3. Mean-variance Optimization

3.1. **Mean-variance equivalence.** In this section we prove that when asset returns can be assumed to follow an elliptical distribution, then even if the investor's utility function is very complicated, the problem of maximizing expected utility can be reduced to maximizing a multivariate quadratic function. For elliptical distributions, the optimal expected-utility portfolio is the mean-variance portfolio.

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

$$u(w) = -\frac{\exp(-\kappa w)}{\kappa}$$

where $\kappa > 0$ is some positive scalar. The distinguishing feature of CARA utility functions is constant absolute risk aversion: $A(w) = \kappa$ for all w.

A particularly easy special case to analyze is when u is exponential and $\tilde{w}$ is normally distributed. In this case the Arrow-Pratt "approximation" is exact. Indeed, supposing $\tilde{w}$ has mean μ and variance σ^2, then

$$E[u(\tilde{w})] = u\left(\mu - \frac{1}{2}\kappa\sigma^2\right) \tag{3.1}$$

The relation (3.1) means that maximizing $E[u(\tilde{w})]$ is equivalent to maximizing

$$\mu - \frac{1}{2}\kappa\sigma^2,$$

since u is monotone.

Note that constant *absolute* risk aversion is a reasonable preference for a hedge fund who is optimizing a portfolio over the next few days and does not anticipate any large capital changes (such as subscriptions or redemptions) over the same period of time. Hence over the range of possible values of w that we're talking about, they'll have roughly the same aversion to (say) one million dollars of volatility.

Note – in the preceding argument, we have not assumed the actual utility function is quadratic. Quadratic functions don't make sense as utility functions, since it implies that beyond some level more wealth is somehow worse.

Let $\boldsymbol{h} \in \mathbb{R}^n$ denote the portfolio holdings, measured in dollars or an appropriate numeraire currency, at some time t in the future. Let $\boldsymbol{h}_0$ denote the current portfolio. let $\boldsymbol{r} \in \mathbb{R}^n$ denote the return over the interval $[t, t+1]$. Hence $\boldsymbol{r} \in \mathbb{R}^n$ is an n-dimensional vector whose i-th component is

$$r_i = p_i(t+1)/p_i(t) - 1$$

where $p_i(t)$ is the i-th asset's price at time t (adjusted for splits or capital actions if necessary). Hence the (one-period) wealth random variable is $\tilde{w} = \boldsymbol{h}'\boldsymbol{r}$ and the expected-utility maximizer chooses $\boldsymbol{h}$ to satisfy

$$\boldsymbol{h}^* = \operatorname{argmax} \mathbb{E}[u(\tilde{w})] \tag{3.2}$$

Definition 3.1. The underlying asset return distribution, $p(\boldsymbol{r})$, is said to be *mean-variance equivalent* if, for any increasing utility function u, there exists some constant $\kappa > 0$ (where κ depends on u) such that

$$\boldsymbol{h}^* = \operatorname{argmax} \{\mathbb{E}[\tilde{w}] - (\kappa/2)\mathbb{V}[\tilde{w}]\} \tag{3.3}$$

where $\boldsymbol{h}^*$ is defined by (3.2)

Which distributions, then, are mean-variance equivalent? We showed above that the normal distribution is; this is easy. Tobin (1958) conjectured that for $n > 2$, any two-parameter distribution is mean-variance equivalent, but this was shown to be false by Feldstein (1969). Many distributions, including heavy-tailed distributions such as the multivariate Student-t, are also mean-variance equivalent, but not all two-parameter families.

Let's now assume we are in the situation of mean-variance equivalence and so we'd like to choose h to maximize the mean-variance quadratic form

$$\mathbb{E}[\boldsymbol{h}'\boldsymbol{r}] - \frac{1}{2}\kappa\mathbb{V}[\boldsymbol{h}'\boldsymbol{r}] = \boldsymbol{h}'\boldsymbol{\mu} - \frac{1}{2}\kappa\, \boldsymbol{h}'\boldsymbol{\Omega}\boldsymbol{h} \tag{3.4}$$

where $\kappa > 0$ is the risk-aversion (we'll stick with this convention for the rest of the course).

