APT – looking for the “factors”

November 24, 2021

Contents

1 the Arbitrage Pricing Theory

A little about the APT.

The usual way of presenting the APT

1.1 One factor models

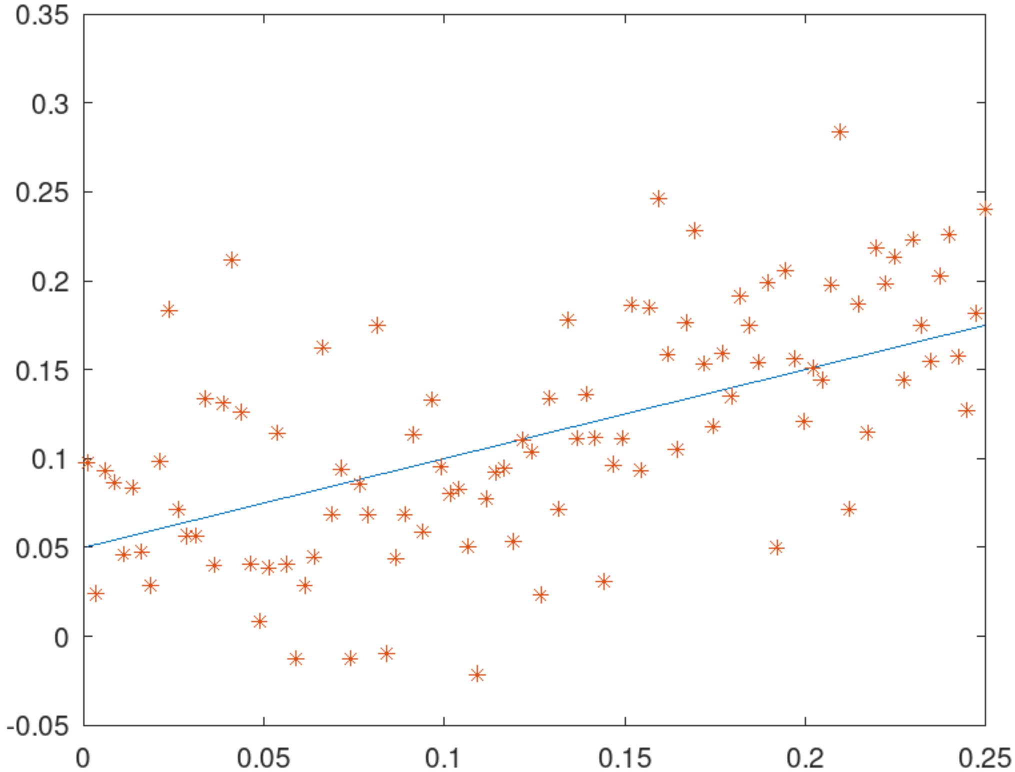
Return on a stock only depend on one variable.

Example: dividend yield

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

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

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



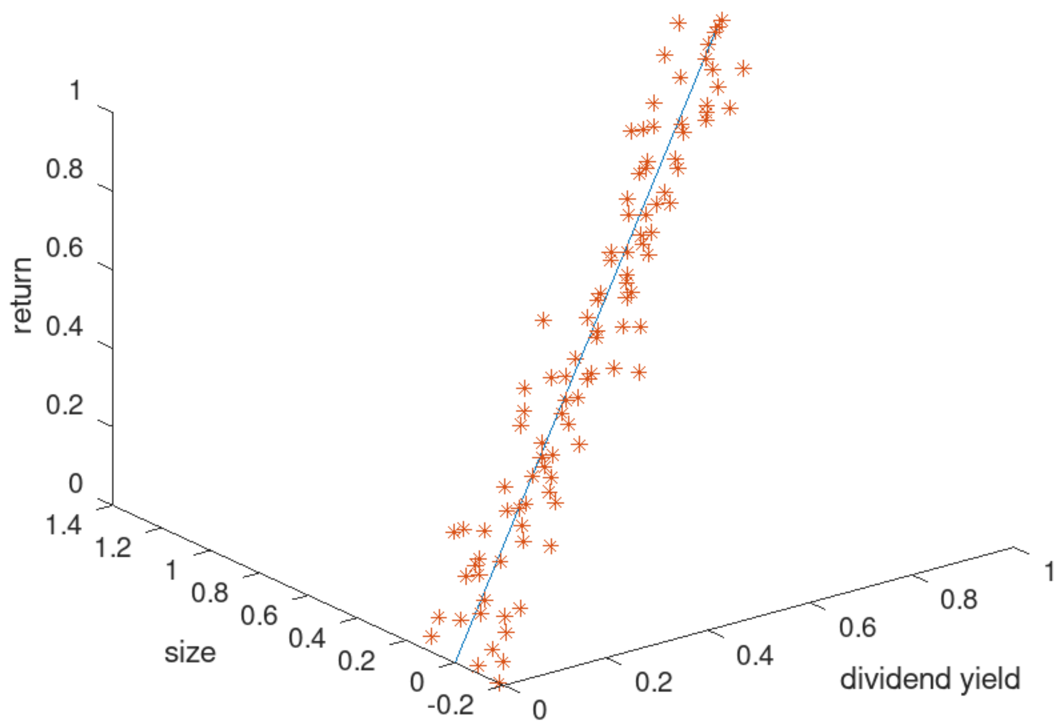
1.2 2 factor models

Expected return function of two variables

Example

- Dividend yield
- Size

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



1.3 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:

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.

   $$e_{it}$$

   is an error term.

