Ec2723, Asset Pricing I
Class Notes, Fall 2005

Choice Under Uncertainty

John Y. Campbell[1]

First draft: July 30, 2003
This version: September 15, 2005

[1]Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA. Email john_campbell@harvard.edu.

There is a basic contrast between:

- *Ordinal utility* $\Upsilon(x)$ is invariant to monotonic transformations, so $\Upsilon(x)$ is equivalent to $\Theta(\Upsilon(x))$ for any strictly increasing Θ.
- *Cardinal utility* $\Psi(x)$ is invariant to positive affine (aka linear) transformations, so $\Psi(x)$ is equivalent to $a + b\Psi(x)$ for any $b > 0$.

In finance we rely heavily on von Neumann-Morgenstern utility theory which says that choice over lotteries, satisfying certain axioms, implies maximization of the expectation of a cardinal utility function, defined over outcomes.

Sketch of von Neumann-Morgenstern theory

Define states $s = 1...S$. Define outcomes x_s in a set X. Probabilities p_s of the different outcomes then define lotteries. When $S = 3$, we can draw probabilities in 2 dimensions (since $p_3 = 1 - p_1 - p_2$). We get the "Machina triangle".

We define a compound lottery as one which determines which primitive lottery we are given. For example a compound lottery L might give us lottery L^a with probability α, and lottery L^b with probability $(1 - \alpha)$. Then L has the same probabilities over the outcomes as $\alpha L^a + (1 - \alpha)L^b$.

We define a preference ordering $\succeq$ over lotteries. A person is indifferent between lotteries L^a and L^b, $L^a \sim L^b$, iff $L^a \succeq L^b$ and $L^b \succeq L^a$.

Continuity axiom: For all L^a, L^b, L^c s.t. $L^a \succeq L^b \succeq L^c$, there exists a scalar $\alpha \in [0, 1]$ s.t.

$$L^b \sim \alpha L^a + (1 - \alpha)L^c.$$

This implies the existence of a preference functional defined over lotteries, i.e. indifference curves on the Machina triangle.

Independence axiom: $L^a \succeq L^b \Rightarrow \alpha L^a + (1 - \alpha)L^c \succeq \alpha L^b + (1 - \alpha)L^c$.

This implies that the preference functional is linear in probabilities (indifference curves on the Machina triangle are straight lines), which means that we can define a scalar u_s for each outcome x_s s.t.

$$L^a \succeq L^b \Rightarrow \sum_{s=1}^{S} p_s^a u_s \geq \sum_{s=1}^{S} p_s^b u_s.$$

That is, L^a delivers higher expected utility.

Allais paradox

This famous paradox challenges the von Neumann-Morgenstern framework. Draw from an urn with 100 balls, labelled 0–99.

	0	$1-10$	$11-99$
L^a	50	50	50
L^b	0	250	50
M^a	50	50	0
M^b	0	250	0

People often say that $L^a \succeq L^b$ but $M^b \succeq M^a$, even though balls 11–99 must be irrelevant to the decision given the independence axiom.

There is a debate over the significance of this paradox. Either the independence axiom is incorrect, or people are easily misled (but can be educated).

Abandoning the independence axiom can lead to other problems. For example, suppose that $L^a \succ L^b$ and $L^a \succ L^c$, but counter to the independence axiom $L^d = 0.5L^b + 0.5L^c \succ L^a$. Then you would pay to switch from L^a to L^d, but once the uncertainty in the compound lottery L^d is resolved, you would pay again to switch back to L^a. This is called the "Dutch book" problem. It arises from the fact that the lottery structure can be costlessly altered, for given underlying probabilities, to exploit people with preferences that violate the independence axiom.

Reducing the dimension of utility

Finance theory generally works with low-dimensional arguments of the utility function. Proceeding from greater to lesser generality,

- Multiple goods, dates, and states (ordinal utility)
- Multiple goods and dates, taking expectation over states (cardinal utility)
- Multiple dates only (one-good simplification)
- Time-separable utility, adding up over dates ($U(C_1) + \delta U(C_2) + ...$)
- Single-date utility $U(C_1)$, which is equivalent to utility defined over wealth $U(W_1)$.

