Lecture 3

Jussi Klemelä

September 29, 2009

1 Formalization of Portfolio Selection

We describe portfolio selection using mathematical terminology. We denote the current time point t_0 and we will make rebalancing of the portfolio at the predetermined time points $t_0, t_0+1, \ldots$. We denote with

$$S_t = (S_t^1, \ldots, S_t^d) \in (0,\infty)^d, \qquad t = t_0, t_0+1, \ldots,$$

the vectors of prices of assets that are available in portfolio selection. The assets can be stocks, commodities, currencies, or bonds. The collection of available assets can also include the bank account.

1.1 Wealth Process

Let

$$U_t = \left(\frac{S_t^1}{S_{t-1}^1}, \ldots, \frac{S_t^d}{S_{t-1}^d} \right), \qquad t = 1, 2, \ldots,$$

be the market vectors of price relatives. The jth component of the market vector at time t expresses the ratio of a closing price of asset j to the closing price of the previous period. This ratio is the factor by which capital invested in the jth asset increases during the trading period t. The portfolio vectors are denoted by

$$b_{t_0}, b_{t_0+1}, \ldots \in \mathbf{R}^d.$$

The elements of a portfolio vector express the proportion of wealth invested in an asset, and not the absolute amount of wealth. We assume that

$$\sum_{j=1}^d b_t^j = 1.$$

In the case of long only portfolios we will assume that $b_t^j \geq 0$ for $j = 1, \ldots, d$, but in the case of shorting negative values for the elements of the portfolio vectors are allowed.

We assume to have initial wealth $W_{t_0} > 0$ which we want to distribute among the available basic assets. The wealth process $W_{t_0}, W_{t_0+1}, \ldots$ of a portfolio can be calculated using the recursive formula

$$W_{t+1} = W_t \cdot b_t^T U_{t+1}, \qquad t = t_0, t_0 + 1, \ldots. \tag{1}$$

This equation follows because the strategy is self-financing and there is no consumption; wealth W_{t+1} is obtained from wealth W_t only through the changes in asset prices and through the changes in wealth allocation. After $(t - t_0)$-trading periods the strategy achieves the wealth

$$W_t = W_{t_0} \prod_{i=t_0}^{t-1} b_i^T U_{i+1}.$$

The initial wealth W_{t_0} is a fixed number, but $W_{t_0+1}, W_{t_0+2}, \ldots$ are real valued random variables. We can write also

$$W_t = W_t(b_{t_0}, \ldots, b_{t-1}),$$

which shows the dependence of the wealth on the portfolio vectors.

1.2 Portfolio Types

The problem of statistical portfolio selection is to choose the portfolio vectors by statistical methods in an optimal way using available historical data. Depending on the portfolio type there are different constraints on the portfolio vectors

$$b_{t_0}, b_{t_0+1}, \ldots \in \mathbf{R}^d.$$

1.2.1 Long Only Portfolio

In the case of long only portfolios we assume that for all t,

1. $\sum_{j=1}^d b_t^j = 1$,
2. $b_t^j \geq 0$ for $j = 1, \ldots, d$.

The jth component of a portfolio vector denotes the proportion of the investor's capital invested in asset j. Since the portfolio weights sum to one, it is assumed that we are always fully invested. However, one of the assets can be a bank account.

1.2.2 Long Only Portfolio with Leveraging

In the case of leveraging we have to distuinguish the bank account from the other assets. The usual long only portfolio may be such that one of the assets is the bank account but that is not necessary. In the case of leveraging we have to include the bank account because we may need to borrow money (shorting the bank account is interpreted as borrowing). Let S_1 be the bank account. Let $0 < L < \infty$ be the allowed maximum leveraging ratio. The portfolio vectors satisfy

1. $\sum_{j=1}^d b_t^j = 1$,
2. $b_t^j \ge 0$ for $j = 2, \ldots, d$,
3. $\sum_{j=2}^d b_t^j \le L$.