   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, Roll, and Ross (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

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 examplified 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.

2 APT estimation illustrated on Norwegian data

2.1 The classical "factor fishing" studies

Exercise 1.

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.

Solution to Exercise 1.

Reading in the data

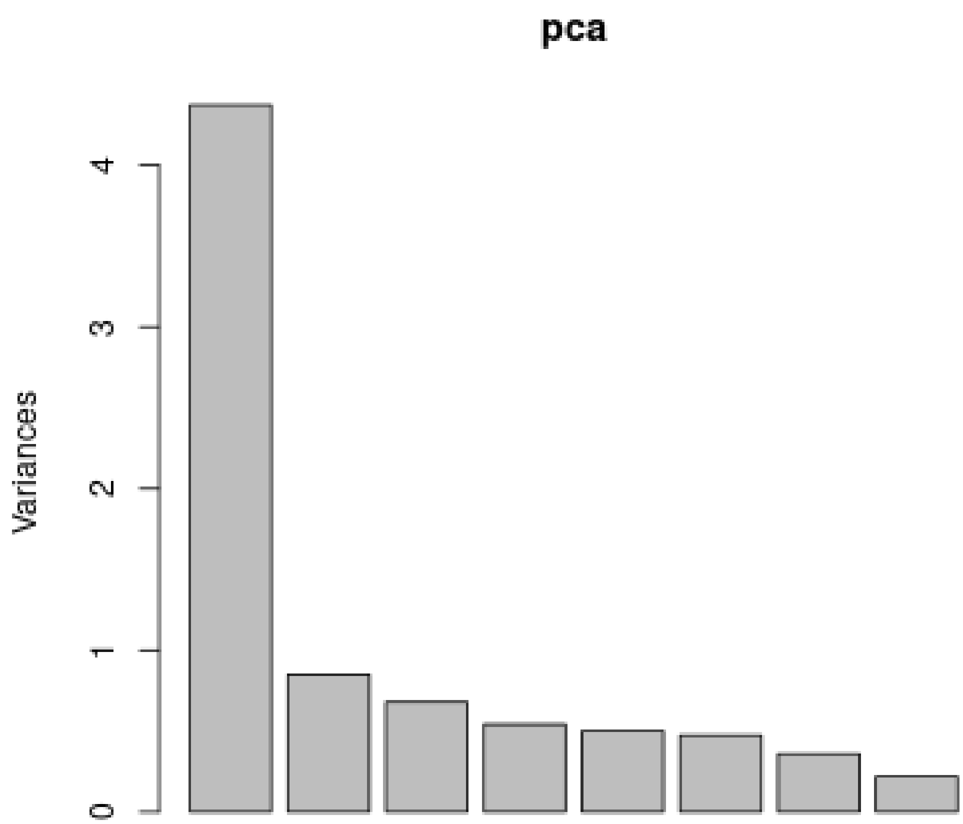
```
rets <- read.table("../../../data/norway/stock_market_portfolios/industry_portfolios_monthly_ew.txt",header=TRUE,se
# last 2  industries are not complete
indret <- rets[,2:9];
```

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"

```
> fact <- factanal(indret,1)
> fact

Call:
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
```

