A. Static (Marginal) Analysis

A.I. Descriptive Statistics

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

Introduction

To conduct data analysis, we need *summary statistics*.

They will be an *average* of the observations or their functions in a given period of time.

Here dynamic aspects are not taken into account.

This means that the fact that available information evolves through time is not taken into account.
$\Rightarrow$ This is essentially a *static* or marginal analysis.

Preliminaries

We consider n financial assets with prices $p_{i,t}$ at date t.

The return of asset i is computed as

$$y_{i,t} = \frac{p_{i,t} - p_{i,t-1}}{p_{i,t-1}} \text{ or } y_{i,t} = \ln \frac{p_{i,t}}{p_{i,t-1}}.$$

The time horizon is one period, say one day.

In a vectorial notation, we get

$$y_t = \begin{pmatrix} y_{1,t} \\ \vdots \\ y_{i,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = (y_{1,t}, ..., y_{i,t}, ... y_{n,t})' .$$

Mean and Covariance

The *mean* is defined by $m = E\left[Y_t\right]$.

It is a vector of size $n \times 1$ with components equal to asset mean return: $m_i = E\left[Y_{i,t}\right]$.

It measures the average *location* of the random vector.

The *covariance matrix* is defined by

$$\Sigma = V\left[Y_t\right] = E\left[\left(Y_t - E\left[Y_t\right]\right)\left(Y_t - E\left[Y_t\right]\right)'\right].$$

It is a matrix of size $n \times n$ with diagonal elements equal to the variances

$$\sigma_i^2 = E\left[Y_{i,t}^2\right] - E\left[Y_{i,t}\right]^2,$$

and off-diagonal elements equal to the covariances

$$\sigma_{ij} = E\left[Y_{i,t} Y_{j,t}\right] - E\left[Y_{i,t}\right] E\left[Y_{j,t}\right].$$

Estimation of the Mean

The variance and covariance measure the *dispersion* and the *linear dependence* between the components of the random vector.

To estimate an expectation we only need to replace E by an empirical average. Hence we get the *empirical mean*

$$\hat{m} = \frac{1}{T}\sum_{t=1}^{T} y_t\,,$$

which is a vector of size $n \times 1$ with components equal to the *average returns*

$$\hat{m}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}\,,$$

Estimation of the Covariance

and the *empirical covariance matrix*

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \left[(y_t - \hat{m}) (y_t - \hat{m})' \right]$$

which is a matrix of size $n \times n$ with diagonal elements equal to the *empirical variances*

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}^2 - \hat{m}_i^2,$$

and off-diagonal elements equal to *empirical covariances*

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^{T} y_{i,t} y_{j,t} - \hat{m}_i \hat{m}_j.$$

Shape Parameters - Skewness

Other summaries are also available. They are called the skewness and the kurtosis and are shape parameters.

The *skewness* summarizes asymmetry:

$$g_{i1} = \frac{E\left[(Y_{it} - m_i)^3\right]}{\sigma_i^3} \text{ (Fischer coefficient)}$$

if > 0,
then we get *positive* (or right) *asymmetry.*
if $= 0$,
then we get *symmetry.*
if < 0,
then we get *negative* (or left) *asymmetry.*

Shape Parameters - Kurtosis

The *kurtosis* summarizes the thickness of tails:

$$b_i = \frac{E\left[(Y_{it} - m_i)^4\right]}{\sigma_i^4} \text{ (Pearson coefficient)}$$

if > 3,
then fatter tails than normal distribution (*leptokurtic*)
if $= 3$,
then normal tails (*mesokurtic*)
if < 3,
then thinner tails than normal distribution (*platykurtic*)

We often work with *excess kurtosis*:

$$g_{i2} = \frac{E\left[(Y_{it} - m_i)^4\right]}{\sigma_i^4} - 3 \text{ (Fischer coefficient)}$$

Empirical Estimators

Again estimators are obtained by substituting an empirical average for the expectation.

Skewness

$$\hat{g}_{i1} = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{m}_i)^3}{\hat{\sigma}_i^3}$$

Kurtosis

$$\hat{b}_i = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{m}_i)^4}{\hat{\sigma}_i^4}$$

$$\hat{g}_{i2} = \hat{b}_i - 3$$

Issues

If the sample is too short, then the *bias* may be large.

It is preferable to use instead the following estimators

$$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} \left[(y_t - \hat{m}) (y_t - \hat{m})' \right]$$

$$\hat{g}_{i1} = \frac{\frac{T}{(T-1)(T-2)} \sum_{t=1}^{T} (y_{it} - \hat{m}_i)^3}{\hat{\sigma}_i^3}$$

$$\hat{g}_{i2} = \frac{\frac{T(T+1)}{(T-1)(T-2)(T-3)} \sum_{t=1}^{T} (y_{it} - \hat{m}_i)^4}{\hat{\sigma}_i^4} - 3 \frac{(T-1)^2}{(T-2)(T-3)}$$