1. Introduction

1.1. **Prices, returns, holdings, and portfolios.** The goal of this course is to develop a quantitative approach to portfolio management. Such an approach is absolutely necessary when managing a systematic trading strategy, but the mathematical tools we will develop are useful for analyzing any portfolio, including portfolios built by fundamental managers.

Some fundamental analysts are more adept at analyzing companies than they are at building portfolios, and there are many aspects of portfolio construction that are simply too complex to be performed manually by humans. Hence ideas of fundamental managers are perhaps best monetized using mathematically sophisticated *portfolio construction* techniques like the ones we'll cover. All of the techniques we will discuss are actually used in most of the largest investment banks, asset management firms, and hedge funds in the world today.

Our topic is *portfolio theory*, interpreted quite broadly as the study of any set of holdings involving one or more assets. The assets involved could be financial assets, such as shares in corporations, or they could be non-financial real assets such as gold bars or land. An important subfield of portfolio theory is the study of optimization: how to form portfolios that are optimal with respect to the investor's utility function, but we will also develop tools to study portfolios which weren't necessarily constructed optimally.

Changes in portfolio holdings are called *trades*. Although it's not always emphasized in classical studies of portfolio theory, (eg. all of the foundational work on "Modern Portfolio Theory (MPT)" ignores trading cost completely!) trading itself is a complex process with many decision variables that need to be considered and potentially optimized. We will discuss how trades are executed in continuous, liquid markets later in the course, and the topic of market microstructure is a full-semester course on its own. For now, suffice it to say that all trading is costly, and the costs from trading can be a meaningful part of the ultimate profitability of a given trading strategy or fund.

We must walk before we can run. Hence our very first forays into the world of portfolio management will be concerned with establishing a coherent notation and a coherent language for describing the fundamental actors in our story: prices, returns, holdings, and portfolios.

Loosely, or in conversation, one sometimes speaks of "the price" of an asset, but in continuous limit order book markets (such as most of the world's equity markets in developed countries) there are always at least two prices.

Definition 1.1. A *bid price* is the highest price that a buyer is willing to pay for the security. The *ask price*, or offer, is the lowest price a seller will accept for the security. The *midpoint price* is the mathematical midpoint between the best bid and offer, $(p_{\text{bid}} + p_{\text{ask}})/2$. These prices are rarely the same: the ask is usually higher than the bid. If you are buying a stock, you pay the ask price. If you sell a stock, you receive the bid price. The difference between the two prices is called the *spread*.

In the absence of some form of predictability in the dynamics of the limit order book, a round-trip transaction of buying a share and selling it subsequently (or selling then buying to cover) must at least cost the spread.

Other kinds of trading costs include borrow fees if one is selling short, financing costs on the long positions, transaction taxes in some markets, and the market impact due to one's own trading, among others. One implication is that strategies with high turnover must also have extremely accurate forecasts to overcome the very-high transaction costs which exist.

Let all of the assets we are going to consider for our portfolio be indexed by $i = 1, \ldots, n$. They do not need to be stocks, but in some cases we will use equities as convenient examples, or use language derived from the equity markets.

Asset prices at time t will be denoted p_t with p_t^i the price for the i-th asset. The *return* of the i-th asset over the interval $[t, t+1]$ will be denoted

$$r_{t+1}^i = p_{t+1}^i / p_t^i - 1. \tag{1.1}$$

As mentioned above, what to use as the price is a subject of some subtlety. One could use the midpoint price as in (1.1), or one could use the last trade price before some cutoff time (such as the market close time). One could also use the volume-weighted average price (VWAP) over some interval. Moreover, one could convert both of the prices in (1.1) to some other currency. There are hence many definitions of *return* appropriate for different purposes.

The prices p_t^i may sometimes need to be adjusted in order that the returns (1.1) represent the total return from holding the asset. For example, in the equity markets, stock splits and dividends are common occurrences.

Definition 1.2. A *stock split* is an issue of new shares in a company to existing shareholders in proportion to their current holdings. A *dividend* is a sum of money paid by a company to its shareholders out of its profits or reserves.

When a stock split occurs, typically after trading hours, the price instantaneously adjusts before the next trading session begins to reflect the dilution factor. A similar effect occurs for dividends on the ex-dividend date. If there are stock splits or dividends between t and $t+1$, the effect of those should be included in the return variable, for example by adjusting the price before the split to be in the same units as price after the split before calculating (1.1). Once this adjustment has been made, the quantity (1.1) is called *total return*.

