

In this section we set up the formal framework, specify a concrete example, and perform various preliminary computations for this example. Later sections will then offer formal definitions of investor preferences and refer back to the example from the present section.

Risky Investments: We consider an investor with a single-period time horizon. The investor has the choice between a number of mutually exclusive investments, $A, B, C, \ldots$. The payoffs of these investments depend on the state of the world. When deciding on his investment, the investor does not know which of the various possible states of the world will come about: payoffs are *risky*. However the investor knows the relative likelihood (probability) of each state and hence is capable of computing the expected value, variance and probability distribution of each asset.

The Choice Problem: At the beginning of the period the investor chooses one of the available investments. He is then tied to the particular investment he has chosen and his end-of-period income will consist entirely of the payoffs from his chosen investment. How large will his income be? That depends on the state of the world during the period, which cannot be foreseen at the beginning of the period but which can be analysed in terms of likelihoods (probability distributions).

When comparing any two of the various available investments, the investor needs to ask himself: which of these two would I rather have? In answering this question, the investor will take account of the knowledge he has about payoff profiles of these investments. We consider the following concepts in assessing a payoff profile: expected payoffs, variance of payoffs, worst-case payoffs, and "dispersion" of payoffs. Later we will use these concepts to declare precise definitions of risk aversion that reflect the choices of a "reasonable" investor.

Example: By way of example, we consider an economy with *eight* states of the world. For simplicity, all eight states are equally likely, so that the probability of state i is $f_i = \frac{1}{8}$ (for $i = 1, 2, \ldots 8$). We consider the following six mutually exclusive investments:

$$A = \begin{pmatrix} 9 \\ 9 \\ 9 \\ 9 \\ 9 \\ 9 \\ 9 \\ 9 \end{pmatrix};\ B = \begin{pmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 9 \\ 9 \\ 9 \\ 9 \end{pmatrix};\ C = \begin{pmatrix} 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \end{pmatrix};\ D = \begin{pmatrix} 14 \\ 13 \\ 12 \\ 11 \\ 9 \\ 8 \\ 7 \\ 6 \end{pmatrix};\ E = \begin{pmatrix} 18 \\ 16 \\ 14 \\ 12 \\ 8 \\ 6 \\ 4 \\ 2 \end{pmatrix};\ F = \begin{pmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 3 \end{pmatrix}.$$

Casual inspection of these six investments reveals that A and C are risk-free, all others are risky. We will be particularly interested in studying and comparing D, E and F.

State-wise dominance: We say that investment X *state-wise dominates* investment Y if X pays more than Y in at least one of the states and not less than Y in any other state. Investment X then is the *dominating* investment and Y "is dominated" (by X). We shall write $X \geq Y$ for shorthand.

In our example, B dominates A, since it pays 2 extra in states 1–4 and the same as A in states 5–8. Likewise, C dominates A. No other investment dominates any other. For instance, D does not dominate E since it pays less than E in states 1-4. In general, if two investments have the same expected payoffs, neither will dominate the other, since any extra payoff in one of the states must be compensated by shortfalls in some of the other states.

Probability Distributions: To determine how risky a portfolio or asset is, we will need to know the *probability distributions* of its payoffs.

An asset or portfolio will result in n possible financial outcomes (sums of money or returns), $x_1, x_2, \ldots, x_n$. We shall denote the probability of outcome

x_i by f_i. If the portfolio is risky then for at least one value x_i, the probability satisfies $0 < f_i < 1$. In this case we can think of the outcome x as a *random variable* taking on the values $x_1, x_2, \ldots, x_n$ so that

$$f_i = \text{Prob}(x = x_i).$$

From the example we see that n can be less than the number of states. So for example portfolio F has $n = 2$ with $x_1 = 3$, $x_2 = 11$.[1] The *probability mass function* $f(x_i)$ of an investment assigns to each possible payoff x_i of the investment its relative likelihood or probability. The value $f_i = f(x_i)$ of the mass function is the likelihood that the investment will have a payoff of x_i pounds. We find the value of $f(x_i)$ by summing over the likelihoods of all those states where the investment pays exactly x_i pounds. In our example, for all six investments, their state-specific payoffs lie somewhere between 1 and 20, and thus we may consider the relevant range of x to be the interval $[1, 20]$, for all mass functions. For the risk-free investment A, payoff $x = 9$ occurs always and all other payoffs occur never, and thus we have

$$f^a(9) = 1; \quad f^a(x) = 0 \text{ for all other } x.$$

Likewise for C:

$$f^c(10) = 1; \quad f^c(x) = 0 \text{ for all other } x.$$

What about the mass functions of the risky investments? For B, we find that it pays 9 in four out of eight states, 11 in the other four states. Since all states are equally likely, this means that

$$f^b(9) = \tfrac{1}{2}, \quad f^b(11) = \tfrac{1}{2}, \quad f^b(x) = 0 \text{ for all other } x.$$

[1] There is an implicit assumption here that states do not matter in themselves and it is only the payoffs that matter. Two assets that have the same probability distribution even when the payoffs occur in different states are said to be equally in distribution and the assumption that investors will rank assets that are equal in distribution equally is called the assumption of probabilistic sophistication.

Likewise for D, E and F. We have:

$$f^d(x) = \tfrac{1}{8} \text{ for } x = 6, 7, 8, 9, 11, 12, 13, 14; \quad f^d(x) = 0 \text{ for all other } x.$$

$$f^e(x) = \tfrac{1}{8} \text{ for } x = 2, 4, 6, 8, 12, 14, 16, 18; \quad f^e(x) = 0 \text{ for all other } x.$$

$$f^f(x) = \tfrac{1}{8} \text{ for } x = 3; \quad f^f(x) = \tfrac{7}{8} \text{ for } x = 11.$$

In some circumstances we will also think of x as a continuous variable taking on any value in some range $[a, b]$. In that case $f(x)$ is a continuous function and is known as a *probability density function*.

Probability Mass Function: It is instructive to draw the histograms for the mass functions of the various investments, with payoffs on the horizontal axis and probabilities on the vertical axis. Here we show C with D, D with E, and D with F.

<u>Exercise</u>: Draw the probability mass functions.

Cumulative Probability Functions: The cumulative probability function gives the probability that the payoff is less than or equal to a particular value. We shall denote the cumulative probability by

$$F_i = F(x_i) = \text{Prob}(x \leq x_i).$$

For a discrete distribution we have

$$F(x_i) = \sum_{j=1}^{i} f_j$$

and for a continuous distribution we have

$$F(x) = \int_a^x f(t)\, dt.$$

The cumulative probability distribution $F(x)$ is the area below the probability density function $f(t)$ below the value of x.

