1 The Correlation Structure of Security Returns

To use the mean variance approach we need to estimate $n$ components of the expected return vector $r$ and $n(n+1)/2$ components of the variance covariance matrix $V$. Analysts following a security can estimate the expected return and the variance of the security, but have trouble estimating the covariance with other securities. To estimate the covariance terms in $V$ requires a very long history and an assumption that returns form a stationary process. Often there is not enough data, and the returns are surely not stationary. What should we do then if we want to use mean variance analysis?

There are two possible directions we can take. First, instead of viewing the problem of finding portfolios of securities, we can restrict our attention to the problem of finding portfolios of asset classes. This greatly reduces the number of estimates to a manageable size. Alternatively we can use an index model to estimate the covariance of securities. In this lecture we explore using historical data to estimate the correlation structure of a small number of asset classes and the use of the single index model to estimate the correlation structure of securities.

We wish to insert here a caveat concerning the use of historical data. Although historical data is widely used to estimate model parameters and future outcomes, the use of historical data has limitations because the model in question may be wrong and because things may change in ways not anticipated by the model.

2 Asset Class Covariances

Analysts frequently utilize historic returns to estimate the covariances among future returns. If the data were drawn from a stationary distribution we would like to use as long a history as possible to maximize the accuracy of our estimates. If, however, the distribution is non stationary we may want to give more weight to recent data than to old data. Some analysts do this by limiting the historic data to 60 months assigning the same weight to each observation. Here we focus on a smoothing average procedure utilized in a number of asset allocation models that assume that the future is more likely to be like the recent past than the distant past.

In exponential smoothing each historic observation is assigned a weight, say $C\alpha^h$ where $\alpha \in (0, 1)$ and $h = 1, 2, \ldots, H$ is the age of the data. Here

$$C = \frac{1 - \alpha}{\alpha(1 - \alpha^H)}$$

is selected so that the sum of the weights is equal to one. As $\alpha \to 1$, the weights tend to $1/H$ giving equal weight to all the data. Conversely, as $\alpha \to 0$ all the weights becomes concentrated in the most recent period.

The choice of $\alpha$ is often selected in terms of the half-life of the data. If the weight of the observation is reduced by half each $k$ periods then $\alpha^k = 1/2$ resulting in

$$\alpha = 2^{-1/k}.$$ 

Let $p$ be an $H$-dimensional vector with components $p_h = C\alpha^h$ and notice that $p$ is a probability vector (it is nonnegative and its components add to one).

Let $R$ be an $n \times H$ matrix of realized returns where $R_{ih}$ is the return of asset class $i$ realized $h$ periods ago. Given $R$ we can estimate $r$ the vector of expected returns by

$$r = Rp.$$ 

Let $e$ the $H$-dimensional vector of ones. Then the vector of return deviations is given by

$$D = R - re'$$

and the covariance matrix by

$$V = D\text{diag}(p)D'.$$
where $\text{diag}(p)$ is a diagonal matrix with the vector $p$ in the diagonal. Notice that $V$ is an $n \times n$ matrix, and that
\[
V_{ij} = \sum_{k=1}^{H} (R_{kh} - r_i)p_h(R_{jh} - r_j).
\]

See the excel spreadsheet “Smoothing Covariance” for a numerical example where $r$ and $V$ are estimated from historical data.

3 Single Index Model

The single index model (SIM) can be used to estimate $r$ and $V$ for either asset classes or for securities. We will refer to securities throughout the analysis, but the results are readily applicable for the case of asset classes. The single index models postulates that the return of a security can be modeled as
\[
R_i = \alpha_i + \beta_i R_m + \epsilon_i
\]
where $\alpha_i$, and $\beta_i$ are constants, $R_m$ is the return of a market index and $\epsilon_i$ is a random variable with mean 0 and variance $\tau_i^2$. Let $\overline{R}_m$ and $\sigma_m^2$ denote the mean and variance of $R_m$. If the $\alpha_i$s, $\beta_i$s, and $\tau_i^2$s are estimated via regression analysis it turns out that $\text{Cov}(\epsilon_i, R_m) = 0$, so we can and do assume without the loss of generality that this holds. The key assumption of the single index model is
\[
\text{Cov}(\epsilon_i, \epsilon_j) = 0 \quad \text{for all} \quad i \neq j.
\]