One must be very careful with stock splits. If a stock price goes from 100 to 1000, the company is more likely to consider a split, because higher share prices make the stock less accessible to individual retail investors. Perhaps the most famous example of this is Apple, but there are many others. If one could look into the future and observe the set of stocks that were going to split, one would have a very profitable (fictitious) strategy. Hence the variable $1/p_t$, where p_t is adjusted to be in units of today's shares, is a (spurious, un-implementable) forecast. It isn't implementable because it has a *lookahead bias*, which is defined as some way of looking into the future in a backtest.

Anecdote: I once asked the CEO of our company why a certain portfolio manager was fired. The reply was: the P&L was straight down, and the backtest P&L was straight up, even when the backtest was run over the same period during which the live P&L had been straight down. Such stories abound – no quantitative trading strategy ever goes live without a good-looking backtest, and the majority of them fail when run in live. Some of the most common reasons are lookahead bias, overfitting, and misunderstanding of market impact.

Just as a length measurement might have units of inches or centimeters, it is useful to think of a stock price p_t as being in units of "shares on day t," while "shares on day $t+1$" are in different units, and must be converted into units of "shares on day t" before any mathematical operation combining the two (such as division, to compute the return) could meaningfully be attempted. A sophisticated programmer could structure code in such a way as to render such meaningless comparisons impossible.

We will also sometimes work with *log-relative returns* or "logrels" for short, which are

$$\log(1 + r^i_{t+1}) = \log p^i_{t+1} - \log p^i_t.$$

These are convenient primarily because the convert a compounded return over many periods, such as

$$\prod_{t=1}^{T}(1 + r^i_t)$$

into a sum. Logrels are something it makes sense to add, while returns can be added only in an approximate calculation. For small r, one has $\log(1 + r) \approx r$ using the Taylor series, so we will sometimes model returns as linear functions of other variables.

Definition 1.3. The notation r_{t+1} without a superscript denotes the entire $n \times 1$ column vector of asset returns, typically referred to as a "cross-section" of asset returns.

When $t+1$ is in the future, r_{t+1} is a random variable with values in $\mathbb{R}^n$. We do not know the probability density function (p.d.f.) of this random variable. Indeed, much of the field of empirical finance is dedicated to developing methods which can help to discover information, knowable at time t, that is relevant to the multivariate density of r_{t+1}. Any structure inherent in the random process driving r_{t+1} potentially gives rise to analogous structure in the portfolio's return.

Definition 1.4. Throughout this course, we will let h^i_t denote our holdings[1] in the i-th security at time t. The number h^i_t will always have units either of US dollars, or another appropriate currency converted to dollars. The portfolio at time t is $h_t \in \mathbb{R}^n$.

Note that a stock portfolio is, in actuality, a collection of holdings of *shares*, not dollar amounts as per the previous definition. To convert shares into the dollar values of the holdings, one must multiply by a price. Hence $h^i_t = s^i_t p^i_t$. As the prices change, so do the values of h^i_t even in the absence

[1]Mnemonic: h is the first letter in *holdings*.

of any trading. The change in the price is equivalently expressed as a return as per (1.1).

Then $h_t \cdot r_{t+1}$ is the portfolio's (paper) profit and loss (abbreviated P&L) over the interval $[t, t+1]$, where the dot denotes a scalar product or "dot product" of the two vectors. Information (eg. from a statistical model's forecast) about the variance-covariance matrix $\mathrm{cov}(r_{t+1})$ potentially gives information about $\sigma^2(h_t)$, the portfolio variance:

$$\sigma^2(h_t) := V[h_t \cdot r_{t+1}] = h_t' \, \mathrm{cov}(r_{t+1}) h_t \, .$$

Minimum volatility funds seek to minimize $\sigma^2(h_t)$ while remaining fully invested. This is harder than it sounds, and we will discuss this problem in some detail later.

The vector r_{t+1}^{rf} denotes the relevant vector of *risk-free rates*. If the asset returns one is modeling are local currency returns, one must be careful to have the correct risk-free rate for each currency!

Definition 1.5. *Excess returns* are the asset returns minus the appropriate risk-free rate of return:

$$\underset{[n \times 1]}{r_{t+1} - r_{t+1}^{\mathrm{rf}}}$$

where for concreteness, we will often indicate the matrix dimensions of each matrix or vector below it.

