F500 Empirical Finance
Lecture 6: Multifactor Pricing Models

Oliver Linton
obl20@cam.ac.uk

June 25, 2019

Outline

1. Linear Factor Model
2. Arbitrage Pricing Theory
3. Diversification and Pervasiveness
4. Multivariate Tests of the Multibeta Pricing Model with Observed Factors
    1. Macro Factor Models
    2. Fama French Factors
5. Characteristic based models
6. Statistical Factor Models

Reading: Linton (2019), Chapter 8

The Population Linear Factor model

Assume that asset returns are generated by the linear factor model (LFM). For asset $i \in \{1, ..., N\}$,

$$R_i = \alpha_i + \sum_{k=1}^{K} b_{ik} f_k + \varepsilon_i$$

such that

- $f_1, \ldots, f_K$ are random "common factors"
- b_{ik} are factor loadings (sensitivity of the return on asset i to factor k)
- ε_i are random shocks containing idiosyncratic risk (as opposed to systematic risk of the economy-wide factors), and at least

$$E\varepsilon_i = 0 \quad ; \quad \mathrm{var}(\varepsilon_i) = \sigma_{\varepsilon i}^2 < \infty.$$

$$\mathrm{cov}(f_k, \varepsilon_i) = 0.$$

A *(unit cost) portfolio* is $w_1, \ldots, w_N$ such that

$$\sum_{i=1}^{N} w_i = 1 \text{ or } w^\intercal i = 1$$

An (zero cost) *arbitrage portfolio* is $w_1, \ldots, w_N$ such that

$$\sum_{i=1}^{N} w_i = 0 \text{ or } w^\intercal i = 0$$

A portfolio that is *hedged against factor risk* (e.g., market neutral) is such that

$$\sum_{i=1}^{N} w_i b_{ik} = 0 \text{ for all } k \text{ or } w^\intercal B = 0$$

A *well-diversified portfolio* is such that

$$\sum_{i=1}^{N} w_i^2 \approx 0 \text{ (as } N \to \infty) \; w^\intercal w \simeq 0$$

Arbitrage Pricing Theory

Ross (1976), Chamberlain and Rothschild (1983, *Econometrica*). Large N economy.

Consider the well diversified, arbitrage portfolio p, that is hedged against factor risk

$$\begin{aligned} R_p &= \sum_{i=1}^{N} w_i R_i \\ &= \sum_{i=1}^{N} w_i \alpha_i + \sum_{k=1}^{K} \sum_{i=1}^{N} w_i b_{ik} \cdot f_k + \sum_{i=1}^{N} w_i \varepsilon_i \\ &\approx \sum_{i=1}^{N} w_i \alpha_i \\ &\approx 0 \end{aligned}$$

otherwise you make money for nothing.

- Let $B = (b_{ik})$ be the $N \times K$ matrix of factor loadings. An arbitrage portfolio that is hedged against all factor risk satisfies

$$w^\intercal i = 0 \text{ and } w^\intercal B = 0,$$

i.e., w is in the null space of (i, B) ($N \times (K+1)$ matrix). Many such portfolios exist by standard linear algebra

- Since the vector α is orthogonal to w it must lie in the space spanned by (i, B), (using $(A^\perp)^\perp = A$).
- Therefore for some constants $\rho, \theta_1, \ldots, \theta_k$ we have

$$\alpha_i = \rho + \sum_{k=1}^{K} b_{ik}\theta_k, \quad \alpha = \rho i + B\theta$$

- It follows that for any asset i

$$\mu_i = E(R_i) = \alpha_i + \sum_{k=1}^{K} b_{ik} E(f_k) = \rho + \sum_{k=1}^{K} b_{ik} (E(f_k) + \theta_k)$$

i.e., $\mu = (\mu_1, \ldots, \mu_N)^\intercal \in \mathrm{span}(i, B)$ - there exists constants $\lambda_0, \lambda_1, \ldots, \lambda_K$ such that

$$\mu_i = \lambda_0 + \sum_{k=1}^{K} b_{ik} \lambda_k,$$

where $\lambda_k = E(f_k) + \theta_k$ are risk premia associated with the k^{th} factor.

- In matrix notation

$$\mu = \lambda_0 i + B\lambda$$

- If the assets are traded factors and there is a risk free rate, then $\lambda_0 = R_f$ and $\lambda_k = E(f_k) - R_f$

$$\mu_i - R_f = \sum_{k=1}^{K} b_{ik}(E(f_k) - R_f)$$

- The CAPM corresponds to the case where $K = 1$ and f_1 is the return on the market portfolio.
- The APT theory doesn't say what the factors are.

When does diversification work?

- First, averaging reduces variance provided correlation is not perfect.
- Consider a portfolio with return

$$R_w = wX + (1-w)Y, \quad w \in \mathbb{R}$$

$$V(w) = \mathrm{var}(R_w) = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\rho_{XY}\sigma_X\sigma_Y.$$

- Clearly (setting $w = 0$ or $w = 1$)

$$\min_w V(w) \leq \min\{\sigma_X^2, \sigma_Y^2\} \leq \max\{\sigma_X^2, \sigma_Y^2\} \leq \max_w V(w)$$

  with strict inequality if and only if $\rho_{XY} \neq +1.$
- Suppose that $\sigma_X^2 = \sigma_Y^2 = 1,$ then (solving $dV(w)/dw = 0$) we have

$$w_{opt} = \frac{1}{2} \quad ; \quad V(w_{opt}) = \frac{1}{2}(1+\rho_{XY}) \leq 1$$

- Suppose that factor model holds with

$$E(\varepsilon\varepsilon^{\intercal}) = \Omega_{\varepsilon} \quad ; \quad \Sigma_f = E(ff^{\intercal}).$$

- Portfolio variance is the sum of two terms

$$\mathrm{var}(w^{\intercal}R) = \overbrace{w^{\intercal}B\Sigma_f B^{\intercal}w}^{\text{common}} + \overbrace{w^{\intercal}\Omega_{\varepsilon}w}^{\text{idiosyncratic}}.$$

