Risk and Portfolio Management with Econometrics, Courant Institute, Fall 2009
http://www.math.nyu.edu/faculty/goodman/teaching/RPME09/index.html
Jonathan Goodman[1], goodman@cims.nyu.edu

Section 1, Mean variance analysis

All models are wrong. Some models are useful.
– George Box, Statistician

1 Risk and return

The *buy side* of the financial services industry is about deciding how to invest resources. An *investor* is a person who has (or an institution that has) an *endowment* to *invest* for the purpose of increasing the endowment. This investor could be an individual with retirement savings, a university with an endowment, a pension fund manager, a company with capital, or a professional manager investing in behalf of clients.

The investor, or the buy side professional acting on the investor's behalf, has a tradeoff between *expected return* and *risk*. He or she makes investment decisions without knowing which ones will yield the greatest return. Of course, we try to predict the returns of various investments. But we also model the *uncertainty* in our forecasted returns. An investor may hesitate to invest in a scheme that is likely to make him or her wealthy but has some chance of bankrupting him or her instead.

The *theory of investment* frames these issues systematically using probability models and optimization. The *probability model* is a description of the future as a random variable. We cannot predict the future state of the world, but we try to predict or model the probability distribution it is taken from. The optimization problem is to choose the best possible investment (or investment strategy) in light of the potential gains and risks as quantified in the probability model.

This class will follow the historical development of this subject and start with probability models and portfolio selection criteria that now are believed to be naive and dangerous.[2] Nevertheless, it will serve as a basis for more refined probability models and selection criteria that prudent investment institutions use today.

It is possible to discuss these issues in a philosophical way, but I prefer to begin by formulating specific if unrealistic models. The drawbacks of these simple models will guide the more sophisticated models and selection criteria that are the main object of this course.

[1]Disclaimer: The author of these notes holds some opinions that differ from those of some of the authors of the standard texts and some investment practitioners. He sometimes indicates areas of disagreement, but the reader should understand that some of the opinions expressed in these notes are not universally shared by all investors.

[2]The popular book *The Black Swan*, by Nassim Taleb (Fellow of the program in financial mathematics at Courant) makes this point forcefully.

2 Mean Variance analysis

Classical portfolio theory, as it existed in about 1975, had two parts. The first was mean variance analysis as a way to allocate assets in a world of risk and return. The investor was assumed to know the risks and the returns of all available investments, but his or her choices were assumed not effect the market, neither prices nor uncertainties in returns. The second part explored what the world would be like if every investor used mean variance analysis, using identical estimates market parameters. We comment on this *CAPM* model briefly below.

A simple multi-asset investment problem is as follows. We have n risky assets. Let R_i denote the *return*[3] on asset i. We suppose our total wealth to be invested is 1, in some units. We will *allocate* w_i to asset i. Assuming that the investment is linear (not true in many business decisions, but nearly true in investments), the total return will be

$$R \ = \ \sum_{i=1}^{n} w_i R_i \ . \tag{1}$$

The total *expected return* is

$$\overline{\mu}_R \ = \ E\left[R\right] \ = \ \sum_{i=1}^{n} w_i E\left[R_i\right] \ = \ \sum_{i=1}^{n} w_i \mu_i \ , \tag{2}$$

where $\mu_i = E\left[R_i\right]$ is the expected return from asset i. We postpone a longer discussion of some of the ways (1) and (2) can be wrong. But one of the major ways is that (2) implicitly assumes that the numbers μ_i are known. We must take (1) subject to the *endowment constraint*

$$\sum_{i=1}^{n} w_i \ = \ 1 \ . \tag{3}$$

Since the w_i sum to unity, they often are called *weights* in the *portfolio allocation* problem. However, we do not always constrain the w_i to be positive, so some of the asset weights can be negative.

The most naive possible portfolio allocation problem would be to choose the w_i subject to the constraint (3) in a way that maximizes the expected return (2). This leads to trivial results. If there are no constraints on the w_i other than (3), there is no optimum (except in very special cases). We put a large positive weight on the asset with the largest μ_i and compensating negative weights on assets with smaller expected return. Such allocations are illegal or anyway impossible for most investors.