<u>Exercise</u>: Draw the cumulative probability functions.

Expected Payoffs: As a first measure of investment performance, we are interested in the *average* payoffs (expected payoffs) of an investment. The expected payoffs of the risk-free investments are simply their fixed payoffs:

$$\mathrm{E}[A] = 9; \quad \mathrm{E}[C] = 10.$$

For the other investments, we recall that all states are equally likely, hence we merely need to sum over all payoffs and divide by the number of states:

$$\mathrm{E}[B] = \frac{11+11+11+11+9+9+9+9}{8} = 10;$$
$$\mathrm{E}[D] = \frac{14+13+12+11+9+8+7+6}{8} = 10;$$
$$\mathrm{E}[E] = \frac{18+16+14+12+8+6+4+2}{8} = 10;$$
$$\mathrm{E}[F] = \frac{11+11+11+11+11+11+11+3}{8} = 10.$$

Thus, all investments except A have the same average payoffs of 10. On this measure, the five investments B–F are all equally desirable.

Deviations: How large are the deviations of actual payoffs from the mean? For the risk-free investments (A and C), deviations are zero in all states. For the others, we have:

$$\Delta B = \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}; \ \Delta D = \begin{pmatrix} +4 \\ +3 \\ +2 \\ +1 \\ -1 \\ -2 \\ -3 \\ -4 \end{pmatrix}; \ \Delta E = \begin{pmatrix} +8 \\ +6 \\ +4 \\ +2 \\ -2 \\ -4 \\ -6 \\ -8 \end{pmatrix}; \ \Delta F = \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ -7 \end{pmatrix}.$$

A quick calculation shows that each of these deviations has an average of zero;

$$\mathrm{E}[B] = \mathrm{E}[D] = \mathrm{E}[E] = \mathrm{E}[F] = 0,$$

confirming that on average, the positive and negative deviations cancel each other out.

Casual inspection of these deviations shows that E tends to have much larger deviations than the other investments. F had mostly small positive deviations but one fairly large negative one (in the last state), whereas D has variable deviations that are smaller than those of E but never as large as the largest (negative) one from F.

Variance: Variance is the average of the squared deviations, and as such it is an important summary measure of risk.

$$\Delta B^2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}; \ \Delta D^2 = \begin{pmatrix} 16 \\ 9 \\ 4 \\ 1 \\ 1 \\ 4 \\ 9 \\ 16 \end{pmatrix}; \ \Delta E^2 = \begin{pmatrix} 64 \\ 36 \\ 16 \\ 4 \\ 4 \\ 16 \\ 36 \\ 64 \end{pmatrix}; \ \Delta F^2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 49 \end{pmatrix}.$$

Hence,

$$\mathrm{Var}[B] = \mathrm{E}[\Delta B^2] = \frac{1+1+1+1+1+1+1+1}{8} = \frac{8}{8} = 1;$$

$$\mathrm{Var}[D] = \mathrm{E}[\Delta D^2] = \frac{16+9+4+1+1+4+9+16}{8} = \frac{60}{8} = 7.5;$$

$$\mathrm{Var}[E] = \mathrm{E}[\Delta E^2] = \frac{64+36+16+4+4+16+36+64}{8} = \frac{240}{8} = 30;$$

$$\mathrm{Var}[F] = \mathrm{E}[\Delta F^2] = \frac{1+1+1+1+1+1+1+49}{8} = \frac{56}{8} = 7.$$

This we find that *on average*, as expected E has the largest deviations, but F has a smaller average deviation than D, despite having one very large negative deviation in the last state. Based on this measure, investment F is the most desirable of the last three investments, closely followed by D, and then much later by E.

Worst-case Payoffs:

If your life income depends exclusively on one particular investment, you will not just be interested in the *average* payoffs and *average* risk of that investment but you will also be interested in the *actual* payoffs in the various states. Ultimately, you will receive one of these actual payoffs, rather than the average payoff. In particular, you will want to to know: What is the *lowest possible payoff* of my investment? This is the “worst-case” payoff. Worst-case payoffs are the payoffs which have the largest negative deviation from the mean.

In our example, the worst-case payoffs are 9 for A, 9 for B, 10 for C, 6 for D, 2 for E, 3 for F. We note that for a risk-free investment, worst-case payoffs are always equal to average payoffs, since for risk-free investments, payoffs are the same in all states and all deviations are zero.

3 Measures of Risk

We now want to ask the question whether risk can be quantified and whether the riskiness of various investments can be compared. We shall give a qualified and subtle yes to this question. It is possible to make comparisons between the riskiness of some pairs of portfolios but not others. Thus we shall have a partial ordering of portfolios according to their riskiness. Before proceeding we shall consider various measures of risk which have been considered in the literature.

Roy's Safety First Rule

Roy in a classic paper[2] proposed the following measure of risk

$$\rho_R = \text{Prob}(x \leq \hat{x}).$$

This gives the probability that the outcome is less than some critical or disaster level $\hat{x}$ and thus the risk measure ρ_R measures the probability of disaster. Roy's risk measure is also known as the *shortfall probability.* The safety-first principle is then to choose a portfolio that minimises this probability of disaster. For a continuous distribution the Roy measure is simply the cumulative probability at $\hat{x}$,

$$\rho_R = F(\hat{x}) = \int_{x_{min}}^{\hat{x}} f(x)\, dx.$$

An upper bound can be found for this probability. Letting

$$\mu = \int x f(x)\, dx$$

denote the mean of the distribution and

$$\sigma^2 = \int (x - \mu)^2\, dx$$

denote the variance of the distribution, it can be shown by Chebycheff's inequality that

$$\rho_R = F(\hat{x}) \leq \frac{\sigma^2}{(\mu - \hat{x})^2}.$$

Roy suggested that if the entire distribution is not known then choosing the investment portfolio which minimised the ratio $\sigma^2/(\mu - \hat{x})^2$ would be desirable. Or equivalently the portfolio which maximised the square-root of the inverse. Taking $\hat{x}$ to be the risk-free payoff r, Roy would suggest choosing the portfolio which maximised $(\mu - r)/\sigma$ would be desirable. This later is

[2] Andrew D. Roy, "Safety first and the holdings of assets", *Econometrica*, Volume 20, No. 3 (July 1952), pp.431-449.

just the reward to variability or *Sharpe ratio* introduced by Sharpe (1966)[3] which is widely used by professional investors.