Must utility be bounded?

Strictly speaking, a von Neumann-Morgenstern utility function must be bounded because of the technique used to derive the utility function from the underlying axioms.

In addition, an unbounded utility function leads to the *super St. Petersburg paradox*.

The St. Petersburg paradox concerns a coin-tossing gamble that pays $\$2^n$ if the first head occurs on the nth toss. The expected payoff of this gamble is

$$\sum_{n=1}^{\infty} 2^{-n} 2^n = \sum_{n=1}^{\infty} 1$$

which is infinite. Does this mean that you would pay an infinite amount to play this game? No, because the expected utility of the payoff can be finite. For example, with log utility

$$\sum_{n=1}^{\infty} 2^{-n} \log(2^n) = \sum_{n=1}^{\infty} 2^{-n} n \log(2) = 2\log(2)$$

which is finite.

For an unbounded utility function, U^{-1} exists everywhere so we can design a gamble which pays $W_n = U^{-1}(2^n)$ if the first head occurs on the nth toss. Then expected utility is infinite. Arrow made much of this problem, but today it is usually ignored. It is standard practice to work with unbounded utility functions.

Jensen's Inequality and risk aversion

Consider a random variable $\widetilde{z}$ and a function f.

Definition: f is *concave* iff for all $\lambda \in [0,1]$ and values a, b,

$$\lambda f(a) + (1-\lambda) f(b) \leq f(\lambda a + (1-\lambda) b).$$

If f is twice differentiable, then concavity implies that $f'' \leq 0$.

Jensen's Inequality: $\mathrm{E} f(\widetilde{z}) \leq f(\mathrm{E}\widetilde{z})$ for all possible $\widetilde{z}$ iff f is concave.

Definition: an agent is *risk averse* if he dislikes all zero-mean risk at all levels of wealth. That is, for all w_0 and risk $\widetilde{x}$ with $\mathrm{E}\widetilde{x} = 0$,

$$\mathrm{E}u(w_0 + \widetilde{x}) \leq u(w_0).$$

This is equivalent to

$$\mathrm{E}u(\widetilde{z}) \leq u(\mathrm{E}\widetilde{z}),$$

where $\widetilde{z} = w_0 + \widetilde{x}$. Thus risk aversion is equivalent to concavity of the utility function.

Measuring risk aversion

A natural measure of risk aversion is f'', scaled to avoid dependence on the units of measurement for utility. Absolute risk aversion A is defined by

$$A = \frac{-f''}{f'}.$$

Note that in general this is a function of the initial level of wealth.

Comparing risk aversion

Let u_1 and u_2 have the same initial wealth. u_1 is more risk-averse than u_2 if u_1 dislikes all lotteries that u_2 dislikes, regardless of the common initial wealth level.

Define $\phi(x) = u_1(u_2^{-1}(x))$. What are the properties of this function?

1. $u_1(z) = \phi(u_2(z))$, so $\phi(.)$ turns u_2 into u_1.
2. $u_1'(z) = \phi'(u_2(z))u_2'(z)$, so $\phi' = u_1'/u_2' > 0$.
3. $u_1''(z) = \phi'(u_2'(z))u_2''(z) + \phi''(u_2(z))u_2'(z)^2$, so

$$\phi'' = \frac{u_1'' - \phi' u_2''}{{u_2'}^2} = \frac{u_1'}{{u_2'}^2}(A_2 - A_1).$$

4. Consider a risk $\widetilde{x}$ that is disliked by u_2, that is a risk s.t. $\mathrm{E}u_2(w_0 + \widetilde{x}) \leq u(w_0)$. We have

$$\mathrm{E}u_1(w_0 + \widetilde{x}) = \mathrm{E}\phi(u_2(w_0 + \widetilde{x})) \leq \phi(\mathrm{E}u_2(w_0 + \widetilde{x})) \leq \phi(u_2(w_0)) = u_1(w_0)$$

for all $\widetilde{x}$ iff ϕ is concave or equivalently $\phi'' \leq 0$. But then from 3) we must have $A_1 \geq A_2$.