But even if possible, most investors would find such extreme choices too risky. Although the expected return is a large positive number, there also is a

[3] If you buy at price P_1 and sell at price P_2, the return is the dimensionless number $R = (P_2 - P_1)/P_1$.

substantial probability of a large negative return. Most investors would want to balance the expected return of an investment with the risk involved.

Variance is one measure of the risk of an asset allocation. Suppose the R_i have variances and covariances:

$$\sigma_i^2 \;=\; \sigma_{ii} \;=\; \text{var}\left[R_i\right] \;=\; E\left[\left(R_i - \mu_i\right)^2\right] \;, \tag{4}$$

and

$$\sigma_{ij} \;=\; \text{cov}\left[R_i, R_j\right] \;=\; E\left[\left(R_i - \mu_i\right)\left(R_j - \mu_j\right)\right] \;. \tag{5}$$

The the variance of the total return in (1) is

$$\sigma_R^2 \;=\; \text{var}[R] \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} \;=\; w^t \Sigma w \;, \tag{6}$$

where we use matrix notation for the last equality: w is a column vector whose components are the w_i, w^t is the row vector that is the transpose of w, and Σ is the *covariance matrix* whose entries are the variances and covariances (4) and (5).

The *risk averse* investor chooses the weights w_i to make the expected return (2) large and the risk, as measured by the variance (6) small. There is no universal risk/return tradeoff that holds for all investors. Some will tolerate more risk for the sake of higher expected return, while others will tolerate lower expected return for the sake of less risk. This class will cover many technical ways to specify the risk/return tradeoff, but all of them have disadvantages.

We cannot identify a universal optimal portfolio, but we can identify many portfolios as sub-optimal. In the context of mean/variance analysis, a portfolio is called *inefficient* if it is possible to increase the expected return without increasing the variance, or (which is the same thing), if it is possible to decrease the variance without decreasing the expected return. A portfolio is *efficient* if it is not inefficient. The set of efficient portfolios forms the *efficient frontier* in w space, the space of all portfolios that satisfy the wealth constraint (3).

2.1 Market assumptions

The mean and variance analysis we do here uses many simplifying assumptions and approximations. A complete list would degenerate into philosophy, so here are just a few.

- The matrix Σ and the vector μ are known exactly.
- The investor can purchase any amount of any asset, either positive or negative. Having a negative amount of a risky asset is *short selling*, or simply *shorting*. Having a negative amount of the risk free asset is called *borrowing*. There is no restriction that the number of shares owned should be an integer.

- The investor is a *price taker*, which means that the investor may purchase any amount of the asset, and nothing the investor does will effect the asset price. The price per share is independent of the amount purchased.

- The price is the same for long and short positions. This really is a combination of previous points.

- There are no *transaction costs*, which means that if an investor first buys w worth of any asset then immediately sells it back, the net change of wealth of the investor is zero. If there were transaction costs, this round trip would have a non-zero net cost to the investor. This also is implicit in the above points.

None of these is exactly true in actual markets. Market parameters are not known exactly. There are costs associated with borrowing beyond paying the risk free rate. There are limits to the investor's ability to borrow, limits that have become more severe in the past year.

Even small investors are not pure price takers. For example, if the *ask price* for a given security is X, that means that someone has an outstanding offer to sell shares of that security at the price X. In a typical situation, there is a relatively small number of shares available at the ask price, as few as 100 shares. If the investor wishes to purchase 200 shares, he or she will have to buy the second 100 at a slightly higher price. This is called *moving the market.* The total price per share for the 200 shares will be slightly more than X.

There usually is a gap between the lowest ask price and the highest *bid* price, the price at which there is an outstanding offer to buy. The difference between these is the *bid-ask spread.*[4] If an investor would buy then immediately sell a share, he or she would lose the bid-ask spread, even if the market did not move (the bid and ask prices did not change) and there were no other transaction costs.

An asset is *liquid* if it is easy to buy or sell. It is rare that an asset is totally illiquid – cannot be bought or sold at any price. More commonly, the degree of liquidity of a market traded asset is measured by its bid-ask spread and how much transactions move the market. Very liquid assets are, for example, market index futures. Less liquid are individual corporate bonds.