Key result: *Can show that with many assets, portfolio can have zero (idiosyncratic) variance under some very weak conditions.*

Definition

We say that the idiosyncratic error is diversifiable if $\lim_{N\to\infty} w^\intercal w = 0$ implies that

$$\lim_{N\to\infty} w^\intercal \Omega_\varepsilon w = 0.$$

We consider the diagonal case first. Suppose that $\Omega_\varepsilon = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$. Then

$$\mathrm{var}(w^\intercal \varepsilon) = w^\intercal \Omega_\varepsilon w = \sum_{i=1}^{N} w_i^2 \sigma_i^2.$$

We have

$$\sum_{i=1}^{N} w_i^2 \sigma_i^2 \leq \max_{1\leq i\leq N} \sigma_i^2 \sum_{i=1}^{N} w_i^2,$$

and it suffices that $\sigma_i^2 \leq c < \infty$ for all i.
So if all the variances are bounded then clearly diversification works for well balanced portfolios.

The assumption of uncorrelated errors is considered a bit strong and is stronger than is needed for the diversification property to hold as we next show.

Theorem

Suppose that

$$\Omega_\varepsilon = D^{1/2}\Psi D^{1/2},$$

where $D = \mathrm{diag}\{\sigma_1, \ldots, \sigma_N\}$ *is a diagonal matrix with bounded entries and* Ψ *is a correlation matrix with* $\lambda_{\max}(\Psi) \leq c < \infty$ *(as* $N \to \infty$*). Then,* $\lim_{N\to\infty} w^\intercal w = 0$ *implies that*

$$\lim_{N\to\infty} w^\intercal \Omega_\varepsilon w = 0.$$

The proof of this result is immediate. Letting $\widetilde{w} = D^{1/2}w$, we have

$$w^\intercal \Omega_\varepsilon w = \widetilde{w}^\intercal \widetilde{w} \frac{\widetilde{w}^\intercal \Psi \widetilde{w}}{\widetilde{w}^\intercal \widetilde{w}} \leq \lambda_{\max}(\Psi) \sum_{i=1}^{N} w_i^2 \sigma_i^2 \to 0, \text{ as } N \to \infty.$$

There are several models of cross sectional correlation that are in use. First, Connor and Koraczyck (1983) assumed that there is **some ordering** of the cross section such that the process ε_i is alpha mixing (Chapter 2). Suppose that

$$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = \rho^{|j-i|}$$

for some ρ with $|\rho| < 1$. Then

$$\Omega_\varepsilon = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \ddots & \ddots & \\ \rho^2 & \ddots & 1 & \rho & \rho^2 \\ \vdots & \ddots & \rho & 1 & \rho \\ \rho^{N-1} & \cdots & \rho^2 & \rho & 1 \end{bmatrix},$$

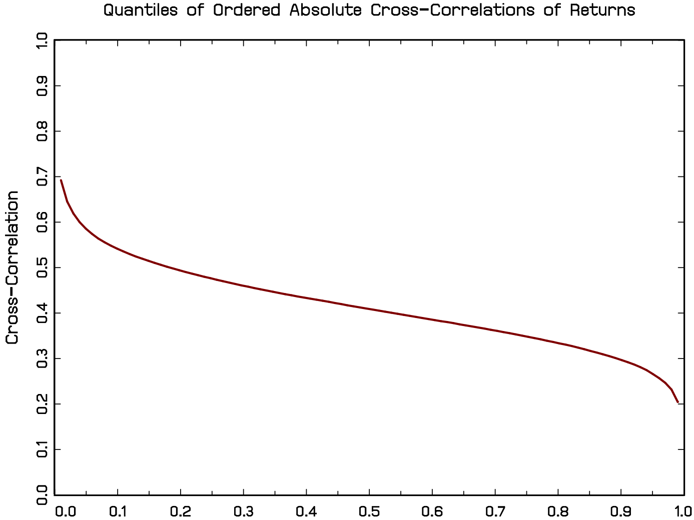
which has bounded eigenvalues.

In fact

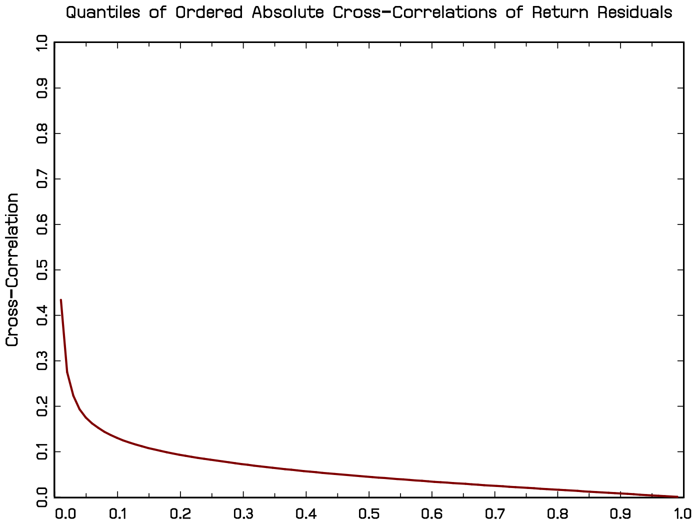
$$
\begin{aligned}
\frac{1}{N} i^{\intercal} \Omega_{\varepsilon} i &= \frac{1}{N}\left[N + 2(N-1)\rho + 2(N-2)\rho^2 + \ldots\right] \\
&\to 1 + 2\sum_{j=1}^{\infty} \rho^j \\
&= \frac{1+\rho}{1-\rho}.
\end{aligned}
$$

So that the equally weighted portfolio (of ε) has zero variance for large N. So diversification eliminates all (idiosyncratic) risk.

In practice, the raw correlations are high. In Figure xx below we show the distribution of $|\mathrm{corr}(R_i, R_j)|$ of S&P500 daily returns over the period 2000-2010.