The 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
```

3 Testing the Arbitrage Pricing Theory - complete development

This is a more complete discussion of empirical testing of the APT.

This discussion of the APT follows the APT chapter in the Finance Handbook, with additional material.

For most of this lecture the classical APT analysis due to Ross (1976). At the end look at how this body of work fit into our basic asset pricing framework.

4 Introduction

Classical APT: Powerful intuition, due to Ross: The *must* be several sources of risk in the economy causing comovements in the very large cross sections provided by the worlds financial markets.

APT a result of trying to formalize this intuition.

Basic assumption: Returns a *linear* function of a small number (k) of pervasive "factors"

The original APT is due to Ross (1976), although the theory was not satisfactorily put into a General Equilibrium framework until the work of Connor (1984).

There are two major types of proofs of an APT-like relation. The first, which is the one that Ross used, relies on noting that as the number of assets in the economy goes to infinity, by a law of large numbers argument, all arbitrage opportunities must disappear. This proof relies on the possibility of increasing the number of assets without limit.

The intuition is as follows: As we increase the number of assets, all idiosyncratic sources of risk can be diversified away. The only relevant source of risk in the limit is the risk that is caused by the factors, and this is therefore the only one that is priced.

Another way of getting an APT-like relation is done in a finite economy, and called an *approximate APT*. It was developed by Chamberlain and Rothschild (1983) and Ingersoll (1984). Intuitively, it puts a bound on the amount of arbitrage that can be done, not due to the factors. If we have a finite, but large number of assets, we will have an approximate APT relation.

For some background on the APT, see Connor (1989). A textbook derivation of the APT is in (Huang and Litzenberger, 1988, pages 103–16) or (Ingersoll, 1987, Chapter 7).

If the APT assumptions hold, we will have a linear pricing relationship for all stocks i.

$$E[r_i] = \lambda_0 + b_{i1}\lambda_1 + \cdots + b_{ik}\lambda_k$$

We often interpret λ_j as the return on a *factor-mimicking portfolio*.

If there is a risk free asset with return R_f, we have $R_f = \lambda_0$.

In matrix form

$$r = E[r] + Bf + \varepsilon$$

Where

$$E[f] = 0$$

$$E[\varepsilon] = 0$$

$$E[f\varepsilon'] = 0$$

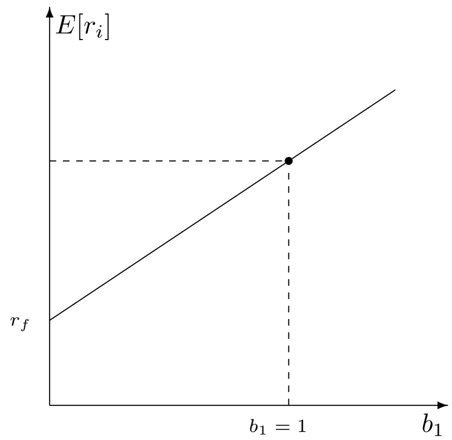
and

$$B = E[ff']^{-1}E[(r - E[r])f'] \quad \text{(projection)}$$

r is n-vector of returns.

If we have one factor and a risk free asset, the APT relation look very much like the CAPM:

$$E[r_i] - r_f = (E[\lambda_1] - r_f)\, b_i$$



If the APT holds, all assets map on the line in the above figure.

The assumption of a factor structure implies the following about the covariance matrix of asset returns.

$$\begin{aligned}\Sigma &= E[(r - E[r])(r - E[r])'] \\ &= E[(Bf + \varepsilon)(Bf + \varepsilon)'] \\ &= E[Bff'B'] + E[\varepsilon f'B'] + E[Bf\varepsilon'] + E[\varepsilon\varepsilon'] \\ &= BE[ff']B' + E[\varepsilon f']B' + BE[f\varepsilon'] + E[\varepsilon\varepsilon'] \\ &= BE[ff']B' + E[\varepsilon\varepsilon']\end{aligned}$$

Let $V = E[\varepsilon\varepsilon']$.

The decomposition

$$\Sigma = BE[ff']B' + V$$

has the following interpretation:

- The first term is the systematic risk.
- The second term is the idiosyncratic risk.

5 Exact factor models

In an exact factor model all idiosyncratic risk is uncorrelated with each others. Implication: V is diagonal.

The assumption of a diagonal V allows us to see why V is the "diversifiable" part of risk.

Let ω be a portfolio.

The idiosyncratic part of the portfolio variance is

$$\omega' V \omega = \sum \omega_i^2 \sigma_i^2 \leq (\max_i \sigma_i^2) \sum \omega_i^2$$

Take the example of an equally weighted portfolio, $\omega_i = \frac{1}{n}\ \forall\ i$.

$$\sum \left(\frac{1}{n}\right)^2 = n\frac{1}{n^2} = \frac{1}{n} \to 0 \text{ as } n \to \infty$$

For a well diversified portfolio,

$$\sum \left(\omega^2\right) \to 0 \text{ as } n \to \infty$$

As long as σ_i^2 is bounded,

$$\omega' V \omega \leq \max_i \sigma_i^2 \sum \omega_i^2 \to 0 \quad \text{as} \quad n \to \infty$$

The part of a portfolios variance due to the idiosyncratic part will disappear as the portfolio becomes well diversified.

6 Choice of rotation.

We now have assumed.

$$r = E[r] + Bf + \varepsilon \tag{1}$$

$$\Sigma = BE[ff']B' + V \tag{2}$$

$$V \quad \text{diagonal}$$

Is this enough to determine B?

No, there is an indeterminacy here. Suppose we are given B and f satisfying (1) and (2) above. Consider a constant, nonsingular matrix L. Define.

$$B^* = BL$$

$$f^* = L^{-1} f$$

Claim: B^* and f^* will also satisfy equation (1) and (2) above.
Proof:

$$\begin{aligned} r &= E[r] + B^* f^* + \varepsilon \\ &= E[r] + (BL)(L^{-1} f) + \varepsilon \\ &= E[r] + B(LL^{-1}) f + \varepsilon \\ &= E[r] + Bf + \varepsilon \end{aligned}$$

