Quantitative Issue

2019. 4. 15

팩터 모델의 구조와 해설

Factor Models for Asset Returns

김동영, CFA Analyst dy76.kim@samsung.com 02 2020 7839
원동은 Research Associate de.won@samsung.com 02 2020 7982

신뢰에 가치로 답하다 삼성증권

Contents

I. 소개

팩터(요인) 모델이란 개별 자산 수익률이 여러 가지 요인들에 의해서 결정된다고 보는 모델이다. 팩터 모델은 포트폴리오 구성, 리스크 관리, 성과 평가 등 다양한 분야에서 사용되고 있다. 이런 팩터 모델의 활용은 해외뿐 아니라 국내에서도, 미들 오피스뿐 아니라 프론트 오피스에서도, 패시브뿐 아니라 액티브에서도 점점 대중화되는 추세다.

본 자료는 Eric Zivot의 Factor Models for Asset Returns 자료를 기초로 하여, 팩터 모델의 실제 구조에 대해 자세히 해설하려는 목적으로 작성되었다.[1]

이 자료가 과학적인 포트폴리오 운용에 조금이나마 도움이 되기를 기원한다.

[내용 Summary][2]

1. 요인 모델은 거시경제 요인 모델, 펀더멘털 요인 모델, 통계적 요인 모델로 나눌 수 있다.
2. 자산들의 기대수익률과 공분산 정보는, 요인 모델을 통해서 축약해서 계산될 수 있다.
3. 거시경제 요인 모델은 관측 가능한 거시경제 변수가 요인으로 사용된다.
4. 펀더멘털 요인 모델은 Size, Value, Momentum과 같은 관찰 가능한 자산 고유의 특성을 사용하는 모델이다.
5. 통계적 요인 모델은 통계 기법인 요인 분석 기법 및 주성분 분석 기법을 사용하는 모델이다.
6. 보통의 거시경제 요인 모델은 개별 자산의 전기간 수익률, 요인 실현값의 전기간 수익률이 주어져 있고, 시계열 회귀분석을 통해서 요인 베타를 추정한다(시계열 회귀분석 방식).
7. 펀더멘털 요인 모델은 BARRA 접근법(횡단면 회귀분석 방식)과 Fama-French 접근법(시계열 회귀분석 방식)으로 나뉜다.
8. BARRA 접근법은 한 시점의 전체 자산들의 수익률, 전체 자산들의 요인 베타가 주어져 있고, 횡단면 회귀분석을 통해서 해당 시점의 요인 실현값을 추정한다.
9. Fama-French 접근법은 요인별로 분위 포트폴리오를 만들어 명시적인 요인 실현값을 먼저 계산한다. 개별 자산의 전기간 수익률, 만들어낸 요인 실현값의 전기간 수익률이 주어져 있는 상태에서, 시계열 회귀분석을 통해서 요인 베타를 추정한다.

---

[1] 이후로는 'factor model' 용어를 '요인 모델'로 표기함. 자료의 수식에는, 행렬과 벡터를 bold체로 표시하는 일반적인 수식 표기법이 적용되어 있음

[2] 회귀분석의 행렬 접근법 (7p), 요인 모델 공분산 행렬 산식 (12p), 회귀분석의 OLS 추정법 증명 (18p) 내용에도 주목

1-1. 자산 수익률용 요인 모델(factor model for asset returns)의 활용법

- 리스크와 리턴을, 설명 가능한 요소들과 설명 불가능한 요소들로 분해
- 비정상 수익률(abnormal return: 초과수익률의 개념)의 측정치를 생성
- 수익률의 공분산 구조를 묘사
- 특정한 스트레스 시나리오에서의 수익률 예측
- 포트폴리오 위험 분석을 위한 체계 제공

1-2. 요인 모델의 3가지 형태

1. 거시경제 요인 모델 (macroeconomic factor model)
   (a) 요인이 관찰 가능한 경제 및 금융 시계열 데이터임
2. 펀더멘털 요인 모델 (fundamental factor model)
   (a) 요인이, 관찰 가능한 자산의 특성에서부터 시작함
   ※ 모델에 따라 횡단면 회귀분석 방식도 있고, 시계열 회귀분석 방식도 있음
3. 통계적 요인 모델 (statistical factor model)
   (a) 요인은 관찰되지 않으며 자산 수익률에서 추출됨

통계적 요인 모델은 통계 기법인 요인 분석(factor analysis)과 주성분 분석(principal component analysis)을 사용하는 모형을 말한다. 통계적 요인 모델은 각 자산의 과거 수익률 정보만을 input 으로 사용한다는 점에서 한계가 존재한다.