Unknown liquidity is a form of risk as important as unknown μ and Σ. Published prices typically are only the best bid and ask prices. It may be hard to know in advance how many shares will be available at the bid and ask prices, or how much a given size of transaction will move the market. For example, suppose $X(t)$ is the time varying price of a given asset. A *stop loss* strategy is to sell asset at price X_0 the first time $X(t) = X_0$. A lack of liquidity may make this impossible. Many institutional investors suffered large losses in the

[4]Small price differences often are called *spreads.* For example, the difference in a bond price for a default free bond and a bond that may default is the *credit spread.* The difference between the return on short dated US Treasuries and LIBOR rates with the same maturities is the *TED* spread – Treasury Euro-Dollar.

Black Friday market decline in 1987 (20% drop in one day) as stock prices fell discontinuously and bid offers disappeared completely.

2.2 Lagrange multipliers

The method of *Lagrange multipliers* is a mathematical technique for finding the maximum or minimum of a function subject to constraints. In the simplest case, there are n variables, $x_1, \cdots, x_n$ that form the components of a vector $x \in R^n$. The problem is to find x that maximizes a smooth function $f(x)$ over all x that satisfy the constraint $g(x) = c$. The gradients ∇f and ∇g have to point in the same direction at the optimal x. Otherwise, it is possible to change x in say that g is constant but f increases (draw a picture to see this). The geometrical condition about the gradients is expressed as requiring that there is a constant, traditionally called λ (for Lagrange), so that

$$\nabla f \;=\; \lambda \nabla g \; . \tag{7}$$

An interesting consequence of this is that we get the same kind of condition if we seek to maximize g with a constraint on the value of f: $\nabla g = \mu \nabla f$. The gradients still have to align, regardless of which is the *objective function* and which is the constraint.

If there are m constraints $g_k(x) = c_k$, then there is a separate Lagrange multiplier for each constraint:

$$\nabla f \;=\; \sum_{k=1}^{m} \lambda_k \nabla g_k \; . \tag{8}$$

In general, a collection of unknown variables is determined by the same number of equations. In the present case, the $n+m$ unknown variables are the n original x_i and the m Lagrange multipliers λ_k. The equations are the m equations $g_k(x) = c_k$ and the system of n equations (8).

As a minor technical aside, we remark that calculus is not needed for the problem of maximizing a quadratic function subject to linear constraints. All the results may be found by algebraic methods – completing the square. To illustrate this, consider the problem of finding the unconstrained maximum of $f(x) = b^t x - \frac{1}{2} x^t A x$, where A is a symmetric $n \times n$ matrix. We could calculate $\nabla f = b - Ax$ and find $x = A^{-1} b$ by setting $\nabla f = 0$. On the other hand, an n dimensional version of completing the square would be the identity (reader: check this):

$$f(x) \;=\; -\frac{1}{2} \left(x - A^{-1} b \right)^t A \left(x - A^{-1} b \right) \;-\; \frac{1}{2} \, b^t A^{-1} b \; .$$

Since A is positive definite, the smallest value of the right hand side is achieved when the first term is zero, i.e. $x - A^{-1} b = 0$. In the same way, neither Lagrange multipliers nor calculus are really needed to derive the results below. We use the calculus approach because it is clearer and simpler.

2.3 The two fund theorem

An efficient portfolio minimizes the variance (6) with the expected return (2) and total wealth (3) fixed. This forms a constrained optimization problem, with $f(w) = \sum_{ij} w_i w_j \sigma_{ij}$, $g_1(w) = \sum_i w_i \mu_i$, and $g_2(w) = \sum_i w_i$. Because the numbers σ_{ij} form a symmetric matrix ($\text{cov}\,[R_i, R_j] = \text{cov}\,[R_j, R_i]$), we have (check this in the 2×2 case, $f(w_1, w_2) = w_1^2\sigma_{11} + w_1 w_2 \sigma_{12} + w_2 w_1 \sigma_{21} + w_w^2 \sigma_{22}$, if it is not clear)

$$\partial_{w_i} f(w) \;=\; 2\sum_{j=1}^{n} \sigma_{ij} w_j$$

The Lagrange multiplier condition (8) with $m = 2$ constraints for this case becomes

$$\sum_{j=1}^{n} \sigma_{ij} w_j \;=\; \lambda_1 \mu_i + \lambda_2 \; . \tag{9}$$