5. The *risk premium* π is the amount one is willing to pay to avoid a pure (zero-mean) risk. It solves

$$\mathrm{E}u(w_o + \widetilde{x}) = u(w_0 - \pi).$$

Defining $z = w_0 - \pi$ and $\widetilde{y} = \pi + \widetilde{x}$, this can be rewritten as

$$\mathrm{E}u(z + \widetilde{y}) = u(z).$$

Now define π_2 as the risk premium for agent 2, and define z_2 and $\widetilde{y}_2$ accordingly. We have

$$\mathrm{E}u_2(z_2 + \widetilde{y}_2) = u_2(z_2).$$

If u_1 is more risk-averse than u_2, then

$$\mathrm{E}u_1(z_2 + \widetilde{y}_2) \leq u_1(z_2),$$

which implies $\pi_1 \geq \pi_2$.

6. Consider a risk that may have a nonzero mean μ. It pays $\mu + \widetilde{x}$ where $\widetilde{x}$ has zero mean. The *certainty equivalent* C^e satisfies

$$\mathrm{E}u(w_o + \mu + \widetilde{x}) = u(w_0 + C^e).$$

This implies that

$$C^e(w_0, u, \mu + \widetilde{x}) = \mu - \pi(w_0 + \mu, u, \widetilde{x}).$$

Thus if u_1 is more risk-averse than u_2, then $C_1^e \leq C_2^e$.

In summary, the following statements are equivalent:

- u_1 is more risk-averse than u_2.
- u_1 is a concave transformation of u_2.
- $A_1 \geq A_2$.
- $\pi_1 \geq \pi_2$.
- $C_1^e \leq C_2^e$.

The Arrow-Pratt approximation

Consider a pure risk $\widetilde{y} = k\widetilde{x}$, where k is a scale factor. Write the risk premium as a function $g(k)$: $g(k) = \pi(w_{0,}u, k\widetilde{x})$ satisfies

$$\mathrm{E}u(w_o + k\widetilde{x}) = u(w_0 - g(k)).$$

Note that $g(0) = 0$.

Now differentiate w.r.t. k:

$$\mathrm{E}\widetilde{x}u'(w_o + k\widetilde{x}) = -g'(k)u'(w_0 - g(k))$$

which implies that $g'(0) = 0$.

Differentiate w.r.t. k a second time:

$$\mathrm{E}\widetilde{x}^2u''(w_o + k\widetilde{x}) = g'(k)^2u''(w_0 - g(k)) - g''(k)u'(w_0 - g(k)),$$

which implies that

$$g''(0) = \frac{-u''(w_0)}{u'(w_0)}\mathrm{E}\widetilde{x}^2 = A(w_0)\mathrm{E}\widetilde{x}^2.$$

Now take a Taylor approximation of $g(k)$:

$$g(k) \approx g(0) + kg'(0) + \frac{1}{2}k^2g''(0).$$

Substituting in, we get

$$\pi \approx \frac{1}{2}A(w_0)k^2\mathrm{E}\widetilde{x}^2 = \frac{1}{2}A(w_0)\mathrm{E}\widetilde{y}^2.$$

The risk premium is proportional to the *square* of the risk. This property of differentiable utility is known as *second-order risk aversion.* It implies that people are approximately risk-neutral with respect to small risks.

We also find that

$$C^e \approx k\mu - \frac{1}{2}A(w_0)k^2\mathrm{E}\widetilde{x}^2,$$

so a positive mean has a dominant effect for small risks.

