the Arbitrage Pricing Theory

A little about the APT.
The usual way of presenting the APT

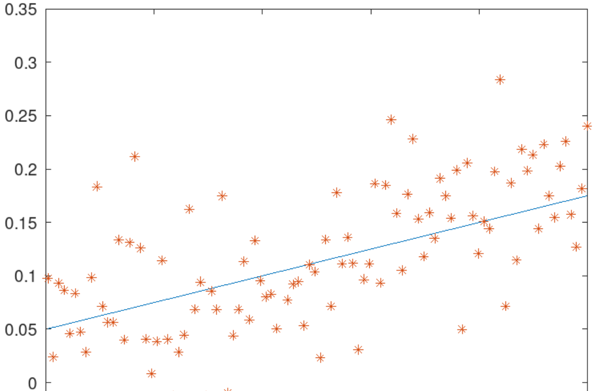
One factor models

Return on a stock only depend on one variable.
Example: dividend yield

$$E[r_i] = \lambda_{0i} + \lambda_{1i}(\text{dividend yield})$$

Another example: return on the stock market (stock market index)

$$E[r_i] = \lambda_{0i} + \lambda_{1i}(E[r_m])$$



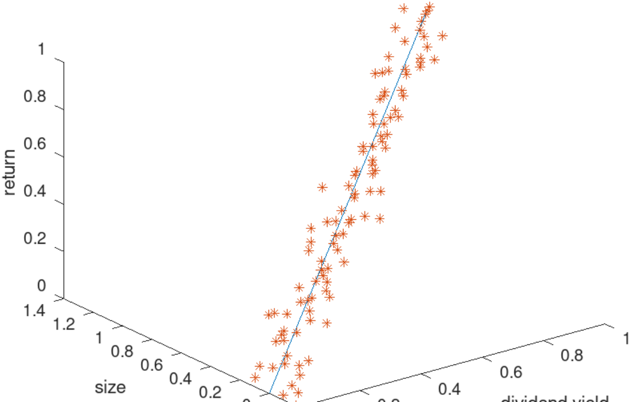
2 factor models

Expected return function of two variables

Example

- Dividend yield
- Size

$E[r_i] = \lambda_{0i} + \lambda_{1i}(\text{dividend yield}) + \lambda_{1i}(\text{size})$



M factor models

In general, may be more than two "factors" influencing returns. If we have M factors, can write the pricing equation as

$$E[r_i] = \lambda_{0i} + \lambda_{1i} f_1 + \lambda_{2i} f_2 + \cdots + + \lambda_{Mi} f_M$$

With matrix algebra, we can rewrite this as

$$E[r_i] = \lambda_i \mathbf{F}$$

where

$$\lambda_i = \begin{bmatrix} \lambda_{0i} \\ \lambda_{1i} \\ \vdots \\ \lambda_{Mi} \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix}$$

if we add an error term, we can write this as

$$r_i = \lambda_i \mathbf{F} + e_i$$

How to go about testing this?
Restriction imposed by theory: Returns a linear function of "factors."
Problem: Factors not identified.
Approaches to testing the APT:

Approach 1 Prespecify the "factors:"
Can then observe the outcomes of the factors, F_t, at dates t:

$$r_{it} = \lambda_i \mathbf{F_t} + e_{it}$$

Here

$$\lambda_i = \begin{bmatrix} \lambda_{0i} \\ \lambda_{1i} \\ \vdots \\ \lambda_{Mi} \end{bmatrix}$$

is a fixed set of constants, to be estimated, whereas

$$r_{it}$$

and

$$\mathbf{F} = \begin{bmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{Mt} \end{bmatrix}$$

are observed data.

If we have specified the “factors.”

Estimation of this is a simple, well known exercise. It is a multivariate regression, and can be estimated similarly to the CAPM test.

The only problem is then the choice of “factors.” The typical choices are to take macroeconomic variables believed to be important. Well known examples include

- Stock market returns
- Level of interest rates
- Shape of the term structure
- Consumption
- ...

