FIN 539 MATHEMATICAL FINANCE
Lecture 1: Decision Theory and Static Choice Problems

Philip H. Dybvig
Washington University in Saint Louis

Decision-theoretic Approach to Investing

The course takes the decision-theoretic approach to choice problems. Decision theory is good for thinking about choices conceptually and includes techniques for solving problems.

The main focus of this course is to study continuous-time models of investments. We will study two common approaches to analyzing these problems: (1) dynamic programming and the Bellman equation and (2) the one-shot approach using the stochastic discount factor. Both of these approaches work by reducing our analysis to a single-period problem, so it is useful to study single-period problems first. Single-period problems are also useful in practice.

- What is decision theory and why is it useful?
- Choice problems and Kuhn-Tucker conditions
- single-period investment problems
- Fundamental Theorem of Asset Pricing

What is decision theory?

Decision theory is the study of choice problems. It can be used to decide what we need to do and can be used to predict what agents in the economy will do. We will focus on the part of decision theory that models decision-making in terms of choice problems. We solve a choice problem, we have a *choice variables* that we pick, an *objective function* to be maximized or minimized, and *constraints* (usually equalities or inequalities) that must be satisfied.

Decision theory also provides a useful framework for organizing our thoughts when we face decisions, whether or not we plan to formulate and solve them analytically. I like to think about the *information* we have available, the *choice variables*, the *objective function*, and any *constraints* we will face. *Game theory* also incorporates strategic decisions by agents who anticipate the actions of other agents, while *general equilibrium* looks at the implications at a market level of agents who optimize but are not strategic.

A mean-variance choice problem

Choose risky portfolio proportions $\theta \in \Re^N$ to
maximize $\mu'\theta - \gamma_R\theta'V\theta/2 - \gamma_T(\theta-\theta_B)'V(\theta-\theta_B)/2 - C'|\theta-\theta_0|$
subject to:
$\mathbf{1}'\theta = 1$ (fully invested)
$(\forall n = 1, ..., N)\theta_n \geq 0$ (long only)

vector of choice variables: θ
objective function: $\mu'\theta - \gamma_R\theta'V\theta/2 - \gamma_T(\theta-\theta_B)'V(\theta-\theta_B)/2 - C'|\theta-\theta_0|$
equality constraint: $\mathbf{1}'\theta = 1$
inequality constraints: $(\forall n = 1, ..., N)\theta_n \geq 0$

Parameters:
μ: $N \times 1$ vector of excess returns
V: $N \times N$ covariance matrix of excess returns
$\gamma_R, \gamma_T \geq 0$: scalar levels of aversion to risk and tracking error
θ_B: benchmark portfolio
θ_0: initial portfolio
C: vector of proportional transaction costs

Choice theory versus (or plus) AI

AI is great when (1) there is a lot of data and (2) the choice environment is static. Example: chess (the rules are constant and there is an essentially unlimited supply of games, can play more if needed)

Choice theory is great for (1) changing environments and (2) not so much data. Example: mortgage-backed securities leading up to the financial crisis (the terms for subprime mortgages were different than for any available data)

In the future, the best practitioners will be able to use both, sometimes in combination. As an example of how to combine choice theory and AI, we can use AI to form predictions of the mean vector and covariance matrix to serve as inputs for an optimization model. And, we can use choice theory to formulate an economic model of the predicted impact on the mean and covariance matrices of what is new about the current circumstances (e.g. the pandemic). The forward-looking analysis from theory supplements the backward-looking information in the AI.

Tiny math review: calculus with matrices

Let $f : \Re^N \to \Re$ and $x \in \Re^N$. Then $\frac{df(x)}{dx} = \nabla f(x) = f'(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots \frac{\partial f(x)}{\partial x_N}\right)$

$$f''(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_N} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_N \partial x_1} & \frac{\partial^2 f(x)}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_N^2} \end{bmatrix}$$

Let $a \in \Re^N$ be a column vector of constants, and let $A \in \Re^{N \times N}$ and $B \in \Re^{N \times K}$ be a matrices of constants. Then[1]

$$\frac{d(a'x)}{dx} = \frac{d(x'a)}{dx} = a, \frac{d(B'x)}{dx} = B', \text{ and } \frac{d(x'B)}{dx} = B$$

$$\frac{d(x'Ax)}{dx} = (A + A')x \quad (= 2Ax \text{ if } A \text{ is symmetric})$$

[1]Some people use different conventions about rows versus columns of derivatives (and I may not be consistent).