By contrast the idiosyncratic errors have smaller correlations but they are still significant. In Figure xx we show the $|\mathrm{corr}(\varepsilon_i, \varepsilon_j)|$ (market model residuals)



- In conclusion, the diversification arguments can hold more generally even in the presence of correlation between the error terms and large variance terms.
- If diversification works we obtain that the portfolio variance is dominated by the common components, i.e.,

$$w^\intercal \Omega_\varepsilon w \approx 0 \quad \Longrightarrow \mathrm{var}(w^\intercal R) \approx \overbrace{w^\intercal B \Sigma_f B^\intercal w}^{\text{common}}.$$

- We next consider some empirical approaches to measuring diversification.

Solniks diversification curve

Given a set of assets the variance of the equally weighted portfolio is

$$\operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} R_i\right) = \frac{1}{N}\overline{\sigma}_i^2 + \overline{\sigma}_{ij}$$

$$\overline{\sigma}_i^2 = \frac{1}{N}\sum_{i=1}^{N} \operatorname{var}(R_i) \quad ; \quad \overline{\sigma}_{ij} = \frac{2}{N(N-1)}\sum_{j=i+1}^{N}\sum_{i=1}^{N-1} \operatorname{cov}(R_i, R_j)$$

Cross covariances more important than own variances when N is large. How does this vary with N? Take a subsample of m assets, but which subsample?

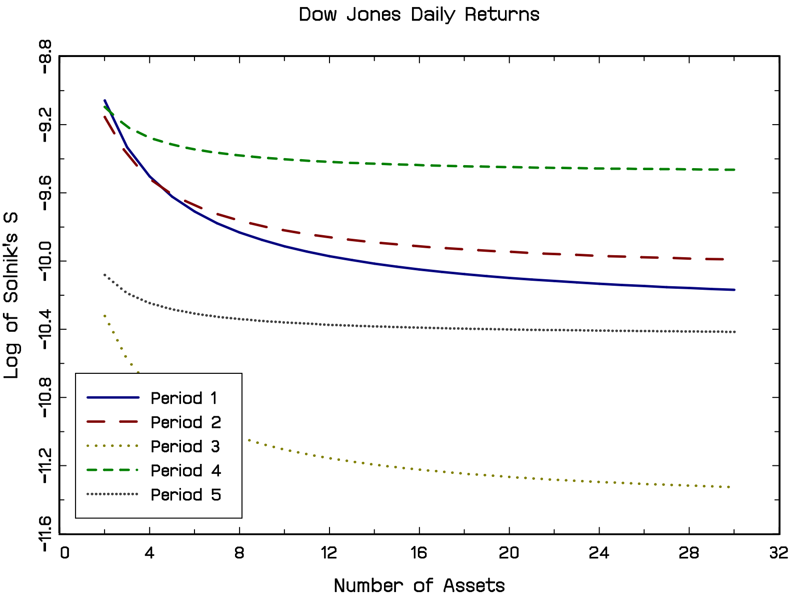
Definition

(Solnik) Sample variance $S(m)$ of a randomly selected equally weighted portfolio of m assets for $m = 1, 2, \ldots, N$

Can show that

$$S(m) = \frac{1}{m}\overline{\sigma}_i^2 + \left(1 - \frac{1}{m}\right)\overline{\sigma}_{ij} \rightarrow \overline{\sigma}_{ij} \text{ as } m \rightarrow \infty$$

(Log of) Solniks curve for 5 subperiods



Global Minimum Variance portfolio

Suppose that Σ is the covariance matrix of returns. For weights w_{GMV} with $i = (1, \ldots, 1)^\intercal \in \mathbb{R}^m$ we have

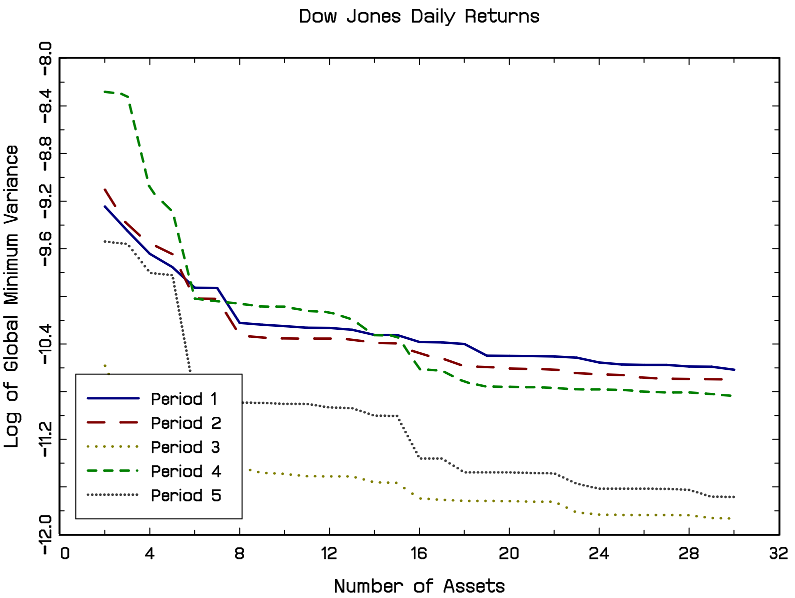
$$w_{GMV} = \frac{\Sigma^{-1} i}{i^\intercal \Sigma^{-1} i}$$

and we achieve variance

$$\sigma^2_{GMV}(m) = \frac{1}{i^\intercal \Sigma^{-1} i}$$

Compute this for assets $R_1, \ldots, R_L$ with $L = 2, \ldots, N$

Log of σ^2_{GMV} for five different subperiods



When are factors Pervasive?

