Ch. 6. Multifactor Pricing Models

- Empirical evidence indicates that CAPM beta does not explain the cross-section of expected returns.
  - Momentum
  - Anomalies (January, size, book/market effects, for example)
- What to do?
  - Dynamic, equilibrium models or multi–factor models?
  - This chapter, multi–factor approach.

How to Proceed?

Two approaches.

1. *Ad–hoc* theorizing.
2. Curve fitting.

What is the difference in the two approaches?

Ad–hoc Theorizing

The factors are specified in advance, but not on the basis of some economic model. The "theoretical" procedure is as follows:

- Run the CAPM and find some anomalies.
- Incorporate "factors" that explain those anomalies into an empirical asset pricing model.
- Re–run the model and see if it now explains the cross–section of expected returns.
- Fama French 3 Factor pricing model is an example of this.

Chen, Roll, and Ross (1985)

This is a slightly different example of ad–hoc theorizing .

- Authors argue that in selecting factors, consider
  1. factors that influence future discount rates
  2. factors that influence future cash flows
- The "model" is based on intuitive analysis and empirical investigation, and consists of five factors.

The factors

1. yield spread between long and short interest rates
2. expected inflation
3. unexpected inflation (past realizations)
4. industrial production growth
5. yield spread between high– and low–grade corporate bonds.

Diagnostics are performed to determine whether consumption growth or oil prices add any explanatory power.

Problems with *ad–hoc* theorizing

- Not completely data–driven, but close.
- Not clear how this advances economic theory, although it has the potential to do so.
- How would one use this model to measure performance? You don't get points for increasing portfolio returns by taking advantage of known anomalies.

Curve Fitting

Rather than being only partly empirical, this approach is completely empirical.

- No attempt is made to identify the "factors" in advance
- The plan is to identify the factors "in sample" and then use the factors out of sample to explain the cross–section of asset returns.
- This is done using factor analysis.

Problems with identifying the factors

- number of factors?
- what do the factors represent?
- is there any way to tell what the factors really represent?
- what procedure (principal components or factor analysis) should be used to get the factors?
- does the procedure affect the explanatory power?

Our goal

1. Present a brief summary of the APT
2. Discuss how the unknown factors of the APT might be identified using principal component analysis
3. We will not discuss the differences between factor analysis and principal component analysis. You should learn this on your own.

Next, we turn to the classical specification of the APT.

The APT

- Multiple (unknown) factors.
- Identification of the market portfolio is not required.
- Not clear how many factors, however.
- The number of factors can either be specified *ex–ante*, or determined using a test *ex–post*.

Assumptions

- Competitive markets.
- The return generating process for asset returns follows a factor structure:

$$
\begin{aligned}
R_i &= a_i + \mathbf{b_i}'\mathbf{f} + \epsilon_i && (1)\\
E[\epsilon_i|\mathbf{f}] &= 0 && (2)\\
E[\epsilon_i^2|\mathbf{f}] &= \sigma_i^2 \leq \sigma^2 < \infty && (3)
\end{aligned}
$$

where

Definitions

- R_i is the return on asset i
- a_i is the intercept of the factor model
- $\mathbf{b_i}$ is a Kx1 vector of factor sensitivities for asset i.
- ϵ_i is a disturbance term

The System

$$\mathbf{R} = \mathbf{a} + \mathbf{Bf} + \epsilon \quad (4)$$
$$E[\epsilon|\mathbf{f}] = 0 \quad (5)$$
$$E[\epsilon\epsilon'|\mathbf{f}] = \Sigma \quad (6)$$

where

Definitions

- **R** is Nx1 vector of asset returns
- **a** is Nx1 vector of intercepts
- $\mathbf{B} = [b_1, ...b_N]'$ is a NxK matrix of factor sensitivities
- ϵ is Nx1

Further Assumptions

- The disturbance term for a large, well–diversified portfolio vanishes.
- Such a portfolio has a large number of stocks with portfolio weights of (approximately) $1/N$.

