III.2 .

Alternatives to Utility

2. Alternatives to Utility

- Maximizing long-term growth
- Stochastic Dominance
- Other risk measures

3.1 Maximising long-term growth

- Learning Objectives
- Formalisation
- Geometric means versus Log utility
- Kelly's result and Samuelson's objection
- Questions

Learning objectives

- formulate the problem of maximising the long term growth of a portfolio,
- discuss how geometric means can be used to maximize long term growth rates,
- relate geometric means to log utility,
- state Kelly's theorem
- illustrate the differences between maximising long term growth and maximising expected growth for a fixed date long in the future,
- solve problems involving portfolio selection for long term growth.

Long-term Growth

- Mean-variance analysis and utility theory are just two approaches to choosing portfolios.
- If we change criteria, we will get different portfolios.
- Suppose we adopt as our criterion the requirement that the investment should do best in the very long run. In other words, we want to maximize the expected long-term growth rate.

The crucial phrase is long-term rate

$\Downarrow$

- We are not looking to win for any fixed time-horizon, but instead
- we want to adopt a strategy that will do best if we wait for an arbitrarily long amount of time.

Growth across several periods

- Each period we will put our entire portfolio into a portfolio that returns a random variable r_j.
- The returns are assumed to have the same distribution each period and to be independent of each other.

OBS: this is quite a big assumption.

- In other words, we assume the return variables r_j are i.i.d.
- If we start with 1, our wealth after N periods is therefore

$$(1 + r_1)(1 + r_2) \ldots (1 + r_N).$$

- The expected value after N periods is therefore

$$\mathbb{E}((1 + r_1)(1 + r_2) \ldots (1 + r_N)).$$

Average growth rate

- The average growth over N periods is

$$r_g = ((1+r_1)(1+r_2)\dots(1+r_N))^{1/N} - 1.$$

- Clearly, we have

$$(1+r_g)^N = (1+r_1)\dots(1+r_N).$$

- We can identify $1+r_g$, as the geometric mean of the numbers, $1+r_j$.

Long term growth rate

Suppose, for simplicity, returns can take only a finite discrete set of values.

- For N very large, the fraction of times each value is taken is its probability so if the possible values are

$$s_j \text{ with probability } p_j \text{ for } j = 1, \ldots, k,$$

  then for N large we have that the total growth converges to

$$(1 + s_1)^{Np_1}(1 + s_2)^{Np_2} \ldots (1 + s_k)^{Np_k},$$

- To get the average growth we take $1/N$ power and subtract one, and so it converges to

$$\mathrm{r_g} = \quad (1 + s_1)^{p_1}(1 + s_2)^{p_2} \ldots (1 + s_k)^{p_k} - 1.$$

  It is this quantity that we must maximize.

Using logs

- So the problem reduces to

$$\max \quad (1+s_1)^{p_1}(1+s_2)^{p_2}\dots(1+s_k)^{p_k}$$

- We can re-express the maximization problem using logs. Since log is increasing it is enough to maximize

$$\begin{aligned}\max \quad & \log\left((1+s_1)^{p_1}(1+s_2)^{p_2}\dots(1+s_k)^{p_k}\right)\\ &=\sum_{i=1}^{k} p_i \log(1+s_i),\\ &=\mathbb{E}(\log(1+r))\end{aligned}$$

Objective

- We have shown that to maximize the long-term growth rate, we must find the portfolio that maximizes

$$\mathbb{E}(\log(1+r)),$$

and this gives a long term growth rate of

$$e^{\mathbb{E}(\log(1+r))} - 1.$$

- It is important to realize that this portfolio need not be mean-variance efficient nor utility maximizing, and generally will be neither.

Geometric means and log utility

- In fact, if one has a log utility function and our initial wealth is W_0 then our expected utility at the end of the year will be

$$\begin{aligned} \mathbb{E}(\log(W_0(1+r))) =& \mathbb{E}(\log(W_0)) + \mathbb{E}(\log(1+r)), \\ =& \log(W_0) + \mathbb{E}(\log(1+r)). \end{aligned}$$

OBS: So, maximising the log utility is the same as maximising the geometric mean.

OBS: Arithmetic and geometric means are different, and maximising gives different answers.

Example: applying the geometric mean

Consider a risky investment:

return R	-0.2	-0.1	0	0.1	0.2
probabilities	0.1	0.2	0.3	0.3	0.1

and that you need to decide how much of your wealth you invest in it.

Problem: What proportion x of your wealth to put in the risky investment?

- Assume that there are no short-selling restrictions
- Consider there is no interest, i.e what you decide to keep in cash, $1 - x$ has zero return.
- We use the same proportions for every period.