- A key assumption in the sequel is that all the included factors are **pervasive**, which is to say they each affect many assets returns. It is saying that all the factors play an important role in explaining the returns of the assets, essentially, nearly all assets are affected in some way by the factors.
  - For example when $K = 1$ we might just require that the number of $b_i \neq 0$, denoted r, is a large fraction of the sample.
- We next give a formal definition of pervasiveness. wlog set $\Sigma_f = I_K$ in the sequel

Definition

We say that the factors are strongly influential or pervasive when

$$\frac{1}{N} B^{\intercal} B \to M,$$

where M is a finite and strictly positive definite matrix.

We have the following result

Theorem

For all weighting sequences w such that $B^{\intercal} w \neq 0$ and such that $i^{\intercal} w = 1$, we have as $N \to \infty$

$$w^{\intercal} B B^{\intercal} w \geq w^{\intercal} w \times \lambda_{\min}(B^{\intercal} B) \geq \lambda_{\min}(M) > 0.$$

- If the diversification condition is also satisfied, then this says that for such portfolios, the variance of returns is dominated by the common factor, and this term cannot be eliminated. The common component is not diversifiable.
- Of course there are also portfolios for which $B^{\intercal}w = 0$, and these have already been introduced in Chapter xx, and are called hedge portfolios. It is a classical result in linear algebra that the subspace of $\mathbb{R}^N$ of hedge portfolios is of dimension $N - K$, and so its complement is of dimension K.
- The APT tells us that those well diversified hedge portfolios do not make any money over their cost.

- There are some concerns about whether the condition holds or at least whether all factors are pervasive, when there are multiple factors (factor zoo). If this condition is not satisfied, then some of the approximations we employ below are no longer valid.

Definition

(Onatskiy (2012)) We say that the factors are weakly influential when

$$B^{\intercal} B \rightarrow D,$$

where D is a diagonal matrix.

- Under this condition, the contribution of the common components to the variance of the portfolio is of the same magnitude as that of the idiosyncratic components. This will affect some econometric testing and estimation discussed below.

- For example, suppose that B are actually observed (see characteristic based factor model section below) and f_t are estimated by a cross sectional regression of returns on B, then the necessary condition for consistency (under iid errors ref) is that as $N \to \infty$

$$\lambda_{\min}(B^{\intercal} B) \to \infty,$$

which is intermediate between strong and weak.

- We may find that some factors are strongly influential, whereas others are not, so that in the multifactor case the matrix B may be rank deficient.
- One way of modelling this is to say that $B = (B_1, B_2)$, where B_1 are strong factors, whereas B_2 are weakish factors that are small in the sense that $\sqrt{T} B_2$ satisfies the strong factor condition but B_2 does not, Bryzgalova (2015).

The Econometric Model

- We suppose we observe the panel of returns $\{Z_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T\}$.
- The K-factor model (for returns or excess returns) is

$$Z_{it} = \alpha_i + \sum_{j=1}^{K} b_{ij} f_{jt} + \varepsilon_{it}$$

$$Z_t = \alpha + Bf_t + \varepsilon_t,$$

where ε_{it} is an idiosyncratic error term with

$$E(\varepsilon_t | f_1, \ldots, f_T) = 0 \quad ; \quad E(\varepsilon_t \varepsilon_t^\intercal | f_1, \ldots, f_T) = \Omega_\varepsilon$$

- Can be justified from multivariate normality of returns and the factors.

Different types of Factor Models

There are three different types of factor models

1. Observable factor models (f_t are observed, b_{ij} are unknown quantities)
    1. The factors are returns to traded portfolios (specifically, the returns on portfolios formed on the basis of security characteristic such as size, B/M).
    2. The factors are macro variables such as yield spread etc.
2. Statistical factor models (f_{jt} and b_{ij} are unknown quantities)
3. Characteristic based models (b_{ij} are observed characteristics such as industry or country and f_{jt} are unknown quantities)

In each case there are slight differences in cases where there is a risk free asset and in cases where there are not.

Multivariate Tests of the Multibeta Pricing Model with Observable Traded factors and Risk Free Rate

Multivariate tests are very similar to those for the CAPM, but with β replaced with the matrix B. First suppose that there is a risk-free asset and that the factor returns F are observable. The log likelihood function of the data $Z_1, \ldots, Z_T$ conditional on $Z_{K1}, \ldots, Z_{KT}$ is

$$\begin{aligned}\ell(\alpha, B, \Omega_\varepsilon) &= c - \frac{T}{2} \log \det \Omega_\varepsilon \\ &\quad - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \alpha - BZ_{Kt})^\intercal \Omega_\varepsilon^{-1} (Z_t - \alpha - BZ_{Kt}).\end{aligned}$$

Letting

$$\widehat{\mu}_f = \frac{1}{T}\sum_{t=1}^{T} Z_{Kt} \qquad ; \qquad \widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Z_t$$

$$\widehat{\Sigma}_f = \frac{1}{T}\sum_{t=1}^{T}\left(Z_{Kt} - \widehat{\mu}_f\right)\left(Z_{Kt} - \widehat{\mu}_f\right)^{\intercal}$$

$$\widehat{\Sigma}_{Zf} = \frac{1}{T}\sum_{t=1}^{T}\left(Z_t - \widehat{\mu}\right)\left(Z_{Kt} - \widehat{\mu}_K\right)^{\intercal}$$

The unrestricted MLE of α, B are the equation by equation OLS, and Ω_ε is the covariance matrix of residuals:

$$\widehat{\alpha} = \widehat{\mu} - \widehat{B}\widehat{\mu}_f \qquad ; \qquad \widehat{B} = \widehat{\Sigma}_{Zf}\widehat{\Sigma}_f^{-1}.$$

$$\widehat{\Omega}_\varepsilon = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \widehat{\alpha} - \widehat{B}Z_{Kt})(Z_t - \widehat{\alpha} - \widehat{B}Z_{Kt})^{\intercal}$$