Many classical results in finance are derived under simplistic assumptions, which we know aren't exactly satisfied in reality. The most common such assumption is normality; asset returns are *never* normally distributed, and yet, it's probably the most common of the various assumptions in finance that are made to aid in computational tractability. A normal distribution is entirely specified if you know its first two moments (mean vector and covariance matrix). Under normality assumptions, one investment can be said to *strictly dominate* another one if it has both higher return and lower risk, where *risk* here is simply the standard deviation of the investment's return.

Many sources I've come across, including published articles in prestigious academic journals, are confused about normal distributions and when normality is needed as an assumption. A very common misconception is that mean-variance optimization assumes normality of returns – it doesn't! Neither does the capital asset pricing model.

It makes sense to consider returns above the relevant risk-free rate when evaluating risky investments. A risky investment with expected return below or equal that of the risk-free investment would be strictly dominated by the risk-free investment. It is for this reason that the Sharpe ratio, defined below, is defined using excess returns:

Definition 1.6. In 1966, William F. Sharpe developed the *Sharpe ratio*:

$$S = \frac{\mathbb{E}[r - r_f]}{\sqrt{\mathrm{var}[r]}}.$$

where r is a scalar-valued random variable denoting the return of an investment. The universal convention is that, whatever frequency the returns are sampled at (daily, weekly, monthly, ...), Sharpe ratio calculations always use annualized returns and annualized volatility.

Here $\mathbb{E}[\,]$ denotes the expectation value of a random variable. As such it does not directly refer to past returns; it is an aspect of a model or a forecast of the future. To emphasize this, the phrase *ex ante* (from Latin ex 'from, out of' + ante 'before.') is often used. One could, of course, also compute the statistic S on past (realized) returns, giving an *ex post* or *realized* Sharpe ratio. The Sharpe ratio is closely related to the Student's t-test for the null hypothesis of whether the true mean is zero.

In testing the null hypothesis that the population mean is equal to zero, one uses the statistic $t = \overline{x}/(s/\sqrt{n})$ where $\overline{x}$ is the sample mean, s is the sample standard deviation, and n is the sample size. If the sample is of daily returns, and if $n = 252$ (the number of trading days in a year), then this equals the Sharpe ratio, but as the sample size grows, the t-statistic also increases (we are increasingly sure that the mean isn't zero), but the Sharpe ratio simply converges to its asymptotic value.

When applied to the returns of a mutual fund, the Sharpe ratio is not the ultimate measure of a manager's skill. This is because most mutual funds are constrained in some way which entails that they cannot simply seek absolute return over the risk-free rate. For example, it is almost impossible for a sufficiently diversified long-only fund to realize a positive Sharpe ratio over a period when the market is down substantially, no matter how great the manager's skill.

Definition 1.7. A *benchmark* is a standard against which the performance of a security, mutual fund or investment manager can be measured. Usually, broad market and market-segment stock and bond indexes are used for this purpose.

In a year like 2008, it's almost impossible for a well-diversified long-only US equity portfolio to have done well in absolute terms, regardless of the skill of the manager. Hence a manager's skill is closely related to performance *relative to the appropriate benchmark.* For this reason, one often considers the information ratio (IR), which is the same formula as the Sharpe ratio, but with the risk-free rate being replaced by an appropriate benchmark. Some managers do in fact have the risk-free rate as their benchmark; these are so-called *absolute return* strategies, but this is more typical for hedge funds than for mutual funds.

1.2. **Utility.** This part of the lecture takes inspiration from Chapter 1 of Eeckhoudt, Gollier, and Schlesinger (2005), which itself follows Pratt (1964) and Bernoulli (1954). Seminal contributions to the material of this lecture include Pratt (1964), Arrow (1963) and De Finetti (1952).

Nicolas Bernoulli submitted five problems to the mathematician Pierre Rémond de Montmort in 1713. These problems are reproduced as an appendix to the second edition of Montmort, 1714 "L'analyse sur les jeux de hazard", p. 402. The last of these problems runs as follows.