Example: applying the geometric mean

We tabulate the returns R and $log(1+R)$ for some x.

probabilities	0.1	0.2	0.3	0.3	0.1
x	returns R				
0.5	-0.1	-0.05	0	0.05	0.1
0.76	-0.152	-0.076	0	0.076	0.152
1	-0.2	-0.1	0	0.1	0.2
2	-0.4	-0.2	0	0.2	0.4
x	$\log(1+R)$ for varying R				
0.5	-0.1054	-0.0513	0.0000	0.0488	0.0953
0.76	-0.1649	-0.0790	0.0000	0.0733	0.1415
1	-0.2231	-0.1054	0.0000	0.0953	0.1823
2	-0.5108	-0.2231	0.0000	0.1823	0.3365

Example: applying the geometric mean

We compute to get

x	$\mathbb{E}(\log(1+R))$	Long-term g-rate	expected return
0.5	0.003373	0.003379	0.005
0.76	0.003829	0.003836	0.0076
1	0.003439	0.003445	0.01
2	-0.007368	-0.007341	0.02

- The bigger x is the bigger the expected return is.
- However, the average long-term growth rate is maximized when $x = 0.76$.
- When $x = 2.00$, the long-term growth rate turns negative, so we will *eventually* end up down a lot of money.

Kelly's theorem

Theorem

Given two investment strategies with annual return rates r and s. Let the wealth after j years be W_j^r, W_j^s. Suppose

$$\mathbb{E}(\log(1+r)) > \mathbb{E}(\log(1+s)),$$

then with probability 1 *there will be an N such that*

$$j > N \implies W_j^r > W_j^s.$$

- Kelly's theorem says that if you wait long enough, the investment with higher expected log return will win.
- However, the theorem does not say anything about N.
- So whilst if you adopt r, you will win, you may have to wait an arbitrarily long amount of time to see your winnings.

Samuelson's objection

Samuelson objected to the Kelly argument in the following way.

- The argument that we should use the geometric mean relied on the law of large numbers.
- With probability one, the fraction of draws that take a given value converge to the probability of that value, as N tends to infinity.
- This is a purely a statement about behaviour at infinity, NOT about any finite N.

Q: What happens if we consider N finite?

Maximising for a large N

- Suppose we fix a finite N.
- Returns from year to year should be independent (given a reasonable level of market efficiency,) so our wealth after N years will be $(1+r_1)(1+r_2)\dots(1+r_N)$, with each r_j distributed the same as r and independent.
- Since the random variables are independent, the expectation is

$$(\mathbb{E}(1+r))^N.$$

This means that to maximize expected wealth, we should maximize

$$\mathbb{E}(1+r),$$

rather $\mathbb{E}(\log(1+r))$.

Kelly versus Samuelson

- Since one statement deals with a fixed time horizon, and the other with behaviour at infinity, they are not contradictory.
- Kelly says to maximize long term gains, we must maximize $\mathbb{E}(\log(1 + r))$.
- Samuelson says to maximize expected gains for any fixed time horizon, we should maximize $\mathbb{E}(1 + r)$.

Extreme example

We illustrate the issues with an extreme example. We have a choice of two investments:

- Investment r has a fixed return of 1 percent.
- Investment s has a return of -100% with probability 0.5 (i.e, lose all money) and 300% with probability 0.5.

We compute:

$$\begin{aligned}
\mathbb{E}(1+r) =& 1+r = 1.01, \\
\mathbb{E}(1+s) =& 0.5 \times 4 = 2 \qquad \Rightarrow \text{Samuelson recommends } s \\
\mathbb{E}(\log(1+r)) =& \log(1+r) = 0.00995 \qquad \Rightarrow \text{Kelly recommends } r \\
\mathbb{E}(\log(1+s)) =& -\infty
\end{aligned}$$

Extreme example

After N years,

- The expected values are $\mathbb{E}(1+r)^N = 1.01^N$, and $\mathbb{E}(1+s)^N = 2^N$.
- The actual values are $(1+r)^N = 1.01^N$ and

$$\begin{aligned}(1+s)^N &= 4^N \quad \text{with probability } 2^{-N},\\ (1+s)^N &= 0 \quad\ \ \text{with probability } 1-2^{-N}.\end{aligned}$$

- So if we wait long enough investment s will have value zero with probability 1, but for any fixed time horizon, it will win in expectation terms.
- Note that r wins in probability one but not with certainty; it is possible to get an infinite string of heads when tossing a coin, but it will only happen with probability zero.

Example: Losing it all

Suppose two investments X and Y are such that X has a non-zero probability of losing everything and Y does not.

What is the geometric mean for X?

$$\mathbb{E}(\log(1 + r_X)) = -\infty,$$

so the geometric mean is $e^{-\infty} - 1 = 0 - 1 = -1$.