Relative risks

Define a multiplicative risk by $\widetilde{w} = w_0(1+k\widetilde{x}) = w_0(1+\widetilde{y})$. Define $\widehat{\pi}$ as the share of wealth one would pay to avoid this risk:

$$\widehat{\pi} = \frac{\pi(w_0, u, w_0k\widetilde{x})}{w_0}.$$

Then

$$\widehat{\pi} \approx \frac{1}{2}w_0A(w_0)k^2\mathrm{E}\widetilde{x}^2 = \frac{1}{2}R(w_0)\mathrm{E}\widetilde{y}^2,$$

where $R(w_0) = w_0A(w_0)$ is the *coefficient of relative risk aversion.*

Decreasing absolute risk aversion

The following conditions are equivalent:

- π is decreasing in w_0.
- $A(w_0)$ is decreasing in w_0.
- $-u'$ is a concave transformation of u, so $-u'''/u'' \geq -u''/u'$ everywhere. The ratio $-u'''/u'' = P$ has been called *absolute prudence* by Kimball, who relates it to the theory of precautionary saving.

Decreasing absolute risk aversion (DARA) seems intuitively appealing. Certainly we should be uncomfortable with increasing absolute risk aversion.

Tractable utility functions

Almost all applied theory and empirical work in finance uses some member of the class of linear risk tolerance (LRT) or hyperbolic absolute risk aversion (HARA) utility functions. These are defined by

$$u(z) = \zeta(\eta + \frac{z}{\gamma})^{1-\gamma},$$

defined over z s.t. $\eta + z/\gamma > 0$.

For these utility functions, we get

$$T(z) = \frac{1}{A(z)} = \eta + \frac{z}{\gamma},$$

$$A(z) = \left(\eta + \frac{z}{\gamma}\right)^{-1},$$

$$P(z) = \frac{\gamma + 1}{\gamma}\left(\eta + \frac{z}{\gamma}\right)^{-1},$$

$$R(z) = z\left(\eta + \frac{z}{\gamma}\right)^{-1}.$$

Important special cases of HARA utility:

- *Quadratic* utility has $\gamma = -1$. This implies increasing absolute risk aversion and the existence of a bliss point at which $u' = 0$. These are important disadvantages, although quadratic utility is tractable in models with additive risk.
- *Exponential or constant absolute risk averse* (CARA) utility is the limit as $\gamma \rightarrow -\infty$. To obtain constant absolute risk aversion A, we need

$$-u''(z) = Au'(z).$$

Solving this differential equation, we get

$$u(z) = \frac{-\exp(-Az)}{A},$$

where $A = 1/\eta$. This form of utility is tractable with normally distributed risks because then utility is lognormally distributed.

- *Power or constant relative risk averse* (CRRA) utility has $\eta = 0$ and $\gamma > 0$. Relative risk aversion $R = \gamma$. For $\gamma \neq 1$,

  $$u(z) = \frac{z^{1-\gamma}}{1-\gamma}.$$

  For $\gamma = 1$,

  $$u(z) = \log(z).$$

  Power utility is appealing because it implies stationary risk premia and interest rates even in the presence of long-run economic growth. Also it is tractable in the presence of multiplicative lognormally distributed risks.

- A negative η represents a *subsistence level.* Rubinstein has argued for this model, but economic growth renders any fixed subsistence level irrelevant in the long run. Models of habit formation have time-varying subsistence levels which can grow with the economy.

Rabin critique

Matthew Rabin has criticized utility theory on the ground that it cannot explain observed aversion to small gambles without implying ridiculous aversion to large gambles. This follows from the fact that differentiable utility has second-order risk aversion.