$$\begin{aligned} \Sigma &= BE[ff']B' + V \\ &= B(LL^{-1})E[ff'](LL^{-1})B' + V \\ &= (BL)E[(L^{-1}f)(f'L)(L^{-1}B') + V \\ &= B^* E[f^* f^{*\prime}]B^{*\prime} + V \end{aligned}$$

There is thus an infinite number of equivalent pairs (f, B) that satisfies the relations (1) and (2) above.

The econometrician can choose a restriction on f and B that is convenient.

Examples include

1. Impose $E[ff'] = I_k$ ($k \times k$ identity matrix).

2. Eigenvector decomposition.

   Define the square root inverse matrix $V^{-\frac{1}{2}}$ as a matrix satisfying $V^{-\frac{1}{2}} V^{-\frac{1}{2}} = V^{-1}$.

   Scale the covariance matrix by pre and postmultiplying with $V^{-\frac{1}{2}}$.

   $$\begin{aligned} \Sigma^* &= V^{-\frac{1}{2}} \Sigma V^{-\frac{1}{2}} \\ &= V^{-\frac{1}{2}} (BE[ff']B' + V) V^{-\frac{1}{2}} \\ &= J \Lambda J + I_k \end{aligned}$$

   where J is the $k \times n$ matrix of the first k eigenvectors, and Λ contains eigenvalues squared along the diagonal.

3. Observable economic factors.

   Suppose we can observe economic shocks g which influence the comovements (f).

   g is equivalent to f if there is a nonrandom matrix L such that

   $$g = L^{-1} f$$

   If this is the case, choose the obvious rotation

   $$f^* = g$$

   and we can write the APT assumed relation

   $$r = E[r] + B^* g + \varepsilon$$

   Using observable economic variables has obvious utility in interpretation.

4. Factor mimicking portfolios.

   Any set of k well–diversified portfolios has de-meaned returns approximately equivalent to a rotation of the factors: Factor mimicking portfolios.

7 Approximate factor models

The assumption of a diagonal V is stricter than necessary. It is possible to find an APT-type relation as long as V satisfies certain restrictions.

Essentially, need the risk associated with the ε terms to be "diversifiable," but the risks associated with the f's to be "pervasive"

The condition is a limiting one, it has to hold as $n \to \infty$.

Example: Industry and sector model, correlations within sector, not correlation without: Block diagonal V.

8 Conditional versions

It is of course possible to assume conditional versions of the basic APT relation

$$r_t = E_{t-1}[r_t] + B_{t-1} f_t + \varepsilon_t$$

where B_t is the *conditional* projection

$$B_{t-1} = E_{t-1}[(r_t - E_{t-1}[r_t]) f_t' E_{t-1}[f_t f_t']]$$

and

$$\Sigma_{t-1} = B_{t-1} \Omega_{t-1} B_{t-1}' + V_{t-1}$$
$$\Omega_{t-1} = E_{t-1}[f_t f_t']$$
$$V_{t-1} = E_{t-1}[\varepsilon_t \varepsilon_t']$$

In this general form this is not identifiable, additional structure is necessary.

Example: GARCH

9 Derivation of the pricing restriction

Will not go through the full APT derivation, only consider the simplest possible case.

Suppose there is no idiosyncratic risk, and there is no arbitrage

$$r = E[r] + Bf$$

Claim: Expected returns satisfy

$$E[r] = \lambda_0 i^n + B\lambda$$

where
λ_0 is a constant,
i^n is a $n-$vector of ones,
λ is a $k-$vector of factor risk premia.

Proof

Perform the projection of $E[r]$ on i^n and B.

$$E[r] = \lambda_0 i^n + B\lambda + \eta$$

From the general properties of projection residuals we know

$$i^n \eta = 0$$

$$B\eta = 0$$

Think of η as a *portfolio* of assets.

The portfolio has zero cost since

$$i^n \eta = 0$$

By no arbitrage, the expected return on this portfolio has to equal zero

The portfolio return is found as

$$\eta' E[r] = \eta' \lambda_0 i^n + \eta' B\lambda + \eta'\eta = 0 + 0 + \eta'\eta$$

Since this has to equal zero (by no arbitrage),

$$0 = \eta E[r] = \eta'\eta$$

The only way this can equal zero is if $\eta \equiv 0$

Thus, we have shown that

$$E[r] = \lambda_0 i^n + B\lambda$$

QED

The general case(with idiosyncratic variance) is shown similarly.

APT can also be derived in a GE framework, and dynamic models has been proposed.

10 Empirical analysis of the APT

Testable implication of the APT

$$E[r] = i^n \lambda_0 + B\lambda$$

Rewrite this in a regression form by substituting for $E[r]$ from $r = E[r] + Bf + \varepsilon$:

$$E[r] = r - Bf - \varepsilon$$

Substitute, get

$$E[r] = r - Bf - \varepsilon \ = i^n \lambda_0 + B\lambda$$

Regression form

$$r = i^n \lambda_0 + B(\lambda + f) + \varepsilon$$

If we need to be explicit about time:

$$r_t = i^n \lambda_{0,t-1} + B(\lambda_{t-1} + f_t) + \varepsilon_t$$

The problems in estimation can be grouped into two

- Similar problems to CAPM tests (eg: errors in variables)
- Unique problems in estimating APT due to the factor structure. (eg: indeterminacy problem)

10.1 Relation to CAPM tests

Recall can split CAPM tests into two types:

1. Cross section (Like Fama and MacBeth (1973)).

   Conditionally on B, estimate λ. (Past data on returns, estimate B, use B to estimate λ (rolling regressions)

2. Time Series (Like Black, Jensen, and Scholes (1972)).

   Conditionally on $\lambda_{0,t-1}$ and $(\lambda_{t-1} + f_t)$, estimate B in time series.

In both these settings the same problems occur as in the CAPM tests.

10.2 Unique problems in testing APT

- Identifying factors

  Solutions

  - Factor Analysis Roll and Ross (1980)
  - Principal Components Connor and Korajczyk (1988)
  - Observable economic variables Chen et al. (1986)