The geometric mean of Y will be greater than -1 since

$$\mathbb{E}(\log(1 + r_Y)) > -\infty.$$

As we would expect, Y wins in the very long term since eventually X will hit zero.

OBS: The geometric mean approach says that if we want to win in the long-term then we should take a fair amount of risk but not so much that we can lose everything.

Levels of riskiness and long term wins

The geometric approach tells us that if we whant to win we should:

-"take a fair amount of risk"

-but not so much risk that we can loose it all.

Many investors beieve the geometric mean approach has a very high risk level, even though you will win eventually.

Most investors would not like to win after they are dead.

Theory questions

1. Define the geometric mean return of an asset.
2. What does Kelly's theorem say?
3. Which utility function is equivalent to maximizing long-term utility?
4. If we have iid returns every year and we want to maximize expected return for precisely 1,000 years away what quantity should be maximize?
5. If we have iid returns every year and we want to maximize returns in the very long term, what quantity should we maximize?

3.2 Stochastic Dominance

- Learning objectives
- Dominance
- First order stochastic dominance
- Second order stochastic dominance
- Third order stochastic dominance
- Questions

Learning objectives

- define dominance, first order stochastic dominance, and second order stochastic dominance,
- relate dominance and efficiency,
- use stochastic dominance to show that investments are preferred by certain classes of rational investors,
- solve problems using first, second and third order stochastic dominance.

Motivation

We have developed various methods of comparing investments. These include

- mean-variance efficiency,
- safety first criteria,
- expected utility,
- geometric means.

There are other methodologies...

- We now introduce the dominance approach which requires only very weak assumptions on the investor, but strong assumptions on the investments.

OBS: With SD we can get away from the fact that mean-variance analysis penalizes upside variance as well as downside variance.

Dominance

- Suppose, we have portfolios with returns R_X and R_Y with the same initial value W_0, and suppose that always

$$W_X \leq W_Y,$$

at the end of the investment period.

  - One would never prefer X to Y.
  - We can say that Y dominates (or is dominant to) X.

- Suppose we now add on the hypothesis that

$$\Pr(W_X < W_Y) > 0.$$

  - Clearly any investor who prefers more to less would prefer Y to X.
  - We then say that Y is strictly dominantes X.

Dominance versus efficiency

Although Y is preferred to X, it need not be more (mean-variance) efficient than X.

- X returns 0 always,
- Y returns 0 with probability 0.99,
- Y returns 100 with probability 0.01.

Investment Y is higher in both mean and variance, so no portfolio is efficinet relative to the other.

First-order stochastic Dominance (FOSD)

- Suppose we have two portfolios with returns Y and Z.
- Suppose they have the same cumulative distribution functions for their returns, i.e., for all real numbers a,

$$\Pr(R_Y \leq a) = \Pr(R_Z \leq a).$$

- We would not be able to distinguish them using any of our methodologies, since they all are based purely on functionals of our estimates of their probability distributions. $\Longrightarrow$ We would be indifferent between them.
- If Y is strictly dominant to X, and we are indifferent between Y and Z, then we should prefer Z to X.
- This is the idea behind stochastic dominance.
- We do not even need Y to exist, merely that such a hypothetical Y could exist is enough.

The definition of FOSD

Definition

Z has first-order stochastic dominance (FOSD) over X if

$$\Pr(R_X \leq a) \geq \Pr(R_Z \leq a), \quad \textit{for all } a,$$
$$\Pr(R_X \leq b) > \Pr(R_Z \leq b), \quad \textit{for some } b.$$

OBS: Note that this says nothing about $\Pr(R_X \leq R_Z)$.

Example 1: FOSD

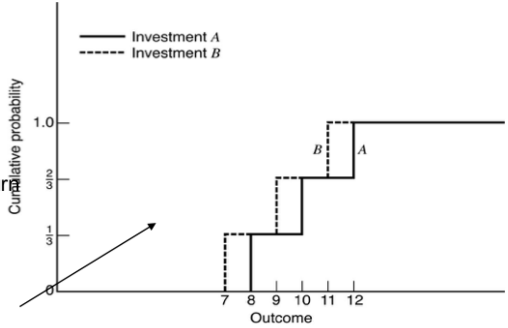
Investment A		Investment B	
Outcome	Probability	Outcome	Probability
12	1/3	11	1/3
10	1/3	9	1/3
8	1/3	7	1/3

First Order Stochastic Dominance

The accumulated probability associated to each possible return is always higher in investment B

A dominates B

(first order dominance)



Example 2: FOSD

Suppose we have

Probability	Jetstar	Exxon
0.2	15	10
0.2	14	10
0.2	13	12
0.2	12	14
0.2	10	15

$\Rightarrow$

R	Jetstar	Exxon
9	0	0
10	0.2	0.4
11	0.2	0.4
12	0.4	0.6
13	0.6	0.6
14	0.8	0.8
15	1	1
16	1	1

- Clearly, there is no simple relationship between the returns of the two companies.
- The cumulative distribution function of Exxon is always at least as big as Jetstar's and sometimes bigger.
- We have FOSD of Jetstar over Exxon!