Results

Given the structure, Ross (1976) shows that the absence of arbitrage in large economies implies that

$$\mu \approx \mathbf{1}\lambda_0 + \mathbf{B}\lambda_K \tag{7}$$

- μ is Nx1 vector of expected returns
- λ_0 is the model zero–beta parameter, and equals the risk–free rate if such a rate exists
- λ_K is a Kx1 vector of factor risk–premia
- $\mathbf{B} = [b_1, ...b_N]'$ is, again, a NxK matrix of factor sensitivities

Why approximate?

- This relationship holds exactly only as the number of assets $\rightarrow$ infinity.
- In finite samples, the assets can be arbitrarily mispriced.
- This is what we mean when we say that we cannot really test the model without imposing additional restrictions.
- There would be no way to reject it.

Approximate vs. exact factor structures

- Won't discuss distinction here.
- The difference is that one holds in finite samples, and the other holds only as the number of assets becomes infinite.

Relevant papers

- For an easy derivation of the APT, see Huberman (1982)
- For an equilibrium derivation of the APT that features an exact factor structure, see Connor (1984)
- For a discussion of the difference between an exact and an approximate factor structure, see Chanberlain and Rothschild (1983)
- For a discussion of deviations from exact factor pricing in an approximate factor environment, see Dybvig (1985), and Grinblatt and Titman (1985)
- See chapter 8 to learn how the exact factor pricing emerges in an ICAPM framework.

Estimation and Testing

The procedure is identical to that in chapter 5.

- Again, start with asset returns conditional on factor realizations that are cross–sectionally multivariate normal and IID through time.
- This a strong assumption, and it amounts to the errors in the regression being multivariate normal.
- There can still be some serial dependence through the factors, although we will defer our exploration of factor selection until later, section 6.4.

Four cases

1. factors are portfolios of traded assets and a risk–free asset exists.
2. factors are portfolios of traded assets and a risk–free asset does not exist.
3. factors are not portfolios of traded assets.
4. factors are portfolios of traded assets and the factor portfolios span the M-V frontier of risky assets. We won't do this one.

The text explores ML estimation. GMM estimation is also possible.

Corrected LR Test Statistic

$$J = -\left(T - \frac{N}{2} - K - 1\right)\left[log|\hat{\Sigma}| - log|\hat{\Sigma}^*|\right] \tag{8}$$

- The starred estimate of the residual covariance matrix is the constrained model.
- This statistic $\sim \chi^2(R)$ where R is the number of restrictions imposed.

Portfolios as factors with a risk–free asset

Write in excess return form.

- $\mathbf{Z_t}$ is Nx1 vector of excess returns for the N assets.

Then a K-factor model is:

$$\mathbf{Z_t} = \mathbf{a} + \mathbf{B}\mathbf{Z_{Kt}} + \epsilon_t \quad (9)$$

$$E[\epsilon_t] = 0 \quad (10)$$

$$E[\epsilon_t \epsilon_t'] = \Sigma \quad (11)$$

$$E[\mathbf{Z_{Kt}}] = \mu_K \quad (12)$$

$$E[(\mathbf{Z_{Kt}} - \mu_K)(\mathbf{Z_{Kt}} - \mu_K)'] = \Omega_K \quad (13)$$

$$Cov[\mathbf{Z_{Kt}}, \epsilon_t'] = \mathbf{0}_{K\times N} \quad (14)$$

Our test: $\mathbf{a} = \mathbf{0}$.

Solution Procedure

- Follows directly from chapter 5
- $$J_1 = \frac{T-N-K}{N}\left[1 + \hat{\mu}_K' \hat{\Omega}_K^{-1} \hat{\mu}_K\right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \quad (15)$$

  where

  $$\hat{\Omega}_K^{-1} = \frac{1}{T}\sum_{t=1}^{T} (\mathbf{Z_{Kt}} - \mu_K)(\mathbf{Z_{Kt}} - \mu_K)'$$

  and

  $$J_1 \sim F_{N,T-N-K}$$

- This test is better than the asymptotic test under multivariate normality. Without multivariate normality, use GMM instead.

No risk–free asset

Identical to the tests of the Black CAPM described in section 5.

Macroeconomic Variables as Factors

See text Section 6.2.3.