  - Factor mimicking portfolios

- Determining the number of factors.

  - Factor analytic framework (likelihood ratio like tests)
  - Principal Components framework (properties of eigenvalues)

Will concentrate on two examples

- Factor analytic framework of Roll and Ross (1980)
- Observable economic variables Chen et al. (1986)

10.3 Factor analysis

Let us return to the expression for the asset covariance matrix.

$$\Sigma = BE[ff']B' + V$$

In the discussion of "rotations," we saw that one possibility was to impose

$$E[ff'] = I$$

as a restriction

With this restriction we can rewrite the covariance matrix as

$$\Sigma = BB' + V$$

where, with a strict factor model, V is diagonal.

From time series of returns, can find the sample estimate of Σ, $\hat{\Sigma}$.

If we assume multivariate normality of f and ε, B and V can be estimated using Maximum Likelihood.

See Roll and Ross (1980) or Rao (1996)

Thus, from the factor analytic approach, estimates of B "pops out" almost as if by magic, even if we only observed the returns.

However, the factor analytic approach does not allow us to *identify* the "factors" f.

The ML framework of factor analysis included a likelihood ratio test for the sufficiency of k factors.

This type of analysis was used by Roll and Ross to get sample estimates of B.

10.4 Observable economic variables

An alternative way to get sample estimates of B was performed in Chen et al. (1986). Given observable variables g, factor sensitivities B can be estimated by a simple regression.

$$r_t = Bg + e$$

giving estimates of B, $\hat{B}$.

11 The cross sectional setup

Given the $n \times k$ matrix $\hat{B}$, the APT implies the following regression.

$$r_t = i^n F_{0,t-1} + \hat{B}F_t + u_t$$

With a large crossection of assets at time t, the sample parameters $\hat{F}_{o,t-1}$ and $\hat{F}_t$ should satisfy

$$\hat{F}_{o,t-1} \to \lambda_{0,t-1}$$

$$\hat{F}_t \to \lambda_{t-1} + f_t$$

If we repeat this procedure for a large number of dates t, and take averages of the estimated coefficients,

$$\bar{F}_t \to \bar{\lambda}$$

the unconditional "factor loading," since $E[f] = 0$

If the risk premia are stationary, $\bar{\lambda}$ estimates the true risk premium λ.

We do have the usual errors in variables problem. The matrix B is estimated (just like β in Fama and MacBeth (1973)) The usual approach to reducing the EIV problem is to group assets into portfolios.

This should give enough background to see what RR and CRR test.

They investigate properties of λ_0 and λ.

12 Empirical implementation: Factor Analysis

12.1 Some notes on "Factor analysis."

Let us first look at how we go about estimating the "factors."

We first mention that in the formulation of the APT above, the factors are not uniquely determined. To be specific:

Suppose $E[\mathbf{r}] = \mathbf{B}'\mathbf{f}$ holds. Just by definition, we can always find a matrix $\mathbf{G}$ such that

$$\mathbf{G}\mathbf{G}' = \mathbf{I},$$

which means that

$$\begin{aligned} E[\mathbf{r}] &= \mathbf{B}'\mathbf{f} \\ &= \mathbf{B}'(\mathbf{G}\mathbf{G}')\mathbf{f} \\ &= (\mathbf{B}'\mathbf{G})(\mathbf{G}'\mathbf{f}) \end{aligned}$$

We can therefore always scale $\mathbf{B}$ and $\mathbf{f}$ up or down by replacing them with $\mathbf{BG}$ and $\mathbf{G}'\mathbf{f}$. This fact is used in the factor estimation. This is often termed that the APT is invariant to factor rotations.

Consider the APT relation written as

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{f}_t + \mathbf{e}_t$$

We assume that

$$E[\mathbf{e}_t] = \mathbf{0}$$

$$E[\mathbf{f}_t] = \mathbf{0}$$

$$E[\mathbf{e}_t\mathbf{f}_t] = \mathbf{0}$$

We also assume a diagonal covariance matrix for the errors:

$$E[\mathbf{e}_t\mathbf{e}_t] = \boldsymbol{\Psi}$$

The method we use to do estimation of the factors is called "factor analysis." Factor analysis is a maximum likelihood based method. It assumes multivariate normality. It uses the covariance matrix in the estimation. Let us therefore start by considering the covariance matrix

$$\begin{aligned} V &= E[(\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})'] \\ &= E[(\mathbf{B}\mathbf{f}_t + \mathbf{e}_t)(\mathbf{B}\mathbf{f}_t + \mathbf{e}_t)'] \\ &= E[\mathbf{B}\mathbf{f}_t\mathbf{f}_t'\mathbf{B}'] + E[\mathbf{e}_t\mathbf{e}_t'] \\ &= \mathbf{B}E[\mathbf{f}_t\mathbf{f}_t']\mathbf{B}' + E[\mathbf{e}_t\mathbf{e}_t'] \\ &= \mathbf{B}\boldsymbol{\Lambda}\mathbf{B}' + \boldsymbol{\Psi} \end{aligned}$$

We define $E[\mathbf{f}_t\mathbf{f}_t'] = \boldsymbol{\Lambda}$. This is where the indeterminacy of the factors enter. Since we can replace $\mathbf{B}$ with $\mathbf{BG}$ and $\mathbf{f}$ with $\mathbf{Gf}$, this means that

$$\mathbf{B}\boldsymbol{\Lambda}\mathbf{B}' = (\mathbf{B}\mathbf{G}')(\mathbf{G}\boldsymbol{\Lambda}\mathbf{G}')(\mathbf{G}\mathbf{B}')$$

for any $\mathbf{G}$ such that $\mathbf{G}\mathbf{G}' = \mathbf{I}$. What we do is to pick $\mathbf{G}$ such that $\mathbf{G}\boldsymbol{\Lambda}\mathbf{G}'$ is an identity matrix. By the existence of $\mathbf{G}$ we are fully justified in imposing $\boldsymbol{\Lambda} = \mathbf{I}$, and we can write the covariance matrix as