This is, in fact, the problem proposed by Markowitz (1952), more than 10 years before Arrow (1963) and Pratt (1964), but Markowitz did not develop the full theory of risk aversion in order to arrive at (3.4). Markowitz won the Nobel Memorial Prize in Economic Sciences in 1990 while a professor of finance at Baruch College of the City University of New York. The solution is clearly $\boldsymbol{h}^* = (\kappa\boldsymbol{\Omega})^{-1}\boldsymbol{\mu}$, but we don't know $\boldsymbol{\Omega}$ or $\boldsymbol{\mu}$.

Note that (3.4) requires a vector and a matrix:

$$\boldsymbol{\mu} = E[\boldsymbol{r}] \in \mathbb{R}^n, \quad \text{and} \quad \boldsymbol{\Omega} = V[\boldsymbol{r}] \in S^n_{++}, \tag{3.5}$$

where S^n_{++} denotes the space of positive-definite $n \times n$ real matrices. Both of these concern the Christmas yet to come, so lie squarely in the domain of the ghost of Christmas future, not the ghosts of Christmas past or present. Like any information about the future, good estimates are very hard to come by. An eigenvector of $\boldsymbol{\Omega}$ with eigenvalue zero would be a long-short portfolio with zero variance. I don't think this animal exists in nature (aside from cases where the universe contains redundant assets), so the above definitely needs to be S^n_{++} and not just S^n_{+}.

For US equities $n \approx 1500$ to 3000 depending on whether our strategy includes small caps. In this range, estimating $\boldsymbol{\Omega}$ directly using sample covariances of the equity returns is pure lunacy. (This is obvious from the dimension of the parameter space, not a profound or controversial statement.) With T historical periods, one has nT data points and $n(n+1)/2$ free parameters in the full covariance matrix, so $2T/(n+1)$ data points per parameter. If $n = 2500$ we need 10 years' history to get 2 data points per parameter! Let B be a $T \times n$ matrix having the stock return time series as columns. De-mean each column. Then the covariance matrix is $\boldsymbol{\Omega} \propto B'B \in S^n_+$. The rank of $B'B$ is at most T, so if $T < n$ it is impossible that $\boldsymbol{\Omega}$ is invertible.

Definition 3.2. A utility function $u : \mathbb{R} \to \mathbb{R}$ is called *standard* if it is increasing, concave, and continuously differentiable.

These properties make economic sense. Even great philanthropists have increasing utility of wealth in their investment portfolio – they would prefer to be able to do more to end hunger, disease etc. Concavity corresponds to risk-aversion as we showed previously. Finally, if the utility function is not continuously differentiable, it implies that there is a certain particular level of wealth for which one penny above that is very different than one penny below.

Definition 3.3. Let ℓ denote a lottery, and let w_ℓ denote the (random) final wealth associated to lottery ℓ. For two scalars $m \in \mathbb{R}$ and $s > 0$, let $L(\mu, \omega)$ denote the space of lotteries ℓ under which $\mathbb{E}[w_\ell] = \mu$ and $\mathbb{V}[w_\ell] = \omega^2$. We say *expected utility is a function of mean and variance* if $\mathbb{E}[u(w_\ell)]$ is the same for all $\ell \in L(\mu, \omega)$. This means that the function $\hat{U}$ defined by

$$\hat{U}(\mu, \omega) := \{\mathbb{E}[u(w_\ell)] : \ell \in L(\mu, \omega)\}$$

is single-valued; the right-hand side is always a single number.

Assuming all lotteries correspond to holding portfolios of risky assets, then Definition 3.3, like Definition 3.1, is a property of the asset return distribution $p(\boldsymbol{r})$; some distributions have this property, and some do not.