That is, we allow negative values for the portfolio weight of the bank account, but the other portfolio weights are nonnegative. The value of the portfolio weight of the bank account is determined by

$$b_t^1 = 1 - \sum_{j=2}^d b_t^j,$$

since the sum of all weights have to be one. The maximum allowed leveraging ratio gives a limit to the amount which can be borrowed. Indeed,

$$1 - L \le b_t^1 \le 1,$$

and we need that $L > 1$ to make the borrowing possible.

1.2.3 Portfolio Allowing Shorting

Selling a stock short means that we sell a stock that we do not own. Typically the stock which is sold short is first borrowed from somebody who owns the stock. If the stock is sold without first borrowing it, the short selling is called *naked short selling*. Selling the bank account short is interpreted as borrowing. Short selling changes the character of the portfolio: a short position on a stock has an unlimited downside risk, but only a limited upside potential. In contrast, a long position on a stock can loose only the invested capital but has an unlimited upside potential.

In the case when shorting is allowed the elements of portfolio vectors can take negative values. We make a constraint to the amount of sorting by requiring that the coefficients lie in a L_1 ball. When there are no risk free rate we assume that the portfolio vectors satisfy conditions

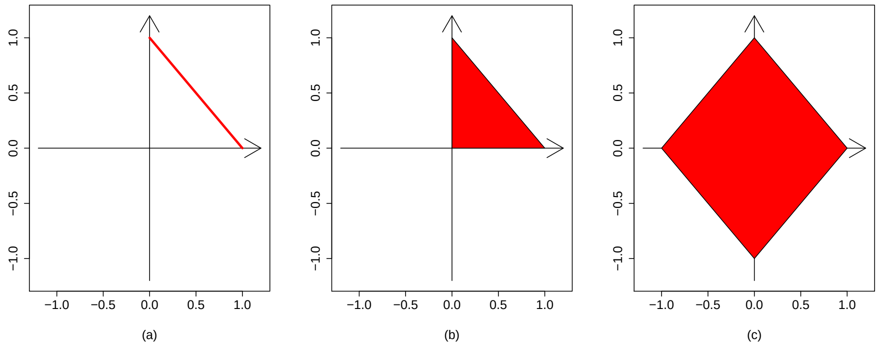

Figure 1: *Regions of Portfolio Weights* (a) The region of long only weights without bank account, (b) the region of long only weights with bank account included, and (c) the region of portfolio weights when shorting is allowed.

1. $\sum_{j=1}^{d} b_t^j = 1$,
2. $\sum_{j=1}^{d} \left| b_t^j \right| \le L$,

where $L > 0$ is a shorting constraint.

When S^1 is the asset with a risk free rate of return, we assume that the portfolio vectors satisfy conditions

1. $\sum_{j=1}^{d} b_t^j = 1$,
2. $\sum_{j=2}^{d} \left| b_t^j \right| \le L$,

where $L > 0$ is a shorting and leveraging constraint. When $b_t^1 < 0$, then the portfolio is leveraged.

Figure 1 shows the regions where portfolio weights are allowed to be. In panel (a) we show the case of long only portfolio with two basic assets. In panel (b) we show the case of three basic assets, one of them being the bank account. In panel (c) we show the case of two basic assets when shorting is allowed.

1.3 Investment Strategy

The sequence $b_{t_0}, b_{t_0+1}, \ldots \in \mathbf{R}^d$ of portfolio vectors will typically be chosen using available relevant information. We mention two hypotheses about what constitutes relevant information in portfolio selection.

1. The relevant information used in choosing the portfolio vector b_t includes typically the vector time series of previous price relatives $Z_t = (U_1, \ldots, U_t)$. Since $U_t \in \mathbf{R}^d$, $Z_t \in \mathbf{R}^{dt}$.