This may be formulated as a matrix equation. Let $\mu = (\mu_1, \ldots, \mu_n)^t$ be a vector of expected returns and $\mathbf{1} = (1, \ldots, 1)^t$ be the vector of all ones. Then (9) may be written $\Sigma w = \lambda_1 \mu + \lambda_2 \mathbf{1}$, or (we will see why the matrix Σ should be invertible)

$$w \;=\; \lambda_1 \Sigma^{-1} \mu \;+\; \lambda_2 \Sigma^{-1} \mathbf{1} \; . \tag{10}$$

A *fund*, f, (short for *mutual fund*) is a linear combination of assets 1, …, n with asset i getting *weight* f_i. Investing amount λ in fund f means that the amount you buy of asset i is λf_i. The formula (10) implies the *two fund theorem*: any efficient portfolio is a linear combination of two mutual funds, $f_1 = \Sigma^{-1}\mu$, and $f_2 = \Sigma^{-1}\mathbf{1}$. Every investor uses the same two funds, but allocate between them differently depending on their risk preferences. Different investors have different Lagrange multipliers, λ_1 and λ_2, but every investor has the same f_1 and f_2.

2.4 A risk free asset and the mutual fund theorem

This analysis simplifies if we assume that there is one asset that has zero risk. This is the same as saying that the asset pays a known return, which we call r. The right side of (10) may not make sense, given that Σ is singular – the variance of the risk free asset being zero. The above analysis assumed that Σ^{-1} exists. We have to do it over in a slightly different way if there is a risk free asset.

Let w_0 be the amount invested in the risk free asset, and w_k the amount invested in risky asset i, for $i = 1, \ldots, n$. The total wealth constraint (3) now is

$$w_0 \;=\; 1 - \sum_{i=1}^{n} w_i \; . \tag{11}$$

The total expected return is (use (11) to eliminate w_0)

$$\overline{\mu}_R \;=\; E\left[R\right] \;=\; w_0 r \;+\; \sum_{i=1}^{n} w_i \mu_i \;=\; r \;+\; \sum_{i=1}^{n} w_i \left(\mu_i - r \right) \;. \tag{12}$$

The quantity $\mu_i - r$ is the *excess return* of asset i, the amount by which the expected return on the risky asset exceeds the risk free rate. Excess returns need not be positive.

The variance of the portfolio is still given by (6) because the risk free part makes no contribution to the variance. The difference here is that $\sum_{i=1}^{n} w_i \neq 1$. Minimizing (6) with (12) as a constraints leads to

$$\sum_{j=1}^{n} \sigma_{ij} w_j \;=\; \lambda \left(\mu_i - r \right) \;. \tag{13}$$

In vector notation, vector of risky asset weights ($w = (w_1, \ldots, w_n)^t$ is the vector consisting only of risky asset weights) in the mutual fund is

$$w \;=\; \lambda \, \Sigma^{-1} \left(\mu \;-\; r\mathbf{1} \right) \;. \tag{14}$$

Note that this is a linear combination of the two mutual funds in the two fund theorem (10). The wealth constraint (11) in vector form is

$$w_0 \;=\; 1 - \mathbf{1}^t w \;=\; 1 \;-\; \lambda \mathbf{1}^t \Sigma^{-1} \left(\mu \;-\; r\mathbf{1} \right) \;. \tag{15}$$

The formula (15) explains λ as a parameter that represents the level of *risk aversion* of the investor. For $\lambda = 0$, the investor tolerates no risk, and puts all of his or her endowment in the risk free asset. As λ increases from zero (assuming the coefficient of λ is positive), the investor starts to put some assets in the risky mutual fund. When λ is very large, w_0 is negative, which indicates borrowing. An investor who is willing to tolerate high levels of risk will borrow at the risk free rate to invest in risky assets that have a higher expected return. Amplifying the return on an investment by borrowing is called *leverage* and is discussed a little more below.