FOSD theorem

Theorem

If portfolios X and Y have returns R_X and R_Y and the investor has a utility function U with $U'(s) > 0$ for all s, and the cumulative probabilities satisfy

$$\Pr(R_X \leq a) \leq \Pr(R_Y \leq a) \text{ for all } a,$$

$$\Pr(R_X \leq b) < \Pr(R_Y \leq b) \text{ for some } b,$$

then X will be preferred to Y.

The proof

Notation:

We will write

$$F_X(a) = \Pr(R_X \leq a),$$
$$F_Y(a) = \Pr(R_Y \leq a).$$

We will also assume the distributions are continuous and

$$F_X(a) = \int_{-\infty}^{a} f_X(s) ds.$$

So

$$f_X(a) = F_X'(a),$$

and f_X is the density of R_X. We make the analogous assumptions for Y.

The proof

Simplifying assumptions:

For simplicity, we assume that $f_X(s) = f_Y(s) = 0,$ for $|s| \geq K$, This implies

$$F_X(-K) = F_Y(-K) = 0,$$

and

$$F_X(K) = F_Y(K) = 1.$$

We will use integration by parts, for any u, v :

$$\int_a^b u(s)v'(s)ds = u(b)v(b) - u(a)v(a) - \int_a^b u'(s)v(s)ds.$$

The proof

If the initial wealth is W_0, if we invest all our money X, after a year we have

$$W_0(1 + R_X).$$

The expected utility of investing in X is therefore

$$\begin{aligned} \mathbb{E}(U(W_0(1 + R_X))) &= \int_{-K}^{K} U(W_0(1 + s)) f_X(s) ds, \\ &= \int_{-K}^{K} U(W_0(1 + s)) \frac{d}{ds} F_X(s) ds. \end{aligned}$$

The proof

Integration by parts:

Our hypotheses are on F_X and U' so we need to move the derivative onto U.

We Integrate by parts, with $u(s) = U(W_0(1+s))$, and $v(s) = F_X(s)$. Since

$$u'(s) = W_0 U'(W_0(1+s)),$$

we get

$$\int_{-K}^{K} U(W_0(1+s)) \frac{d}{ds} F_X(s) ds$$

$$= U(W_0(1+K)) - \int_{-K}^{K} W_0 U'(W_0(1+s)) F_X(s) ds$$

where we used that $F_X(K)=1$ and $F_X(-K)=0$

The proof

Doing the same for Y and subtracting, we get

$$\mathbb{E}(U(X)) - \mathbb{E}(U(Y)) = \\ - \int_{-K}^{K} W_0 U'(W_0(1+s))(F_X(s) - F_Y(s))ds.$$

Since the derivative of U is positive and $F_X(s) - F_Y(s)$ is non-positive and sometimes negative, we have that the difference in expected utilities is positive and we are done. ∎

FOSD and means

Suppose we apply our result to the simplest increasing utility function:

$$U(W) = W.$$

If X first-order stochastically dominates Y, we have by our theorem

$$\mathbb{E}(1 + R_X) = \mathbb{E}(U(1 + R_X)) > \mathbb{E}(U(1 + R_Y)) = \mathbb{E}(1 + R_Y).$$

Note: The inequality is strict.

- This means that FOSD $\implies$ a greater expected value.
- Turning this round, if two portfolios have the same expected return, FOSD cannot help us distinguish them.

Q: What is the risk profile of an investor with the above utility function?

FOSD vs mean variance efficiency

Not all portfolios that has FOSD are mean variance efficient.

Second-order stochastic dominance (SOSD)

- The first-order stochastic dominance preference theorem only assumes that the investor prefers more to less.
- It does not assume risk aversion.
- It therefore cannot help us to take risk into account when choosing investments.

$$\Downarrow$$

- It is unlikely that investments satisfy such a strong hypothesis.
- We can weaken the assumption on the returns at the cost of strengthening the assumption on utility.