To understand Rabin's critique, consider a gamble that wins \$11 with probability 1/2, and loses \$10 with probability 1/2. With diminishing marginal utility, the utility of the win is at least $11u'(w_0+11)$. The utility cost of the loss is at most $10u'(w_0-10)$. Thus if a person turns down this gamble, we must have $10u'(w_0-10) > 11u'(w_0+11)$ which implies

$$\frac{u'(w_0+11)}{u'(w_0-10)} < \frac{10}{11}.$$

Now suppose the person turns down the same gamble at an initial wealth level of $w_0 + 21$. Then

$$\frac{u'(w_0 + 21 + 11)}{u'(w_0 + 21 - 10)} = \frac{u'(w_0 + 32)}{u'(w_0 + 11)} < \frac{10}{11}.$$

Combining these two inequalities,

$$\frac{u'(w_0 + 32)}{u'(w_0 - 10)} < \left(\frac{10}{11}\right)^2 = \frac{100}{121}.$$

If this iteration can be repeated, it implies extremely small marginal utility at high wealth levels, which would induce people to turn down apparently extremely attractive gambles.

Responses:

1. As we increase wealth, a person will continue to turn down a given absolute gamble indefinitely only if absolute risk aversion is constant or increasing. Rabin's most extreme results assume this, and can be understood as a critique of constant or increasing absolute risk aversion.

2. Aversion to small risks probably results from some other aspect of human psychology besides declining marginal utility of wealth. For example, people may be averse to any loss, regardless of size. (This would explain the Allais paradox and is a feature of Kahneman and Tversky's prospect theory.) But this does not mean that we should abandon utility theory for studying large risks in financial markets.

Comparing risks

We would like to compare the riskiness of different distributions. Three possible notions of increasing risk:

1) Something that all concave utility functions dislike.

2) More weight in the tails of the distribution.

3) Added noise.

The classic analysis of Rothschild and Stiglitz shows that these are all equivalent. They are *not* equivalent to higher variance. Consider random variables $\widetilde{X}$ and $\widetilde{Y}$ which have the same expectation.

1) $\widetilde{X}$ is weakly less risky than $\widetilde{Y}$ if no individual with a concave utility function prefers $\widetilde{Y}$ to $\widetilde{X}$:

$$E\left[u(\widetilde{X})\right] \geq E\left[u(\widetilde{Y})\right]$$

for all concave $u\,(.)$. $\widetilde{X}$ is less risky (without qualification) if there is *some* concave $u(.)$ which strictly prefers $\widetilde{X}$.

Note that this is a partial ordering. It is not the case that for any $\widetilde{X}$ and $\widetilde{Y}$, either $\widetilde{X}$ is weakly less risky than $\widetilde{Y}$ or $\widetilde{Y}$ is weakly less risky than $\widetilde{X}$. We can get a complete ordering if we restrict attention to a smaller class of utility functions than the concave, such as the quadratic.

2) $\widetilde{X}$ is less risky than $\widetilde{Y}$ if the density function of $\widetilde{Y}$ can be obtained from that of $\widetilde{X}$ by applying a mean-preserving spread (MPS). An MPS $s(x)$ is defined by

$$s\,(x) = \begin{pmatrix} \alpha \text{ for } c < x < c+t \\ -\alpha \text{ for } c' < x < c'+t \\ -\beta \text{ for } d < x < d+t \\ \beta \text{ for } d' < x < d'+t \\ 0 \text{ elsewhere} \end{pmatrix}$$

where $\alpha, \beta, t > 0$; $c + t < c' < d - t < d + t < d'$; and $\alpha\,(c' - c) = \beta(d' - d)$, that is, "the more mass you move, the less far you can move it."

An MPS is something you add to a density function $f(x)$. If $g(x) = f(x) + s(x)$, then (i) $g(x)$ is also a density function, and (ii) it has the same mean as $f(x)$.

(i) is obvious because $\int s\,(x)\,dx =$ area under $s(x) = 0$.