2.5 The efficient frontier

The set of all efficient portfolios forms the *efficient frontier*. Any investor using mean variance analysis[5] would choose some portfolio on the efficient frontier. Such investors choose efficient portfolios that suit their risk preferences. The mutual funds f_1 and f_2 depend only on the market, and not the individual investor. The investor must choose λ_1 and λ_2 to satisfy the wealth constraint. After that, there remains a one parameter family of choices.

The reader is warned that not every solution of the Lagrange multiplier equations is a minimum, or even a *local minimum*. Even for a single function

[5]Such investors sometimes are called *rational*, but there are many rational reasons not to use mean/variance analysis, see below.

of one variable, $f(x)$, not every solution of $f'(x) = 0$ is a local minimum. It is possible to be a local maximum, for example. This would correspond to the largest possible variance compatible with a given wealth constraint and expected returns (reader: Can this happen?).

The efficient frontier, the set of efficient portfolios, form as straight line in risk/return space. We illustrate this in the case where there is a risk free asset. The expected return (12), in vector notation, is

$$\overline{\mu}_R \;=\; E\left[R\right] \;=\; r + (\mu - r\mathbf{1})^t\, w \;=\; r \;+\; \lambda\, (\mu - r\mathbf{1})^t\, \Sigma^{-1}\, (\mu - r\mathbf{1})\ . \qquad (16)$$

As discussed above, $\lambda = 0$ puts all the assets in the risk free fund and therefore returns r. Increasing λ increases the expected return, proportionate to λ. The expected return goes to infinity as $\lambda \to \infty$.

We find the variance by substituting (14) into (6) and using $\Sigma^{-1}\Sigma = I$:

$$\sigma_R^2 \;=\; \mathrm{var}[R] \;=\; \lambda^2\, (\mu - r\mathbf{1})^t\, \Sigma^{-1}\, (\mu - r\mathbf{1})\ . \qquad (17)$$

The standard deviation is the square root of the variance:

$$\sigma_R \;=\; \mathrm{sd}[R] \;=\; \lambda \sqrt{(\mu - r\mathbf{1})^t\, \Sigma^{-1}\, (\mu - r\mathbf{1})}\ .$$

This also is proportional to λ. The value $\lambda = 0$ has zero risk.

The *Sharpe ratio* is a quantity that measures the relation between risk and return in a mutual fund. It is the ratio of the excess return and the standard deviation. In the present case, it is

$$\mathrm{SR} \;=\; \frac{\overline{\mu}_R \,-\, r}{\sigma_R} \;=\; \sqrt{(\mu - r\mathbf{1})^t\, \Sigma^{-1}\, (\mu - r\mathbf{1})}\ . \qquad (18)$$

The Sharpe ratio of a portfolio of stocks measures the *skill* of the portfolio manager. You can increase the expected return without using skill, but simply by using *leverage* – borrowing at the risk free rate to invest in risky assets. The Lagrange multiplier, λ measures this leverage. The ratio (18) is designed explicitly to remove λ from the equation.

2.6 Portfolio β

Let R be the return on the optimal mean variance mutual fund. In the case where there is a risk free asset, some linear algebra (below) gives the relationship

$$\mu_i - r \;=\; \beta_i\, (\overline{\mu}_R - r)\ , \qquad (19)$$

where

$$\beta_i \;=\; \frac{\mathrm{cov}[R_i, R]}{\sigma_R^2}\ . \qquad (20)$$

Before giving the derivation, we comment on the formulas themselves. The mutual fund has assets with a variety of returns, some high and some low. The formula (19) says that the excess return of an asset is proportional to the overall excess return of the mutual fund, the proportionality constant being the covariance of that asset with the overall mutual fund. The linear algebra below shows that the weights in the mutual fund are adjusted to make this true.

The ratio (20) is the *beta* of asset i to the overall mutual fund. Some assets in the optimal fund may have negative beta to the fund. It might seem that if $\mu_i - r < \overline{\mu}_R - r$, it would be better to leave asset i out of the portfolio. However, a negative beta means that asset i is anti-correlated to the mutual fund as a whole. This means that asset i tends on average to decrease when the mutual fund increases. Including asset i therefore reduces the variance – the risk – of the mutual fund more than it reduces the expected return.