SOSD theorem

Theorem

If portfolios X and Y have returns R_X and R_Y, and the investor has a utility function U with

$$U'(s) > 0, \qquad U''(s) < 0,$$

for all s, and the cumulative probabilities satisfy

$$\int_{-\infty}^{a} \Pr(R_X \leq s)ds \leq \int_{-\infty}^{a} \Pr(R_Y \leq s)ds \text{ for all } a,$$

$$\int_{-\infty}^{b} \Pr(R_X \leq s)ds < \int_{-\infty}^{b} \Pr(R_Y \leq s)ds \text{ for some } b,$$

then X will be preferred to Y.

FOSD *versus* SOSD

- FOSD makes assumptions about cumulative distribution functions and requires the investor to:
  - prefer more to less.
- SOSD makes assumptions about the integral of the cumulative distribution functions and requires the investor to:
  - prefer more to less, and
  - to be risk averse.

OBS: So, FOSD requires a stronger assumption on assets but needs weaker assumptions on investors (utility fcns).

FOSD *versus* SOSD

If X is first-order stochastic dominant to Y, then it is also second-order stochastic dominant.

$$FOSD \implies SOSD$$

This is easy to prove: simply integrate!

$$\Pr(R_X \leq a) \leq \Pr(R_Y \leq a) \text{ for all } a \text{ and,}$$

$$\Pr(R_X \leq b) < \Pr(R_Y \leq b) \text{ for some } b.$$

OBS: So, SOSD is a weaker condition than FOSD.

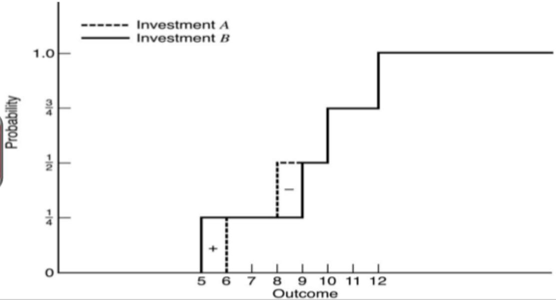
Using SOSD

- When dealing with discrete random variables, we must replace integrals with sums.
- We simply have to compute the sums of the sums of the probability that each value is taken, provided the values are uniformly spaced.
- If they are not uniformly spaced, we have to use a finer subdivision to make them uniformly spaced.

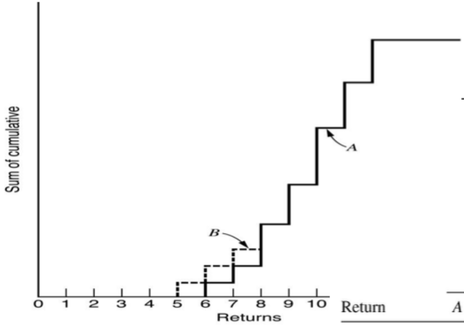
Example 3: SOSD

A		B	
Outcome	Probability	Outcome	Probability
6	1/4	5	1/4
8	1/4	9	1/4
10	1/4	10	1/4
12	1/4	12	1/4

Second Order Stochastic Dominance



Example 3: SOSD



Second Order Stochastic Dominance

Graphic Representation

The **sum of accumulated probabilities** associates to each return is always higher in investment B

⇓

A dominates B

(Second order stochastic dominance)

Return	Cumulative Probability A	Cumulative Probability B	Sum of Cumulative Probability A	Sum of Cumulative Probability B
4	0	0	0	0
5	0	1/4	0	1/4
6	1/4	1/4	1/4	1/2
7	1/4	1/4	1/2	3/4
8	1/2	1/4	1	1
9	1/2	1/2	1 1/2	1 1/2
10	3/4	3/4	2 1/4	2 1/4
11	3/4	3/4	3	3
12	1	1	4	4

SOSD and means

Q: If X SOSD Y, what can we say about their mean returns?

We compute

$$\begin{aligned}\mathbb{E}(R_X) &= \int_{-K}^{K} s\ f_X(s)ds, \\ \text{\{integrating by parts\}} \qquad &= F_X(K)K - \int_{-K}^{K} F_X(s)ds, \\ &= K - \tilde{F}_X(K).\end{aligned}$$

We assumed that $\tilde{F}_X(K) \leq \tilde{F}_Y(K)$, so

$$\mathbb{E}(R_X) \geq \mathbb{E}(R_Y).$$

<-- Note: The inequality is not strict.

SOSD and means

- This means that we can use second order stochastic dominance to distinguish between investments with the same mean.
- However, it will never tell us to prefer an investment with lower expected return because it is less risky.
- This is inevitable since we have made no assumptions about how risk averse the investor is. They may have a tiny risk aversion or a huge one.