Tiny math review: eigenvalues and eigenvectors

Let A be a square $(N \times N)$ matrix. Then λ is an eigenvalue of A with associated eigenvector $x \neq \mathbf{0}$ if

$$Ax = \lambda x.$$

and the set of eigenvalues is the same (including multiplicity) as the set of solutions to the polynomial equation

$$\det(A - \lambda I) = 0,$$

where $\det(\cdot)$ indicates the determinant. In our applications, A is often symmetric: in finance, A is likely to be a covariance matrix or a matrix of second partial derivatives. If A is symmetric, then A has N eigenvalues, all real (not complex numbers), and we can choose the corresponding eigenvectors to be an orthonormal basis of $\Re^N$. (Recall that x_1, ..., x_N are orthonormal if $x_i' x_j = 1$ for $i = j$ and $= 0$ otherwise.)

Tiny math review: positive and negative definiteness

The symmetric square $(N \times N)$ matrix is said to be positive definite if $(\forall x \neq 0) x'Ax > 0$, positive semi-definite if $(\forall x \neq 0) x'Ax \geq 0$, negative definite if $(\forall x \neq 0) x'Ax < 0$, or negative semi-definite if $(\forall x \neq 0) x'Ax \leq 0$. Definiteness is equivalent to signs of eigenvalues:

- A is positive definite iff all its eigenvalues are positive.
- A is positive semi-definite iff all its eigenvalues are positive or zero.
- A is negative definite iff all its eigenvalues are negative
- A is negative semi-definite iff all its eigenvalues are negative or zero

A covariance matrix is always positive semi-definite, and if nonsingular, positive definite. If the second derivative matrix of a function is negative semi-definite everywhere, the function is concave, and if it is negative definite everywhere strictly concave. Similarly, if the second derivative matrix of a function is positive semi-definite everywhere, the function is convex, and if positive definite everywhere, strictly convex.[2]

[2]These are "almost" equivalences for twice differentiable functions; some people define concavity and strict concavity this way.

Eigenvalues and covariance matrices

A symmetric matrix V is a covariance matrix for some random variables if and only if V is positive semi-definite. If V is not positive semi-definite, the V has a negative eigenvalue, say $\lambda_i < 0$ with associated eigenvector x_i. Then $x_i' V x_i = x_i' \lambda_i x_i = \lambda_i |x_i|^2 < 0$, which cannot be.

Sometimes clients will give you the covariance matrix they want you to use, but it is not positive semi-definite. This is a problem for simulation, since no simulated random variables will conform to a covariance matrix that is not positive semi-definite, and it is a problem with optimization because the first-order conditions (usually) will not give you an actual optimum and even if it does give a optimum (thanks to some constraints) it will not be reasonable. Usually the source of the impossible covariance matrix is computing different elements of the matrix in different sample periods. My *ad hoc* fix of this is to replace each eigenvalue λ_i with $\hat{\lambda}_i = \max(\lambda_i, \kappa)$, where κ is a small positive number. This will repair the impossible part of the covariance matrix without causing too much disruption to the rest.

Simulating random variables using eigenvalues and eigenvectors

To simulate from positive semi-definite V, let λ_1, ... λ_N be the eigenvalues with associated orthonormal eigenvectors x_1,...x_N. We can take some N-vector ε of independent unit normal variables, $\varepsilon \sim N(\mathbf{0}, I_{N\times N})$, and construct the N-vector y of random variables as

$$y = X\Gamma\varepsilon$$

where X is a matrix whose ith column is the ith eigenvector of V and Γ is a diagonal matrix whose ith diagonal element is $\sqrt{\lambda_i}$. Then the covariance matrix of the y's is $\mathrm{E}[yy'] = \mathrm{E}[X\Gamma\varepsilon\varepsilon'\Gamma'X'] = X\Gamma I_{N\times N}\Gamma'X' = X\Lambda X'$, where $\Lambda = \Gamma\Gamma'$ is the diagonal matrix with the eigenvalues in order on the diagonal. However, we will show that $X\Lambda X' = V$. Letting $\mathbf{1}_i$ be the ith unit vector (the N-vector with $\mathbf{1}_i = 1$ and $\mathbf{1}_j = 0$ for $j \neq i$), we have that $(\forall i)X\Lambda X'x_i = X\Lambda\mathbf{1}_i = X\lambda_i\mathbf{1}_i = \lambda_i x_i$, and because a matrix is determined by its product with all elements of a basis. Therefore, a symmetric matrix V is a covariance matrix if and only if it is postive semi-definite.

Matrix square root, repairs

Note that $S \equiv X\Gamma$ is one of many square roots of the matrix V, defined by $V = SS'$.