> *Peter tosses a coin and continues to do so until it should land "heads" when it comes to the ground. He agrees to give Paul one ducat if he gets "heads" on the very first throw, two ducats if he gets it on the second, four if on the third, eight if on the fourth, and so on, so that with each additional throw the number of ducats he must pay is doubled. Suppose we seek to determine the value of Paul's expectation.*

— Bernoulli (1954) quoting Montmort, 1714

An important breakthrough in the analysis of decisions under risk was achieved in 1738 when Daniel Bernoulli wrote a Latin manuscript entitled: "Exposition of a New Theory on the Measurement of Risk." (translated in Bernoulli (1954)) Its main purpose is to show that two people facing the same lottery may value it differently. This idea was quite novel at the time, since famous scientists before Bernoulli (among them Pascal and Fermat) had argued that the value of a lottery should be equal to its mathematical expectation and hence identical for all people, independently of their risk attitude. In order to justify his ideas Bernoulli uses three examples. One of them, the "St. Petersburg paradox" is the problem proposed by Nicolas Bernoulli to Montmort in 1713.

The celebrity of the St. Petersburg paradox has overshadowed the other two examples given by Daniel Bernoulli that show that, most of the time, the value of a lottery is not equal to its mathematical expectation. One of these two examples, which presents the case of an individual named "Sempronius," anticipates the central contributions that will be made to risk theory about 230 years later by Arrow, Pratt and others.

> Sempronius owns goods at home worth a total of 4000 ducats and in addition possesses 8000 ducats worth of commodities in foreign countries from where they can only be transported by sea. However our daily experience teaches us that of two ships one perishes.

In modern-day language, we would say that Sempronius faces a risk on his wealth. This wealth may represented by a lottery $\tilde{x}$, which takes on a value of 4000 ducats with probability 1/2 (if his ship is sunk), or 12000 ducats with probability 1/2. Its mathematical expectation is given by

$$E[\tilde{x}] = \frac{1}{2}4000 + \frac{1}{2}12000 = 8000$$

Now Sempronius has an ingenious idea. Instead of trusting all his 8000 ducats of goods to one ship he now trusts equal portions of these commodities to two ships. Assume that the ships follow independent but equally dangerous routes. If both ships perish (probability $(1/2)^2 = 1/4$), he would end up with his sure wealth of 4000 ducats. Similarly, both ships will succeed with probability 1/4, in which case his final wealth amounts to 12000 ducats. Finally, if only one ship succeeds to download the commodities safely, the final wealth would amount to 8000 ducats. The probability of this event is 1/2 because it is the complement of the other two events.

$$E[\tilde{y}] = \frac{1}{4}4000 + \frac{1}{2}8000 + \frac{1}{4}12000 = 8000$$

Hence, according to Bernoulli and to modern risk theory, the mathematical expectation of a lottery is not an adequate measure of its value. Bernoulli suggests a way to express the fact that most people prefer $\tilde{y}$ to $\tilde{x}$: a lottery should be valued according to the "expected utility" that it provides. Instead of computing the expectation of the monetary outcomes, we should use the expected utility of the wealth. This relationship is characterized by a utility function u, which for every wealth level x tells us the level of "satisfaction" or "utility" $u(x)$ attained by the agent with this wealth.

Although the function u transforms the objective result x into a perception, this transformation is assumed to exhibit some basic properties of rational behavior. For example, the function should be increasing in x. Bernoulli argues that if the utility u is not only increasing but also concave, then lottery $\tilde{y}$ will have higher expected utility than lottery $\tilde{x}$, in accordance with intuition. The concavity of the relationship between wealth x and utility u is quite natural. It implies that the marginal utility of wealth is decreasing with wealth: one values a one-ducat increase in wealth more when one is poorer than when one is richer.

Definition 1.8. An agent is *risk-averse* if, at any wealth level w, she dislikes every lottery with an expected payoff of zero: $\forall\, w, \forall\, \tilde{z}$ with $E[\tilde{z}] = 0$,

$$E[u(w + \tilde{z})] \leq u(w).$$

Any lottery $\tilde{z}$ with non-zero expected payoff can be decomposed into its expected payoff $E[\tilde{z}]$ and a zero-mean lottery $\tilde{z} - E[\tilde{z}]$. Thus a risk-averse agent always prefers receiving the expected outcome of a lottery with certainty, rather than the lottery itself. This implies that, for any lottery $\tilde{z}$ and for any initial wealth w,

$$E[u(w + \tilde{z})] \leq u(w + E[\tilde{z}])$$

Proposition 1.1. An agent is risk-averse, ie. inequality

$$E[u(w + \tilde{z})] \leq u(w + E[\tilde{z}])$$

holds for all w and $\tilde{z}$, if and only if $u()$ is concave.

The above Proposition is a rewriting of the famous Jensen's inequality. Consider any real-valued function ϕ. Jensen's inequality states that $E\phi(\tilde{y}) \leq \phi(E[\tilde{y}])$ for any random variable $\tilde{y}$ if and only if ϕ is concave.