If Definition 3.3 does *not* hold for a given distribution, then there isn't much hope for mean-variance to hold either. Intuitively, if Definition 3.3 does *not* hold then $\mathbb{E}[u(w_\ell)]$ must depend on something apart from $\mathbb{E}[w_\ell]$ and $\mathbb{V}[w_\ell]$ so it should be easy to construct a counterexample where the right-hand side (3.3) is sub-optimal because of this "extra term."

Definition 3.4. An *indifference curve* is a level curve of the surface $\hat{U}$, or equivalently a set of the form $\hat{U}^{-1}(c)$.

The intuition behind the terminology of Def. 3.4 is that the investor is indifferent among the outcomes described by the various points on the curve.

Tobin (1958) assumed that expected utility is a function of mean and variance, and showed mean-variance equivalence as a consequence. Unfortunately, Tobin's proof was flawed – it contained a derivation which is only valid for elliptical distributions. The flaw in Tobin's proof, and a counterexample, was pointed out by Feldstein (1969). After presenting a correct proof, we will discuss the flaw.

Recall that for a scalar-valued random variable X, the *characteristic function* is defined by

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right],$$

If the variable has a density, then the characteristic function is the Fourier transform of the density. The characteristic function of a real-valued random variable always exists, since it is the integral of a bounded continuous function over a finite measure space.

Generally speaking, characteristic functions are especially useful when analyzing moments of random variables, and linear combinations of random variables. Characteristic functions have been used to provide especially elegant proofs of some of the key results in probability theory, such as the central limit theorem.

If a random variable X has moments up to order k, then the characteristic function ϕ_X is k times continuously differentiable on $\mathbb{R}$. In this case

$$\mathbb{E}[X^k] = (-i)^k \phi_X^{(k)}(0).$$

If ϕ_X has a k-th derivative at zero, then X has all moments up to k if k is even, but only up to $k-1$ if k is odd, and

$$\phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$$

If $X_1, \ldots, X_n$ are independent random variables, then

$$\phi_{X_1+\cdots+X_n}(t) = \phi_{X_1}(t) \cdots \varphi_{X_n}(t).$$

Definition 3.5. An $\mathbb{R}^n$-valued random variable $\boldsymbol{x}$ is said to be *elliptical* if its characteristic function, defined by $\phi(\mathbf{t}) := \mathbb{E}[\exp(i\,\mathbf{t}'\boldsymbol{x})]$ takes the form

$$\phi(\mathbf{t}) = \exp(i\,\mathbf{t}'\boldsymbol{\mu})\psi(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}) \tag{3.6}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the vector of medians, and $\boldsymbol{\Omega}$ is a matrix, assumed to be positive definite, known as the *dispersion matrix*. The function ψ does not depend on n. We denote the distribution with characteristic function (3.6) by $\mathcal{E}_\psi(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

The name "elliptical" arose because the isoprobability contours are ellipsoidal. If variances exist, then the covariance matrix is proportional to $\boldsymbol{\Omega}$, and if means exist, $\boldsymbol{\mu}$ is also the vector of means.

Eq. (3.6) does not imply that the random vector $\boldsymbol{x}$ has a density, but if it does, then the density must be of the form

$$f_n(\boldsymbol{x}) = |\boldsymbol{\Omega}|^{-1/2}\, g_n\left[(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right] \tag{3.7}$$

Eq. (3.7) is sometimes used as the definition of elliptical distributions, when existence of a density is assumed. In particular, (3.7) shows that if $n = 1$, then the transformed variable $z = (x - \mu)/\sqrt{\Omega}$ satisfies $z \sim \mathcal{E}_\psi(0, 1)$.

The multivariate normal is the most well-known elliptical family; for the normal, one has $g_n(s) = c_n \exp(-s/2)$ (where c_n is a normalization constant) and $\psi(T) = \exp(-T/2)$. Note that g_n depends on n while ψ does not. The elliptical class also includes many non-normal distributions, including examples which display heavy tails, and are therefore better suited to modeling asset returns. For example, the multivariate Student-t distribution with ν degrees of freedom has density of the form (3.7) with

$$g_n(s) \propto (\nu + s)^{-(n+\nu)/2} \tag{3.8}$$

and for $\nu = 1$ one recovers the multivariate Cauchy.