- The multivariate tests of $\alpha = 0$ (the APT restrictions in this case) are similar to those for the CAPM. The F test

$$F = \frac{(T-N-K)}{N}(1+\widehat{\mu}_f^\intercal \widehat{\Sigma}_f^{-1}\widehat{\mu}_f)^{-1}\widehat{\alpha}^\intercal \widehat{\Omega}_\varepsilon^{-1}\widehat{\alpha}$$

  is exactly distributed as $F(N, T-N-K)$ under the normality assumption.
- In the absence of normality the statistic is asymptotically (sample size T gets large) chi-squared with N degrees of freedom.
- Version without risk free rate, use LR test and compare with $\chi^2(N-1)$

Multifactor Pricing Tests with Macro Factors and Risk free rate

- Suppose instead that the observed factors are macroeconomic shocks rather than portfolio returns. Then the expected returns associated with the factors have to be estimated as additional free parameters rather like the mean of the zero beta portfolio.
- In this case, the APT restrictions are

$$\alpha = B\gamma, \text{ for some } \gamma \in \mathbb{R}^K$$

There are $N - K$ restrictions. Likelihood ratio test easiest here.

Macroeconomics factors

(Chan et al. (1985) and Chen et al. (1986)).

$$R_{it} = \alpha_i + b_i^\intercal f_t + \varepsilon_{it}$$

$$f_t = m_t - E_{t-1} m_t \text{ or } f_t = m_t - m_{t-1},$$

where m_t are (nonstationary) macroeconomic variables and f_t are "surprises" or differences. Monthly data. 1958-1984

1. The percentage change in industrial production (led by one period)
2. A measure of unexpected inflation
3. The change in expected inflation
4. The difference in returns on low-grade (Baa and under) corporate bonds and long term government bonds (junk spread)
5. The difference in returns on long-term government bonds and short term Tbills (Term spread)

- They estimate b_i by time series regressions and then do cross sectional regressions on $\widehat{b}_i$ to estimate factor risk premia; 20 portfolios are used on the basis of firm size at the beginning of the period
- They find that average factor risk premia are statistically significant over the entire sample period for industrial production, unexpected inflation, and junk. They also include a market return but find it is not significant
- Recent developments: Macroeconomic policy endogeneity to level of asset prices. Event study approach.

Fama and French Factors

- Fama and French (1993) "measure" f directly. They construct (6) double sorted portfolios formed on 2 size and 3 book to market.
  - That is, first sort the stocks according to their size (at the given date) and divide into two: large size and small size.
  - Then sort each of these groups according to the BTM and divide each into three further groups. The sorting could equally be done the other way round, and in both ways one obtains six groups of stocks

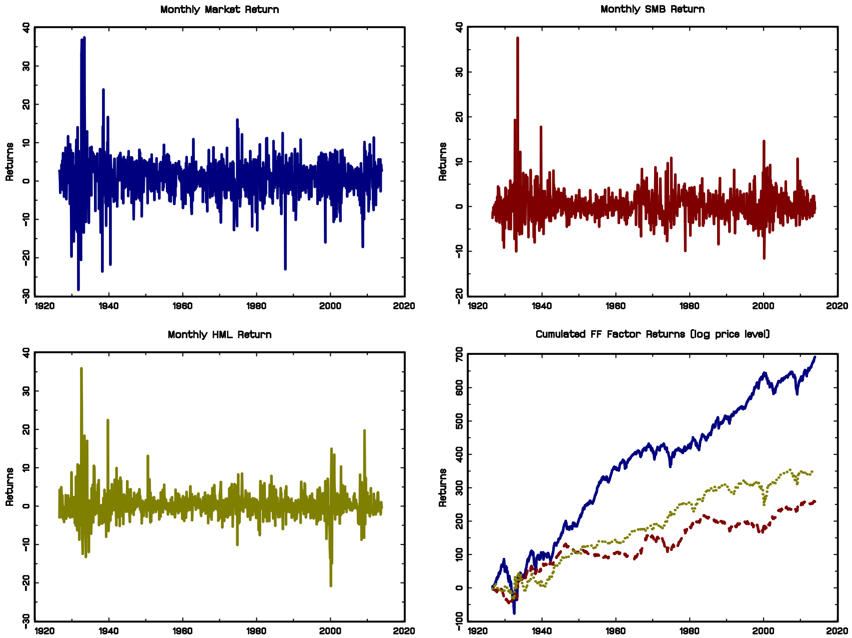
BTM/Size	1	2
1	high/large	high/small
2	medium/large	medium/small
3	low/large	low/small

- These are then equally weighted in within each group to produce a portfolio that measures large size and high BTM (denoted (1,1)), etc.
- The size factor return is proxied by the difference in return between a portfolio of low-capitalization stocks and a portfolio of high-capitalization stocks, adjusted to have roughly equal book-to-price ratios (SMB)

$$\frac{1}{3}\{(1,1)-(1,2)+(2,1)-(2,2)+(3,1)-(3,2)\}.$$

- They are zero net investment portfolios.

- The value factor is proxied by the difference in return between a portfolio of high book-to-market stocks and a portfolio of low book-to-market stocks, adjusted to have roughly equal capitalization (HML)
- The market factor return is proxied by the excess return to a value-weighted market index (MKT)
- Data on the factors is available from Ken French web site http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.h


Monthly Market Return
Monthly SMB Return
Monthly HML Return
Cumulated FF Factor Returns (log price level)
Returns
1920
1940
1960
1980
2000
2020

- Then form 25 portfolios (on size and value characteristics too!)

$$Z_{pt} = \alpha_p + b_{p,SMB} SMB_t + b_{p,HML} HML_t + b_{p,MKT} MKT_t + \varepsilon_{pt},$$

where $p = 1, \ldots, 25$ and $t = 1, \ldots, 60$.

  - They do not reject the APT restrictions in their sample (although they do reject CAPM where SMB and HML are dropped).
  - They explain well the size and value anomalies of the CAPM.