One way to measure the degree of risk aversion of an agent is to ask how much she is ready to pay to get rid of a zero mean risk $\tilde{z}$. The answer to this question will be referred to as the risk premium Π associated to that risk. Mathematically, the risk premium is defined by

$$E[u(w + \tilde{z})] = u(w - \Pi)$$

One very convenient property of the risk premium, is that it is measured in the same units as wealth. This allows us to define how much wealth a given risk is worth.

Definition 1.9. The *certainty equivalent* e of risk $\tilde{z}$ is the sure increase in wealth that has the same effect on welfare as having to bear risk $\tilde{z}$, i.e.,

$$E[u(w+\tilde{z})] = u(w+e)$$

When $E[\tilde{z}] = 0$ one has $e = -\Pi$ which makes sense given the interpretation of Π as the amount one would pay to remove (with certainty!) the risk. A direct consequence of Proposition 1.1 is that the risk premium is nonnegative when u is concave, i.e. when the agent is risk-averse.

The risk premium is a complicated nonlinear function of the distribution of $\tilde{z}$, of initial wealth w and of the utility function u. We can attempt to approximate the risk premium using the Taylor series for $u()$:

$$\begin{aligned} u(w-\Pi) &\approx u(w) - \Pi u'(w) \\ E[u(w+\tilde{z})] &\approx u(w) + \frac{1}{2}\sigma^2 u''(w) \end{aligned}$$

Combining these, one has

$$\Pi \approx \frac{1}{2}\sigma^2 A(w) \quad \text{where} \quad A(w) := -\frac{u''(w)}{u'(w)} \tag{1.2}$$

Under risk aversion, function A is positive. It would be zero or negative respectively for a risk-neutral or risk-loving agent. $A(\cdot)$ is hereafter referred to as the *degree of absolute risk aversion* of the agent.

Eq (1.2) from the previous page:

$$\Pi \approx \frac{1}{2}\sigma^2 A(w) \quad \text{where} \quad A(w) := -\frac{u''(w)}{u'(w)}$$

is known as the *Arrow-Pratt approximation*, as it was developed independently by Arrow (1963) and Pratt (1964). The cost of risk, as measured by the risk premium, is approximately proportional to the variance of its payoffs. Thus, the variance might appear to be a good measure of the degree of riskiness of a lottery. Kenneth Arrow was Professor of Economics at Harvard and the youngest person ever to receive the Nobel Prize in Economics. He graduated from City College (CUNY), Class of 1940.

Definition 1.10. An agent v is said to be *more risk-averse* than another agent u with the same initial wealth if the risk premium of any risk is larger for agent v than for agent u.

Proposition 1.2. The following three conditions are equivalent:
(a) Agent v is more risk-averse than agent u,
(b) For all w, $A_v(w) \geq A_u(w)$.
(c) Function v is a concave transformation of function u, meaning:

$$\exists\, \phi \;\text{ with }\; \phi' > 0 \;\text{ and }\; \phi'' < 0 \;\text{ such that }\; v(w) = \phi(u(w))$$

We leave the proof of the last proposition as an exercise.

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.

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.

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

$$E[u(\tilde{w})] = u\left(\mu - \frac{1}{2}\kappa\sigma^2\right) \tag{1.3}$$

The relation (1.3) means that maximizing $E[u(\tilde{w})]$ is equivalent to maximizing

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

since u is monotone.

Do not confuse the direction of logical inference; we have shown that normality implies mean-variance equivalence, not the converse!

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.

We henceforth assume that volatility in the wealth random variable $\tilde{w}$ comes from trading financial assets (as opposed to the stormy seas our friend Sempronius had to worry about!)

Let $h \in \mathbb{R}^n$ denote the portfolio holdings, measured in dollars or an appropriate numeraire currency, at some time t in the future. Let h_0 denote the current portfolio. let $R \in \mathbb{R}^n$ denote the return over the interval $[t, t+1]$. Hence $R \in \mathbb{R}^n$ 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 $\tilde{w} = h'R$ and the expected-utility maximizer chooses $\boldsymbol{h}$ to satisfy

$$\boldsymbol{h}^* = \operatorname{argmax} \mathbb{E}[u(w)] \tag{1.4}$$

Definition 1.11. The underlying asset return distribution, $p(R)$, is said to be *mean-variance equivalent* if, for any increasing utility function u, there exists some constant $a > 0$ (where a depends on u) such that

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

where $\boldsymbol{h}^*$ is defined by (1.4).