   Sometimes it is thought that the historical stock prices contain all relevant information and we do not need any external fundamental data. In this case we use only the information in the past asset prices to choose the portfolio.

2. The relevant information can include also some external information about the state of the economy or about the state of the companies whose stocks are in the portfolio. For example, Z_t can contain macroeconomic information like default spreads and term spreads and Z_t can contain information about the companies like dividend yield and earnings. Also, Z_t can contain information about the past development of assets which do not belong to the available investment universum, like various stock and bond indices.

   In this more general setting we assume that $Z_t \in \mathbf{R}^{k(t)}$, where $k(t) \in \{1, 2, \ldots\}$ is some function of time index t.

We call Z_t the state vector. When $Z_t \in \mathbf{R}^{k(t)}$, an investment strategy $B = (b_{t_0}, b_{t_0+1}, \ldots)$ is a sequence of functions

$$b_t : \mathbf{R}^{k(t)} \to \mathbf{R}^d, \qquad t = t_0, t_0 + 1, \ldots.$$

We shall write $b_t = b_t(Z_t)$. When the portfolio choice is time invariant we can write $b = b(Z_t) = b_t(Z_t)$.

2 Comparison of Portfolio Selection Strategies

2.1 Comparison of Distributions

2.1.1 Connection to Distribution Selection

We shall later discuss several different strategies of choosing the portfolio and we need a method for ranking these strategies. The comparison of portfolio selection strategies reduces to the comparison of probability distributions in the following way. We assume that at the current time t_0 we have available wealth $W_{t_0} \in \mathbf{R}$ and a collection of trading strategies.[1] We want to use the

[1]In practise we need that $W_{t_0} > 0$, but some strategies require only a small initial wealth. For example, we can borrow funds for investing, or we can write options. These

initial wealth W_{t_0} to our advantage to obtain an increased wealth at a later time T. For two different trading strategies we obtain final wealths $W_T^{(1)}$ and $W_T^{(2)}$ at time T. Wealths $W_T^{(i)}$ are random variables, whose distributions are unknown. We have now random variables

$$U_T^{(i)} = \frac{W_T^{(i)}}{W_{t_0}}, \qquad i = 1, 2,$$

and we choose the trading strategy i which is such that the distribution of $U_T^{(i)}$ is better in some sense. We have to use statistical methods to get a hold on the distributions of $U_T^{(i)}$.

2.1.2 Statistical Distribution Selection

Since the distributions of $U_T^{(i)}$ are unknown, we shall use the following approach to choose the investment strategy. We consider portfolio selection where the rebalancing is made at the equally spaced time points $1, \ldots, t_0, t_0 + 1, \ldots$. We shall choose the portfolio strategy which always optimizes the wealth at the next time of rebalancing. That is, we consider the case $T = t_0 + 1$. We have available the prices $S_0, \ldots, S_{t_0}$, where t_0 is the current time. The strategies obtain the portfolio vectors $b_0^{(i)}, \ldots, b_{t_0-1}^{(i)}$ and the wealth processes $W_0^{(i)}, \ldots, W_{t_0}^{(i)}$, where the initial wealth can be taken to be equal to one. Then we can calculate the wealth relatives

$$\frac{W_t^{(i)}}{W_{t-1}^{(i)}}, \qquad t = 1, \ldots, t_0.$$

These wealth relatives can be used to estimate the distribution of

$$U^{(i)} = \frac{W_{t_0+1}^{(i)}}{W_{t_0}}, \qquad i = 1, 2.$$

2.1.3 Examples of Distribution Selection

The selection of the best trading strategy requires that we can make an ordering among the univariate random variables U (or an ordering among their distributions). Figure 2 illustrates the comparison of distributions. Panel (a) shows two densities of distributions which are easy to compare; the densities have the same shape but the other dominates the other, because its

strategies require some initial wealth only to give collateral to the bank or to the option exchange. The amount of initial wealth has an influence on the collection of available investment strategies.