One problem with Roy's measure is that it looks only at probabilities and not the size of the loss. Thus suppose that there are two portfolios and take $\hat{x} = 0$ so that any negative payoff is considered a disaster. Suppose portfolio A has a payoff of -100 with probability 0.01 and that all other payoffs are positive. Similarly suppose that portfolio B has a payoff of -1 with probability 0.02 and all other payoffs are positive. Roy's index is higher for portfolio B though arguably portfolio A is potentially more disastrous.

Value at Risk

Value at Risk (VaR) is the inverse of Roy's risk measure. Whereas Roy's measure asks what is the probability of a loss below a particular level $\hat{x}$, VaR determines the value $\hat{x}$ such that the portfolio will suffer a loss of more that $\hat{x}$ with a given probability. It is usual to fix the probability of loss at 1% or 5% and calculate the return distribution say for one day so that the VaR measures the minimum loss likely to be suffered on 1 or 5 of the next 100 trading days. Using the cumulative distribution function we have

$$\alpha = F(VaR)$$

or

$$VaR = F^{-1}(\alpha).$$

As with Roy's measure there are some problems with the use of VaR as a measure of risk. It only considers the distribution in the lower tail and will therefore ignore differences in the distribution above the VaR threshold. The VaR measure is not *sub-additive* which means that the it may be possible

[3] William Forsyth Sharpe, "Mutual Fund Performance," *Journal of Busines*, Volume 39, No. 1 (January 1966), pp.119-138.

Prob	A	B	A+B
0.03	-9	1	-8
0.94	1	1	2
0.03	1	-9	-8
VaR(5%)	-1	-1	8

Table 1: VaR is not sub-additive

to combine two assets or portfolio and increase the VaR. Table 1 shows two portfolios A and B and calculates the VaR at the 5% level for each. Since the probability of any loss is 3%, the maximum value of the portfolio in the lowest 5% of the distribution is 1 in both cases. Thus VaR, which is the *loss* at risk is negative for both portfolios. However combining the two portfolios produces a 6% probability of a loss of 8 and thus the VaR at the 5% level is 8. This shows that combining two portfolios can increase the VaR significantly.[4] Finally as we shall show the VaR can be inconsistent with second-order stochastic dominance.

Domar-Musgrave Risk Index

Domar and Musgrave[5] proposed a measure of risk that measures the expected value of returns less than the returns on a risk-free investment. Letting r denote the return on a risk-free asset, the Domar-Musgrave risk index is given by

$$\rho_{DM} = -\int_{x_{min}}^{r} (x - r) f(x)\, dx$$

in the case of a continuous return or

$$\rho_{DM} = -\sum_{x_i \leq r} (x - r) f(x_i)$$

[4]This is not an untypical example as the assets A and B can be interpreted as payoffs to written call and put options.

[5]Evsey C. Domar and Richard A. Musgrave, "Proportional income taxation and risk-taking", *Quarterly Journal of Economics*, Volume 58 (1944), pp.387-422.

in the discrete case. Calculating the risk index for the three portfolios D, E and F from our example we find that

$$\rho_{DM}[D] = \frac{5}{4}; \qquad \rho_{DM}[E] = \frac{5}{2}; \qquad \rho_{DM}[F] = \frac{7}{8}.$$

It is possible to generalise the Domar-Musgrave risk index to any particular target return $\hat{x}$ which may or may not be the risk-free return. In this case the measure is also known as *expected shortfall* as it measures the average return that falls short of some target return $\hat{x}$. That is expected shortfall is measured by

$$\rho_{ES} = -\int_{x_{min}}^{\hat{x}} (x - \hat{x}) f(x)\, dx$$

in the case of a continuous return or

$$\rho_{ES} = -\sum_{x_i \leq \hat{x}} (x - \hat{x}) f(x_i)$$

in the discrete case.

There are some difficulties with the Domar-Musgrave index or expected shortfall as a measure of risk too. Consider a case where the target is $\hat{x} = 0$ and suppose that some portfolio A has a return of -500 with probability 0.1 and portfolio B has a return of -100 with probability 0.5 and that all other payoffs are positive (an hence do not affect the index). Then the risk index or expected shortfall for both portfolios is the same, $\rho_{ES}[A] = \rho_{ES}[B] = 50$ but perhaps many investors would regard portfolio A with the potentially much larger loss as more risky.

Conditional Value at Risk

Conditional Value of Risk (CVaR) was introduced by Rockafellar and Uryasev[6] to overcome some of the problems that occur with using VaR as a risk

[6] See for example R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions", *Journal of Banking & Finance*, 2002.

measure. The conditional value at risk measure is related to the Domar-Musgrave index.[7] Like the VaR the CVaR fixes a probability of loss and then takes the VaR as the target for the generalised Domar-Musgrave index. Thus in the continuous case we have

$$CVaR = -\int_{x_{min}}^{VaR_\alpha} (x - VaR_\alpha) f(x)\, dx.$$

This is simply (minus) the expected value of the tail of the distribution below the VaR level.

Variance and Semi-variance

Markowitz developed a mean-variance framework for portfolio choice[8] in which he argued that investors in choosing between portfolios with equal expected value should choose the portfolio with the minimum variance.[9][10] This work was later used by Sharpe and Lintner in developing the Capital Asset Pricing Model (CAPM). As we have seen variance is a measure of average deviation and weighs both positive and negative variations equally. In his 1959 book, *Portfolio Selection*, Markowitz suggested the idea of looking at semi-variance which only considers deviations below some critical threshold. Letting the threshold be $\hat{x}$, the semi-variance risk measure is given by

$$\rho_{SV} = \sum_{x_i \leq \hat{x}} (x - \hat{x})^2 f(x_i)$$

[7] There are several variants of CVaR which were developed by different authors in different literatures such as *mean excess loss*, *tail value at risk*, *mean shortfall* which have only relatively minor differences.

[8] Harry M. Markowitz, "Portfolio selection", *Journal of Finance*, Volume 7, No. 3, (March 1952), pp.77-91.

[9] Fisher in 1906 had also suggested using variance as a measure of risk.