Contents

II. 요인 모델 구조

2-1. 요인 모델 구조 (Factor Model Specification)

자산 수익률 기반의 다요인 모델의 세 가지 형태는 다음의 일반 형식을 가진다.

$$
\begin{aligned}
R_{it} &= \alpha_i + \beta_{1i} f_{1t} + \beta_{2i} f_{2t} + \cdots + \beta_{Ki} f_{Kt} + \varepsilon_{it} \\
&= \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{f}_t + \varepsilon_{it} \qquad (1)
\end{aligned}
$$

여기서,

R_{it} = 자산 i(i=1,…,N)의 t시점(t=1,…,T)에서의 단순 수익률 (실질 혹은 무위험이자율 대비 초과수익률 형태)

f_{kt} = k번째 공통 요인 (k=1,…,K)

β_{ki} = k번째 요인에 대한 자산 i의 요인 적재값 혹은 요인 베타

ε_{it} = 자산의 개별 요인

위 식은 특정 자산 혹은 주식의 수익률을, 여러 자산에 공통적으로 영향을 끼치는 공통 요인의 수익률과 그에 대한 노출도 등으로 분해할 수 있다는 뜻이다.

예를 들어, 삼성전자 주식의 수익률을 1) GDP, 원달러 등 거시경제 요인들의 선형 결합으로 볼 수도 있고, 2) Value 팩터와 Momentum 팩터 등 펀더멘털 요인들의 선형 결합으로 볼 수도 있다.

1) 거시경제 요인 모델 예제: $R_{삼성전자} = \beta_{삼성전자,GDP} \cdot f_{GDP} + \beta_{삼성전자,원달러} \cdot f_{원달러} + \varepsilon_{삼성전자}$

이 때 $\beta_{삼성전자,GDP}$ = 삼성전자의 GDP 베타가 1, f_{GDP} = GDP 성장률이 3%,

$\beta_{삼성전자,원달러}$ = 삼성전자의 원달러 베타가 0.5, $f_{원달러}$ = 원달러 변화율이 -2%,

$\varepsilon_{삼성전자}$ = 삼성전자 개별 요인 수익률 1%라고 할 때,

삼성전자 수익률 3%는 $[R_{삼성전자} = 1 \times 3\% + 0.5 \times (-2\%) + 1\%]$와 같이 분해됨

2) 펀더멘털 요인 모델 예제: $R_{삼성전자} = \beta_{삼성전자,Value} \cdot f_{Value} + \beta_{삼성전자,Mom} \cdot f_{Mom} + \varepsilon_{삼성전자}$

이 때 $\beta_{삼성전자,value}$ = 삼성전자의 Value 익스포저가 0.3, f_{Value} = Value 팩터 수익률이 4%,

$\beta_{삼성전자,Mom}$ = 삼성전자의 모멘텀 익스포저가 -0.5, f_{Mom} = 모멘텀 팩터 수익률이 -2%,

$\varepsilon_{삼성전자}$ = 삼성전자 개별 요인 수익률 0.8%라고 할 때,

삼성전자 수익률 3%는 $[R_{삼성전자} = 0.3 \times 4\% - 0.5 \times (-2\%) + 0.8\%]$와 같이 분해됨

이 때 수익률 분해식을 행렬을 이용해서 표현할 수 있다.

보통 행렬 수식에서 prime 변수는 전치행렬을 말한다($\boldsymbol{A}'$= $\boldsymbol{A}$의 전치행렬= $\boldsymbol{A}$의 행과 열을 서로 바꾼 행렬). $\boldsymbol{\beta}_i'$는 자산 i의 각 요인들에 대한 요인 베타의 행벡터를 의미하며, $\boldsymbol{f}_t$는 공통 요인의 열벡터를 의미한다.[3]

즉, 앞의 거시경제 요인 모델 예제에서,

$R_{삼성전자} = \boldsymbol{\beta}_i'\boldsymbol{f}_t + \varepsilon_{it} = 3\% = [1 \quad 0.5]\begin{bmatrix} 3\% \\ -2\% \end{bmatrix} + 1\%$ 로 표현할 수 있다.