One could choose $g_n(s)$ to be identically zero for sufficiently large s, which would make the distribution of asset returns bounded above and below. Thus, one criticism of the CAPM – that it requires assets to have unlimited liability – is not a valid criticism.

Let $\boldsymbol{v} = \mathbf{T}\boldsymbol{x}$ denote a fixed (non-stochastic) linear transformation of the random vector $\boldsymbol{x}$. It is of interest to relate the characteristic function of $\boldsymbol{v}$ to that of $\boldsymbol{x}$. In general,

$$\begin{aligned} \phi_{\boldsymbol{v}}(\mathbf{t}) &= \mathbb{E}[e^{i\,\mathbf{t}'\boldsymbol{v}}] = \mathbb{E}[e^{i\,\mathbf{t}'\mathbf{T}\boldsymbol{x}}] && (3.9) \\ &= \phi_{\boldsymbol{x}}(\mathbf{T}'t) && (3.10) \end{aligned}$$

If we furthermore assume that the random vector $\boldsymbol{x}$ comes from an ellptical distribution, then we may further expand (3.10) as:

$$\begin{aligned}\phi_{\boldsymbol{x}}(\mathbf{T}'t) &= e^{i\,\mathbf{t}'\mathbf{T}\boldsymbol{\mu}}\psi\left(\mathbf{t}'\mathbf{T}\boldsymbol{\Omega}\mathbf{T}'\mathbf{t}\right) &(3.11)\\ &= e^{i\,\mathbf{t}'\mathbf{m}}\psi\left(\mathbf{t}'\Delta\mathbf{t}\right) &(3.12)\end{aligned}$$

where for convenience we define $\mathbf{m} := \mathbf{T}\boldsymbol{\mu}$ and $\boldsymbol{\Delta} := \mathbf{T}\boldsymbol{\Omega}\mathbf{T}'$; therefore $\boldsymbol{v} = \mathbf{T}\boldsymbol{x}$ is also elliptical, with median vector and dispersion matrix given by $\mathbf{m}$ and $\boldsymbol{\Delta}$.

Even with the same function ψ, the functions f_n, g_n appearing in the density (3.7) can have rather different shapes for different n (= the dimension of $\mathbb{R}^n$), as we see in (3.8). However the function ψ does not depend on n. For this reason, one sometimes speaks of an elliptical "family" which is identified with a single function ψ, but possibly different values of $\boldsymbol{\mu}, \boldsymbol{\Omega}$ and a family of functions g_n determining the densities – one such function for each dimension of Euclidean space. We have shown in (3.12) that marginalization of an elliptical family results in a new elliptical of the same family (ie. the same ψ-function); this is an important property which will be used in the next theorem.

Theorem 3.1. If the distribution of $\boldsymbol{r}$ is elliptical, and if u is a standard utility function, then expected utility is a function of mean and variance, and moreover

$$\partial_\mu \hat{U}(\mu,\omega) \geq 0 \quad\text{and}\quad \partial_\omega \hat{U}(\mu,\omega) \leq 0 \tag{3.13}$$

Proof. For the duration of this proof, fix a portfolio with holdings vector $\boldsymbol{h} \in \mathbb{R}^n$ and let $x = \boldsymbol{h}'\boldsymbol{r}$ denote the wealth increment. Let $\mu = \boldsymbol{h}'\mathbb{E}[\boldsymbol{r}]$ and $\omega^2 = \boldsymbol{h}'\boldsymbol{\Omega}\boldsymbol{h}$ denote moments of x. Applying the marginalization property (3.12) with the $1 \times n$ matrix $\mathbf{T} = \boldsymbol{h}'$ yields

$$\phi_x(t) = e^{i\,t\mu}\psi(t^2\omega^2).$$

The k-th central moment of x will be

$$i^{-k}\frac{d^k}{dt^k}\psi(t^2\omega^2)\Big|_{t=0}$$

From this it is clear that all odd moments will be zero, and the $2k$-th moment will be proportional to ω^{2k}. Therefore the full distribution of x is completely determined by μ, ω, so expected utility is a function of μ, ω.