Third-order dominance

- We can go one comparing now integrals of $\tilde{F}_X$ with $\tilde{F}_Y$.
- In discrete time that would mean computing the sums of the sums of the cumulative probabilities.
- We add assumption on utility ARA'(s)<0
- As before lower order stochastic dominans implies higher order

Higher-order dominance

- We can keep on integrating by parts.
- Each time, we get more and more conditions on higher and higher-order derivatives of U.
- We get conditions on iterated integrals of the cumulative distribution functions and so on.

Recall, however,

- when using utility functions we profiled investors up to their absolute and relative risk aversion, only.
- We did not look into higher order derivatives of the utility function U.
- So, any stochastic dominance of order higher than the third would be hard to interpret.

Interpretation summary

Investors prefer more to less

+

Investors are risk averse

+

Investors exhibits decreasing absolute risk aversion

First Order Stochastic Dominance (FOSD)

Second Order Stochastic Dominance (SOSD)

Third Order Stochastic Dominance (TOSD)

For higher orders we loose financial intuition ...

Example 4

Compare X and Y using:

- stochastic dominance,
- long-term growth rate,
- safety criteria of Roy, Kataoka and Telser (with $R_L = 7$, $\alpha = 0.25$, when applicable)
- mean-variance efficiency.

X		Y	
return	probability	return	probability
5	0.1	5	0.1
6	0.3	6	0.1
8	0.1	7	0.1
9	0.2	8	0.3
12	0.3	10	0.1
		11	0.3

Example 4

To tackle this problem, we build a table.

Each row is

- a possible return from either investment, in increasing order,
- its probability for each investment,
- its cumulative probability for each investment,
- the sum of the cumulative probabilities for each investment from higher rows.
- We include $\log(1+R)$, and R^2 to compute expected long-term growth and the variance of R.

Example 4

R	log(1+R)	R^2	f_X	F_X	$\tilde{F}_X$	f_Y	F_Y	$\tilde{F}_Y$
2	0.020	4	0	0	0	0	0	0
3	0.030	9	0	0	0	0	0	0
4	0.039	16	0	0	0	0	0	0
5	0.049	25	0.1	0.1	0	0.1	0.1	0
6	0.058	36	0.3	0.4	0.1	0.1	0.2	0.1
7	0.068	49	0	0.4	0.5	0.1	0.3	0.3
8	0.077	64	0.1	0.5	0.9	0.3	0.6	0.6
9	0.086	81	0.2	0.7	1.4	0	0.6	1.2
10	0.095	100	0	0.7	2.1	0.1	0.7	1.8
11	0.104	121	0	0.7	2.8	0.3	1	2.5
12	0.113	144	0.3	1	3.5	0	1	3.5

Example 4

We can also compute using the table:

mean R_X	8.5	mean R_Y	8.5
variance R_X	6.85	variance R_Y	4.25
$\mathbb{E}(\log(1+R_X))$	0.08129	$\mathbb{E}(\log(1+R_Y))$	0.08140
$\Pr(R_X < 7\%)$	0.4	$\Pr(R_Y < 7\%)$	0.2
$R_L(X)$	6	$R_L(Y)$	7

- SOSD: $Y \succ X$
- long-term growth rate: $Y \succ X$
- safety criteria of Roy, Kataoka and : $Y \succ X$ (where X, Y does not satisfy Telser condition)
- mean-variance efficiency: $Y \succ X$

In this case we get $Y \succ X$. In general this does not need to be the case.

Theory questions

1. What does it mean for an investment to be dominant, first order, second order or third order stochastically dominant over another investment?
2. If X is dominant to Y, must it be more efficient in a mean-variance sense?
3. Under reasonable assumptions which should be stated clearly, prove that a portfolio that FOSD another investment will be preferred.
4. Does FOSD imply SOSD? Justify your answer. What about the other way round?
5. If X FOSD Y, what can we say about their expected returns? What about SOSD? Justify your answers.
6. Financially interpret TOSD.
7. Why stochastic dominance analysis typically stops at third-order dominances?

3.3 Risk Measures

- Learning Objectives
- Value-at-risk (VaR)
- Conditional expected shortfall (CES)
- Connection of risk measures with safety first criteria and utility
- Questions

Learning objectives

- Discuss shortcomings of variance as a risk measure.
- Define value at risk (VAR).
- Find the VAR of simple portfolios.
- Define monotonicity and show that VAR is monotone.
- Discuss the shortcomings of VAR.
- Define an excess and be able to compute with the distribution of excesses.
- Relate kurtosis to VAR.
- Define conditional expected shortfall and expected shortfall.
- Relate utility functions to risk-measures.

Value-at-Risk (VaR)

VaR = Value-at-Risk

- This is the most popular measure used for controlling trading risk in the finance industry.
- Its essence is the idea of how much or more can we lose with some probability, for example, five percent or one percent in one day.
- Capital adequacy rules are sometimes based on 0.03 percent across a year.
- VaR is usually expressed in terms of wealth losses rather than returns.