2-2. 요인 모델의 전제

1. 요인 실현값 $\boldsymbol{f}_t$가 안정적이다

$$E[\boldsymbol{f}_t] = \boldsymbol{\mu}_f$$

$$cov(\boldsymbol{f}_t) = E\left[(\boldsymbol{f}_t - \boldsymbol{\mu}_f)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)'\right] = \boldsymbol{\Omega}_f$$

2. 자산의 개별 오차항인 ε_{it}는 공통 요인 각각의 f_{kt}와 비상관이다

$$cov(f_{kt}, \varepsilon_{it}) = 0\,, \quad for\ all\ k, i, and\ t.$$

3. 오차항 ε_{it}는 시계열 상관이 없으며, 동시점에서 자산 간에 상관이 없다

$$cov(\varepsilon_{it}, \varepsilon_{js}) = \sigma_i^2\,, \quad for\ all\ i = j\ and\ t = s$$

$$= 0\,, \ otherwise$$

[3] 보통의 경우, 열벡터를 V, 행벡터를 V'(프라임) 형태로 표시함

2-3. 표기법

첨자 t를 가진 벡터는 모든 자산의 횡단면을 나타낸다

※ t 시점에서의 N개 전체 자산들의 수익률을 표시

$$\underset{(N\times 1)}{\boldsymbol{R}_t} = \begin{pmatrix} R_{1t} \\ \vdots \\ R_{Nt} \end{pmatrix}, \quad t = 1, \dots, T$$

첨자 i를 가진 벡터는 주어진 자산의 시계열을 나타낸다

※ 자산 i의 T개의 전체 시점에서의 수익률을 표시

$$\underset{(T\times 1)}{\boldsymbol{R}_i} = \begin{pmatrix} R_{i1} \\ \vdots \\ R_{iT} \end{pmatrix}, \quad i = 1, \dots, N$$

모든 자산의 모든 기간에 걸친 행렬 (열=자산들, 행=시점)

$$\underset{(T\times N)}{\boldsymbol{R}} = \begin{pmatrix} R_{11} & \cdots & R_{N1} \\ \vdots & \ddots & \vdots \\ R_{1T} & \cdots & R_{NT} \end{pmatrix}$$

2-4. 횡단면 회귀분석

다요인 모델 식 (1)은 각 자산에 대한 방정식을 쌓아서 시간 t에서의 횡단면 회귀분석 모델로 재작성될 수 있다

$$\underset{(N\times 1)}{\boldsymbol{R}_t} = \underset{(N\times 1)}{\boldsymbol{\alpha}} + \underset{(N\times K)}{\boldsymbol{B}}\underset{(K\times 1)}{\boldsymbol{f}_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t}, \quad t = 1, \dots, T \qquad (2)$$

$$\underset{(N\times K)}{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_N' \end{bmatrix} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{NK} \end{bmatrix}$$

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \boldsymbol{f}_t] = \boldsymbol{D} = diag(\sigma_1^2, \dots, \sigma_N^2)$$

※ diag는 해당 원소를 가지는 대각행렬을 뜻함

[회귀분석의 행렬 접근법]

y변수를 하나의 x변수로 설명하는 단순회귀분석 모형은 다음과 같이 표시된다.

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad i = 1,2,3,\dots,n$$

독립변수가 k개인 다중회귀분석 모형은 다음과 같이 표시된다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + +\beta_k x_{ki} + u_i \quad i = 1,2,3,\dots,n$$

여기서, β_0는 절편, β_1~β_k는 편기울기계수, u_i는 확률적 교란항, i는 i번째 관찰치, n은 샘플의 크기다.

이 식에서 n개의 관찰치를 모두 넣은 연립방정식은 다음과 같다.

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_k x_{k1} + u_1$$

$$y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_k x_{k2} + u_2$$

$$\dots \dots \dots$$

$$y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_k x_{kn} + u_n$$

이의 행렬 형태 표시는 다음과 같이 할 수 있다.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix} + \begin{bmatrix} x_{11} & \cdots & x_{k1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \qquad \text{(A)}$$

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underset{(n\times 1)}{\boldsymbol{\beta_0}} + \underset{(n\times k)}{\boldsymbol{X}} \; \underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{u}}$$

식 (A)가 다중회귀분석을 행렬로 표현한 식이다.

