9 Factor Models

The main message of the CAPM is that the cross-sectional returns of assets can be solely explained by the betas. That is, if we know the betas, we can describe the cross-sectional returns on the assets as

$$R_{it} = \alpha_i + \beta_{iM} R_{Mt} + \varepsilon_{it}$$

where $\alpha_i = (1 - \beta_{iM}) R_f$. The above model could be viewed as a factor model, where R_{Mt} is the common factor, which is independent of the individual asset, but changes across time, and β_{iM} is the factor loading, which is asset dependent, but is invariant over time.

It is quite natural to conject that there are more common factors that has explanatory power of the cross-sectional returns on assets. These factors could be macroeconomic fundamentals, firm or industry related factors, or even statistical factors.

9.1 Factor Models

A general form of the factor model for the return on asset i at time t is

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \cdots + \beta_{im} f_{mt} + \varepsilon_{it},$$

where α_i represents the asset-specific intercept, $f_{jt}, j = 1, 2, \ldots, m$ are the m common factors, β_{ij} is the factor loading for asset i on the j-th factor, and ε_{it} is the specific factor of asset i.

It is usually assumed that for financial asset returns, the factor $f_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ is an m-dimensional weakly stationary process such that $\mathbb{E} f_t = \mu$, and $\mathrm{Cov}(f_t) = \Sigma$. We also assume that the specific factors ε_{it} is white noise, uncorrelated with the common factors, and uncorrelated with other specific factors. That is,

$$\mathbb{E}\varepsilon_{it} = 0,$$

$$\mathrm{Cov}(f_{jt}, \varepsilon_{is}) = 0,$$

and

$$\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \sigma_i^2, & \text{if } i = j \text{ and } t = s, \\ 0, & \text{otherwise.} \end{cases}$$

We may rewrite the model as

$$r_t = \alpha + \beta f_t + \varepsilon_t \tag{9.1}$$

where $r_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})'$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$, β is a matrix whose (i, j)-th entry is β_{ij}, $f_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$, and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. We may further rewrite this model as

$$r_t = \lambda F_t + \varepsilon_t$$

where $\lambda = [\alpha, A]$, and $F_t = (1, f_t')'$.

We may use the relationship

$$\mathbb{E} r_t F_t' = \lambda \mathbb{E} F_t F_t'$$

to identify

$$\lambda = (\mathbb{E} F_t F_t')^{-1} \mathbb{E} r_t F_t'.$$

If the factors are known, that is, if F_t is known for each t, we may estimate the factor model by OLS. The OLS estimator of λ is given by

$$\hat{\lambda} = \left(\frac{1}{T} \sum_{t=1}^{T} F_t F_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} r_t F_t' \right).$$

9.2 Example: The Fama-French Three-Factor Model

Fama and French (1992) considered three factors in determining asset returns: the overall market return, the performance of small stocks relative to large stocks (small minus big, SMB), and the performance of value stocks relative to growth stocks (high minus low, HML).

9.3 Statistical Factors

It is also possible that we do not specify what the factors are but to extract factors statistically from the data. A statistical factor model is of the form

$$r_t - \mu = \beta f_t + \varepsilon_t$$

where μ is the mean of r_t, β is the same as in (9.1), and f_t is basically the factors in (9.1), but demeaned. Notice that in this model, we treat both β and f_t as unknown.

This model is not uniquely identified, because of that fact that $\beta f_t = (c\beta)(1/cf_t)$ for any $c \neq 0$. We need to impose some restrictions on f_t so that it is uniquely identified:

$$\mathbb{E} f_t = 0, \qquad \mathrm{Cov}(f_t) = I_m.$$

Similarly as before, assume that $\mathbb{E}\varepsilon_t = 0, \mathrm{Cov}(\varepsilon_t) = D = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$, and $\mathrm{Cov}(f_t, \varepsilon_t) = 0$.

The covariance matrix of the return r_t is

$$\Omega = \mathrm{Var}(r_t) = \mathrm{Var}(\beta f_t) + \mathrm{Var}(\varepsilon_t) = \beta\beta' + D.$$

To estimate the model, we use principal component analysis. We first obtain the sample covariance matrix of r_t:

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{\mu})(r_t - \hat{\mu})'$$

where

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t.$$

Then we calculate the eigen-pairs $(\hat{\lambda}_1, \hat{v}_1), (\hat{\lambda}_2, \hat{v}_2), \ldots, (\hat{\lambda}_N, \hat{v}_N)$ of $\hat{\Omega}$ so that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq$

$\hat{\lambda}_N$. Then the estimated factor loadings are given by

$$\hat{\beta} = \left[\sqrt{\hat{\lambda}_1}\hat{v}_1, \sqrt{\hat{\lambda}_2}\hat{v}_2, \cdots, \sqrt{\hat{\lambda}_m}\hat{v}_m\right].$$

The estimator of the factors are given by

$$\hat{f}_t = \begin{bmatrix} \hat{v}_1'/\sqrt{\hat{\lambda}_1} \\ \hat{v}_2'/\sqrt{\hat{\lambda}_2} \\ \cdots \\ \hat{v}_m'/\sqrt{\hat{\lambda}_m} \end{bmatrix} r_t.$$

To determine the number of factors in the statistical factor model, we employ the approach developed by Bai and Ng (2002). For a candidate number of factors m, we first estimate the m-factor model, get the loadings and factors, denoted by $\hat{\beta}^m$ and $\hat{f}_t^m$, respectively. Then we calculate

$$V(m) = \ln\left(\frac{1}{NT}\sum_{t=1}^{T}\iota'(r_t - \hat{\mu} - \hat{\beta}^m\hat{f}_t^m)\right) + m\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right).$$

We compute the V value for all candidate numbers of factors, and choose the m that minimizes the V value.