We now prove the inequalities (3.13). Write

$$\hat{U}(\mu,\omega) = \mathbb{E}[u(x)] = \int_{-\infty}^{\infty} u(x)f(x)dx\,. \tag{3.14}$$

Note that the integral is over a one-dimensional variable. Using the special case of eq. (3.7) with $n = 1$, we have

$$f_1(x) = \omega^{-1}g_1[(x-\mu)^2/\omega^2]. \tag{3.15}$$

Using (3.15) to update (3.14), we have

$$\hat{U}(\mu,\omega) = \mathbb{E}[u(x)] = \int_{-\infty}^{\infty} u(x)\omega^{-1} g_1[(x-\mu)^2/\omega^2]\, dx\,.$$

Now make the change of variables $z := (x-\mu)/\omega$ and $dx = \omega\, dz$, which yields

$$\hat{U}(\mu,\omega) = \int_{-\infty}^{\infty} u(\mu+\omega z) g_1(z^2) dz\,.$$

The desired property $\partial_\mu \hat{U}(\mu,\omega) \geq 0$ then follows immediately from the condition from Definition 3.2 that u is increasing.

The case for $\partial_\omega \hat{U}$ goes as follows.

$$\begin{aligned}
\partial_\omega \hat{U}(\mu,\omega) &= \int_{-\infty}^{\infty} u'(\mu+\omega z) z g_1(z^2) dz \\
&= \Big[\int_{-\infty}^{0} + \int_{0}^{\infty}\Big] u'(\mu+\omega z) z g_1(z^2) dz \\
&= -\int_{0}^{\infty} u'(\mu-\omega z) z g_1(z^2) dz \\
&\qquad + \int_{0}^{\infty} u'(\mu+\omega z) z g_1(z^2) dz \\
&= \int_{0}^{\infty} z g_1(z^2)[u'(\mu+\omega z) - u'(\mu-\omega z)] dz
\end{aligned}$$

A differentiable function is concave on an interval if and only if its derivative is monotonically decreasing on that interval, hence

$$u'(\mu+\omega z) - u'(\mu-\omega z) < 0$$

while $g_1(z^2) > 0$ since it is a probability density function. Hence on the domain of integration, the integrand of

$$\int_{0}^{\infty} z g_1(z^2)[u'(\mu+\omega z) - u'(\mu-\omega z)] dz$$

is non-positive, and hence $\partial_\omega \hat{U}(\mu,\omega) \leq 0$, completing the proof of Theorem 3.1.

Recall Definition 3.4 above of indifference curves. Imagine the indifference curves written in the σ,μ plane with σ on the horizontal axis. If there are two branches of the curve, take only the upper one. Under the conditions of Theorem 3.1, one can make two statements about the indifference curves:

(1) $d\mu/d\sigma > 0$ or an investor is indifferent about two portfolios with different variances only if the portfolio with greater σ also has greater μ,
(2) $d^2\mu/d\sigma^2 > 0$, or the rate at which an individual must be compensated for accepting greater σ (this rate is $d\mu/d\sigma$) increases as σ increases.

These two properties say that the indifference curves are convex.