The description of how to do the simulation also tells us how to do the repair when we replace λ_i by $\hat{\lambda}_i \equiv \max(\lambda_i, \kappa)$ for some small $\kappa > 0$ Let $\hat{\Lambda}$ be the diagonal matrix with $\hat{\Lambda}_{ii} = \sqrt{\hat{\lambda}_i}$. Then $\hat{V} = X'\hat{\Lambda}X$ is the repaired covariance matrix.

If we try to use the simulation procedure for a matrix V that is not positive semi-definite, we will have Γ_{ii} is the square root of a negative number and is therefore imaginary! This "helps" since if ε is a real-valued random variable with mean zero and positive variance, $\mathrm{var}(i\varepsilon) = \mathrm{E}[(i\varepsilon)^2] = -\mathrm{var}(\varepsilon)$. However, this is of course total nonsense. Unless your clients insist on complex-valued simulated security returns, you had better find a way to make the covariance matrix positive semi-definite.

Solving the choice problem: Kuhn-Tucker Conditions

Consider the maximization problem:

Choose $x \in \Re^N$ to
maximize $f(x)$
subject to $(\forall i \in \mathcal{E}) g_i(x) = 0$, and
$(\forall i \in \mathcal{I}) g_i(x) \leq 0$.

$x = (x_1, ..., x_N)$ is a vector of *choice variables*.
$f(x)$ is the scalar-valued *objective function*.
$g_i(x) = 0$, $i \in \mathcal{E}$ are *equality constraints*.
$g_i(x) \leq 0$, $i \in \mathcal{I}$ are *inequality constraints*.
$\mathcal{E} \cap \mathcal{I} = \emptyset$

Kuhn-Tucker conditions:

$\nabla f(x^*) = \Sigma_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla g_i(x^*)$
$(\forall i \in \mathcal{I}) \lambda_i \geq 0$
$\lambda_i g_i(x^*) = 0$

Solving the choice problem: some definitions

- A *feasible solution* (or *feasible choice*) x satisfies the constraints but may not maximize the objective function. A problem is said to be *feasible* if it has a feasible solution.
- An *optimal solution* (or *optimal choice*) x is feasible and x has the smallest value of the objective function (largest if maximizing) of all feasible solutions. (Also, commonly called "solution.")
- An *interior solution* is an optimal solution at which no constraints are binding.
- A *corner solution* is an optimal solution at which constraints are binding.
- A *local optimum* is a feasible choice x^* that, for some $\varepsilon > 0$, is optimal in the problem with the additional constraint $\|x - x^*\| \leq \varepsilon$.
- *value* is $f(x^*)$ for optimal x^*

Solving the choice problem: Kuhn-Tucker theorem

The feasible solution x^* is said to be *regular* if the set $\{\nabla g_i(x^*)|g_i(x^*) = 0\}$ is a linearly independent set. In particular, an interior solution is always regular.

This maximization problem is said to be *convex* if the objective is concave ($f''(x)$ negative semidefinite everywhere) and the constraint set is convex. The constraint set is convex if $g_i(x)$ is affine for all $i \in \mathcal{E}$ and $g_i(x)$ is convex ($g_i''(x)$ is positive semidefinite) for all $i \in \mathcal{I}$.

If x^* is regular and f and the g_i's are differentiable, the Kuhn-Tucker conditions are necessary for feasible x^* to be optimal. If the optimization problem is convex, then the Kuhn-Tucker conditions are sufficient for an optimum.

Applied mathematicians commonly look at minimizing problems, while we will solve portfolio choice problems that are formulated as maximimization problems. It is easy to convert from one to the other: simply multiply the objective function by -1.

In-class exercise: Kuhn-Tucker conditions

Choose risky portfolio proportions $\theta \in \Re^N$ to
maximize $\mu'\theta - \gamma_R \theta' V \theta/2 - \gamma_T(\theta - \theta_B)' V (\theta - \theta_B)/2$
subject to: $(\forall n = 1, ..., N)\theta_n \geq 0$ (long only)

Assume that V is positive definite and that γ_R and γ_T are positive.

- Use the Kuhn-Tucker conditions to derive the solution of the problem without the long-only constraint.
- Write down the Kuhn-Tucker conditions for the problem including the long-only constraint.
- Solve the portfolio problem including the long-only constraint under the assumption that the security returns are uncorrelated.