Defining VAR

- If the value of our portfolio today is V_0 and at time t is V_t.
- The losses over the time interval $(0, t)$ can be defined as:

$$L_t = V_0 - V_t \ .$$

- The VaR at probability p for time period t is the value x such that

$$\Pr(L_t > x) = p.$$

- By convention, we always take VaR to be positive or zero. So if at probability p we make money, we will set the VaR to be zero.

OBS: VaR is many times defined in terms of $1 - p$, instead. So, at the 95 percent level or 99 percent level.

Discrete distributions

- The previous definition always works for losses with continuous distribution functions.
- Whenever we have discrete distributions there may not be a level x at which the probability of losing x or more is precisely p.
- However, there will be a level at which the probability jumps across p and we use that instead.
- I.e., in general the VaR at probability p for time period t is the lowest value x satisfying

$$\Pr(L_t > x) \leq p.$$

Example: Value-at-Risk

A portfolio A loses 10 *million with probability* 0.005, *loses* 5 *million with probability* 0.02, *loses* 1 *million with probability* 0.05. *Otherwise, it makes* 1 *million dollars. Find the VAR at* 1 *and* 5 *percent levels.*

We have as possible losses $\{-1, 1, 5, 10\}$ million (negative loss is a gain)

$$
\begin{aligned}
\mathbb{P}(L_t > 10) =& 0 \\
\mathbb{P}(L_t > \ 5) =& 0.005, \\
& \cdots 0.01 \cdots \\
\mathbb{P}(L_t > \ 1) =& 0.025 = 0.02 + 0.005 \\
& \cdots 0.05 \cdots \\
\mathbb{P}(L_t > -1) =& 0.075 = 0.005 + 0.02 + 0.05.
\end{aligned}
$$

The one percent VaR level is thus 5 million.
Similarly, the 5 percent VaR level is 1 million.

VaR: monotonicity

For each proposed risk measure, we can assess its properties:

1. One important property is monotonicity: if two portfolios are of the same value and one portfolio always returns more than the other, then it has less risk.

OBS: A risk measure should be monotone. This property holds for VaR but not for variance.

$$L_{\mathrm{W}} \leq L_{\mathrm{V}} \Rightarrow \mathrm{VaR}_{\alpha}(L_{\mathrm{W}}) \leq \mathrm{VaR}_{\alpha}(L_{\mathrm{V}})$$

VaR: monotonicity

Proof. Consider the values of two portfolios V and W. Suppose

$$W_0 = V_0, \qquad W_t \geq V_t. \qquad \Longrightarrow \qquad -V_t \geq -W_t.$$

Let L^V be the losses for V and L^W for W. We then have

$$L^V = V_0 - V_t \geq W_0 - W_t = L^W.$$

So if

$$\mathbb{P}(L^W > x) = p,$$

then

$$\mathbb{P}(L^V > x) \geq p,$$

So the VaR for V is at least the VaR for W. ■

VaR: insensitivity beyond p

2. One problem with VaR is its insensitivity beyond p. I.e., it does not pick up what happens beyond level p.

For example, asset A is worth 1 and at time t

$$\begin{cases} 1.05 \text{ with probability } 0.9, \\ 0.9 \text{ with probability } 0.1 \end{cases}$$

whereas B is worth 1 and at time t

$$\begin{cases} 1.1 \text{ with probability } 0.99, \\ 0 \text{ with probability } 0.01 \end{cases}$$

If we work at a 5% VaR level, B has no risk – its VAR is negative which we take to be zero. But A has a VAR of 0.1.

VaR: not sub-additive

Sub-additivity:

An intuitive and desirable property of risk measures is that the sum of the risks of two portfolios considered separately should be more than or equal to that of the two portfolios considered together.

So if ta risk-measure is called ω, we should have

$$\omega(X + Y) \leq \omega(X) + \omega(Y).$$

=> We do NOT require equality because, X and Y could be natural hedges that is have negative correlation. An extreme case is $Y = -X$, in that case $X + Y$ has no risk.

VaR: not sub-additive

3. Unfortunately VaR is not sub-additive !

If we needed to show VaR were sub-additive, we needed to show it would hold for all pairs of portfolios. So to show it is not, we need to construct one example in which it fails.

Proof.

Suppose assets C and D are independent and worth 1 initially. Each is worth

$$\begin{cases} 1.1 \text{ with probability } 0.96, \\ 0.5 \text{ with probability } 0.04 \end{cases}$$

What is the VaR at 5% for each individually and for both together?

VaR: not sub-additive

Clearly, each has zero VaR when considered on its own. What about together?