It is easy to see that
\[
\overline{R}_i = \alpha_i + \beta_i \overline{R}_m,
\]
so that the vector of expected returns is given by
\[
r = \alpha + \overline{R}_m \beta.
\]

Under the single index model the returns of securities are correlated only through the return of the market index $R_m$. Indeed, for $i \neq j$, we have
\[
\text{Cov}(R_i, R_j) = \beta_i \beta_j \text{Cov}(R_m, R_m) + \beta_i \text{Cov}(R_m, \epsilon_j) + \beta_j \text{Cov}(R_m, \epsilon_i) + \text{Cov}(\epsilon_i, \epsilon_j)
\]
so
\[
\sigma_{ij} = \beta_i \beta_j \sigma_m^2,
\]
when $j \neq i$. On the other hand, when $j = i$, we have
\[
\text{Cov}(R_i, R_i) = \beta_i \beta_i \text{Cov}(R_m, R_m) + \beta_i \text{Cov}(R_m, \epsilon_i) + \beta_i \text{Cov}(R_m, \epsilon_i) + \text{Cov}(\epsilon_i, \epsilon_i)
\]
so
\[
\sigma_i^2 = \sigma_{ii} = \beta_i^2 \sigma_m^2 + \tau_i^2
\]
when $j = i$.

If we let $T$ denote the diagonal matrix with $\tau_i^2$, $i = 1, \ldots, n$ in the diagonal, and let $\beta$ denote the column vector of betas, then we can write
\[
V = \sigma_m^2 \beta \beta' + T.
\]

For this reason the single-index model is also known as the diagonal model.

To fit the single index model we need to estimate the $\alpha_i$s, $\beta_i$s, the $\tau_i^2$s, $\overline{R}_m$ and $\sigma_m^2$. Altogether we need $3n + 2$ estimates. For example, if $n = 1000$ we need 3,002 estimates. In contrast, the full covariance model requires estimating the mean and variance of each security and the correlation among all securities. In total the full covariance model requires the estimation of $n + n(n + 1)/2$ elements. If $n = 100$ then the number of parameters that need to be estimated is 501,500.
3.1 The Variance of a Well Diversified Portfolio under the SIM

Let \( x_p \) be a portfolio, then
\[
\overline{R}_p = r'x_p = \alpha_p + \beta_p \overline{R}_m
\]
and
\[
\sigma_p^2 = \beta_p^2 \sigma_m^2 + x_p' T x_p,
\]
where \( \alpha_p \equiv \alpha' x_p \) and \( \beta_p \equiv \beta' x_p \).

The above equation decomposes the variance of a security or portfolio \( x_p \) into a market risk term \( \beta_p^2 \sigma_m^2 \) and unique risk \( x_p' T x_p \). The market risk is often called systematic or undiversifiable risk. The unique risk is often called unsystematic risk or diversifiable risk. Indeed, if \( x_p \) is a well diversified portfolio then the unique risk \( x_p' T x_p \) goes to zero as \( n \to \infty \) if each component of \( x_p \) is bounded by \( k/n \) for some constant \( k \) and \( \tau_i \leq s^2 \) for some constant \( s \). Thus, for a well diversified portfolio we have
\[
\sigma_p^2 \approx \beta_p^2 \sigma_m^2.
\]

3.2 Estimating \( \alpha_i \) and \( \beta_i \).

The easiest way to estimate \( \alpha_i \) and \( \beta_i \) is via linear regression. Before we discuss linear regression in detail here is a heuristic derivation under the assumption that the market index is itself a portfolio. This heuristic leads to the same estimates obtained by linear regression.

Recall that for fixed \( i \)
\[
\sigma_{ij} = \beta_i \beta_j \sigma_m^2 + \tau_i^2 I(i = j)
\]
where \( I(i = j) = 1 \) if \( i = j \) and 0 otherwise.

Now assume that \( R_m = \sum_{i=1}^n x_{mi} R_i \), so that the market index is itself a portfolio with weights \( x_{mi}, i = 1, \ldots, n \). It is easy to see that \( \alpha_m = \alpha' x_m = 0 \) and \( \beta_m = \beta' x_m = 1 \) are the only consistent values of \( \alpha_m \) and \( \beta_m \).