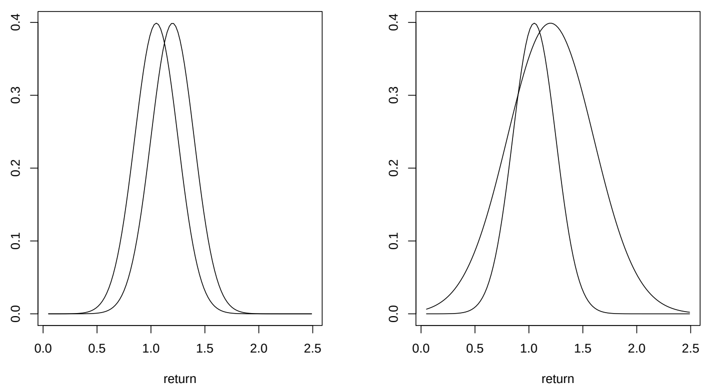

Figure 2: *Comparison of distributions* (a) Two return densities which are easy to compare and (b) two return densities which are difficult to compare.

mode is at 1.2 whereas the mode of the other density is at 1.05. Panel (b) shows two densities which cannot be compared straightforwardly; the mode of the other is at 1.2 but its variance is larger, whereas the mode of the other is at 1.05 but its variance is smaller.

Figure 2 shows an example of *stochastic dominance*. We say that the distribution of $U^{(1)}$ stochastically dominates the distribution of $U^{(2)}$ if $F_2(u) \geq F_1(u)$ for all $u \in \mathbf{R}$, where F_1 is the distribution function of $U^{(1)}$ and F_2 is the distribution function of $U^{(2)}$. [2] If the distribution of $U^{(1)}$ stochastically dominates the distribution of $U^{(2)}$ then $P(U^{(1)} \geq u) \geq P(U^{(2)} \geq u)$ for all $u \in \mathbf{R}$.

3 Characterization of Univariate Distributions

3.1 Basic Statistics

Center of Distribution The basic statistics for characterizing the center of a distribution are the mean and the median. In addition, the mode is sometimes used to characterize the center of a distribution.

[2]Stochastic dominance occurs if and only if the dominant distribution has a higher expected utility for all increasing and continuously differentiable utility functions. Stochastic dominance is also called first order stochastic dominance to distinguish it from the second order stochastic dominance.

1. *Mean* The mean of a distribution of random variable $Y \in \mathbf{R}$ is defined as
$$EY = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy,$$
where $f_Y : \mathbf{R} \to \mathbf{R}$ is the density function of Y. Given a sample $Y_1, \ldots, Y_n$ from the distribution of Y the mean EY can be estimated with the sample mean (the arithmetic mean):
$$\bar{Y}_T = \frac{1}{T} \sum_{i=1}^{T} Y_i.$$

2. *Median* The median of a distribution is the point which divides the probability mass to two equal parts:
$$\mathrm{med}(Y) = \inf \{y : F_Y(y) \geq 1/2\},$$
where $F_Y : \mathbf{R} \to \mathbf{R}$ is the distribution function of Y. When F_y is continuous, then
$$\mathrm{med}(Y) = F_Y^{-1}(1/2)$$
and the median satisfies
$$F_Y(\mathrm{med}(Y)) = 1/2.$$
The sample median is the $[T/2]$-order statistics:
$$\widehat{\mathrm{med}}(Y) = Y_{([T/2])},$$
where we denote with $Y_{(1)}, \ldots, Y_{(T)}$ the order statistics (the sample values $Y_1, \ldots, Y_T$ in increasing order) and $[x]$ is the largest integer $\leq x$.