$$\mathbf{V} = \mathbf{B}\mathbf{B}' + \boldsymbol{\Psi}$$

The estimation of the factors relies on this restriction. For a given number of factors (and hence a given dimension of $\mathbf{B}$), estimation consist of forming a likelihood function with the restricted $\mathbf{V} = \mathbf{BB}' + \mathbf{\Psi}$ as the covariance matrix, and maximises this with respect to the parameters $\mathbf{B}$ and $\mathbf{\Psi}$.[1] The values at which the likelihood function is maximised gives the estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{P}}$.

The final problem in factor analysis is the estimation of the number of factors present. This is done by redoing the analysis above with a different number of factors, and comparing the values of the likelihood functions for different number of factors.

12.2 The analysis in the Roll and Ross (1980) paper

Let us now look at the Roll and Ross (1980) paper. This was the first paper that took the APT to the data. The first problem they had to solve was the estimation of the "factors." This they did by the method of factor analysis shown above.

Now, how can we go about testing the APT given estimated coefficients $\widehat{\mathbf{B}}$?

If the estimated B is true, we know that this relation holds:

$$E[r_i] = \lambda_0 + \lambda_1 b_{i1} + \cdots + \lambda_k b_{ik}$$

where

$$E[\lambda_0] = \mu_i$$
$$E[\lambda_i] = 0 \quad \text{for} \quad i = 1, \cdots, n$$

The obvious thing to do is then to replace b_{ij} with estimates $\hat{b}_{ij}$, run a cross-sectional regression on

$$r_{it} = \lambda_{0t} + \lambda_{1t}\hat{b}_{i1} + \cdots + \lambda_{kt}\hat{b}_{ik} + \epsilon_{it}$$

for the n available assets in cross-section, get estimates $\hat{\lambda}_{it}, i = 0, \cdots, k$ and test whether

$$\mathcal{H}_0: \begin{array}{lcl} \lambda_0 & = & \mu_i \\ \lambda_j & = & 0, \quad j = 1, \cdots, k \end{array}$$

This is what RR tests. They use the usual "rolling regression" method, by estimating $\widehat{\mathbf{B}}_t$ using 5 years worth of data previous to time t, ie looking at $t-61, \cdots, t-1$ to find $\widehat{\mathbf{B}}_t$, and then looking at the cross-section at time t, estimating $\widehat{\boldsymbol{\lambda}}_t$. This is redone for all the periods we have data, and we get a time series of observations $\widehat{\boldsymbol{\lambda}}_t, t = 61, \cdots, T$ which we can use to test the hypotheses above.

There are some econometric corrections, but that is the basic idea.

Data: Portfolios of individual NYSE stocks, randomly picked.

Main results: Seem to be five factors in the data.

Test alternative: Own variance. Not conclusive evidence whether the own variance adds explanatory power.

Final problem: Different factors across groups. For computational reasons, the tests are done on a large number of "groups" of stocks. Problem with the tests: We do not know whether the estimated factors are identical *and* in the same order across groups. The test in the paper can not reject the equality, but there is a controversy about this, see Dhrymes, Friend, and Gultekin (1984), Roll and Ross (1984) and Trzcinka (1986).

[1] The method does not take the max of the normal distribution likelihood function. It can be shown (See Joereskog, 1967 and Joereskog and Goldberger, 1973), that an equivalent method is to take the max of $G = \frac{1}{2}\text{tr}(\mathbf{I} - \mathbf{S}^{-1}\mathbf{\Sigma})$, where $\mathbf{S}$ is the sample covariance matrix, $\mathbf{\Sigma} = \mathbf{BB}' - \mathbf{\Psi}$ and $\text{tr}(\cdot)$ is the trace of the matrix. This optimisation problem is much easier to solve.

13 Implementation: Using prespecified data as factors.

The paper by Chen et al. (1986) takes a very different approach. It can be viewed as a "fishing expedition." The paper specifies a number of macroeconomic variables and sees if these are useful in predicting asset returns.

As such the tests are divorced from the APT in the sense that:

- We can reject the ability of these variables to explain stock returns, but the APT may still be true.
- We may accept these variables as significant, but they may not be the "true" factors, the variables may instead also be influenced by the "true" factors.

This is still an useful exercise in the sense of motivating use of the APT-type valuation, with more than one source of risk explaining returns, but we should be aware that it is not explicitly related to the formal APT.

The motivation of the paper is at a very intuitive level, divorced from any formal theory:

We think about how macroeconomic variables influence the economy's pricing operator: Consider the simplest possible model of stock prices, the constant expected dividend relation:

$$P_{0i} = \frac{E[d_{1i}]}{r_i},$$

where d_1 is the dividend in period 1, and r_i the interest rate for stock i.

Any systematic economic variables that influence d_1 or r_i may change stock prices and hence returns. As examples is mentioned:

r_i: the firms interest rate can be influenced by

- Riskless Rate.
- Risk Premia.
- Marginal utility of wealth.

$d_{i,1}$: The expected dividends can be influenced by