- Subsequently, it all went pear shaped and APT rejected. Additional factors have been proposed in the literature
  - Momentum factor, Carhart (1997).
  - Own-volatility factor, Goyal and Santa Clara (2003).
  - Liquidity, Amihud and Mendelson (1986), Pastor and Stambaugh (2003)
  - Fama and French 5 factor model (2015). RMW (Robust Minus Weak) and CMA (Conservative Minus Aggressive)
  - Factor zoo

Explicit Characteristics-Based Models

- Rosenberg (1974) considered the multifactor regression model where f_t are unknown "parameters" and B is related to observable (time invariant) stock characteristics such as: industry/country (dummy variables). Other possible characteristics include: size and value.
- Suppose that

$$\alpha_i = d_\alpha^\intercal x_i, \qquad b_i = D_b x_i, \ i = 1, \ldots, N$$

where x_i is an observed $J \times 1$ vector of characteristics, and D_b is a $K \times J$ matrix of unknown parameters.

- Substituting into the return equation we obtain

$$Z_{it} = \alpha_i + b_i^\intercal f_t + \varepsilon_{it} = x_i^\intercal \left(d_\alpha + D_b^\intercal f_t \right) + \varepsilon_{it} = x_i^\intercal f_t^* + \varepsilon_{it},$$

where $f_t^* = d_\alpha + D_b^\intercal f_t$ is a $J \times 1$ vector of characteristic specific factors for $t = 1, \ldots, T$.

- For each t estimate f_t^* by the cross-sectional regression

We can write

$$Z_t = Xf_t^* + \varepsilon_t$$

It follows that

$$E(Z_t|X) = X\mu_{f^*}$$
$$\mathrm{var}(Z_t|X) = X\Sigma_{f^*}X + \Omega_\varepsilon,$$

where X is the $N \times J$ matrix of observed characteristics and μ_{f^*} and Σ_{f^*} are the mean and variance of the factors.
Can do mean/variance portfolio choice conditional on characteristics

Statistical Factor Models: Identifying the Factors in Asset Returns

Definition

The statistical factor model for observed returns where neither B nor F are observed. For each time period write

$$Z_t = \alpha + Bf_t + \varepsilon_t,$$

Definition

Strict factor structure - idiosyncratic error is cross-sectionally uncorrelated so

$$E\varepsilon_t\varepsilon_t^\intercal = \Omega_\varepsilon = D,$$

where D is a diagonal matrix (containing the idiosyncratic variances).

- We treat B as fixed quantities and f_t as random variables where $E(f_t) = 0$ without loss of generality. We also assume $\mathrm{cov}(f_t, \varepsilon_t) = 0$.
- Then the population covariance matrix satisfies

$$\Sigma_{N\times N} = \mathrm{var}(Z_t) = B\Sigma_f B^\intercal + D,$$

where $\Sigma_f = \mathrm{var}(f_t)$.

- The RHS has $NK + K(K+1)/2 + N$ parameters which is less than $N(N+1)/2$ on LHS. A big reduction in dimensionality
- However, in this case where factors are unknown there is an identification issue. Can't uniquely identify (B, Σ_f) or (B, f_t)

Identification Issue

- One can write for any nonsingular matrix L,

$$f_t^* = Lf_t \quad ; \quad \Sigma_{f^*} = L\Sigma_f L^\intercal$$

so that

$$Bf_t = BL^{-1}Lf_t = B^* f_t^*$$

$$B\Sigma_f B^\intercal = B^*\Sigma_{f^*} B^{*\intercal}$$

- One solution is to restrict $\Sigma_f = I_K$. In this case,

$$\Sigma = \Sigma(B, D) = BB^\intercal + D.$$

Then B, D are unique (B upto sign and orthonormal transformations)

Estimation

Can be estimated by maximum likelihood factor analysis provided $N << T$ (*small* N *and large* T).

$$\ell(\alpha, B, D) = c - \frac{T}{2} \log \det \Sigma(B, D) - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \alpha)^{\intercal} \Sigma(B, D)^{-1} (Z_t - \alpha).$$

- The MLE for α is still the sample average

$$\widehat{\alpha} = \frac{1}{T} \sum_{t=1}^{T} Z_t.$$

- We then substitute in to obtain

$$\ell(\widehat{\alpha}, B, D) = c - \frac{T}{2} \log \det \Sigma(B, D) - \frac{1}{2} \mathrm{tr} \left(\widehat{\Sigma} \Sigma(B, D)^{-1} \right)$$

$$\widehat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (Z_t - \widehat{\alpha})(Z_t - \widehat{\alpha})^{\intercal}$$

- Total of $NK + N$ parameters in $\Sigma(B, D)$. Solve nonlinear first order conditions (for $i = 1, \ldots, N$ and $k = 1, \ldots, K$)

$$\begin{aligned} \frac{\partial \ell}{\partial b_{ik}}(\widehat{\alpha}, \widehat{B}, \widehat{D}) &= 0 \\ \frac{\partial \ell}{\partial \sigma^2_{\varepsilon i}}(\widehat{\alpha}, \widehat{B}, \widehat{D}) &= 0 \end{aligned}$$

- Iterative nonlinear procedure such as Newton Raphson

- Given $\widehat{B}, \widehat{D},$ the period by period factor realizations can be estimated by cross-sectional regression, i.e., OLS or GLS

$$\widehat{f}_t = (\widehat{B}^\intercal \widehat{B})^{-1} \widehat{B}^\intercal (Z_t - \widehat{\alpha})$$

$$\widehat{f}_t = (\widehat{B}^\intercal \widehat{D}^{-1} \widehat{B})^{-1} \widehat{B}^\intercal \widehat{D}^{-1} (Z_t - \widehat{\alpha})$$

- Consistency requires $\widehat{B}^\intercal \widehat{B} \to \infty$ (pervasive factors).
- Estimated factors. Replacing B with $\widehat{B}$ creates an errors in variables problem, affects standard errors at least

Factor Models and Portfolio Choice

- A value of the factor model is in dimensionality reduction. It is important in portfolio choice and asset allocation, which usually involves an inverse covariance matrix that has to be estimated.
- Note that when

$$\Sigma = BB^{\intercal} + D,$$

we have by the Sherman, Morrison, Woodbury formula

$$\Sigma^{-1} = D^{-1} - D^{-1}B(I_K + B^{\intercal}D^{-1}B)^{-1}B^{\intercal}D^{-1},$$

which only involves inverting the $K \times K$ matrix $I_K + B^{\intercal}D^{-1}B$ and the elements of the diagonal matrix D.