[10] Markowitz doesn't give a single measure of risk but rather traces out an *efficient frontier* of portfolios with minimum variance for a given level of return. Nevertheless his criterion is to choose the portfolio with the minimum variance if the expected returns are the same.

for discrete distributions and

$$\rho_{SV} = \int_{x_{min}}^{\hat{x}} (x - \hat{x})^2 f(x)\, dx$$

for continuous distributions. Often the critical value is taken to be the mean return μ so that only deviations below the mean are taken into account. The reason for preferring the semi-variance is that the variance considers both very high and very low returns equally desirable or undesirable and therefore will choose portfolios that eliminate either extremes. The semi-variance however, concentrates only on losses and therefore portfolios chosen on the basis of semi-variance will focus on eliminating losses.

Exercise: Calculate the semi-variance (using μ as the cut-off value) for the portfolios D, E and F from our example.

Lower Partial Moments

The measures we have consider so far are actually all special cases of the lower partial moments which were suggested by Fishburn[11] as suitable risk measures. The lower partial moments for a continuous distributions are given by

$$\rho_{LPM}^{\beta} = \int_{-\infty}^{\hat{x}} (\hat{x} - x)^{\beta} f(x)\, dx$$

for any $\beta \geq 0$. It is easy to check that

$$\rho_{LPM}^{0} = \int_{x_{min}}^{\hat{x}} (\hat{x} - x)^{0} f(x)\, dx = \int_{x_{min}}^{\hat{x}} f(x)\, dx = F(\hat{x}) = \rho_R$$

$$\rho_{LPM}^{1} = \int_{x_{min}}^{\hat{x}} (\hat{x} - x) f(x)\, dx = -\int_{x_{min}}^{\hat{x}} (x - \hat{x}-) f(x)\, dx = \rho_{DM}$$

$$\rho_{LPM}^{2} = \int_{x_{min}}^{\hat{x}} (\hat{x} - x)^2 f(x)\, dx = \rho_{SV}.$$

[11] Peter C. Fishburn, "Mean-risk analysis with risk associated with below-target returns", *American Economic Review*, Volume 67, No. 2 (March 1977), pp.116-26.

Thus the zeroth lower partial moment (where $\beta = 0$) is Roy's risk index, the first lower partial moment (where $\beta = 1$) is the expected shortfall or Domar-Musgrave measure and the second lower partial moment (where $\beta = 2$) is a semi-variance.

Coherent Risk Measures

We have seen that several different measures of risk have been suggested. There are however advantages and disadvantages to all these different measures and no one measure is universally accepted as the best. In order to describe what might be acceptable for a risk measure a set of criteria were developed by Artzner, Delbaen, Eber, and Heath[12] Consider two portfolios A and B, a measure of risk is called *coherent* if it satisfies the following four properties:

Subadditivity: $\rho(A + B) \leq \rho(A) + \rho(B)$

Monotonicity: if $A \leq B$, then $\rho(A) \geq \rho(B)$

Positive homogeneity: if $\lambda > 0$, then $\rho(\lambda A) = \lambda \rho(A)$

Translation invariance: for any number m, $\rho(A + m) = \rho(A) - m$.

We'll consider the rationale for each of these axioms. Subadditivity has an easy interpretation. If the subadditivity did not hold, then $\rho(A + B) > \rho(A) + \rho(B)$. This would imply, for instance, that in order to decrease risk, a firm might be motivated to break up into different incorporated affiliates or split portfolio into two separate parts. Such an artifact should not really decrease the overall risk. Monotonicity is perhaps obvious. If portfolio A always has lower returns than portfolio B then it ought to be defined as

[12] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. 1999. Coherent measures of risk. Mathematical Finance 9 (November): 203-228.

being more risky. Positive homogeneity is slightly more tricky to interpret. Subadditivity implies $\rho(A) + \rho(A) \leq 2\rho(A)$. Taking $\lambda = 2$ in the positive homogeneity axiom then equivalent to the assumption that $\rho(A+A) \geq 2\rho(A)$. That is saying that a portfolio of $2A$ is more risky than two portfolios of A. Translation invariance implies that the risk $\rho(A)$ decrease by m, by adding a sure return m to a portfolio A. The idea is that a risk measure is a measure of what kind of risk is acceptable. Thus if one adds a certain amount 100 (in every state) to the portfolio then the amount that needs to be added to this new portfolio to make it acceptable must have gone down by 100.

As we have seen some measures of risk are not coherent. For example VaR does not satisfy the assumption of subadditivity. Other measures such as Domar-Musgrave measure or expected shortfall are coherent. Nevertheless both VaR and expected shortfall have both detractors and advocates and no one measure is universally agreed to be better than others.[13]

4 Investor Preferences

The measures of risk we have so far considered associate with a given probability distribution a single number to measure risk. As we have seen there seems no universal agreement on what is the best measure and perhaps different people can have different views of risk simply because of their preferences. A question therefore to ask is whether there are situation is which everyone would agree that invest A was riskier than investment B. To try to answer this question we look at investor's preferences over risky investments. It offers formal definitions of risk aversion that cover many realistic cases of investment choices. We consider pair-wise comparisons of mutually exclusive investments X and Y and ask which of the two investments is better,

[13]Though see Nasim Taleb's invective against VaR at http://www.fooledbyrandomess.com/.

given a list of assumptions of "reasonable" investment behaviour. We use our example from the previous section for illustration.

Complete Preferences: For any pair of investments X and Y, the individual can express a preference for X over Y or visa-versa or is indifferent between the two.

Transitive Preferences: For any triple of investments, X, Y and Z, if X is preferred to Y and Y is preferred to Z then X is preferred to Z.

The idea behind these two assumptions is that investors can make choices. This is obvious for complete preferences. Without this assumption there will be some consequence bundles that the investor cannot compare. If they cannot compare the outcomes then the choice between portfolios leading to these outcomes becomes impossible. It is equally true that transitivity is needed to be able to choose between alternative portfolios. Without transitivity there would be a chain of preferences amongst some three consequences X, Y, Z such that X is preferred to Y and c' is preferred to Z and Z is preferred to X. In these circumstances it is impossible to say which consequence investment is the best.

Non-satiation: We say that an investor is *non-satiated* if he prefers a dominating investment over a dominated one. In most work in theoretical finance, we assume that all investors are non-satiated. This assumption makes sense, since payoffs are made in cash and cash can be spent on a wide range of different consumption goods and any unwanted cash can be discarded at almost no cost ("free disposal"). The assumption of non-satiation is useful when making comparisons of investments that are at different average payoff levels, but it does not let us decide between investments that have the same expected payoffs but different risk profiles. For those cases we need more specific risk assumptions.