(ii) follows from the fact that the "mean" of $s(x)$, $\int xs\,(x) = 0$, which follows from $\alpha(c' - c) = \beta(d' - d)$.

$$\begin{aligned}
\int xs(x) &= \int_{c}^{c\prime+t} x\alpha dx + \int_{c\prime}^{c\prime+t} x(-\alpha)\,dx + \int_{d}^{d+t} x(-\beta)\,dx + \int_{d\prime}^{d\prime+t} x\beta dx \\
&= \alpha\left[\frac{x^2}{2}\right]_c^{c+t} - \alpha\left[\frac{x^2}{2}\right]_{c'}^{c'+t} - \beta\left[\frac{x^2}{2}\right]_d^{d+t} + \beta\left[\frac{x^2}{2}\right]_{d'}^{d'+t} \\
&= \frac{\alpha}{2}\left[(c+t)^2 - c^2 + c'^2 - (c'+t)\right]^2 + \frac{\beta}{2}\left[(d'+t)^2 - d'^2 + d^2 - (d+t)^2\right] \\
&= t\left[\beta(d'-d) - \alpha(c'-c)\right] = 0.
\end{aligned}$$

In what sense is an MPS a spread? It's obvious that if the mean of $f(x)$ is between $c\prime+t$ and d, then $g(x)$ has more weight in the tails. It's not so obvious when the mean of $f(x)$is off to the left or the right. Nevertheless, we can show that $\widetilde{Y}$ with density g is riskier than $\widetilde{X}$ with density f in the sense of 1) above. In this sense the term "spread" is appropriate.

$$\begin{aligned}
& E\left[u\left(\widetilde{X}\right)\right] - E\left[u\left(\widetilde{Y}\right)\right] \\
=\; & \int u(z)\left[f(z) - g(z)\right]dz = -\int u(z)\,s(z)\,dz \\
=\; & -\alpha\int_c^{c+t} u(z)\,d + \alpha\int_{c'}^{c'+t} u(z)\,dz + \beta\int_d^{d+t} u(z)\,dz - \beta\int_{d'}^{d'+t} u(z)\,dz \\
=\; & -\alpha\int_c^{c+t}\left[u(z) - u(z+c'-c)\right]dz + \beta\int_d^{d+t}\left[u(z) - u(z+d'-d)\right]dz \\
=\; & -\alpha\int_c^{c+t}\left[u(z) - u(z+c'-c) - \frac{\beta}{\alpha}\left\{u(z+d-c) - u(z+d'-c)\right\}\right]dz.
\end{aligned}$$

But

$$\frac{\beta}{\alpha} = \frac{(c'-c)}{d'-d}.$$

Also,

$$u(z+h) - u(z) = u'(z^*)h$$

for some z^* between z and $z+h$. Thus

$$u(z) - u(z + c' - c) = (c' - c)\, u'(z_1^*)$$

for some z_1^* between z and $z + c' - c$, and

$$u(z + d - c) - u(z + d' - c) = -(d' - d)\, u'(z_2^*)$$

for some z_2^* between $z + d - c$ and $z + d' - c$. We get

$$E\left[u\left(\widetilde{X}\right)\right] - E\left[u\left(\widetilde{Y}\right)\right] = \alpha(c' - c) \int_c^{c+t} [u'(z_1^*) - u'(z_2^*)] dz > 0,$$

where the inequality follows because $z_1^* < z_2^*$ so $u'(z_1^*) > u'(z_2^*)$.

3) A formal definition of "added noise" is that $\widetilde{X}$ is less risky than $\widetilde{Y}$ if $\widetilde{Y}$ has the same distribution as $\widetilde{X} + \widetilde{\varepsilon}$, where $E[\widetilde{\varepsilon}|X] = 0$ for all values of X. We say that $\widetilde{\varepsilon}$ is a "fair game" with respect to X.

This condition is stronger than zero covariance, $\operatorname{Cov}\left(\widetilde{\varepsilon}, \widetilde{X}\right) = 0$. It is weaker than independence, $\operatorname{Cov}\left(f(\widetilde{\varepsilon}), g\left(\widetilde{X}\right)\right) = 0$ for all functions f and g. It is equivalent to $\operatorname{Cov}\left(\widetilde{\varepsilon}, g\left(\widetilde{X}\right)\right) = 0$ for all functions g.