Now let us compute the covariance
\[
\sigma_{im} = \sum_{j=1}^n \sigma_{ij} x_{mj}
\]
\[
= \sum_{j=1}^n [\beta_i \beta_j \sigma_m^2 + \tau_i^2 I(i = j)] x_{mj}
\]
\[
= \beta_i \beta_m \sigma_m^2 + \tau_i^2 x_{mi}
\]
\[
= \beta_i \sigma_m^2 + \tau_i^2 x_{mi}
\]
\[
\approx \beta_i \sigma_m^2.
\]
The last equation follows from \( \beta_m = 1 \), and the approximation from the fact that \( x_m \) is a well diversified portfolio so \( \tau_i^2 x_{mi}^2 \approx 0 \). Using \( \sigma_{im} = \beta_i \sigma_m^2 \) as an approximation we see that
\[
\beta_i = \frac{\sigma_{im}}{\sigma_m^2}
\]
(1)

If we have data \( R_{ih} \) and \( R_{mh} \) over periods \( h = 1, \ldots, H \) we can estimate \( \beta_i \), via (1), by
\[
\hat{\beta}_i = \frac{\sum_{h=1}^H (R_{ih} - \overline{R}_i)(R_{mh} - \overline{R}_m)}{\sum_{h=1}^H (R_{mh} - \overline{R}_m)^2},
\]
(2)

Now, since
\[
\overline{R}_i = \alpha_i + \beta_i \overline{R}_m
\]
it follows that
\[
\alpha_i = \overline{R}_i - \beta_i \overline{R}_m.
\]
This leads us to estimate \( \alpha_i \) by
\[
\hat{\alpha}_i = \overline{R}_i - \hat{\beta}_i \overline{R}_m.
\] (3)

Both \( \hat{\alpha}_i \) and \( \hat{\beta}_i \) are statistics, i.e., random variables, that are used to estimate \( \alpha_i \) and \( \beta_i \). The values of \( \hat{\alpha}_i \) and \( \hat{\beta}_i \) are random because they depend on the realization of the observations. The estimators \( \hat{\alpha}_i \) and \( \hat{\beta}_i \) are unbiased. Indeed,
\[
E[\hat{\alpha}_i] = \alpha_i,
\]
and
\[
E[\hat{\beta}_i] = \beta_i.
\]

As we shall see next, the estimators (2) and (3) also arise from linear regression.

### 3.2.1 Linear Regression

Suppose we have data \((x_i, y_i)\) and, perhaps after plotting the data, we suspect that \( y \) varies linearly with \( x \). We postulate the linear model
\[
y_i = a + bx_i + \epsilon_i
\]
where \( a \) and \( b \) are unknown constants. Our objective is to find estimates \( \hat{a} \) and \( \hat{b} \) of \( a \) and \( b \) that minimize the sum of square errors. The vector of errors is given by
\[
\epsilon = y - ae - bx
\]
and the sum of squares error, as a function of \( a \) and \( b \) is given by
\[
s(a, b) = \epsilon^T \epsilon = (y - ae - bx)'(y - ae - bx).
\]
The objective is to find \( a \) and \( b \) to minimize \( s(a, b) \). Taking partial derivatives with respect to \( a \) and \( b \) and setting them equal to zero we obtain the system of equations:
\[
a \epsilon' e + b \epsilon' x = \epsilon' y
\]
and
\[
a \epsilon' x + b x' x = y' x.
\]
Before solving this equations we remark that
\[
\epsilon' (y - ae - bx) = 0
\]
indicating that the vector of errors is orthogonal to \( e \). Likewise, we have
\[
x' (y - ae - bx) = 0
\]
indicating that the vector of errors is orthogonal to \( x \). Solving we obtain:
\[
\hat{b} = \frac{\text{Cov}(x, y)}{\text{Cov}(x, x)} = \frac{(x - \bar{x})' (y - \bar{y} e)}{(x - \bar{x})' (x - \bar{x})}.
\]
The estimate of \( a \) is
\[
\hat{a} = \bar{y} - \hat{b} \bar{x}.
\]
Now let us look at the errors. The regression line is \( \hat{a} e + \hat{b} x \) so the errors are
\[
\hat{\epsilon} = y - \hat{a} e - \hat{b} x = y - (\bar{y} - \hat{b} \bar{x}) e - \hat{b} x
\]
\[
= (y - \bar{y} e) - \hat{b} (x - \bar{x} e).
\]
Letting
\[
\text{Var}(\epsilon) = \frac{1}{n} \epsilon^t \epsilon
\]
and
\[
\text{Var}(y) = \frac{1}{n} (y - \bar{y})^t (y - \bar{y})
\]
we can write:
\[
\text{Var}(\epsilon) = \text{Var}(y) - 2 \hat{b} \text{Cov}(x, y) + \hat{b}^2 \text{Var}(x)
\]
\[
= \text{Var}(y) - \frac{\text{Cov}(x, y)^2}{\text{Var}(x)}.
\]
Consequently,
\[
\text{Var}(y) = \text{Var}(\epsilon) + \frac{\text{Cov}(x, y)^2}{\text{Var}(x)}.
\]
Dividing by \text{Var}(y) we obtain
\[
1 = \frac{\text{Var}(\epsilon)}{\text{Var}(y)} + R^2.
\]
where
\[
R^2 = \frac{\text{Cov}(x, y)^2}{\text{Var}(x) \text{Var}(y)}.
\]
Here \(R^2\) is the proportion of the total error explained by the linear model.
Notice that \(R^2 = \rho_{xy}^2\).