Risk Aversion I [Preference for Certainty]: We say that an investor has "preference for certainty" and is *risk averse* if a risk-free investment is preferred over a risky investment of equal expected payoffs.

In our example, if a risk-averse investor is given the choice between C and D, say, he will choose C, since C has the same expected payoffs as D but no risk. We note that this definition of risk aversion does not require us to make judgements between different risky investments; we merely compare one risky investment with one risk-free investment of equal expected payoffs. "Preference for Certainty" is a very plausible assumption that is pretty much universally adopted in finance. However, it is too weak to help us decide on choices between two risky investments, such as D and E. For those choices we need more specific assumptions.

A diagrammatic way to represent preferences is to use indifference curves. Suppose there are just two states. Let the horizontal axis represents the amount of cash in state 1 and the vertical axis represents cash in state 2. Choose a particular investment X. A line is drawn through the point X which corresponds to all the other possible consequences of investments Y which are indifferent to X. This set of points $I(X)$ is called the indifference curve. Indifference curves cannot cross by the assumption of transitivity and because of the assumptions of monotonicity and transitivity the indifference curve is "thin" and downward sloping. All the points above the indifference curve belong to the weak preference set $WP(X)$ which represents all the points that are either preferred to or are indifferent to X.

The slope of the indifference curve shows how much extra is required in state 2 to compensate the individual for a one unit decrease in the amount received in state 1. This amount is called the marginal rate of substitution, $MRS(X)$, is a measure of the marginal rate of substitution at X. The marginal rate of substitution depends on preferences and therefore is likely to vary from one person to another even at the same consequence payoff X. The ratio q_1/q_2 represents how much the market will compensate the individual for reducing the consequent payoff in state 1 by one unit. Decreasing the amount received in state 1 by one unit will save the investor q_1 on his investment. Thus the amount extra that can be had in state 2 is q_1/q_2.

Given our three assumptions little more can be said about indifference curves other than that they do not cross, are thin and are downward sloping. However there are some other features that we might expect preferences over uncertain outcomes to exhibit. Firstly **risk aversion**, that is a desire for more certain rather than a riskier outcomes with the same expected value.[14] Secondly **state independence**, that is that preferences depend on the outcome and not the state of nature per se. Thirdly **separability**, that is that the preferences in one particular state are to some extent independent of the preferences in some other state that didn't actually occur.

Utility

We shall discuss the properties of risk aversion, separability and state independence in greater detail shortly. First we consider another convenient way of representing preferences which is through a utility function. A utility function associates with each possible investment $X = (x_1, x_2)$ a number $v(x_1, x_2)$ such that $v(x_1, x_2) > v(y_1, y_2)$ if and only if X is preferred to $Y = (y_1, y_2)$. The only property of the utility function that is important is that it ranks consequences, assigning a higher number to a preferred consequence. The actual size of the number is unimportant and any transformation of the function v that produces the same ordering of consequences as does v is equally valid as a representation of preferences. Any transformation of v that is an increasing function of v will produce the same ranking. Therefore v is said to be unique only up to a positive monotonic transformation and this representation is referred to as **ordinal utility**.

[14]When we consider models with more than one period we also expect preferences to exhibit **impatience**, that is a desire for consumption earlier rather than later.

To illustrate the ordinal nature of utility consider an example of a preference ranking which is the so-called Cobb-Douglas preferences

$$v(x_1, x_2) = Ax_1^a x_2^b$$

where A, a and b are parameters.[15] Since $\log_e$ is a positive monotonic transformation, the alternative utility function

$$\hat{v}(x_1, x_2) = \log_e(v(x_1, x_2)) = \log_e(A) + a \log_e(x_1) + b \log_e(x_2)$$

is an equally valid representation of the same preferences.

The marginal rate of substitution is calculated by totally differentiating the utility function. Suppose there are just two states, then totally differentiating $v(x_1, x_2)$ gives

$$dv = \frac{\partial v(x_1, x_2)}{\partial x_1} dx_1 + \frac{\partial v(x_1, x_2)}{\partial x_2} dx_2.$$

The partial derivatives are often called marginal utilities as they show how much utility is increased for a small (marginal) increase in the payoff in one state. Thus we define $MU_1(x_1, x_2) = \partial v(x_1, x_2)/\partial x_1$ and $MU_2(x_1, x_2) = \partial v(x_1, x_2)/\partial x_2$. The marginal rate of substitution is calculated along an indifference curve where the utility does not change, i.e. where $dv = 0$. Hence

$$MRS(cx_1, x_2) = -\frac{dx_2}{dx_1} = \frac{\frac{\partial v(x_1, x_2)}{\partial x_1}}{\frac{\partial v(x1, x_2)}{\partial x_1}}.$$

Note that MRS is defined to be a positive number and since the indifference curve is downward sloping, the slope dx_2/dx_1 is multiplied by -1 to give a positive number. As we can see the marginal rate of substitution is the ratio of marginal utilities. In the case of Cobb-Douglas preferences the marginal rate of substitution is

$$MRS(x_1, x_2) = \frac{\frac{a}{x_1}}{\frac{b}{x_2}} = \frac{xc_2}{bx_1}.$$

[15] This Cobb-Douglas functional form was originally applied in production theory where the x_i's are regarded as inputs and v is the output.

Remember that the marginal rate of substitution is slope of the indifference curve and therefore is independent of whether the utility is assigned using the function $v(x_1, x_2)$ or $\hat{v}(x_1, x_2)$.

So far nothing has been said about probabilities in defining preferences. However the probabilities can be illustrated in a diagram. Take some point $X = (x_1, x_2)$ in the diagram and suppose the probabilities of the two states are π_1 and π_2 where $\pi_1 + \pi_2 = 1$. The expected payoff from this investment is

$$\mathrm{E}[X] = \pi_1 x_1 + \pi_2 x_2.$$

The line of all points with the same expected value is called the **fair odds line**. The value $\mathrm{E}[X]$ is illustrated where the fair odds line hits the 45^0 line through the origin. This is the point of certainty where the payoff is the same in both states. The slope of the line is $-\pi_1/\pi_2$. A preference for certainty is reflected through the utility function if $v(E[X], E[X]) > v(x_1, x_2)$ for $x_1 \neq x_2$. Let's consider the implications of a preference for certainty graphically in terms of the indifference curves. Consider the point where the 45^0 line intersects the fair odds line. If there is a preference for certainty, all points along the fair odds line must lie on a lower indifference curve that the indifference curve passing through the intersection of the fair odds line and the 45^0 line. Thus the weak preference set $WP(E[X], E[X])$ must lie above the fair odds line. This means that locally the indifference curve is convex to the origin or that the weak preference set is locally convex. If preferences are smooth then it implies that the marginal rate of substitution at the 45^0 line is equal to the ratio of probabilities

$$MRS(E[X], E[X]) = \frac{\pi_1}{\pi_2}.$$

In this case then, when there is a preference for certainty the indifference curve or preferences do reflect some information about the probabilities of the various states. That is if we did not know the probabilities of the states but knew an individuals preferences, then we could infer the individual's probability assessment of the states by calculating the marginal rate of substitution at the point of certainty. Note that this implies that the marginal rate of substitution along the 45^0 line will be the same irrespective how far out along the 45^0 line we are. The slope of the all indifference curves where they cut the 45^0 line are the same.