The proof of (19) with (20) starts with a simple formula for the covariance in question (see (13)):

$$\mathrm{cov}[R_i, R] \;=\; \sum_{j=1}^{n} \mathrm{cov}[R_i, w_j R_j] \;=\; \sum_{j=1}^{n} \sigma_{ij} w_j \;=\; \lambda\,(\mu_i - r) \;,$$

so

$$\mu_i \,-\, r \;=\; \frac{1}{\lambda}\,\mathrm{cov}[R_i, R] \;.$$

Here, the value of λ is the one that gives the pure mutual fund, that is,

$$\sum_{i=1}^{n} w_i \;=\; \mathbf{1}^t w \;=\; 1 \;.$$

Therefore, the desired (19) and (20) follow from understanding that this value of λ is

$$\lambda_* \;=\; \frac{\sigma_R^2}{\overline{\mu}_R - r} \;. \tag{21}$$

Here is the algebra behind (21). When $\lambda = \lambda_*$, so that $\sum_i w_i = 1$, we have

$$\begin{aligned}\overline{\mu}_R \,-\, r \;&=\; \sum_{i=1}^{n} w_i\,(\mu_i - r) \\ &=\; w^t\,(\mu - r\mathbf{1}) \;.\end{aligned}$$

Using (13) to express $\mu - r\mathbf{1}$ in terms of w gives

$$\overline{\mu}_R \,-\, r \;=\; \frac{1}{\lambda_*}\,w^t \Sigma w \;=\; \frac{1}{\lambda_*}\,\sigma_R^2 \;,$$

which is the formula (21).

2.7 The CAPM interpretation of β

The *Capital Asset Pricing Model* (*CAPM*) is a fanciful speculation of a possible alternate universe in which everyone in the world invests using mean variance analysis using the same μ and Σ. In that universe, everyone uses the same stock weights, though by different total amounts depending on their risk preferences. Therefore, in that world, the total *capitalization* of asset i is proportional to w_i, the weighting in the optimal mutual fund of mean variance analysis. The capitalization of a company is the total value of its outstanding stock. Therefore, the mutual fund is the entire market. If you want to buy it, you just buy each asset in proportion to its total capitalization. Therefore, in the CAPM universe, the mutual fund is called the *market*, and the beta of an asset is the same as its beta *to the market*.

In this case, the beta of a particular stock depends on the covariance of that stock price with the *market portfolio*. Then (19) and (20) become predictions that relate the expected return of a stock to the covariance of that stock to the market. One could test the CAPM model by estimating μ_i and σ_{iM} for each i. Here, σ_{iM} is the covariance of the price of asset i with the "market". According to CAPM, the market portfolio is the universe of all available assets, each weighted its total capitalization. The complete failure of the markets to satisfy CAPM constraints is explored at great length in many ways in the book *Financial Econometrics* by Campbell, Lo, and McKinley.

Why mention this apparently silly model of the world? Because it influences the way people think about markets. For example, there are many investment funds (hedge funds mostly) that claim to be "beta neutral". This means that their returns are uncorrelated with some overall market index. In the Gaussian world, this would mean that the fund returns are independent of the market index. If there were many such funds, and if they were independent of each other, this would be a wonderous investment opportunity. Just put a little money in each one and (through diversification) earn a high return at very low risk. It is highly unlikely that there are many brilliant investors with secret and independent high yield beta neutral investment strategies.

2.8 Criticisms of mean/variance analysis

All models are wrong. Some models are dangerous.

There are many drawbacks to mean variance analysis. Much of this class is devoted to overcoming these drawbacks. They fall into five rough categories:

1. Investors care about more than just mean and variance.
2. The μ and Σ are hard to estimate.
3. They are not a complete description of market returns.
4. Returns are not linear functions of the investment weights.
5. Investment strategies are not simple.

1. Investors care about more than just mean and variance. There are many measures of risk other than variance. For example, the probabilities of low or negative returns, $F(r) = \Pr(R \leq r)$ may important (see (1)). Values of

r for which $F(r)$ is small are *tail probabilities* that prudent investors will want to understand.