3.3 Using The Single Index Model

In practice, the Betas obtained via regression are adjusted to improve the accuracy of the forecast of future Betas. Here we are interested in the use of the SIM in finding optimal portfolios. Assume that
\[
V = \sigma_m^2 \beta \beta^t + T,
\]
where the parameters of \(V\) are obtained by possibly adjusting the outputs of the linear regression.

We are interested in finding the portfolio \(x_q = \frac{V^{-1} q}{\sigma_m^2 \beta y}\) with the highest Sharpe ratio, where, as before, \(q = r - R_f \epsilon\) is the vector of expected excess returns. Recall that we can write \(x_q = y / \epsilon^t y\) where \(V y = q\).

Notice that
\[
V y = \sigma_m^2 \beta y + T y = q
\]
implies
\[
y = T^{-1} (q - \sigma_m^2 \beta y) = T^{-1} (q - C \beta).
\]
where
\[
C = \sigma_m^2 \beta y.
\]
Notice that \(y = \epsilon^t V^{-1} q x_q\), implies that
\[
\beta^t y = \epsilon^t V^{-1} q \beta x_q = \epsilon^t V^{-1} q \beta q,
\]
This gives
\[
C = \sigma_m^2 \beta_q \epsilon^t V^{-1} q.
\]
Notice that \( q_t = \mathbf{q}' x_q = \mathbf{q}' V^{-1} q / \mathbf{d}' V^{-1} q \), and that Also \( \sigma_q^2 = \mathbf{x}_q' V x_q = \mathbf{q}' V^{-1} q / (\mathbf{d}' V^{-1} q)^2 \). Dividing we obtain \( \frac{\sigma_q^2}{\sigma_q^2} = \mathbf{d}' V^{-1} q \), so we have

\[
C = \sigma_m^2 \beta_q \frac{\overline{R}_q - R_f}{\sigma_q^2},
\]

where \( \overline{R}_q = q_t + R_f \) is the expected return of the tangency portfolio. Then

\[
y_k = \frac{q_t - C \beta_i}{\tau_i^2} \quad i = 1, \ldots, n,
\]

and

\[
x_q = \frac{y}{\mathbf{c}' y}.
\]

From this we observe that \( x_i \) and \( y_k \) have the same sign provided that \( \sum_{j=1}^n y_j > 0 \). Under this assumption we see that \( x_i > 0 \) if and only if \( q_t > C \beta_i \). This means that security \( i \) is held long if and only if its excess return \( q_t \) is larger than \( C \) times \( \beta_i \).

Suppose now that two securities, say \( i \) and \( j \), have the same excess-return, the same Beta, and are both held long in the portfolio. It is easy to see that if \( \tau_j^2 = 2 \tau_i^2 \) then you would want to hold twice as much of security \( i \) than you would of security \( j \), that is \( \mathbf{c}_i' x_q = 2 \mathbf{c}_j' x_q \). This follows since \( y_k = \frac{q_t - C \beta_i}{\tau_i^2} \) implies \( y_j = 0.5 y_k \).