A classical reference is Chen et al. [1986].

Once these variables have been specified, we are back in the familiar multiple regression framework.
But there is a theoretical problem: We do not know that the "proxies" we have picked are the "true" factors.
If we reject the model, may be a specification error, the "proxies" ar not important in predicting returns, but there are others, *true* factors that would predict returns

Approach 2: This is a case for *not* trying to prespecify the "factors," instead to try to estimate the "factors" from the data. This approach is exemplified in the Roll and Ross [1980] paper. To estimate

$$r_{it} = \lambda_i \mathbf{F_t} + e_i$$

Problem: Neither λ_i nor $\mathbf{F_t}$ is actually observed, only r_{it}. RR therefore use the technique of *factor analysis*.

Assumption: r_i normally distributed, with independent errors

$$E[e_i e_i'] = \sigma_i^2$$

$$E[e_i e_j'] = 0$$

Then the technique of factor analysis will estimate λ_i for each stock.

The factor analytic method will produce estimates of λ_i for a given dimension of F (A given number of factors)

To estimate the number of factors, see how much we improve the fit by adding one "factor"

Intuitively, see "how diagonal" the empirical error covariance matrix is.

APT estimation illustrated on Norwegian data

The classical "factor fishing" studies

One way of empirically investigate the APT is to use the multivariate statistics that look for structure in data, either as principal components, or as factor analysis.
Consider the ten industry portfolios at the Oslo Stock Exchange. Investigate the apparent number of *pervasive factors* in these returns.
Hint: The time series for some of the industries are not complete.

Reading in the data

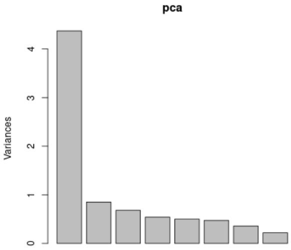
```
rets <- read.table("../../../data/norway/stock_market_portf
# last 2  industries are not complete
indret <- rets[,2:9];
```

Doing a principal compoents analysis

```
> pca <- prcomp(indret,scale=TRUE)
> pca
Standard deviations:
[1] 2.0969921 0.9407977 0.8604766 0.7288378 0.6717544 0.6626867 0.5706448
[8] 0.4794333

Rotation:
                        PC1         PC2         PC3        PC4          PC5
X10.Energy.ew.    -0.3864563  0.06337547 -0.36337603 -0.2074789 -0.435870975
X15.Material.ew.  -0.3170406  0.29774228  0.64109403  0.2413182 -0.052037510
X20.Industry.ew.  -0.4140517  0.08234816 -0.06698331 -0.2455308 -0.327474277
X25.ConsDisc.ew.  -0.3708609  0.05665170  0.32268859  0.3627879 -0.034986634
X30.ConsStapl.ew. -0.3699629  0.12925201 -0.11899131 -0.2497945  0.831313297
X35.Health.ew.    -0.2161791 -0.86698741  0.33926954 -0.2780799  0.007162377
X40.Finan.ew.     -0.3902916  0.17331253 -0.08987029 -0.2687152 -0.009983089
X45.IT.ew.        -0.3243866 -0.31453452 -0.45934976  0.7034800  0.087248865
                          PC6         PC7         PC8
X10.Energy.ew.     0.29204240 -0.22383050  0.58836475
X15.Material.ew.   0.53712227  0.22353064  0.02060562
X20.Industry.ew.   0.01015819 -0.27042081 -0.75930091
X25.ConsDisc.ew.  -0.64036864 -0.40776349  0.21434240
X30.ConsStapl.ew.  0.14286230 -0.23898072  0.03552341
X35.Health.ew.     0.04980636  0.05155643  0.06306435
X40.Finan.ew.     -0.41380689  0.75009497  0.05807301
X45.IT.ew.         0.14801781  0.19525155 -0.14936630
```

Here it is useful to see a plot of the percentage contributions of the principal components.