In case you are wondering how one might calculate $d\mu/d\sigma$ along an indifference curve, we may assume that the indifference curve is parameterized by $\lambda \to (\mu(\lambda), \sigma(\lambda))$ and differentiate both sides of

$$\mathbb{E}[u(x)] = \int u(\mu + \sigma z) g_1[z^2]\, dz\,.$$

with respect to λ. By assumption the left side is constant (has zero derivative) on an indifference curve. Hence

$$0 = \int u'(\mu + \sigma z)(\mu'(\lambda) + z\sigma'(\lambda)) g_1(z^2)\, dz$$

Hence

$$\frac{d\mu}{d\sigma} = \frac{\mu'(\lambda)}{\sigma'(\lambda)} = -\frac{\int_{\mathbb{R}} z\, u'(\mu + \sigma z) g_1(z^2)\, dz}{\int_{\mathbb{R}} u'(\mu + \sigma z) g_1(z^2)\, dz}$$

If $u' > 0$ and $u'' < 0$ at all points, then the numerator $\int_{\mathbb{R}} z\, u'(\mu + \sigma z) g_1(z^2)\, dz$ is negative, and so $d\mu/d\sigma > 0$.

The proof that $d^2\mu/d\sigma^2 > 0$ is similar (exercise).

What, exactly, fails if the distribution $p(\boldsymbol{r})$ is not elliptical? The crucial step of this proof assumes that a two-parameter distribution $f(x; \mu, \sigma)$ can be put into "standard form" $f(z; 0, 1)$ by a change of variables $z = (x - \mu)/\sigma$. This is not a property of all two-parameter probability distributions; for example, it fails for the lognormal.

One can see by direct calculation that for logarithmic utility, $u(x) = \log x$ and for a log-normal distribution of wealth,

$$f(x; m, s) = \frac{1}{sx\sqrt{2\pi}} \exp(-(\log x - m)^2/2s^2)$$

then the indifference curves are not convex. The moments of x are

$$\mu = e^{m+s^2/2}, \quad \text{and} \quad \sigma^2 = (e^{m+s^2/2})^2(e^{s^2} - 1)$$

and, with a little algebra one has

$$\mathbb{E}u = \log \mu - \frac{1}{2}\log(\sigma^2/\mu^2 + 1) \tag{3.16}$$

One may then calculate $d\mu/d\sigma$ and $d^2\mu/d\sigma^2$ along a parametric curve of the form

$$\mathbb{E}u = \text{constant} \tag{3.17}$$

and see that $d\mu/d\sigma > 0$ everywhere along the curve, but $d^2\mu/d\sigma^2$ changes sign.

Theorem 3.1 implies that for a given level of median return, investors always dislike dispersion. We henceforth assume, unless otherwise stated, that the first two moments of the distribution exist. In this case, (for elliptical distributions) the median is the mean and the dispersion is the variance, and hence the underlying asset return distribution is mean-variance equivalent in the sense of Def. 3.1. We emphasize that this holds for any smooth, concave utility.

Problem 3.1. Under the conditions of Theorem 3.1, show that $d^2\mu/d\sigma^2 > 0$, or the rate at which an individual must be compensated for accepting greater σ (this rate is $d\mu/d\sigma$) increases as σ increases.

Problem 3.2. Prove (3.16), and then use (3.16) to show that for a parametric curve of the form (3.17), $d\mu/d\sigma > 0$ everywhere along the curve, but $d^2\mu/d\sigma^2$ changes sign. Verify this on a computer by plotting the relevant curves for various values of the parameters m, s.

References

Arrow, Kenneth J (1963). "Liquidity preference, Lecture VI in "Lecture Notes for Economics 285, The Economics of Uncertainty", pp 33-53". In:

Feldstein, Martin S (1969). "Mean-variance analysis in the theory of liquidity preference and portfolio selection". In: *The Review of Economic Studies* 36.1, pp. 5–12.

Markowitz, Harry (1952). "Portfolio selection*". In: *The Journal of Finance* 7.1, pp. 77–91.

Pratt, John W (1964). "Risk aversion in the small and in the large". In: *Econometrica: Journal of the Econometric Society*, pp. 122–136.

Tobin, James (1958). "Liquidity preference as behavior towards risk". In: *The review of economic studies* 25.2, pp. 65–86.