3) is sufficient for 1):

$$\begin{aligned} E\left[U(\widetilde{X} + \widetilde{\varepsilon})|X\right] &\leq U(E\left[\widetilde{X} + \widetilde{\varepsilon}|X\right]) = U(X) \\ &\Rightarrow E\left[U(\widetilde{X} + \widetilde{\varepsilon})\right] \leq E\left[U(\widetilde{X})\right] \\ &\Rightarrow E\left[U(\widetilde{Y})\right] \leq E\left[U(\widetilde{X})\right] \end{aligned}$$

because $\widetilde{Y}$ and $\widetilde{X} + \widetilde{\varepsilon}$ have the same distribution.

In fact, Rothschild-Stiglitz show that 1), 2) and 3) are all equivalent. This is a powerful result because one or the other condition may be most useful in a particular application.

Why are these not equivalent to $\widetilde{Y}$ having greater variance than $\widetilde{X}$? It is obvious from 3) that if $\widetilde{Y}$ is riskier than $\widetilde{X}$ then $\widetilde{Y}$ has greater variance than $\widetilde{X}$. The problem is that the reverse is not true in general. Greater variance is necessary but not sufficient for increased risk. $\widetilde{Y}$ could have greater variance than $\widetilde{X}$ but still be preferred by some concave utility functions if it has more desirable higher-moment properties. This possibility can only be eliminated if we confine attention to a limited class of distributions such as the normal distribution.

Stochastic dominance

The following are definitions.

$\widetilde{X}$ *dominates* $\widetilde{Y}$ if $\widetilde{Y} = \widetilde{X} + \widetilde{\xi}$, where $\widetilde{\xi} \leq 0$.

$\widetilde{X}$ *first-order stochastically dominates* $\widetilde{Y}$ if $\widetilde{Y}$ has the distribution of $\widetilde{X} + \widetilde{\xi}$, where $\widetilde{\xi} \leq 0$. Equivalently, if $F(.)$ is the cdf of $\widetilde{X}$ and $G(.)$ is the cdf of $\widetilde{Y}$, then $\widetilde{X}$ first-order stochastically dominates $\widetilde{Y}$ if $F(z) \leq G(z)$ for every z. First-order stochastic dominance implies that every increasing utility function will prefer $\widetilde{X}$.

$\widetilde{X}$ *second-order stochastically dominates* $\widetilde{Y}$ if $\widetilde{Y}$ has the distribution of $\widetilde{X} + \widetilde{\xi} + \widetilde{\varepsilon}$, where $\widetilde{\xi} \leq 0$ and $\mathrm{E}\left[\widetilde{\varepsilon} | X + \xi\right] = 0$. Second-order stochastic dominance implies that every increasing, concave utility function will prefer $\widetilde{X}$. Increased risk is the special case of second-order stochastic dominance where $\widetilde{\xi} = 0$.

The principle of diversification

Consider n lotteries with payoffs $\widetilde{x}_1$, $\widetilde{x}_2$, ..., $\widetilde{x}_n$ that are independent and identically distributed (iid). You are asked to choose weights α_1, α_2, ..., α_n subject to the constraint that $\sum_i \alpha_i = 1$. It seems obvious that the best choice is complete diversification, with weights $\alpha_i = 1/n$ for all i. The payoff is then

$$\widetilde{z} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{x}_i.$$

To prove that this is optimal, note that the payoff on any other strategy is

$$\sum_i \alpha_i \widetilde{x}_i = \widetilde{z} + \sum_i \left(\alpha_i - \frac{1}{n} \right) \widetilde{x}_i = \widetilde{z} + \widetilde{\varepsilon},$$

and

$$\mathrm{E}\left[\widetilde{\varepsilon} | z\right] = \sum_i \left(\alpha_i - \frac{1}{n} \right) \mathrm{E}\left[\widetilde{x}_i | z\right] = k \sum_i \left(\alpha_i - \frac{1}{n} \right) = 0.$$

Note that diversification increases utility because you are averaging independent risks. Adding independent risks does not have the same effect.