- Portfolios that hedge or mimic the factors are the basic components of various portfolio strategies
  - The mimicking portfolio for a given factor is the portfolio with the maximum correlation with the factor
  - The hedge portfolio is the one that is orthogonal to the loadings of the factor
- If all assets are correctly priced, then each investor's portfolio should be some combination of cash and the mimicking portfolios.

- We can write
$$\widehat{f}_{jt} = \widehat{w}_j^{\intercal} \left(Z_t - \widehat{\alpha} \right),$$
where $\widehat{w}_j = ((\widehat{B}^{\intercal} \widehat{B})^{-1} \widehat{B}^{\intercal})_j$ or $\widehat{w}_j = ((\widehat{B}^{\intercal} \widehat{D}^{-1} \widehat{B})^{-1} \widehat{B}^{\intercal} \widehat{D}^{-1})_j$.
- Can show that these portfolio weights solve the following problem (taking $\Omega = \widehat{D}$ or $\Omega = I$)
$$\min_{w_j} w_j^{\intercal} \ \Omega \ w_j$$
subject to
$$w_j^{\intercal} \widehat{b}_h = 0 \quad h \neq j \quad ; \quad w_j^{\intercal} \widehat{b}_j = 1$$
- The set of portfolios that are hedged against factors h, $h \neq j$ and have unit exposure to factor j is of dimension $N - K$. We are finding the one with smallest idiosyncratic variance. Can normalize the weights $\widehat{w}_j$ to sum to one, so that they are portfolio weights.

Asymptotic Principal Components

- An alternative to maximum likelihood factor analysis is asymptotic principal components (*small* T *and large* N). In the population we can write the $T \times 1$ excess return vector Z_i as

$$Z_i = \alpha_i i_T + F b_i + \varepsilon_i,$$

  where F is $T \times K$, ε_i is $T \times 1$, and b_i is $K \times 1$.
- Assume that F is a fixed matrix (or if random we can condition on its realization) and that b_i are random variables from a common distribution with $E(b_i) = 0$ and

$$E(b_i b_i^\intercal) = \Sigma_b.$$

- Let

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_{\varepsilon i}^2$$

- It follows that

$$
\begin{aligned}
\Psi &= \frac{1}{N}\sum_{i=1}^{N} E\left[\left(Z_i - E(Z_i)\right)\left(Z_i - E(Z_i)\right)^{\intercal} | F\right] \\
&= F\overbrace{\left(\frac{1}{N}\sum_{i=1}^{N} E(b_i b_i^{\intercal})\right)}^{\Sigma_b(N)} F^{\intercal} + \overbrace{\left(\frac{1}{N}\sum_{i=1}^{N} \sigma_{\varepsilon i}^2\right)}^{\sigma^2(N)} I_T, \\
&= F\Sigma_b F^{\intercal} + \sigma^2 I_T.
\end{aligned}
$$

- The right hand side has $TK + K(K+1)/2 + 1$ parameters, which is less than LHS which has $T(T+1)/2$.

Identification problem

- There is an identification problem: for any nonsingular matrix L

$$Fb_i = FLL^{-1}b_i = F^*b_i^*$$

- In this case, assume that (with $\gamma_1 \geq \ldots \geq \gamma_K$)

$$\Sigma_b = \mathrm{diag}\{\gamma_1, \ldots, \gamma_K\} = \Gamma \quad ; \quad F^\intercal F = I_K.$$

$$\Psi = F\Gamma F^\intercal + \sigma^2 I_T,$$

- Then F, Γ, σ^2 are unique (F upto sign)

- For any symmetric $T \times T$ matrix Ψ we have the unique eigendecomposition

$$\Psi = Q\Lambda Q^{\intercal} = \sum_{t=1}^{T} \lambda_t q_t q_t^{\intercal}$$

where eigenvectors Q satisfy

$$QQ^{\intercal} = Q^{\intercal}Q = I_T$$

and eigenvalues

$$\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_T\}$$

ordered from largest to smallest.

- This means that for λ_t, q_t

$$\left(F\Gamma F^{\intercal} + \sigma^2 I_T\right) q_t = \lambda_t q_t$$

- In fact F satisfies this equation

$$\left(F\Gamma F^{\intercal}+\sigma^2 I_T\right)F=F\left(\Gamma+\sigma^2 I_K\right),$$

which means that F contains the K eigenvectors corresponding to the eigenvalues, the diagonal elements of $\Gamma+\sigma^2 I_K$.

- The remaining $T-K$ eigenvectors G are contained in a $T\times T-K$ matrix such that $F^{\intercal}G=0$ with eigenvalues σ^2, i.e.,

$$\left(F\Gamma F^{\intercal}+\sigma^2 I_T\right)G=\sigma^2 G.$$

- In conclusion,
  - the eigenvalues of Ψ are

$$\gamma_1 + \sigma^2 \geq \ldots \geq \overbrace{\gamma_K + \sigma^2 > \sigma^2}^{\text{spike}} = \ldots = \sigma^2.$$

  - F are the eigenvectors corresponding to the K largest eigenvalues of Ψ

$$F = \mathrm{eigvec}_K[\Psi]$$

- This shows unique identification of F. The Γ, σ^2 are also uniquely identified by this.

APC estimates (in the first pass) factor returns rather than factor betas using the above identification argument

1. The sample covariance matrix

$$\widehat{\Psi} = \left(\frac{1}{N} \sum_{i=1}^{N} (Z_{it} - \overline{Z}_t)(Z_{is} - \overline{Z}_s) \right)_{s,t=1}^{T}, \quad \overline{Z}_t = \frac{1}{N} \sum_{i=1}^{N} Z_{it}$$

   which is a $T \times T$ matrix of excess return cross-products.