Finding the Factors

- The previous analysis assumed that the factors were known.
- But suppose one takes a purely data–driven view.
- In that case, we want to extract the factors from the cross–section of asset returns
- The techniques for doing this are
  1. principal components
  2. factor analysis

Factor analysis vs. principal components

- The text discusses different procedures for finding factors from the variance–covariance matrix of returns.
- We will not discuss these here.
- Instead, we will turn to a principal components procedure to estimate the factors.
- The procedure we will use is equivalent to that in Connor and Korajczyk (1986, 1988).
- Once we have completed the discussion here, I encourage you strongly to read those papers.

Principal Components

- Suppose the random variables $\{r_1, ...r_N\}$ have a certain multivariate distribution with mean vector $\mathbf{u}$ and covariance matrix Σ.
- What we want to do is describe each one of these N variables by a small number of "other" variables (which will be linear combinations of the N variables), with a high degree of accuracy.
- This would be trivially true if all variables moved proportionally. Then, a single variable would suffice to describe the behavior of all of the others.

- In this context, the N variables are the asset returns.
- Suppose we have a sample of T observations on our asset returns.
- We can write this as a $T \times N$ data matrix, X

$$X = \begin{bmatrix} r_{11} & ... & r_{1N} \\ ... & ... & ... \\ r_{T1} & ... & r_{TN} \end{bmatrix}$$

Example with the one variable

- This single variable takes on T values, to be arranged in a column vector, $\mathbf{p}$.
- We don't know $\mathbf{p}$ yet, but let's proceed as if we did.
- If all variables moved proportionally, then each column of X equals some scalar multiple of $\mathbf{p}$

$$\mathbf{r_i} = a_i \mathbf{p}$$

where each column vector is $T \times 1$.

- Therefore,

$$\mathbf{X}_{T\times N} = \mathbf{p_{T\times 1}} \mathbf{a'_{1\times N}}$$

or

$$\begin{bmatrix} p_1 \\ \vdots \\ p_T \end{bmatrix} [a_1, ...a_N] = \begin{bmatrix} p_1 a_1 & ... & p_1 a_N \\ ... & ... & ... \\ p_T a_1 & ... & p_T a_N \end{bmatrix}$$

- $\mathbf{a}'$ is an N–element row vector of those scalar multiples. The factor realization, $\mathbf{p}$ is the only thing that changes over time. The "betas" are the same each period in this model.

Normalization

- The product $\mathbf{pa}'$ remains unchanged when $\mathbf{p}$ is multiplied by a constant, c, as long as $\mathbf{a}$ is multiplied by $1/c$.
- So, we impose $\mathbf{p}'\mathbf{p} = \sum_{t=1}^{T} p_t^2 = 1$, and we can find an unique factor (except for sign — $\mathbf{p}$ and $\mathbf{a}$ can be replaced by their negatives).
- Of course, in general, $\mathbf{X} = \mathbf{pa}'$ will not hold exactly. There will be a non–zero matrix of discrepancies:

$$\mathbf{X} - \mathbf{pa}'$$

for whatever vectors $\mathbf{p}$ and $\mathbf{a}$.

Our Job

- Select those vectors **p** and **a** to minimize the sum of squares of all of the NT discrepancies.
- One can verify that the sum of squares of elements a_{ij} of an $m \times n$ matrix A can be written as the trace of $A'A_{n\times n}$ (the sum of its diagonal elements):

$$trace(A'A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$$

Therefore, our objective is to minimize

$$tr[(X - pa')'(X - pa)]$$

$$\begin{aligned} &= tr(X'X) - tr(ap'X) - tr(X'pa') + tr(ap'pa') \\ &= tr(X'X) - 2p'xa + a'a \end{aligned}$$

where use is made of:

$$tr(ap'x) = tr(p'xa) = p'xa$$

and

$$tr(ap'pa') = tr(p'pa'a) = p'pa'a = a'a$$

Our Minimization problem

We are behaving as if we know p.

What we want to do is minimize SSR by choosing our coefficient vector, a.

$$Min_a \; tr(X'X) - 2p'Xa + a'a \tag{16}$$

Solution

- Differentiate with respect to a for a given vector p and set equal to zero.

  $$-2X'p + 2a = 0$$

  so that

  $$a_{N\times 1} = X'_{N\times T} p_{T\times 1}$$

- Now, substitute into our optimand, equation (**??**):

$$\begin{aligned} &= \quad tr(X'X) - 2p'XX'p + p'XX'p \\ &= \quad tr(X'X) - p'XX'p \end{aligned}$$

Solution - continued

- Now that we have substituted in for a, our problem is in terms of observables, X, and factor, p.
- So now, we'll pick the factor to minimize SSR.
- Only the last part is a function of p, so our optimization problem becomes
$$Max_p p'XX'p$$
subject to $p'p = 1$.