여기에서 절편에 해당하는 $\boldsymbol{\beta_0}$를 $\boldsymbol{X}$ 부분에 합칠 수 있다. 이를 적용하면 다음과 같다.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad \text{(B)}$$

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underset{(n\times(k+1))}{\boldsymbol{X}} \; \underset{((k+1)\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{u}}$$

식 (B)가 일반적인 다중회귀분석 모형의 행렬 표현식이다. 이 때 $\boldsymbol{y}$와 $\boldsymbol{X}$가 주어진 데이터이며 이를 가지고 회귀분석을 통해서 $\boldsymbol{\beta}$를 추정하게 된다.

앞의 박스 내용에 있는 횡단면 회귀분석 식 (2)에서는 $\boldsymbol{R}_t$와 $\boldsymbol{B}$가 주어진 데이터이며, 회귀분석을 통해 $\boldsymbol{f}_t$와 $\boldsymbol{\alpha}$를 추정하게 된다.

$$\underset{(N\times 1)}{\boldsymbol{R}_t} = \underset{(N\times 1)}{\boldsymbol{\alpha}} + \underset{(N\times K)}{\boldsymbol{B}}\underset{(K\times 1)}{\boldsymbol{f}_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t}$$

$$\begin{bmatrix} R_{1,t} \\ \vdots \\ R_{N,t} \end{bmatrix} = \begin{bmatrix} \alpha_t \\ \vdots \\ \alpha_t \end{bmatrix} + \begin{bmatrix} \beta_{11,t} & \cdots & \beta_{1K,t} \\ \vdots & \ddots & \vdots \\ \beta_{N1,t} & \cdots & \beta_{NK,t} \end{bmatrix} \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{K,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N,t} \end{bmatrix}$$

여기서, $\begin{bmatrix} R_{1,t} \\ \vdots \\ R_{N,t} \end{bmatrix}$는 t시점에서의 N개 자산 각각의 수익률, $\begin{bmatrix} \beta_{11,t} & \cdots & \beta_{1K,t} \\ \vdots & \ddots & \vdots \\ \beta_{N1,t} & \cdots & \beta_{NK,t} \end{bmatrix}$는 t시점에서의 개별 자산들의 개별 요인들에 대한 베타들이다. $\begin{bmatrix} f_{1,t} \\ \vdots \\ f_{K,t} \end{bmatrix}$는 t시점에서의 K개 요인 실현값이며 회귀분석에 의해서 추정되는 값이다. 즉, 이 식은 앞에서 본 회귀분석의 행렬 표현식과 동일하다.

따라서, 이 식은 t시점에서의 전체 자산들의 수익률을 종속변수로, 요인 베타 행렬을 독립변수로 하고, 요인 실현값을 추정하는 횡단면 회귀분석식에 해당한다.

2-5. 시계열 회귀분석

다요인 모델 식 (1)은 자산 i에 대한 관측치를 쌓음으로써 자산 i에 대한 시계열 모델로 재작성될 수 있다.

$$\underset{(T\times 1)}{\boldsymbol{R}_i} = \underset{(T\times 1)}{\boldsymbol{1}_T}\underset{(1\times 1)}{\alpha_i} + \underset{(T\times K)}{\boldsymbol{F}}\underset{(K\times 1)}{\boldsymbol{\beta}_i} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}_i}, \quad i = 1,\dots,N \tag{3}$$

$$\underset{(T\times K)}{\boldsymbol{F}} = \begin{bmatrix} \boldsymbol{f}_1' \\ \vdots \\ \boldsymbol{f}_T' \end{bmatrix} = \begin{bmatrix} f_{11} & \cdots & f_{K1} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{KT} \end{bmatrix}$$

$$E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'] = \sigma_i^2\boldsymbol{I}_T$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \alpha \\ \vdots \\ \alpha \end{bmatrix} + \begin{bmatrix} x_{11} & \cdots & x_{k1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{kn} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \tag{A}$$

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underset{(n\times 1)}{\boldsymbol{\alpha}} + \underset{(n\times k)}{\boldsymbol{X}}\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{u}}$$

식 (A)가 일반적인 다중회귀분석 모형을 행렬로 표현한 식이다. 이 때 $\boldsymbol{y}$와 $\boldsymbol{X}$가 주어진 데이터이며 이를 가지고 회귀분석을 통해서 $\boldsymbol{\beta}$와 $\boldsymbol{\alpha}$를 추정하게 된다.