Which distributions, then, are mean-variance equivalent? We showed above that the normal distribution is; this is easy. Tobin (1958) conjectured that 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.

Definition 1.12. For a given asset return distribution $p(R)$, we say *expected utility is a function of mean and variance* if for any increasing utility function u, one has $\mathbb{E}[u(w_1)] = \mathbb{E}[u(w_2)]$ whenever $\mathbb{E}[w_1] = \mathbb{E}[w_2]$ and $\mathbb{V}[w_1] = \mathbb{V}[w_2]$.

Definition 1.12, like Definition 1.11, is a property of the asset return distribution $p(R)$; some distributions have this property, and some do not. This holds for every increasing function $u()$ if and only if $w_1 \underset{d}{=} w_2$ whenever $\mathbb{E}[w_1] = \mathbb{E}[w_2]$ and $\mathbb{V}[w_1] = \mathbb{V}[w_2]$.

Chamberlain (1983) has shown that the distributions satisfying Definition 1.12 are precisely the *elliptical* distributions, defined just below.

A multivariate density is said to be *elliptical* if its characteristic function, defined by $\phi_n(\mathbf{x}) := \mathbb{E}[\exp(i\,\mathbf{t}'\mathbf{x})]$ takes the form

$$\phi_n(\mathbf{x}) = \exp(i\,\mathbf{t}'\boldsymbol{\mu})\psi(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}) \tag{1.5}$$

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.* Ingersoll (1987) points out that if variances exist, then the covariances matrix is proportional to $\boldsymbol{\Omega}$, and if means exist, $\boldsymbol{\mu}$ is also the vector of means.

Eqn. (1.5) does not imply that the random vector $\mathbf{x}$ has a density, but if it does, then the density must be of the form

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

This includes many non-normal distributions of interest, including fat-tailed distributions. For example, the multivariate Student-t distribution with ν degrees of freedom is characterized by $g_n(s) \propto (\nu+s)^{-(n+\nu)/2}$, and for $\nu = 1$ one recovers the multivariate Cauchy.

Theorem 1.1. If the distribution of R is elliptical, and if u is differentiable and concave, then there exists a function f such that $\mathbb{E}[u(w)] = f(\mu, v)$ where $\mu := \boldsymbol{h}'\boldsymbol{\mu}$ and $v = \boldsymbol{h}'\boldsymbol{\Omega}\boldsymbol{h}$. Moreover,

$$\frac{\partial}{\partial v} f(\mu, v) < 0.$$

Theorem 1.1 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, $p(\mathbf{x})$, is mean-variance equivalent in the sense of Def. 1.11. We emphasize that this holds for any smooth, concave utility.

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

$$E[h'R] - \frac{1}{2}\kappa V[h'R] = h'E[R] - \frac{1}{2}\kappa\, h'V[R]h \tag{1.6}$$

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 (1.6). 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 $h^* = (\kappa V[R])^{-1} E[R]$, but we don't know $V[R]$ or $E[R]$.

Note that (1.6) requires a vector and a matrix:

$$E[R] \in \mathbb{R}^n, \quad \text{and} \quad V[R] \in S^n_{++}, \tag{1.7}$$

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 $V[R]$ with eigenvalue zero would be a long-short portfolio with zero variance. I don't think this animal exists in nature, aside from contrived artificial cases, 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 $V[R]$ 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 $V[R] \propto B'B \in S^n_+$. The rank of $B'B$ is at most T, so if $T < n$ it is impossible that $\mathbb{V}[R]$ is invertible.

References

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

Bernoulli, Daniel (1954). "Exposition of a new theory on the measurement of risk". In: *Econometrica: Journal of the Econometric Society*, pp. 23–36.

Chamberlain, Gary (1983). "A characterization of the distributions that imply meanVariance utility functions". In: *Journal of Economic Theory* 29.1, pp. 185–201.

De Finetti, Bruno (1952). "Sulla preferibilita". In: *Giornale degli economisti e Annali di Economia*, pp. 685–709.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger (2005). *Economic and financial decisions under uncertainty.* Princeton University Press.

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.

Ingersoll, Jonathan E (1987). *Theory of financial decision making.* Rowman & Littlefield.

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

Montmort, Pierre Rémond de (1714). *Essai d'analyse sur les jeux de hazards. 2.* Jombert.

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.