Solution — continued

The Lagrangian is:

$$L = p'XX'p - \lambda(p'p - 1)$$

So that

$$\frac{d}{dp} = 0 = 2XX'p - 2\lambda p = 0$$

or

$$(XX'_{T\times T} - \lambda I_T)p_{T\times 1} = \mathbf{0} \tag{17}$$

Hence, p is a **characteristic vector** of the $T \times T$ positive semi–definite matrix XX' corresponding to root λ.

Put another way, λ is an eigenvalue, and p is the associated eigenvector.

Which eigenvalue?

To find out which root is to be taken (which eigenvalue), premultiply (**??**) by p':

$$\rightarrow p'XX'p = \lambda p'Ip = \lambda$$

Because we want to maximize $p'XX'p$, we should take the largest root of XX' - we need to find the eigenvector p associated with the largest eigenvalue, λ.

Digression on Eigenvalues and Eigenvectors

- Consider a matrix A, a vector p and a scalar λ such that

$$A_{T\times T}p_{T\times 1} = \lambda p$$

- This is the fundamental equation for the scalar eigenvalue, or characteristic root, λ; and the eigenvector, or characteristic vector, p.
- Notice that the equation is non–linear, since it contains the product of two unknowns, λ, and p.
- But if we could discover λ, then the equation for p alone becomes linear.
- To that end, rewrite the fundamental equation as

$$(A - \lambda I)_{T\times T}p_{T\times 1} = 0_{T\times 1}$$

- $$A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1T} \\ a_{21} & a_{22} - \lambda & \ldots & a_{2T} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & a_{TT} - \lambda \end{bmatrix}$$

  That is, stick λ down the diagonal.

- Therefore, p lies in the nullspace of (is orthogonal to) $A - \lambda I$.
- Key point: For all λ, $p = 0$ satisfies this equation. But the zero vector is useless here. We are interested in solutions for the principal components that are non–zero. That is, the nullspace of $A - \lambda I$ must contain a non–zero vector.
- It turns out that the number λ will be an eigenvalue of A with

corresponding non–zero eigenvector if and only if

$$|A - \lambda I| = 0$$

This is the characteristic equation for the matrix A.

Example

- $$A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$$

- The characteristic equation is

$$0 = det \begin{bmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{bmatrix}$$

or

$$\lambda^2 - \lambda - 2 = 0$$

- Solving for the two distinct eigenvalues: $\lambda = \{-1, 2\}$

- Now, plug in the largest eigenvalue, 2 and multiply by the p vector:

$$(A - \lambda I)p = \begin{bmatrix} 2 & -5 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- The solution for p is any multiple of

$$p = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

- We could do the same thing for $\lambda = -1$. As you can see, the eigenvectors are not uniquely determined. In our principal components problem, we've placed more structure on the solution:

$$p'p = 1$$

- Here,

$$p'p = [5y \;\; 2y]\begin{bmatrix} 5y \\ 2y \end{bmatrix}$$

where we are looking for the y such that $p'p = 1$. This means that

$$25y^2 + 4y^2 = 1$$

$$y = \left(\frac{1}{29}\right)^{1/2}$$

- Hence, our first principal component in our example is

$$p = \begin{bmatrix} 5\left(\frac{1}{29}\right)^{1/2} \\ 2\left(\frac{1}{29}\right)^{1/2} \end{bmatrix}$$

Back to our problem

Next, let's interpret what our coefficient vector, a, means

- Recall our first–order condition again

$$(XX'_{T\times T} - \lambda I_T)p_{T\times 1} = 0$$

Premultiplying this time by X' yields

$$X'(XX'_{T\times T} - \lambda I_T)p_{T\times 1} = 0$$

which implies that

$$(X'X - \lambda I_T)X'p_{T\times 1} = (X'X - \lambda I)a = 0$$

Interpretation of a

- **Hence, the coefficient vector a is a characteristic vector of $X'X$ corresponding to the largest root except that it's not normalized to have unit length**
- If you look at the text, a is what is referred to as the factor sensitivity matrix, B, with a one factor model.
- p was a characteristic vector of XX', a $T \times T$ matrix, and was normalized to have unit length.
- Next, let's examine the relationship between p and a.