시계열 회귀분석 식 (3)에서는 $\boldsymbol{R}_i$와 $\boldsymbol{F}$가 주어진 데이터이며, 회귀분석을 통해 $\boldsymbol{\beta}_i$와 α_i를 추정하게 된다.

$$\underset{(T\times 1)}{\boldsymbol{R}_i} = \underset{(T\times 1)}{\boldsymbol{1}_T}\underset{(1\times 1)}{\alpha_i} + \underset{(T\times K)}{\boldsymbol{F}}\underset{(K\times 1)}{\boldsymbol{\beta}_i} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}_i}$$

$$\begin{bmatrix} R_{i,1} \\ \vdots \\ R_{i,T} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} f_{11} & \cdots & f_{K1} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{KT} \end{bmatrix}\begin{bmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,K} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,1} \\ \vdots \\ \varepsilon_{i,T} \end{bmatrix}$$

여기서, $\begin{bmatrix} R_{i,1} \\ \vdots \\ R_{i,T} \end{bmatrix}$는 자산 i의 전기간에서의 수익률, $\begin{bmatrix} f_{11} & \cdots & f_{K1} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{KT} \end{bmatrix}$는 K개 요인들의 전기간에서의 수익률이다. $\begin{bmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,K} \end{bmatrix}$는 자산 i의 K개 요인들에 대한 요인 베타이며, 회귀분석에 의해서 추정되는 값이다.

이 식은 자산 i의 전기간에서의 수익률에 대한 시계열 회귀분석에 해당한다.

요인 모델 상에서 요인 베타와 요인 실현값을 관리하는 형식은, 횡단면 회귀분석 방식과 시계열 회귀분석 방식으로 나눠진다고 볼 수 있다.

2-6. 다변량 회귀분석 (시계열)

i=1,…,N의 데이터들을 합치게 되면, 시계열 회귀분석 모델 (3)을 다변량 회귀분석으로 표현할 수 있다

$$[\boldsymbol{R}_1, \dots, \boldsymbol{R}_N] = \boldsymbol{1}_T[\alpha_1, \dots, \alpha_N] + \boldsymbol{F}[\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_N] + [\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N]$$

혹은

$$\underset{(T\times N)}{\boldsymbol{R}} = \underset{(T\times 1)}{\boldsymbol{1}_T}\underset{(1\times T)}{\boldsymbol{\alpha}'} + \underset{(T\times K)}{\boldsymbol{F}}\underset{(K\times N)}{\boldsymbol{B}'} + \underset{(T\times N)}{\boldsymbol{E}}$$

$$= \boldsymbol{X\Gamma}' + \boldsymbol{E}$$

여기서, $\underset{(T\times(K+1))}{\boldsymbol{X}} = [\boldsymbol{1}_T \,\vdots\, \boldsymbol{F}]$, $\underset{((K+1)\times N)}{\boldsymbol{\Gamma}'} = \begin{bmatrix}\boldsymbol{\alpha}' \\ \boldsymbol{B}'\end{bmatrix}$

직전의 2-5절 시계열 회귀분석에서의 식 (3)은 자산 i 하나에 대한 시계열 회귀분석을 말한다.

$$\begin{bmatrix} R_{i,1} \\ \vdots \\ R_{i,T} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} f_{11} & \cdots & f_{K1} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{KT} \end{bmatrix}\begin{bmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,K} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,1} \\ \vdots \\ \varepsilon_{i,T} \end{bmatrix}$$

여기서, 절편 벡터를 행렬에 합치면 다음의 식으로 표시할 수 있다(2-4절의 해설에 있는 식 (B)와 동일).

$$\begin{bmatrix} R_{i,1} \\ \vdots \\ R_{i,T} \end{bmatrix} = \begin{bmatrix} 1 & f_{11} & \cdots & f_{K1} \\ \vdots & & \ddots & \vdots \\ 1 & f_{1T} & \cdots & f_{KT} \end{bmatrix}\begin{bmatrix} \alpha_i \\ \beta_{i,1} \\ \vdots \\ \beta_{i,K} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,1} \\ \vdots \\ \varepsilon_{i,T} \end{bmatrix}$$

이 식을 가지고 모든 자산(1,…,N)에 대한 회귀분석식을 옆으로 붙이면, 다변량 회귀분석식이 된다.