Recall the Kuhn-Tucker conditions:

$\nabla f(x^*) = \Sigma_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla g_i(x^*)$
$(\forall i \in \mathcal{I})\lambda_i \geq 0$
$\lambda_i g_i(x^*) = 0$

A complete-markets choice problem

Given initial wealth w_0,
choose state-contingent consumptions c_1, ... c_Ω, to
maximize $\Sigma_{\omega=1}^{\Omega} \pi_\omega u(c_\omega)$
st $\Sigma_{\omega=1}^{\Omega} p_\omega c_\omega = w_0$

The objective is the von Neumann-Morgernstern utility function $\mathrm{E}[u(\cdot)]$, π is the vector of probabilities, and p is the vector of state prices. We can also state this in terms of the state-price density (or stochastic discount factor) $\xi_\omega = p_\omega / \pi_\omega$:

Given initial wealth w_0,
choose state-contingent consumptions c_1, ... c_Ω, to
maximize $\mathrm{E}[u(c_\omega)]$
st $\mathrm{E}[\xi_\omega c_\omega] = w_0$

K-T conditions: $u'(c) = \lambda \xi$.

von Neumann-Morgenstern utility function $E[u(\cdot)]$

- increasing ($u'(c) > 0$ suffices): prefers more to less
- strictly concave ($u''(c) < 0$ suffices): risk averse
  - Jensen's inequality – prefers mean to a risky random payoff
  - first-order conditions sufficient for an optimum
- absolute risk aversion: $A(c) \equiv -u''(c)/u'(c)$ aversion to absolute gambles
- relative risk aversion: $R(c) \equiv -cu''(c)/u'(c)$ aversion to relative gambles
- constant relative risk aversion (CRRA): $c^{1-R}/(1-R)$ or $\log(c)$
- constant absolute risk aversion (CARA): $-\exp(-Ac)/A$
- hyperbolic absolute risk aversion (HARA): $(c-c_0)^{1-R}/(1-R)$, $\log(c-c_0)$, or CARA. (Note: includes CRRA: $c_0 = 0$)
- a positive affine transform $a + bu(c)$ where $b > 0$ gives the same choices

Arrow-Pratt risk aversion measures $A(c)$ and $R(c)$

For "small" absolute gambles with $\mathrm{E}[\epsilon] = 0$, using a second-order Taylor expansion, the break-even risk premium π solves

$$u(c) = [Eu(c + \pi + \varepsilon)] \approx u(c) + u'(c)\pi + \frac{1}{2}u''(c)\mathrm{var}(\varepsilon)$$

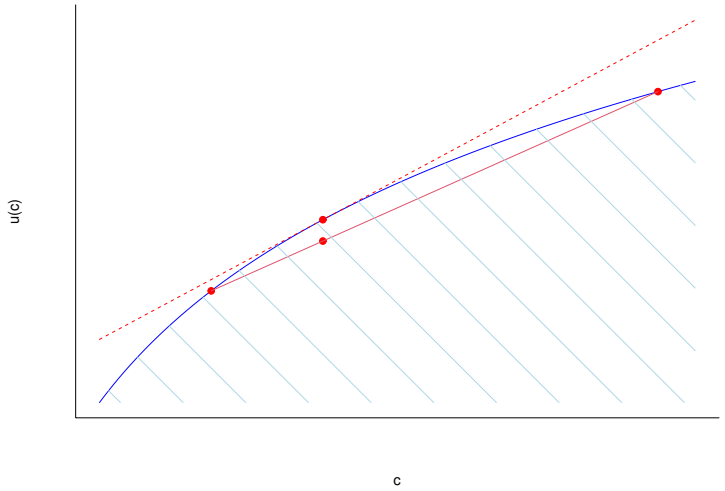
which implies that

$$\pi \approx A(c)\frac{\mathrm{var}(\varepsilon)}{2},$$

where $A(c) \equiv -u''(c)/u'(c)$ as defined in the previous slide.

Similarly, for "small" relative gambles with $\mathrm{E}[\epsilon] = 0$, we have that the break-even risk premium π for which $u(c) = \mathrm{E}[u(c(1 + \pi + \varepsilon))]$ is roughly $\pi \approx R(c)\mathrm{var}(\varepsilon)/2$, where $R(c) \equiv -cu''(c)/u'(c)$ as defined in the previous slide.

von Neumann-Morgenstern utility function (drawing pictures)



Concave: the shaded set $S \equiv \{(c,u)|u < u(c)\}$ below the graph is a convex set: for $x, y \in S$ and $a \in [0,1]$, $ax + (1-a)y \in S$. For an interior point c_0 in the domain of u, $(\exists z)(\forall c)(u(c) \leq u(c_0) + z(c - c_0))$. If u is differentiable at c_0, $z = u'(c_0)$.

For random c, let $c_0 = \mathrm{E}[c]$. Then, $\mathrm{E}[u(c)] \leq \mathrm{E}[u(c_0) + (c - c_0)u'(c_0)] = u(c_0) + (\mathrm{E}[c] - c_0)u'(c_0) = u(c_0)$ (Jensen's inequality). The agent is risk averse.