p and a

- The first–order condition, (**??**) implies that

$$\lambda p = X(X'p) = Xa$$

and hence

$$p = \frac{1}{\lambda} Xa$$

- The vector p thus derived is the best linear description of the X columns in the least squares sense.
- It is known as the first principal component of the N variables represented by X. This will add a subscript to p, a, and λ:

$$p_1, a_1, \lambda_1$$

Other Principal Components

- X is approximated by $p_1 a_1'$, and we're left with the matrix $X - p_1 a_1'$ of deviations.
- Can the matrix of residuals be approximated by another matrix of unit rank, $p_2 a_2'$, so that we obtain

$$p_1 a_1' + p_2 a_2'$$

as a more accurate measure of X?

- Assume $p_2' p_2 = 1$ and $p_1' p_2 = 0$. We get the second for free because $X - p_1 a_a'$ is orthogonal to p_1.
- Now, the procedure is the same as before, except that we replace X by $X - p_1 a_1'$.

- So,

$$Min\ \ tr(X - p_1 a_1' - p_2 a_2')'(X - p_1 a_1' - p_2 a_2')$$

$$= \quad tr(X - p_1 a_1')'(X - p_1 a_1') - 2tr(X - p_1 a_1')'(p_2 a_2')_{orthogonal} + tr(a_2 p_2' p_2 a_2')$$

$$= \quad tr(X - p_1 a_1')'(X - p_1 a_1') - 2a_2' X p_2 + a_2' a_2$$

- Minimization with respect to a_2 implies that

$$a_2 = X' p_2$$

- The function to be minimized is

$$= tr(X - p_1 a_1')'(X - p_1 a_1') - p_2' X X' p_2$$

so that we want to

-

$$Max\, p_2' X X' p_2$$

subject to $p_2' p_2 = 1$ and $p_1' p_2 = 0$

- The Lagrangian is

$$L = p_2' XX' p_2 - \lambda_2 (p_2' p_2 - 1) - \mu p_1' p_2$$

So that

$$\frac{d}{dp2} = 0 = 2XX'p_2 - 2\lambda_2 p_2 - \mu p_1 = 0$$

- Premultiplying by p_1' yields

$$2p_1' XX' p_2 - 0 = \mu p_1' p_1 = \mu$$

- This implies that $\mu = 0$ because from (**??**),

$$XX' p_1 = \lambda_1 p_1$$

- and hence

$$p_1' XX' p_2 = \lambda_1 p_1' p_2 = 0$$

implying that

$$(XX' - \lambda_2 I) p_2 = 0 \qquad (18)$$

- Hence, p_2 is a characteristic vector of XX' corresponding to the root λ_2
- It is orthogonal to p_1, which corresponded to the largest root.
- At the same time, λ_2 should be as large as possible because the objective is to maximize $p_2' XX' p_2$.
- Hence, the second principal component p_2 is a characteristic vector corresponding to the second largest root λ_2.
- We can go on this way to get the r largest principal components.
- The i th component minimizes the sum of squares of the residuals that are left after the earlier components have been extracted.

- This minimization is subject to the constraint that

$$p_i' p_i = 1$$

and

$$p_{i-1}' p_i = 0$$

- This implies that p_i is the characteristic vector of XX' corresponding to the i th largest root, λ_i.
- if you think of the matrix X as a matrix of asset returns, then the p's are the factor portfolios and the a's are the factor sensitivities (the betas) and, well, there you are!

The number of factors?

Is there a statistical test for determining the number of factors?

- Yes, there is.
- Connor and Korajczyk develop an asymptotic test ($N \to \infty$) for the adequacy of K factors under the assumption of an approximate factor structure.
- Their test (intuitively) uses the result that the average cross–sectional variation in returns explained by the $K + 1$st factor approaches zero as the number of assets increases