2. Do the empirical eigendecomposition and take

$$\widehat{F} = \mathrm{eigvec}_K[\widehat{\Psi}]$$

3. Given this estimate of F, the factor betas can be estimated by time-series OLS regression

$$\widehat{b}_i = (\widehat{F}^\intercal \widehat{F})^{-1} \widehat{F}^\intercal (Z_i - \widehat{\alpha}_i i_T)$$

- The main problem with this approach is that it assumes time series homoskedasticity for the idiosyncratic error, which is not a good assumption.
- Jones (2001, JFE) has extended the estimation problem to allow for time varying average idiosyncratic variance.

$$\Psi = F\Gamma F^{\intercal} + D, \quad D = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_T^2\}$$

- Iterative application of APC (like Factor MLE). That is, given first round estimates, calculate the time series heterosekdasticity

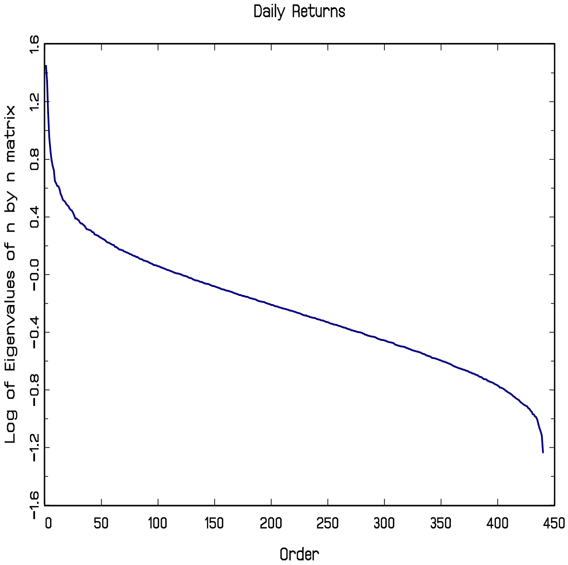
$$\widehat{D} = \mathrm{diag}\{\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_T^2\}$$

and then recompute the factors

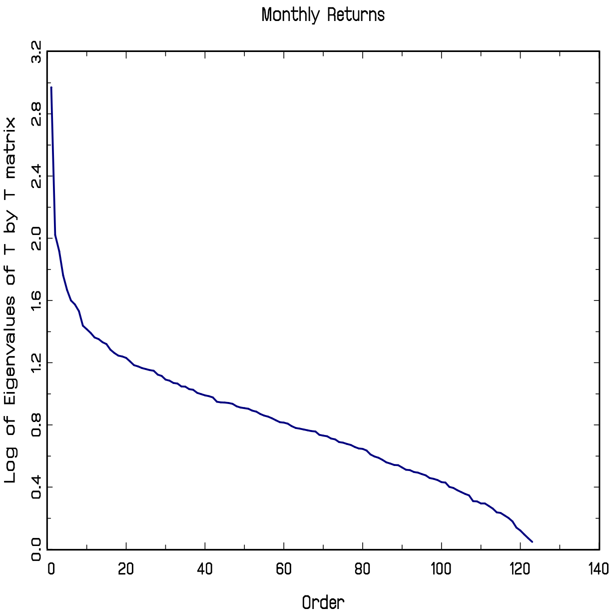
$$\widehat{F} = \mathrm{eigvec}_K[\widehat{\Psi} - \widehat{D}]$$

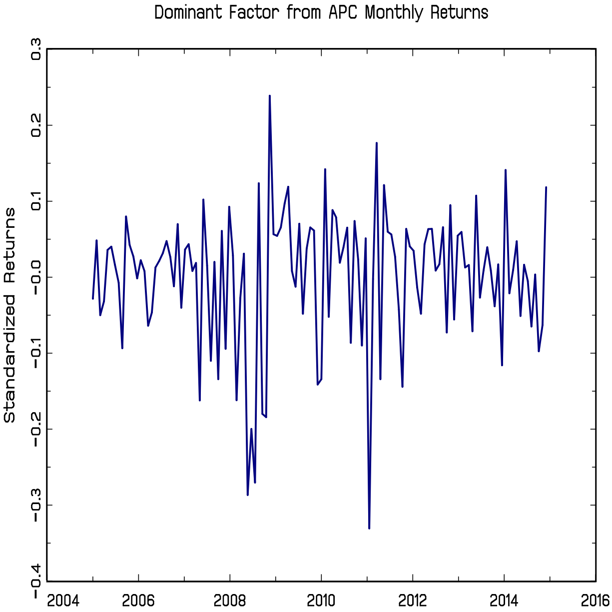
and loadings likewise and iterate

Eigenvalues of $\widehat{\Omega}_T$ for daily SP500 returns ($N = 441,\ T = 2732$)



Eigenvalues of $\widehat{\Psi}_N$ for Monthly SP500 returns ($N = 441$, $T = 124$)





Dominant Factor from APC Monthly Returns
Standardized Returns
0.3
0.2
0.1
-0.0
-0.1
-0.2
-0.3
-0.4
2004
2006
2008
2010
2012
2014
2016

The case where both N and T are large can yield further results

Definition

Bai and Ng (2002). Solve the following

$$\min_{B,F} \sum_{t=1}^{T} \sum_{i=1}^{N} \left\{ Z_{it} - b_i^{\intercal} f_t \right\}^2$$

subject to the identification constraint either that $B^{\intercal} B / N = I_K$ or $F^{\intercal} F / T = I_K$

- The procedure can be understood as iterative least squares (cross-section regression then time series regression then etc). Equivalent (upto normalization) to APC when T is fixed
- They show consistency of this procedure when N and T are large
- They also propose model selection method to determine the number of factors K
- Large literature now on estimating large covariance matrices with shrinkage