Singular Value Decomposition

- Unfortunately, M is not always diagonalizable and even when it is, in general, the diagonalization algorithm is not always stable. This makes it less than ideal method to use in practice, when you need software to "just run", without complaints or exceptions.
- For example, when you design automatic systems, you would then need to write code to deal with each possible complaint and exception, and that increases the software burden significantly.
- Also, we do not always need high powers of M: one power that we often have a very particular interest in is -1, i.e., the inverse, that appears in portfolio optimization as we will see a few slides later.

Singular Value Decomposition: definition

- Singular Value Decomposition of an m by n matrix M is given by an orthogonal m by n matrix U, a diagonal n by n matrix D, and orthogonal n by n matrix V such that

$$\underset{m\times n}{M} = \underset{m\times n}{U}\,\underset{n\times n}{D}\,\overset{n\times n}{V^T} \tag{7}$$

The entries of D, which are non-negative, are called singular values and play a role similar to eigenvalues. This decomposition is very stable: it works in pretty much all practical situations and gives us a lot of information about M.

- For details and excellent discussion, see Ch.2 of Numerical Recipes.

Singular Value Decomposition

- If M is symmetric, it has an orthonormal eigenbasis. If, in addition, (M is a covariance matrix, its eigenvalues are non-negative) Since both SVD and diagonalization are unique (up to ordering), they must coincide. [Why?] Therefore,

$$\sqrt{C} = UD^{1/2}V^T = UD^{1/2}V^{-1} = UD^{1/2}U^{-1}. \tag{8}$$

$V^T = V^{-1}$

- Similarly, the inverse of C

$$C^{-1} = UD^{-1}V^T \tag{9}$$

when C is indeed invertible, but, with SVD, we can go a step further: we leave the zero elements of D at zero and only take reciprocals of the non-zero elements of D. 找主成分

- We can also choose to set the small d_is to zero. This, by the way, is Principal Component Analysis (PCA).

Optimal Portfolio: no constraints

- Suppose we have a history of N securities for the last I days. Let H be the $I \times N$ matrix of their histories.
- Then their covariance matrix

$$C = H^T H \tag{10}$$

H: history of each particular securities.

(We assumed for simplicity of notation and without loss of generality that these histories have zero mean.)

- Let R be the vector of the securities' expected returns. It is natural to want to find the portfolio weights w so that

$$\langle R, w\rangle - a\,\langle w, Cw\rangle \tag{11}$$

is maximized. $\langle R, w\rangle$ above is the expected return of the portfolio and $\langle w, Cw\rangle$ is its variance, a measure of risk. The constant a indicates the amount of our risk aversion.

Optimal Portfolio: no constraints

- If the maximum is attained in the interior, the gradient has to be zero at the maximum. If we take the gradient of (11) with respect to w, we must have

$$R - 2aCw = 0. \tag{12}$$

In optimization theory (12) is called the first order condition. Solving for w, we get:

$$w = \frac{1}{2a} C^{-1} R \tag{13}$$

if C is invertible.

- Well, is it? Let's explore what happens in practice.

Covariance matrix: individual histories

- Suppose we have a large number of securities to model, i.e., the number of columns N of the history matrix H in (10) is large, say, 1000. Does it matter what I is?
- To find out, first recall from linear algebra that rank of H (and of H^T) is at most $\min(I, N)$. It follows that the rank of $C = H^T H$ cannot be greater than $\min(I, N)$.
- Therefore, if we want C to be invertible we must have $\boxed{I > N}$. In other words, we must have at least 1000 observations for *each* instrument if we are to have any hope of forming a portfolio of 1000 instruments: normally, I needs to be much larger than N.
- If you use any other method of maximizing (11), e.g., optimization, you are still trying to arrive at (13) and this non-invertability will still be an issue, although it might manifest itself in some other way.