We can alternatively employ the method of "factor analysis"
First look at the case of one "factor"

```
factanal(x = indret, factors = 1)
Uniquenesses:
   X10.Energy.ew.  X15.Material.ew.  X20.Industry.ew.  X25.ConsDisc.ew.
            0.364             0.641             0.240             0.488
X30.ConsStapl.ew.    X35.Health.ew.     X40.Finan.ew.         X45.IT.ew.
            0.470             0.858             0.369             0.630
Loadings:
                  Factor1
X10.Energy.ew.    0.797
X15.Material.ew.  0.599
X20.Industry.ew.  0.872
X25.ConsDisc.ew.  0.716
X30.ConsStapl.ew. 0.728
X35.Health.ew.    0.377
X40.Finan.ew.     0.795
X45.IT.ew.        0.608
               Factor1
SS loadings      3.941
Proportion Var   0.493

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 81.99 on 20 degrees of freedom.
The p-value is 1.8e-09
```

Then look at two factors

```
factanal(x = indret, factors = 2)
Uniquenesses:
   X10.Energy.ew.  X15.Material.ew.  X20.Industry.ew.  X25.ConsDisc.ew.
            0.184             0.564             0.253             0.338
X30.ConsStapl.ew.    X35.Health.ew.     X40.Finan.ew.        X45.IT.ew.
            0.483             0.853             0.380             0.621
Loadings:        Factor1 Factor2
X10.Energy.ew.     0.854   0.297
X15.Material.ew.   0.276   0.600
X20.Industry.ew.   0.668   0.549
X25.ConsDisc.ew.   0.312   0.752
X30.ConsStapl.ew.  0.506   0.511
X35.Health.ew.     0.219   0.315
X40.Finan.ew.      0.548   0.565
X45.IT.ew.         0.501   0.358
                Factor1 Factor2
SS loadings       2.204   2.123
Proportion Var    0.275   0.265
Cumulative Var    0.275   0.541

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 34.46 on 13 degrees of freedom.
The p-value is 0.00103
```

and 3 factors

```
factanal(x = indret, factors = 3)
Uniquenesses:
   X10.Energy.ew.  X15.Material.ew.  X20.Industry.ew.  X25.ConsDisc.ew.
            0.255             0.552             0.223             0.300
X30.ConsStapl.ew.    X35.Health.ew.     X40.Finan.ew.        X45.IT.ew.
            0.485             0.836             0.376             0.005

Loadings:
                  Factor1 Factor2 Factor3
X10.Energy.ew.    0.764   0.257   0.308
X15.Material.ew.  0.314   0.584
X20.Industry.ew.  0.705   0.479   0.225
X25.ConsDisc.ew.  0.285   0.744   0.257
X30.ConsStapl.ew. 0.513   0.431   0.257
X35.Health.ew.    0.173   0.267   0.251
X40.Finan.ew.     0.575   0.496   0.218
X45.IT.ew.        0.272   0.193   0.940

               Factor1 Factor2 Factor3
SS loadings      1.959   1.729   1.281
Proportion Var   0.245   0.216   0.160
Cumulative Var   0.245   0.461   0.621

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 10.9 on 7 degrees of freedom.
The p-value is 0.143
```

APT ex post (2012)

APT was very popular as a research topic in the eighties and nineties, but has now limited interest.

The empirical studies of the APT are more important because of what they lead to.

- The factor "fishing expecitions" using factor analysis or principal components lead to the use of these methods for empirically reducing a large dataset to a few "factors"
- The Chen Roll and Ross type of analysis, with prespecified, observable factors, lead directly towards Fama French.

Nai fu Chen, Richard Roll, and Stephen Ross. Economic forces and the stock market. *Journal of Business*, 59:383–403, 1986.

Richard Roll and Stephen A Ross. An empirical investigation of the Arbitrage Pricing Theory. *Journal of Finance*, 35:1073–1103, December 1980.