As we have mentioned in the case of financial assets where the payoffs are amounts of money, the label of the state is likely to be unimportant. Thus we can reasonably suppose that how assets are viewed does not change if the states are simply relabelled. That is changing the name of the state from "rain" to "shine" and "shine" to "rain" should not affect how assets are viewed.[16] Indeed it is really the distribution of payoffs that matters and not how states are labelled at all. Thus we make the assumption that individuals are **probabilistically sophisticated**[17] and evaluate consequences simply by their distribution function. Probabilistic sophistication imposes a symmetry on preferences. Thus if points X and Y are on the same fair-odds line and are equal in distribution, i.e. $x_1 = y_2$ and $x_2 = y_1$ then by the assumption the indifference curve through X must also pass through Y.

Expected Utility Maximization: The most common assumption about preferences is that utility is expected utility.

$$v(x_1, x_2) = pu_1(x_1) + (1-p)u_2(x_2)$$

where $u_1(x_1)$ and $u_2(x_2)$ are the von-Neumann Morgenstern utility functions with $u'(\cdot) > 0$ and $u''(\cdot) \leq 0$. Since $u_1(\cdot)$ and $u_2(\cdot)$ are concave functions the indifference curves in (x_1, x_2)-space are convex to the origin and thus are quasi-concave. It is typically assumed that the von-Neumann Morgenstern utility functions do not depend on the state so that $u_1(\cdot) = u_2(\cdot)$.

Certainty Equivalent and Risk Aversion: The certainty equivalent of an investment X is the value of an alternative safe investment X^{ce} which

[16] In other contexts it might. etc.
[17] See Machina and Schmeidler 1992

has the same utility,

$$v(x^{ce}, x^{ce}) = v(x_1, x_2).$$

Graphically the certainty equivalent is the value of the point on the 45^0 line where it is cut by the indifference curve that passes through (x_1, x_2). We can write the certainty equivalent as a function of x_1 and x_2, i.e. $x^{ce}(x_1, x_2)$.

In the case of expected utility the certainty equivalent is given by

$$x^{ce}(x_1, x_2) = u^{-1}(\pi_1 u(x_1) + \pi_2 u(x_2)).$$

Since the inverse function $u^{-1}(\cdot)$ is an increasing monotonic transformation, this function represents the same preferences as the expected utility function and the certainty equivalent can itself be treated as a utility function.

This property of the certainty equivalent is complectly general. Since indifference curves do not cross by the transitivity assumption, the certainty equivalent gives the same preference ranking as the utility function itself. Thus if preferred it is possible to treat the certainty equivalent as the utility function and work with the certainty equivalent. That is maximising the certainty equivalent is equivalent to maximising utility.

Risk Premia: Risk premia are a measure of how far the certainty equivalent is from the expected value. There are two alternatives to measuring this difference: the absolute difference and the relative difference.

The absolute risk premium is the amount that can be taken away from the expected value such that if this amount is received with certainty it is no worse than the original distribution. Consider the investment X, the absolute risk premium $\rho^a(x_1, y_1)$ is defined by

$$v(E[X] - \rho^a(x_1, x_2), E[X] - \rho^a(x_1, x_2)) = v(x_1, x_2).$$

To see how this relates to the certainty equivalent remember that

$$v(x^{ce}(x_1, x_2), x^{ce}(x_1, x_2)) = v(x_1, x_2)$$

so that the absolute risk premium satisfies $\rho^a(x_1, x_2) = E[X] - x^{ce}(x_1, x_2)$ and it is a measure of the willingness to pay for certainty.

An alternative way to measure the willingness to pay for certainty is to measure the proportion[18] that can be taken away from expected value such that if this amount is received with certainty it is no worse than the original distribution. This amount is known as the relative risk premium. The relative risk premium $\rho^r(x_1, x_2)$ and is defined by

$$v\left(\frac{E[X]}{\rho^r(x_1, x_2)}, \frac{E[X]}{\rho^r(x_1, x_2)}\right) = v(x_1, x_2).$$

Since $v(x^{ce}(x_1, x_2), x^{ce}(x_1, x_2)) = v(x_1, x_2)$, the relative risk premium is simply the ratio of the expected value and certainty equivalent

$$\rho^r(x_1, x_2) = \frac{E[X]}{y^{ce}(x_1, x_2)}.$$

Diagrammatically the relative risk premium is the distance between the certainty equivalent and the expected value measured along the 45^0 line.

Arrow-Pratt Theory: Arrow (Aspects of the Theory of Risk Bearing, 1965) and Pratt (Econometrica, 1964) showed in the case of expected utility how the risk premium is related to the curvature of the von-Neumann Morgenstern utility function. Consider the investment X with an expected payoff of $\pi x_1 + (1-\pi)x_2 = E[X] = w$ where w can be interpreted as initial wealth. The expected utility is $\pi u(x_1) + (1-\pi)u(x_2)$. Suppose the individual pays an absolute risk premium of ρ^a to eliminate the risk. Their utility would then be $u(E[X] - \rho^a)$ where ρ^a satisfies the equation

$$\pi u(x_1) + (1-\pi)u(x_2) = u(E[X] - \rho^a).$$

The risk premium is, as before, $\rho^a = w - x^{ce}$. Some calculus using a Taylor expansion can be used to show that

$$\rho^a \approx -\frac{1}{2}\sigma^2 \frac{u''(w)}{u'(w)}$$

[18] we have to assume that x and y are positive in this case.

where $\sigma^2 = \pi x_1^2 + (1-\pi)x_2^2$ is the variance of the portfolio. The function