This is particularly important when returns are not Gaussian. Empirical studies show returns rarely are Gaussian. If[6] $R \sim \mathcal{N}(\mu, \sigma^2)$, then a *three sigma shortfall* is very unlikely:

$$\Pr\left(R \;<\; \mu \,-\, 3\sigma \right) \;\approx\; .5\% \;\; .$$

However, there are non-Gaussian distributions with $\mu = E[R]$ and $\sigma^2 = \mathrm{var}[R]$ so that the same event is twenty times more likely

$$\Pr\left(R \;<\; \mu \,-\, 3\sigma \right) \;\approx\; 10\% \;\; .$$

If a three sigma shortfall has serious consequences, it is prudent to know more about the distribution of R than its mean and variance.

Many modern investment opportunities and strategies amplify the non-Gaussian nature of returns. The simplest example is a stock option that pays $V = S - K$ (S is the stock price and K is the *strike price* of the option) if $S > K$, and $V = 0$ if $S < K$. Even if S may be Gaussian or nearly Gaussian, V is far from Gaussian. A *CDS* (*credit default swap*) is a more extreme example of non-Gaussian returns. The issuer of a CDS receives a steady stream of small fixed payments as long as a certain company makes its contracted interest payments (*coupons*) on its outstanding bonds. If the company defaults (fails to make a coupon payment), the issuer of the CDS must pay the entire principle of the bond. To be sure, CDS instruments may be a thing of the past (in 2009). But there are other ways for an investor to *take a position* that has a high probability of a small positive return and a small probability of a large negative return.

What should you do if you want a systematic quantitative investment procedure? What should you optimize? One possibility is *robust optimization*, that optimizes in a mean/varance sense as above, but with protections such as constraints on the probabilities of large losses. Robust optimization procedures take into account the uncertainty in our statistical estimates of μ and Σ. These are ad-hoc but very practical. More systematic approaches involve optimizing *expected utility*. This has important philosophical backing, but suffers from the drawback that nobody knows what their "utility" is. Both robust optimization and utility approaches are discussed later in this class.

2. The μ and Σ are hard to estimate. An investor using a mean variance needs estimates of μ and Σ. Some will make their own estimates from historical market data. Others will buy estimates from vendors, such as Barra or Bloomberg. Either way, the investor will be using estimated values[7] $\widehat{\mu}$ and $\widehat{\Sigma}$ instead of the actual μ and Σ. Simply using the estimates in mean variance analysis would lead to

$$\widehat{w} \;=\; \lambda_1 \widehat{\Sigma}^{-1}\widehat{\mu} \;+\; \lambda_2 \widehat{\Sigma}^{-1}\mathbf{1} \;. \tag{22}$$

[6] This notation means that R is a Gaussian (normal) randome variable with mean μ and variance σ^2.

[7] Statisticians use $\widehat{\theta}$ to denote a statistical estimate of a parameter θ.

A matrix, Σ, is called *ill conditioned* if relatively small changes in Σ lead to much larger relative changes in Σ^{-1}. A typical large covariance matrix is likely to be ill conditioned. This means that even if $\widehat{\Sigma}$ is close to Σ, still $\widehat{\Sigma}^{-1}$ may be far from Σ^{-1}. Portfolio weights (22) may give a portfolio that has much more risk that it appears to (the left side is actual variance, the right side is estimated variance):

$$\widehat{w}^t \Sigma \widehat{w} \gg \widehat{w}^t \widehat{\Sigma} \widehat{w} .$$

The weights (22) also may give variance much larger than is the best possible:

$$\widehat{w}^t \Sigma \widehat{w} \gg w^t \Sigma w .$$

Decision theory is a branch of statistics that takes into account statistical errors in decision making. Different kinds of errors may have vastly different consequences. For example, a statistical over-estimate of the risk may lead to investments that are slightly too cautious, while a statistical under-estimate would expose an investor to possibly large losses.