$$\begin{bmatrix} R_{1,1} & \cdots & R_{N,1} \\ \vdots & \ddots & \vdots \\ R_{1,T} & \cdots & R_{N,T} \end{bmatrix} = \begin{bmatrix} 1 & f_{11} & \cdots & f_{K1} \\ \vdots & & \ddots & \vdots \\ 1 & f_{1T} & \cdots & f_{KT} \end{bmatrix}\begin{bmatrix} \alpha_1 & \cdots & \alpha_N \\ \beta_{1,1} & \cdots & \beta_{N,1} \\ \vdots & \ddots & \vdots \\ \beta_{1,K} & \cdots & \beta_{N,K} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,1} & \cdots & \varepsilon_{N,1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1,T} & \cdots & \varepsilon_{N,T} \end{bmatrix}$$

$$\underset{(T\times N)}{\boldsymbol{R}} = \underset{(T\times(K+1))}{\boldsymbol{X}}\underset{((K+1)\times N)}{\boldsymbol{\Gamma}'} + \underset{(T\times N)}{\boldsymbol{E}}$$

2-7. 다변량 회귀분석 (횡단면)

또는, t=1,…,T의 데이터를 수집할 경우 횡단면 회귀분석 모델 (2)를 다변량 회귀분석으로 표현할 수 있다

$$[\boldsymbol{R}_1, \dots, \boldsymbol{R}_T] = [\boldsymbol{\alpha}, \dots, \boldsymbol{\alpha}] + \boldsymbol{B}[\boldsymbol{f}_1, \dots, \boldsymbol{f}_T] + [\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T]$$

혹은

$$\underset{(N\times T)}{\boldsymbol{R}'} = \underset{(N\times 1)}{\boldsymbol{\alpha}}\underset{(1\times T)}{\boldsymbol{1}_T'} + \underset{(N\times K)}{\boldsymbol{B}}\underset{(K\times T)}{\boldsymbol{F}'} + \underset{(N\times T)}{\boldsymbol{E}'}$$

$$= \boldsymbol{\Gamma}\boldsymbol{X}' + \boldsymbol{E}'$$

여기서, $\underset{((K+1)\times T)}{\boldsymbol{X}'} = \begin{bmatrix} \boldsymbol{1}_T' \\ \boldsymbol{F}' \end{bmatrix}$, $\underset{(N\times(K+1))}{\boldsymbol{\Gamma}} = [\boldsymbol{\alpha} \vdots \boldsymbol{B}]$

2-8. 기대수익률 분해

$$E[R_{it}] = \alpha_i + \boldsymbol{\beta}_i' E[\boldsymbol{f}_t]$$

- $\boldsymbol{\beta}_i' E[\boldsymbol{f}_t]$= 체계적인 리스크 요인에 의해 설명되는 기대수익률
- $\alpha_i = E[R_{it}] - \boldsymbol{\beta}_i' E[\boldsymbol{f}_t]$= 설명되지 않은 기대수익률 (비정상 수익률)

※ 비정상 수익률은 초과수익률의 개념으로도 볼 수 있음

기대수익률은 2-1절의 식 (1)에서 바로 구해진다.

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{f}_t + \varepsilon_{it} \tag{1}$$

2-9. 공분산 구조

횡단면 회귀분석을 사용할 때,

$$\underset{(N\times 1)}{\boldsymbol{R}_t} = \underset{(N\times 1)}{\boldsymbol{\alpha}} + \underset{(N\times K)}{\boldsymbol{B}}\underset{(K\times 1)}{\boldsymbol{f}_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t}, \quad t = 1, \dots, T$$

다요인 모델의 가정에 따르면, 자산 수익률들의 $(N\times N)$ 공분산 행렬은 다음과 같다

$$cov(\boldsymbol{R}_t) = \boldsymbol{\Omega}_{FM} = \boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}' + \boldsymbol{D} \tag{4}$$

주석: 식 (4)는 다음을 의미한다

$$var(R_{it}) = \boldsymbol{\beta}_i'\boldsymbol{\Omega}_f\boldsymbol{\beta}_i + \sigma_i^2$$

$$cov\left(R_{it}, R_{jt}\right) = \boldsymbol{\beta}_i'\boldsymbol{\Omega}_f\boldsymbol{\beta}_j$$

[요인 모델 상의 자산 공분산 행렬 산식 증명]

기본적인 변수들은 다음과 같다.