- Inflation.
- Production.

This is then viewed as justification for considering the following factors as explanatory variables:

1. `MP`: Monthly growth in industrial production.
2. `UI`: Unexpected inflation.
3. `DEI`: Change in expected inflation.
4. `UPR`: Risk premium, difference low-grade bonds and risk free rate.
5. `UTS`: Term structure.

Consider the regression:

$$R_t = a + b_{mp}\mathtt{MP}_t + b_{dei}\mathtt{DEI}_t + b_{ui}\mathtt{UI}_t + b_{upr}\mathtt{UPR}_t + b_{uts}\mathtt{UTS}_t + e_t$$

Econometric Question: We are interested in testing whether the coefficients above are significant in explaining stock returns.

As stock market returns use two broad indices:

`EWNY` Equally weighted index of NYSE stocks.

VWNY Value weighted index of NYSE stocks.

Econometric Implementation: Fama and MacBeth (1973) type analysis. (rolling regressions).

- At time t
- Run the above regression over previous 5 years. Gives estimates of coefficients a,$b_{mp} \cdots b_{uts}$ above.
- Run a cross-sectional regression on these 6 variables. Gives an estimate of risk premia for each of the 5 "factors,"
- Redo for time $t+1$.

At the end we get a time series of estimated risk premia. Do a t-test of these to see if it is significantly different from zero.

Results:

- Significant:

  MP Production.

  UI Unexpected inflation.

  UPR Risk premium.

- Marginally significant:

  UTS : Term structure.

Test specific alternatives:

- Stock market. Does the stock market add explanatary power to the five variables we consider? Test by adding EWNY or VWNY to the explanatory variables, not significantly added explanatory power.
- Consumption. Add monthly consumption series CG, the consumption growth. Not significant.
- Oil Prices. (OG) Not significant.

Econometric problem using t-test in the "rolling regressions:" We have the usual "errors in variables" problem, the t-tests should be adjusted. See Shanken (1992).

14 Fitting APT in with standard "pricing operator" framework

How does the analysis of the APT find into our general "pricing operator" framework? The basic assumption of a factor structure used in deriving the APT does not fit into our pricing operator framework, but using the expression

$$E_t[m_{t+1}R_{t+1}]$$

allows us to characterize/give intuition about what the APT says.

Suppose the factor model

$$R_{i,t+1} = a_{it} + \sum_{j=1}^{K} b_{ij,t} F_{j,t+1} + u_{i,t+1}$$

describes asset returns.

Suppose that the idiosyncratic part of asset returns is uncorrelated with m_{t+1}, and consider the basic pricing operator

$$E_t[m_{t+1}R_{t+1}] = 1$$

If we let $R_{i,t+1}$ be any return, and $\lambda_{0,t+1}$ be a zero beta return relative to m_{t+1},

$$E[m_{t+1}(R_{it+1} - \lambda_{0,t+1})] = 0$$

$$E[m_{t+1}]\lambda_{0,t+1} = E[m_{t+1}R_{it+1}] = \text{cov}(m_{t+1}, R_{i,t+1}) + E[m_{t+1}]E[R_{it+1}]$$

$$\lambda_{0,t+1} = \frac{\text{cov}(m_{t+1}, R_{i,t+1})}{E[m_{t+1}]} + E[R_{it+1}]$$

$$E[R_{it+1}] = \lambda_{0,t+1} + \frac{\text{cov}(-m_{t+1}, R_{i,t+1})}{E[m_{t+1}]}$$

Substitute the assumed factor structure for R_{it+1} on the RHS:

$$E[R_{it+1}] = \lambda_{0,t+1} + \frac{\text{cov}(-m_{t+1}, R_{i,t+1})}{E[m_{t+1}]}$$

$$= \lambda_{0,t+1} + \frac{\text{cov}(-m_{t+1}, a_{it} + \sum b_{ijt}F_{jt+1} + u_{it+1})}{E[m_{t+1}]}$$

$$= \lambda_{0,t+1} + \frac{\text{cov}(-m_{t+1}, \sum b_{ijt}F_{jt+1})}{E[m_{t+1}]}$$

$$= \lambda_{0,t+1} + \sum b_{ijt}\frac{\text{cov}(-m_{t+1}, F_{jt+1})}{E[m_{t+1}]}$$

This relation allows us to interpret the factors:

A factor which covary positively with the IMRS m_{t+1} has a negative risk premium, whereas one that covary negatively with m has a positive risk premium.

15 Further reading.

For an overview of the econometric problems in "double-pass" estimation of both CAPM and APT, see Shanken (1992).

Some notes on the identification of the number of factors are in Brown (1989).

For a comparison of Principal Components and Factor Analysis, see Shuklar and Trzcinka (1990).

16 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.

A Appendix: Factor Analysis

B Factor Analysis.

As an example of a more complicated Maximum Likelihod problem, wee look at what is termed “factor analysis”

We consider a relation

$$\mathbf{r}_t = \mathbf{B}\mathbf{f}_t + \mathbf{e}_t$$

We assume that

$$E[\mathbf{e}_t] = \mathbf{0}$$

$$E[\mathbf{f}_t] = \mathbf{0}$$

$$E[\mathbf{e}_t\mathbf{f}_t] = \mathbf{0}$$

We also assume a diagonal covariance matrix for the errors:

$$E[\mathbf{ee}] = \mathbf{\Psi}$$

The method we use to do estimation of the factors is called “factor analysis.”

Let us first look at how we go about estimating the “factors.”

We first mention that the factors are not uniquely determined. To be specific:

Suppose $E[\mathbf{r}] = \mathbf{B}'\mathbf{f}$ holds. Just by definition, we can always find a matrix $\mathbf{G}$ such that

$$\mathbf{GG}' = \mathbf{I},$$

which means that

$$\begin{aligned} E[\mathbf{r}] &= \mathbf{B}'\mathbf{f} \\ &= \mathbf{B}'(\mathbf{GG}')\mathbf{f} \\ &= (\mathbf{B}'\mathbf{G})(\mathbf{G}'\mathbf{f}) \end{aligned}$$