A closely related framework is *Bayesian*, in which you assume that Σ actually is a random variable with its own *prior* distribution. The statistical estimate $\widehat{\Sigma}$, is just another random variable with a different distribution, the *posterior* (prior for "before", and posterior for "after", that is, after taking into account the actual data). In this case R is a very complicated function not only of the market returns R_i but also of the market data used in (22). We still could try to minimize variance for a given return, but

$$\text{var}[R] \neq \widehat{w}^t \widehat{\Sigma} \widehat{w} , \tag{23}$$

Unlike classical statistics based on least squares and linear algebra, modern statistics relies on Monte Carlo simulation to estimate things like var$[R]$. This class will involve considerable simulation, as an alternative to simple but naive and incorrect closed form formulas.

3. Σ and μ are not a complete description of market returns. Only a Gaussian is completely described by its mean and variance. A one dimensional non-Gaussian random variable may be represented by its probability density or distribution function, which may be estimated from data. The harder problem is modeling relationships between variables R_i when the R_i are not Gaussian. For Gaussians, the only relationship is covariance. Giving the variances and covariances specifies the joint distribution completely.

Relationships between non-Gaussian random variables may be more subtle. For example, suppose that R_i is the one day return on stock i. It is well known that the apparent correlation between the R_i is much larger on on days in which the overall market has a large loss. A popular approach these days is the *copula* method, which is a way of separating the model of the distributions of individual random variables from the model of the relationships between random variables. These are discussed in this class.

Regime changing models are a different way to model the relationship between random variables. In its simplest form, there may be one binary random variable that is "good economy" with probability p and "bad economy" with probability $1 - p$. In the good economy state, n individual bonds default independently with 1% probability each. In a bad economy, defaults are still independent, but the probability increases to 20%. If $p = 95\%$, the overall default probability is approximately 2% (reader, calculate it exactly).

Suppose a portfolio contains $n = 100$ bonds that may default. It is virtually impossible to have more than ten defaults in the good economy state, so the probability of more then ten defaults is approximately the probability of being in the bad economy state, which is 5%. If the 100 defaults were independent with 2% probability each (the independent default model with the same default probability), the probability of more then ten defaults is a thousand times smaller than this. Going from the regime changing model to independent defaults changes the probability of more than ten defaults from a small but important 5% to a totally negligible .004%.

The choice of models has a large impact on our ability to estimate the parameters. Suppose, for example, we have 20 observations of the defaults of 100 bonds. In the independent default model, that is 2000 observations of an event that has probability about 2%. This will mean (see later in the class) that we can estimate the default probability somewhat accurately but not precisely (the expected number of defaults is ten). However, in the regime changing model, there is a good chance that in 20 observations we never see a regime change (this has 50% probability). Our 20 observations are not nearly enough to estimate the regime changing model. If we believe in regime change, we must find a fundamentally different way to fit it to data, or find far more data.

Regime changing models raise a larger issue: what about the market actually is *stationary* and what is changing in time? The regime changing model has the default probability changing in time. Other market parameters also could change, μ and Σ in particular. In fact, our NYU colleague Robert Engle received a Nobel Memorial Prize in Economics partly for modeling the way Σ might change in time (*ARCH, Auto-Regressive Conditional Heteroskedasticity* – skedasticity being another word for variance). Chapter 3 of Attilio Meucci's book is about what parameters might be stationary, and how to tell from data.

4. Returns are not linear functions of the investment weights. In real life, investment returns are not linear. If you want a startup company to succeed, you need to give enough resources. Otherwise it will fail. Half the necessary resources will not lead to half success. More subtle nonlinearities come from *market frictions* such as transaction costs. Suppose there are 100 shares of a stock for sale at \$12.34/share, 200 more for sale at \$12.36/share, and another 100 at \$12.37/share. The price for 200 shares is more than double the price for 100. Yes, only a little more, but small costs like this have a major impact on high frequency trading strategies. If you are a large investor and word gets out that you want a million shares, the price can jump a lot. Courant Math Finance Fellow Robert Almgren and co-founder of the program Neil Chriss have written important papers on this issue. Conversely, a smaller investor may be able to profit from discovering that a large investor is trying to acquire a large stock position.

5. Investment strategies are not simple. Most institutional investment strategies are far more sophisticated than simply choosing n weights for the available assets. Finding the optimal strategy of a certain form may be impossible because there are too many possibilities to do a systematic search.