If the tangency portfolio is the market index then \( R_q = R_m, \sigma_q^2 = \sigma_m^2 \), and \( \beta_q = \beta_m = 1 \). Then,

\[
C = \frac{\overline{R}_m - R_f}{\beta_m} = \overline{R}_m - R_f = q_m
\]

and security \( i \) is held long if and only if

\[
q_t > \beta_i q_m.
\]

Now the rule is to hold long all securities with excess returns larger than their Betas multiplied by

the excess return of the market index.

See example on page 191 of textbook, and excel spreadsheet “singleindex” for a detailed example.

4 The Constant Correlation Models

The constant correlation model assumes \( \sigma_{ij} = \rho \sigma_i \sigma_j \) for all \( i \neq j \), that is, it assumes that there is a constant correlation among the securities. This leads to a parsimonious model where only \( 2n + 1 \) parameters need to be estimated. Under this model

\[
V = \Lambda + \rho \sigma' \sigma
\]

where \( \sigma \) is the \( n \times 1 \) column vector of security standard deviations and \( \Lambda \) is a diagonal matrix with \( \Lambda_{ii} = (1 - \rho) \sigma_i^2 \) for all \( i = 1, \ldots, n \).

Here again we are concerned with finding implications about the tangency portfolio \( x_q \).

Notice that \( V y = q \) implies

\[\Lambda y = q - \rho \sigma \sigma' y = q - C \sigma,\]

where \( C = \rho \sigma' y \), so \( y = \Lambda^{-1} (q - C \sigma) \).

Premultiplying \( \Lambda^{-1} (q - C \sigma) \) by \( \sigma' \), and using the definition of \( C \) to solve for \( \sigma' y \), we obtain

\[
\sigma' y = \sigma' \Lambda^{-1} (q - C \sigma) = \sigma' \Lambda^{-1} q - \rho \sigma' y \sigma' \Lambda^{-1} \sigma.
\]

Solving for \( \sigma' y \) we find

\[
C = \frac{\rho \sigma' \Lambda^{-1} q}{1 + \rho \sigma' \Lambda^{-1} \sigma}.
\]
Expanding we find
\[ \sigma' \Lambda^{-1} q = \sum_{i=1}^{n} \frac{\sigma_i q_i}{(1 - \rho) \sigma_i^2} = \sum_{i=1}^{n} \frac{q_i}{(1 - \rho) \sigma_i}. \]

On the other hand,
\[ \sigma' \Lambda^{-1} \sigma = \sum_{i=1}^{n} \frac{\sigma_i^2}{(1 - \rho) \sigma_i^2} = \frac{n}{1 - \rho}. \]

We conclude that
\[ C = \frac{\rho \sum_{i=1}^{n} S_i}{1 - \rho + n \rho}, \]
where \( S_i = \frac{q_i}{\sigma_i}, \) is the Sharpe ratio of security \( i. \) Multiplying and dividing by \( n \) we see that
\[ C = \frac{n \rho}{n \rho + (1 - \rho) \bar{S}}, \]
where \( \bar{S} \) is the average Sharpe ratio. Thus, \( C \leq \bar{S}. \)

Now
\[ y_i = \frac{q_i - C \sigma_i}{(1 - \rho) \sigma_i^2} = \frac{S_i - C}{(1 - \rho) \sigma_i} \quad i = 1, \ldots, n, \]
and
\[ x_q = \frac{y}{\sigma y}. \]

It is easy to see that \( x_q \) and \( y_i \) have the same sign under the assumption that \( \sum_{i} y_j > 0. \) Under this assumption, security \( i \) is held long if and only if
\[ S_i > C. \]

Now suppose two securities, say \( i \) and \( j, \) have the same Sharpe ratio, and both are held long in the tangency portfolio. If \( \sigma_i = 2 \sigma_j, \) then \( y_j = 2 y_i, \) and hence \( x_{q,j} = 2 x_{q,i} \) since \( x_q \) is proportional to \( y. \)

In tests done by Elton and Gruber the constant correlation model outperformed both single and multi-index models and the differences were almost always statistically significant. Moreover, using the constant correlation model often led to a significant increase in expected returns for a given risk value. A more disaggregated averaging model would attempt to estimate the mean correlation within and between groups of stocks, e.g., between industries. This model also outperformed single and multi-index models, but the comparison with the overall mean model is somewhat ambiguous.