We can therefore always scale $\mathbf{B}$ and $\mathbf{f}$ up or down by replacing them with $\mathbf{BG}$ and $\mathbf{G}'\mathbf{f}$. This fact is used in the factor estimation. This is often termed the invariance to “factor rotations.”

Factor analysis is a maximum likelihood based method. It assumes multivariate normality. It uses the covariance matrix in the estimation. Let us therefore start by considering the covariance matrix

$$\begin{aligned} V &= E[\mathbf{rr}'] \\ &= E[(\mathbf{Bf}+\mathbf{e})(\mathbf{Bf}+\mathbf{e})'] \\ &= E[\mathbf{Bff}'\mathbf{B}'] + E[\mathbf{ee}'] \\ &= \mathbf{B}E[\mathbf{ff}']\mathbf{B}' + E[\mathbf{ee}'] \\ &= \mathbf{B\Lambda B}' + \mathbf{\Psi} \end{aligned}$$

We define $E[\mathbf{ff}'] = \mathbf{\Lambda}$. This is where the indeterminacy of the factors enter. Since we can replace $\mathbf{B}$ with $\mathbf{BG}$ and $\mathbf{f}$ with $\mathbf{Gf}$, this means that

$$\mathbf{B\Lambda B}' = (\mathbf{BG}')(\mathbf{G\Lambda G}')(\mathbf{GB}')$$

for any $\mathbf{G}$ such that $\mathbf{GG}' = \mathbf{I}$. What we do is to pick $\mathbf{G}$ such that $\mathbf{G\Lambda G}'$ is an identity matrix. By the existence of $\mathbf{G}$ we are fully justified in imposing $\mathbf{\Lambda} = \mathbf{I}$, and we can write the covariance matrix as

$$\mathbf{V} = \mathbf{BB}' + \mathbf{\Psi}$$

The estimation of the factors relies on this restriction. For a given number of factors (and hence a given dimension of $\mathbf{B}$), estimation consist of forming a likelihood function with the restricted $\mathbf{V} = \mathbf{BB'} + \mathbf{\Psi}$ as the covariance matrix, and maximises this with respect to the parameters $\mathbf{B}$ and $\mathbf{\Psi}$.[2] The values at which the likelihood function is maximised gives the estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{P}}$.

The final problem in factor analysis is the estimation of the number of factors present. This is done by re-doing the analysis above with a different number of factors, and comparing the values of the likelihood functions for different number of factors.

References

Fisher Black, Michael Jensen, and Myron Scholes. The capital asset pricing model, some empirical tests. In Michael C Jensen, editor, *Studies in the theory of capital markets*. Preager, 1972.

Stephen J Brown. The number of factors in equity returns. *Journal of Finance*, 44(5):1247–62, 1989.

Gary Chamberlain and Michael Rothschild. Arbitrage and mean variance analysis on large asset markets. *Econometrica*, 51: 1281–1304, 1983.

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

Gregory Connor. A unified beta pricing theory. *Journal of Economic Theory*, 34:13–31, 1984.

Gregory Connor. Notes on the arbitrage pricing theory. In Sudipto Bhattacharya and George M Constantinides, editors, *Theory of valuation*, pages 298–309. Rowman and Littlefield, 1989.

Gregory Connor and Robert A Korajczyk. A test for the number of factors in an approcimate factor model. Working Paper 51, Northwestern University, June 1988.

Phoebus J Dhrymes, Irwin Friend, and N Bulent Gultekin. A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory. *Journal of Finance*, 39:323–46, June 1984.

Eugene F Fama and J MacBeth. Risk, return and equilibrium, empirical tests. *Journal of Political Economy*, 81:607–636, 1973.

Chi-fu Huang and Robert H Litzenberger. *Foundations for financial economics*. North–Holland, 1988.

Jonathan Ingersoll. Some results in the theory of Arbitrage Pricing. *Journal of Finance*, 39:1021–39, 1984.

Jonathan Ingersoll. *Theory of financial decision making*. Rowman, 1987.

Karl G Joereskog. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4):443–82, 1967.

Karl G Joereskog and Arthur S Goldberger. Factor analysis by generalized least squares. *Psychometrika*, 37(3):243–60, 1972.

C Radhakrishna Rao. Principal components and factor analysis. In G S Maddala and C R Rao, editors, *Handbook of Statistics*, volume 14 Statistical Methods in Finance, chapter 16, pages 489–505. North–Holland, 1996.

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

Richard Roll and Stephen A Ross. A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory: A reply. *Journal of Finance*, 39:1073–1103, June 1984.

Stephen A Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976.

Jay Shanken. On the estimation of beta-pricing models. *Review of Financial Studies*, 5(1):1–34, 1992.

Ravi Shuklar and Charles Trzcinka. Sequential tests of the Arbitrage Pricing Theory. A comparison of principal components and maximum likelihood factors. *Journal of Finance*, 45:1541–64, 1990.

Charles Trzcinka. On the number of factors in the arbitrage pricing model. *Journal of Finance*, 41:347–367, 1986.

[2] In practice, the factor analytic method does not take the max of the normal distribution likelihood function. It can be shown (See Joereskog, 1967 and Joereskog and Goldberger, 1973), that an equivalent method is to take the max of $G = \frac{1}{2}\text{tr}(\mathbf{I} - \mathbf{S^{-1}\Sigma})$, where $\mathbf{S}$ is the sample covariance matrix, $\mathbf{\Sigma} = \mathbf{BB'} - \mathbf{\Psi}$ and $\text{tr}(\cdot)$ is the trace of the matrix. This optimisation problem is much easier to solve.