3. *Mode* The mode is the point which maximizes the density function of Y:
$$\mathrm{mode}(Y) = \mathrm{argmax}_{y \in \mathbf{R}} f_Y(y).$$
Typically the mode is used only for unimodal distributions (a unimodal distribution is such that it has only one local maximum). A mode can be estimated by finding a maximizer of a density estimate:
$$\widehat{\mathrm{mode}}(Y) = \mathrm{argmax}_{y \in \mathbf{R}} \hat{f}_Y(y),$$
where $\hat{f}_Y : \mathbf{R} \to \mathbf{R}$ is an estimate of the density function f_Y.

Spread of Distribution The spread of a distribution can be measured by variance, other moments, lower and upper partial moments, quantiles, shortfall, and absolute shortfall.

1. *Variance* The basic characterizations of the spread of a distribution are the variance and the standard deviation. The variance of a random variable $Y \in \mathbf{R}$ is defined as the second centered moment:

$$\mathrm{Var}(Y) = E(Y - EY)^2 = \int_{-\infty}^{\infty} (y - EY)^2 \, f_Y(y) \, dy,$$

where EY is the mean of the distribution of Y and $f_Y : \mathbf{R} \to \mathbf{R}$ is the density function of Y. The standard deviation is the square root of the variance:

$$\mathrm{std}(Y) = \sqrt{\mathrm{Var}(Y)}\,.$$

Given a sample $Y_1, \ldots, Y_T$ from the distribution of Y the sample variance is

$$\widehat{\mathrm{Var}}(Y) = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \bar{Y}_T)^2,$$

where $\bar{Y}_T$ is the sample mean. The sample standard deviation is the square root of the sample variance:

$$\widehat{\mathrm{std}}(Y) = \sqrt{\widehat{\mathrm{Var}}(Y)}\,.$$

2. *Centered Moments* The variance can be extended to other centered moments

$$E|Y - EY|^k,$$

where $k = 1, 2, \ldots$. The sample centered moments are

$$\frac{1}{T}\sum_{t=1}^{T} |Y_t - \bar{Y}_T|^k,$$

where $\bar{Y}_T$ is the sample mean.

3. *Partial Moments* The centered moments take contribution from the left and the right tails of the distribution. The lower partial moments take contribution only from the left tail of the distribution:

$$\mathrm{LPM}_{\tau,k}(Y) = E(\tau - Y)_+^k = \int_{-\infty}^{\tau} (\tau - y)^k f_Y(y) \, dy, \qquad (2)$$

where $(x)_+ = \max\{x, 0\}$, $k = 0, 1, 2, \ldots$, and $\tau \in \mathbf{R}$ is a target rate. For example, when $k = 0$, then

$$\mathrm{LPM}_{\tau,0}(Y) = P(Y \le \tau)$$

is the probability that Y is smaller than τ. The upper partial moment is defined as

$$\mathrm{UPM}_{\tau,k}(Y) = E(Y - \tau)_+^k = \int_\tau^\infty (y - \tau)^k f_Y(y)\, dy. \tag{3}$$

The sample lower partial moment is

$$\widehat{\mathrm{LPM}}_{\tau,k}(Y) = \frac{1}{T} \sum_{i=1}^{T} (\tau - Y_i)_+^k. \tag{4}$$

For $k = 0$ we estimate the probability content of the tail:

$$\widehat{\mathrm{LPM}}_{\tau,0}(Y) = \frac{N(\tau)}{T},$$

where

$$N(\tau) = \#\left\{Y_i : i = 1, \ldots, T,\ \ Y_i \le \tau\right\}. \tag{5}$$

4. *Conditional Moments* The lower conditional moments are the moments conditioned on the left tail of the distribution:

$$\mathrm{LCM}_{\tau,k}(Y) = E\left[(\tau - Y)^k \mid \tau - Y \ge 0\right],$$

where $k = 0, 1, 2, \ldots$ and $\tau \in \mathbf{R}$ is a target rate. The sample lower conditional moment is

$$\widehat{\mathrm{LCM}}_{\tau,k}(Y) = \frac{1}{N(\tau)} \sum_{i=1}^{T} (\tau - Y_i)_+^k, \tag{6}$$

where $N(\tau)$ is defined in (5). Note that in (4) the sample size is the denominator but in (6) we have divided with the number of observations in the left tail.