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

is called the **Arrow-Pratt** coefficient of absolute risk aversion. It is a measure of the curvature of the utility function. By monotonicity $u'(w)$ is positive, so the risk premium is positive, and the Arrow-Pratt coefficient of absolute risk aversion positive, if the second derivative of the utility function $u''(w)$ is negative, which indicates that the utility function is concave. The larger the coefficient $A(w)$ the greater the risk premium, the greater will the individual pay to avoid the lottery and the more risk averse is the individual. It is well known that the only von-Neumann Morgenstern utility function which has a constant coefficient of absolute risk aversion is the negative exponential function

$$u(w) = -\exp^{-aw}.$$

It is also possible to define the Arrow-Pratt coefficient of relative risk aversion

$$R(w) = wA(w)$$

that corresponds to the relative risk premium. The von Neumann Morgenstern utility functions that exhibit constant relative risk aversion are of the power form

$$u(w) = \frac{w^{(1-r)}}{(1-r)}$$

with $u(w) = \log_e(w)$ when $r = 1$. Again it is easily checked that $R(w) = r$.

5 Stochastic Dominance

First-order stochastic dominance: To examine stochastic dominance of two investments we look at the cumulative probability distribution functions.

Thus if two investments X and Y have cumulative distribution functions F and G respectively we say that X *first-order stochastically dominates* Y if

$$F(x) \leq G(x)$$

for all possible values of x (with strict inequality for at least one x). Remember that $F(x) = \text{prob}(X \leq x)$ so if $F(x) \leq G(x)$ then the probability of getting an outcome worse than x is always lower for X than for Y no matter which x we choose. We should think that most investors will prefer X to Y and one can show this is the case if the utility function is monotonic. Thus all investors with monotonically increasing preferences will agree that X is better than Y.

It is important to remember however that many investments will not be able to be ranked by first order stochastic dominance. For these investments $F(x) < G(x)$ for some x but $G(x) < F(x)$ for other values of x. Such investments are *non-comparable* according to first-order stochastic dominance. Thus first-order stochastic dominance can only be used to find a *partial ordering* of the set of possible investment opportunities.

Second-order stochastic dominance: For two investments X and Y with distribution functions F and G we say that X *second-order stochastically dominates* Y if

$$\int_{x_{min}}^{x} F(t)\,dt \leq \int_{x_{min}}^{x} G(t)\,dt$$

for all possible values of x (with strict inequality for at least one x). We can illustrate this condition by drawing the two cumulative distribution functions. The net area between the two curves evaluated at some point x is the difference

$$\int_{x_{min}}^{x} G(t) - F(t)\,dt.$$

taking into account positive and negative areas (negative if $G < F$) and this sum is required to be positive for all x. Thus we must start with a positive area and if the curves ever cross the sum of the positive areas must always outweigh the negative areas that follow.

It can be shown that all investors who are risk-averse expected utility maximisers will agree that X is better than Y if X second-order stochastically dominates Y. The intuition is that with risk-aversion marginal utility is declining and thus higher weight is placed on lower return outcomes. If X second-order stochastically dominates Y then any positive area where $F < G$ is greater than any negative area where $G < F$ and therefore higher weight is put on the positive areas.

Thus we can conclude that every risk averse expected utility maximiser will agree that X is preferable to Y if X second-order stochastically dominates Y. Again not all investments will be comparable according to second-order stochastic dominance so this again implies only a partial ordering of the set of all possible investments.

Stochastic Dominance and Variance: (1) If investment X stochastically dominates investment Y, then X will have a lower variance and a better worst-case payoff than Y. (2) However, it is possible for X to have a lower variance than Y without X stochastically dominating Y; lower variance is a *necessary* but not a *sufficient* condition for stochastic dominance.

Exercise: Compare portfolios D and E using the stochastic dominance criteria (first-order and second-order).

6 Increases in Risk

Consider the following two assets (or portfolios) A and B with payoffs

$$A = \begin{pmatrix} 7 \\ 4 \\ 6 \\ 0 \end{pmatrix} \qquad B = \begin{pmatrix} 5 \\ 5 \\ 3 \\ 3 \end{pmatrix}$$

where the four states occur with probabilities $(\frac{1}{6}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4})$ respectively. Which of these two assets is the riskier? The expected value of these assets are $E[A] = E[B] = 4$. The variance of the assets are $Var[A] = \frac{13}{2}$ and $Var[B] = 1$ and it does seem reasonable to say that A is riskier than B as the payoffs are more spread out.

Consider the two assets C and D with payoffs

$$C = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 11 \end{pmatrix} \qquad D = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 4 \\ 4 \\ 4 \\ 4 \end{pmatrix}$$

where all states occur with probability $\frac{1}{8}$. Again $E[C] = E[D] = 4$ but $Var[C] = 7$ and $Var[D] = 4$. In this case it seems much more arguable to conclude that C is riskier than D. Asset C has a constant payout except in one state, whereas D has an equal chance of no payout or a slightly higher payout.

A natural way to define the concept of A is "less risky than" B is to suppose that B is the same as A plus some extra risk or "noise". To make it clear that we are dealing with assets with a random payoff, we shall write $\tilde{A}$, $\tilde{B}$ etc. to indicate that the payoff to the assets is random.

To make this definition precise we have to consider what it means for two assets A and B to be equivalent and how "noise" can be defined. First consider the two assets E and F which have the following payoffs

$$E = \begin{pmatrix} 7 \\ 4 \end{pmatrix} \qquad F = \begin{pmatrix} 4 \\ 7 \end{pmatrix}$$

where both states are equally probable. These two assets are not the same, since they pay out different amounts in different states. Nevertheless the two assets are equally risky. In both cases there is an equal chance of a payoff of either 4 or 7. In this case we write $\tilde{E} \stackrel{d}{=} \tilde{F}$ and say that the two assets are **equal in distribution**. This is the same as saying that investors are probabilistically sophisticated and care only about the distribution of the payoffs and not the states themselves.

We have already seen that combining one asset with another can reduce the variance of the portfolio if the assets are negatively correlated. Therefore in defining "noise" we want to consider a situation that is uncorrelated with an asset X. In fact we shall use a concept slightly stronger than uncorrelated, namely **mean independence**. We say that a random variable $\tilde{\epsilon}$ is mean independent of $\tilde{X}$ if the the expected value of $\tilde{\epsilon}$ is the same for every possible value of the random variable $\tilde{X}$. That is $E_\epsilon[\epsilon|X] = E_\epsilon[\epsilon]$ for all possible values that X can take.