Covariance matrix: factors

- Some people try to avoid the problem on the previous slide by using factors. Let's look at this approach more closely. First we need some notation: let F be the matrix of histories of the chosen factors. The matrix F has the same length as H, but is narrower. Let L be the matrix of factor loadings. Every column of L holds the regression coefficients of the corresponding column of H on the factors. The matrix L, therefore, is as wide as H and its height equals the width of F.
- Regression of H onto F will not be exact, of course: it will have residuals. We gather them in a matrix Z, which will have the same shape as H.
- In summary,

$$H = FL + Z \tag{14}$$

(L) residual (proj on F) reminder. (f) FL Z

Covariance matrix: factors

- Note that on the previous slide we did not specify what the factors are or how many of them we are looking at: for now, the analysis is completely generic.
- We have

$$\begin{aligned} C &= H^T H = (FL+Z)^T(FL+Z) \\ &= (Z^T + L^T F^T)(FL+Z) \\ &= Z^T FL + Z^T Z + L^T F^T FL + L^T F^T Z \\ &= Z^T Z + L^T F^T FL \end{aligned} \tag{15}$$

$F^T Z = Z^T F = 0$ Z is orthogonal to F

To see why the last equality holds, recall that regression of a column of H on the factors is a projection of that column onto the space spanned by the columns of F. That means that the residual must be perpendicular to the span of the columns of F. But that's just another way of saying that $Z^T F = 0$.

H residual e proj(H)

Covariance matrix: factors

- Let's look at (15) again:

$L^T F^T = (FL)^T$

$$\begin{aligned} C &= Z^T Z + L^T F^T FL \\ H^T H &= Z^T Z + (FL)^T FL \end{aligned} \tag{16}$$

- Different choices of factors will lead to different decompositions. For example, if we choose no factors at all we will end up in the situation on slide 10 with $Z = H$.
- We can use just one factor, e.g., the market factor. That will give us CAPM.
- People often use Fama-French factors: market, size, book.
- Another common choice is to use sector indices as factors.
- And let's not forget Principal Component Analysis (PCA).

Covariance matrix: factors and rank

- It would be natural to think that the need for long history that we saw on slide 10 is gone when we use just a few factors.
- Let us look again at (16), dispassionately:
  1. The rank of F cannot exceed the number of factors. Therefore, the same must be true of FL.
  2. The rank of F cannot be larger that the rank of H, since normally it makes no sense to choose more factors than individual instruments.
  3. Recall that $Z = H - FL$. Therefore, the rank of Z cannot exceed the rank of H.
  4. This means that C in (16) is not invertible unless Z has more rows than columns, i.e., unless we have more history than the number of instruments, *regardless of how we pick the factors.* observations > securities

Covariance matrix: caution!

- What could possibly make C invertible?
- Suppose that we have the confidence (something the trading community does not lack) that we picked our factors so beautifully that the residuals are pure noise. Since they are noise, they must all be mutually orthogonal. That would make $Z^T Z$ in (16) diagonal. And, since each diagonal element is a variance of something, it must be positive. Z is residual
- That would make C invertible, because it would be a sum of a strictly positive diagonal matrix and of a low-rank one.
- But the diagonal matrix, the one that makes the inversion possible, and dominates the portfolio, is made out of pure noise. By construction!
- Are we sure we want to use this method for our portfolio weights?

Covariance matrix: caution!

- There has to be a better way...
- Indeed, there are several.
- Anything is better than inverting noise.
- One thing you can do that's crude but effective is to replace $Z^T Z$ with a diagonal matrix of variances of the individual instruments. This works in the sense of giving a reasonable and reasonably stable portfolio, but has no logical justification, since it completely disregards the fact that Z is a matrix of residuals and simply plops down another matrix in its place.
- Another, more sophisticated, and also effective method is to follow Ledoit, *etal*, who blend C with a specially constructed diagonal matrix and do so with justification from statistics.

Covariance matrix: PCA

- In the methods mentioned above, the factors were picked "by hand", i.e., based on economic intuition, but with the side effect of running into mathematical difficulties.
- We could ignore economics and instead focus on math. Then we can pick the factors based on the the fraction of total variance that they hold.
- We do it by diagonalizing C so that its eigenvalues are in decreasing order: $C = UDU^{-1} = UDU^T$ for some diagonal D and orthonormal U.

Covariance matrix: PCA

- Now consider the matrix HU. We can think of columns of U as loadings on the instruments which are columns of H. In other words, columns of HU are some sort of factors. coefficient
- Let's compute the covariance matrix of these factors:

$$(HU)^T HU = U^T H^T HU = U^T CU = D. \tag{17}$$

$UDU^T = C$

In other words, the factors that are the columns of HU are orthogonal.

- This method is called Principal Component Analysis. The factor that corresponds to the top eigenvalue is called the first principal component and it has the highest variance of all possible factors. The second principal component has the second highest variance, and so on.