5. *Quantiles* The median can be extended to other quantiles:

$$\mathrm{Q}_p(Y) = \inf\{y : P(Y \le y) \ge p\}, \tag{7}$$

where $0 < p < 1$. When $p = 0.5$, then $\mathrm{Q}_p(Y)$ is the median of Y. In the case of a continuous distribution function

$$P\left(Y \le \mathrm{Q}_p(Y)\right) = p.$$

The sample quantile is the $[pT]$-order statistics:

$$\widehat{Q}_p(Y) = Y_{([pT])}, \tag{8}$$

where we denote with $Y_{(1)}, \ldots, Y_{(T)}$ the order statistics (the sample values $Y_1, \ldots, Y_T$ in increasing order) and $[x]$ is the largest integer $\le x$.

6. *Shortfall and Absolute Shortfall* The absolute shortfall is defined as a risk measure where the expectation is taken only over the left tail, when the left tail is defined as the region which is to the left of a quantile of the distribution. Thus the absolute shortfall is defined as

$$\mathrm{ASF}_p = -E\left[Y I_{(-\infty, Q_p(Y)]}(Y)\right]$$

The shortfall is defined as

$$\mathrm{SF}_p = Q_p(Y) - E\left[Y I_{(-\infty, Q_p(Y)]}(Y)\right].$$

For $\tau = Q_p(Y)$ and $k = 1$ the lower partial moment is close to the shortfall:

$$\mathrm{LPM}_{Q_p(Y),1} = Q_p(Y) P\left(Y \le Q_p(Y)\right) - E[Y I_{(-\infty, Q_p(Y)]}(Y)],$$

where $P\left(Y \le Q_p(Y)\right) \approx p$. The sample absolute shortfall is

$$\widehat{\mathrm{ASF}}_p = -\frac{1}{T} \sum \left\{ Y_i : i = 1, \ldots, T,\ Y_i \le \widehat{Q}_p(Y) \right\}, \tag{9}$$

where $\widehat{Q}_p(Y)$ is the sample quantile defined in (8). The sample shortfall is

$$\widehat{\mathrm{SF}}_p = \widehat{Q}_p(Y) - \widehat{\mathrm{ASF}}_p. \tag{10}$$

7. *Expected Shortfall* The expected shortfall is the negative conditional expectation under the condition that the random variable is less than a quantile. Thus the expected shortfall is defined as

$$\mathrm{ESF}_p = -E\left[Y \mid Y \le Q_p(Y)\right].$$

We can write

$$\mathrm{ESF}_p = -E[Y \mid Y \le -\mathrm{VaR}_p(Y)]$$

and the term tail conditional VaR can be used to denote the expected shortfall. Note that a natural modification of the expected shortfall is the lower conditional moment of order $k = 1$ and target rate $\tau = Q_p(Y)$:

$$\mathrm{LCM}_{Q_p(Y),1} = Q_p(Y) - E[Y \mid Y \le Q_p(Y)].$$

An estimate of the expected shortfall is

$$\widehat{\mathrm{ESF}}_p = -\frac{1}{N(\widehat{Q}_p(Y))} \sum \left\{ Y_i : i = 1, \ldots, T, \;\; Y_i \le \widehat{Q}_p(Y) \right\}, \tag{11}$$

where $\widehat{Q}_p(Y)$ is the sample quantile defined in (8) and $N(\tau)$ is defined in (5).

4 Examination

Possible questions in the examination:

5) Define (the population) variance, lower partial moments, quantiles, absolute shortfall, and shortfall.

6) Let $Y_1, \ldots, Y_T$ be a sample. Use the sample to define estimates of variance, lower partial moments, quantiles, absolute shortfall, and shortfall.