Consider the the asset B that has a payoff of 5 with probability $\frac{1}{2}$ and a payoff of 3 with probability $\frac{1}{2}$. Suppose that ϵ has the following values

$$\epsilon = \begin{cases} 2 & \text{with probability } \frac{1}{3} \\ -1 & \text{with probability } \frac{2}{3} \end{cases} \qquad \text{if } B = 5$$

and

$$\epsilon = \begin{cases} 3 & \text{with probability } \frac{1}{2} \\ -3 & \text{with probability } \frac{1}{2} \end{cases} \qquad \text{if } B = 3.$$

It is easy to calculate that $E_\epsilon[\epsilon|B = 5] = E\epsilon[\epsilon|B = 3] = 0$. Thus the mean of ϵ is the same independent of the value of B. Since the mean is the same independent of B, $E_\epsilon[\epsilon|B] = E_\epsilon[\epsilon]$ and it can be checked that

$$E_\epsilon[\epsilon] = \frac{1}{2}E_\epsilon[\epsilon|B = 5] + \frac{1}{2}E\epsilon[\epsilon|B = 3] = 0.$$

Mean independence is weaker than independence but is stronger than uncorrelated. The implications are

$$\text{Independence} \Rightarrow \text{Mean Independence} \Rightarrow \text{Uncorrelated}$$

Remember that the two variables X and ϵ are uncorrelated when $cov(X, \epsilon) = E[X \cdot \epsilon] - E[X] \cdot E[\epsilon] = 0$. To see that mean independence implies zero covariance note

$$E[X \cdot \epsilon] = E_X[E_\epsilon[X \cdot \epsilon | X]] = E_X[X \cdot E_\epsilon[\epsilon | X]] = E_X[X \cdot E_\epsilon[\epsilon]] = E_X[X] \cdot E_\epsilon[\epsilon].$$

Here the first equality holds by the Law of iterated expectation; the second equality holds as the expectation over ϵ is taken for each value of X; the third equality holds by mean independence, $E_\epsilon[\epsilon|X] = E[\epsilon]$; and the last equality again holds as expectation over X is taken for each value of ϵ.

In the example ϵ is not independent of B because the distribution function for ϵ depends upon B. However B and ϵ are uncorrelated. This can be easily checked as

$$B \cdot \epsilon = \begin{pmatrix} 6 \\ -3 \\ 3 \\ -3 \end{pmatrix}$$

with probabilities $(\frac{1}{6}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4})$ which gives $E[B \cdot \epsilon] = 0$. In this case the two variables B and ϵ are said to be **orthoganal**. Since $E[\epsilon] = 0$, the covariance is zero and B and ϵ are uncorrelated.

In comparing the risk of two assets or portfolios X and Y we wish to compare variability or deviation from the mean rather than the levels of the variable. Therefore we compare the deviations $X - E[X]$ and $Y - E[Y]$. Therefore the standard definition for “riskier than” is as follows:

Risk: The asset or portfolio $\tilde{Y}$ is riskier than the asset or portfolio $\tilde{X}$ if there is a random variable $\tilde{\epsilon}$ such that

$$\tilde{Y} - E[\tilde{Y}] \stackrel{d}{=} \tilde{X} - E[\tilde{X}] + \tilde{\epsilon} \qquad \text{where} \qquad E[\tilde{\epsilon}|\tilde{X}] = E[\tilde{\epsilon}] = 0.$$

The definition says that the deviation of the random variable $\tilde{Y}$ from its mean is equal in distribution to the deviation of the random variable $\tilde{X}$ from its mean plus some *noise* $\tilde{\epsilon}$. It turns out that this definition is exactly the same as second-order stochastic dominance but with the additional assumption that the means of the two distributions are the same.

Consider our example again with asset B and the random variable ϵ. We can calculate $B + \epsilon$ as follows:

$$B + \epsilon = \begin{cases} 7 & \text{with probability } \frac{1}{3} \\ 4 & \text{with probability } \frac{2}{3} \end{cases} \qquad \text{if } B = 5$$

and

$$B + \epsilon = \begin{cases} 6 & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases} \qquad \text{if } B = 3.$$

Since $B = 5$ or $B = 3$ with equal probability we have

$$B + \epsilon = \begin{cases} 7 & \text{with probability } \frac{1}{6} \\ 4 & \text{with probability } \frac{1}{3} \\ 6 & \text{with probability } \frac{1}{4} \\ 0 & \text{with probability } \frac{1}{4} \end{cases} .$$

This shows that $B + \epsilon$ and A have the same distribution function and thus are equal in distribution. We can say that A is riskier than B because it is obtained from B by adding some "noise".

It turns out that we cannot compare C and D using this definition. That is there is no noise ϵ such that $C \stackrel{d}{=} D + \epsilon$ or $D \stackrel{d}{=} C + \epsilon$. We say that C and D are incomparable in terms of risk. Thus we simply have to accept that not all assets or portfolios can be compared in terms of their riskiness.

But all asset including C and D can be compared in terms of variance. This implies that in general variance is not a good measure of risk. Indeed that it is an over strong measure of risk and a larger variance does not always correspond to what we mean by larger risk. However the opposite is true:

$$\text{Greater Risk} \Rightarrow \text{Greater Variance}$$

Thus we can conclude that a riskier asset has the greater variance. To see this we can compute the variance of Y, when Y is riskier than X as follows:

$$
\begin{aligned}
var(Y) = var(Y - E[Y]) &= var(X - E[X] + \epsilon) \\
&= var(X - E[X]) + var(\epsilon) + 2cov(X - E[X], \epsilon) \\
&= var(X) + var(\epsilon) \\
&> var(X).
\end{aligned}
$$

Here the first line follows because Y and $X + \epsilon$ are equal in distribution; the second line follows from the rule for variances and the third line follows because X and ϵ are uncorrelated. Finally since ϵ is noise, its variance is positive so $var(X) + var(\epsilon) > var(X)$. Thus greater variance is a necessary condition for greater risk but it is not sufficient.

7 Conclusion

There have been many measures of risk that have been proposed. Many measures that have been proposed are not coherent. Other measures which are coherent but are complete run into the problem that not all risk averse investors would necessarily agree on which investment is riskier. If however, we accept a partial measure of risk that compares some but not all portfolios of investment, then second order stochastic dominance of the equivalent (when the means are the same) measure of increases in risk will give a measure on which all risk averse investors will agree.