Initial value is 2. Final values

$$\begin{cases} 2.2 \text{ with probability } 0.96^2, \\ 1.1 + 0.5 = 1.6 \text{ with probability } 2 * 0.96 * 0.04, \\ 1 \text{ with probability } 0.04^2. \end{cases}$$

The VaR is

$$2 - 1.6 = 0.4 > 0 + 0.$$

So sub-additivity fails. ■

VaR: excesses

- Suppose we are market risk managers and we monitor a portfolio with 5% daily VaR (i.e. for the time period of a day) for 220 trading days.
- We would expect the daily losses to exceed the VaR level a number of times if the VaR is correct.
- Concretely, we would expect excesses about 5% of the days

$$220 \times 0.05 = 11.$$

- If there are a lot more we should be worried.
- If there are a lot less our estimation of VAR is probably too conservative and we should be worried too!
- If we monitor at level p for N periods, we can compute the probability distribution of excesses as it is simply binomial.
- The probability of zero excesses in the example above is $(1 - 0.05)^{220}$.

VaR: Normal distribution

- One easy to estimate VaR is to assume asset values are normal or log-normal.
- Suppose the portfolio losses have mean μ and standard deviation σ for the time period t.
  One can then just read the VaR off a normal distribution table.
  Let z denote a normal random variable with mean 0 and variance 1, and then let $\Phi(\cdot)$ be the cumulative distribution.
  Let
  $$z_{(1-p)} = \Phi^{-1}(1-p)$$
  with p the VaR-level, then the VaR is the value x,
  $$x = \mu + \sigma z_{(1-p)}.$$

Normal dist VaR is easy to calculate but can be dangerous...

Example VaR

Compute the daily 99% VaR of a portfolio of 1 million EUR whose daily returns are normally distributed with mean 0.03% and standard deviation of 1%.

From the standard normal table

$$N(-2.33) = 0.01$$

Thus the 99% Value-at-Risk is

$$\text{VaR} = V(0)(2.33\sigma - \mu) = 1000000(0.0233 - 0.003) = 23000\text{kr}.$$

VaR: fat tails

- Distributions in finance often have fat tails, that is the probability of being far from the mean is greater than that for a normal distribution with the same mean and variance.
- This can often be summarized by looking at the kurtosis or fourth moment:

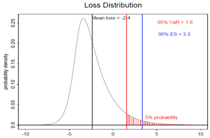
$$\frac{\mathbb{E}((X-\mu)^4)}{\mathrm{Var}(X)^2}.$$

  For a normal distribution, this is equal to three. For fat-tailed distributions it will be higher.
- A normal approximation to a fat-tailed distribution will lead to VaR numbers that are too low, as the probability of large moves is **underestimated**.

Conditional expected shortfall (CES)

- Because of the short-comings of VaR, one alternative that has been suggested is CES, also known as "tail-VaR".
- Here we take the expected losses given that we are in the worst part of the distribution.
- So the CES at level p for a time period t is

$$\mathbb{E}(L_t | L_t > \text{VAR}_p(L_t)) \; .$$



McNeil, Frey and Embrechts 2005

- This has a number of nice properties including sub-additivity, that is the CES of the sum of two portfolios is less than or equal to the sum of the two CESs.
- We have also easy expressions if we rely in the Normal distribution.

OBS: Although mathematicians prefer CES, regulators prefer VAR for historical reasons, ... up to recently!

Short fall

A simplification of CES is to simply use shortfall below a given fixed level, x, so we take

$$\mathbb{E}(L_t|L_t > x).$$

- This has the virtue of simplicity without penalising upside risk.
- Its big virtue is that it is easy to explain.

Risk-measures and utility

We have previously seen that mean-variance efficiency analysis corresponds to using quadratic utility functions (or a 2nd order Taylor approximations of other utility functions).

Q: What about other risk measures?

- Short fall – a discontinuous utility function that doesn't look at anything above x.
- Semi-variances – a utility function that is linear above the expectation and quadratic below it.
- VAR – does not naturally correspond to a utility function.
- CES – does not naturally correspond to a utility function.

Theory questions

- Define value at risk.
- What does it mean for a risk-measure to be sub-additive? Prove or disprove that each of VAR and variance is sub-additive.
- What does it mean for a risk-measure to be monotone? Prove or disprove that each of VAR and variance is monotone.
- What does it mean for a distribution to be fat-tailed? How will the VAR of such a distribution compare to that of a normal distribution?
- What is a VAR excess? What form does the distribution of the number of excess over a fixed period of time take?
- If we change the size of a loss below the VAR level, what effect will it have on the VAR?
- For each of the following risk-measuares, discuss how they relate to utility functions: short fall, semi-variance, VAR, CES.