$$\boldsymbol{R}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$$

$$E[\boldsymbol{R}_t] = \boldsymbol{\alpha} + \boldsymbol{B}\cdot E[\boldsymbol{f}_t] \text{ (2-8절)}$$

$$E[\boldsymbol{f}_t] = \boldsymbol{\mu}_f, \quad cov(\boldsymbol{f}_t) = E\left[(\boldsymbol{f}_t - \boldsymbol{\mu}_f)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)'\right] = \boldsymbol{\Omega}_f \text{ (2-2절)}$$

$$cov(f_{kt}, \varepsilon_{it}) = 0, \quad for\ all\ k, i, and\ t \quad \text{(2-2절)}$$

$$E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\boldsymbol{f}_t] = \boldsymbol{D} = diag(\sigma_1^2, \dots, \sigma_N^2) \qquad \text{(2-4절)}$$

이 때, 자산의 $(N\times N)$ 공분산 행렬($\boldsymbol{\Omega}_{FM}$: 요인 모델 공분산)은 다음과 같다.

$$cov(\boldsymbol{R}_t) = E[(\boldsymbol{R}_t - E[\boldsymbol{R}_t])(\boldsymbol{R}_t - E[\boldsymbol{R}_t])']$$

$$= E[(\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t - \boldsymbol{\alpha} - \boldsymbol{B}\cdot E[\boldsymbol{f}_t])(\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t - \boldsymbol{\alpha} - \boldsymbol{B}\cdot E[\boldsymbol{f}_t])']$$

$$= E[(\boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\mu}_f) + \boldsymbol{\varepsilon}_t)(\boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\mu}_f) + \boldsymbol{\varepsilon}_t)']$$

$$= E[(\boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\mu}_f) + \boldsymbol{\varepsilon}_t)((\boldsymbol{f}_t - \boldsymbol{\mu}_f)'\boldsymbol{B}' + \boldsymbol{\varepsilon}_t')]$$

$$= E[\boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\mu}_f)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)'\boldsymbol{B}'] + E[\boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\mu}_f)\boldsymbol{\varepsilon}_t'] + E[\boldsymbol{\varepsilon}_t(\boldsymbol{f}_t - \boldsymbol{\mu}_f)'\boldsymbol{B}'] + E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t']$$

(여기서, $cov(f_{kt}, \varepsilon_{it}) = 0$ 이므로 $cov(\boldsymbol{f}_t, \boldsymbol{\varepsilon}_t) = E[(\boldsymbol{f}_t - \boldsymbol{\mu}_f)\boldsymbol{\varepsilon}_t'] = 0$임. 따라서 2, 3항은 0이 됨.)

$$= E[\boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\mu}_f)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)'\boldsymbol{B}'] + E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t']$$

$$= \boldsymbol{B}\cdot E[(\boldsymbol{f}_t - \boldsymbol{\mu}_f)(\boldsymbol{f}_t - \boldsymbol{\mu}_f)']\cdot\boldsymbol{B}' + E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t']$$

$$= \boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}' + \boldsymbol{D}$$

2-10. 포트폴리오 분석

$\boldsymbol{w} = (w_1, \dots, w_N)$를 포트폴리오 비중($w_i$=자산 i의 부의 비율)의 벡터라고 하자. $\boldsymbol{R}_t$가 단순 수익률의 $(N \times 1)$ 벡터라고 하면, 포트폴리오 수익률은 다음과 같다.

$$R_{p,t} = \boldsymbol{w}'\boldsymbol{R}_t = \sum_{i=1}^{N} w_i R_{it}$$

포트폴리오 요인 모델:

$$\boldsymbol{R}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t \;\; \Rightarrow$$

$$\text{포트폴리오 수익률 } R_{p,t} = \boldsymbol{w}'\boldsymbol{\alpha} + \boldsymbol{w}'\boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{w}'\boldsymbol{\varepsilon}_t = \alpha_p + \boldsymbol{\beta}_p'\boldsymbol{f}_t + \varepsilon_{p,t}$$

$$\text{여기서, } \alpha_p = \boldsymbol{w}'\boldsymbol{\alpha}, \;\; \boldsymbol{\beta}_p' = \boldsymbol{w}'\boldsymbol{B}, \;\; \varepsilon_{p,t} = \boldsymbol{w}'\boldsymbol{\varepsilon}_t$$

$$\text{포트폴리오 분산 } var(R_{p,t}) = \boldsymbol{\beta}_p'\boldsymbol{\Omega}_f\boldsymbol{\beta}_p + var(\varepsilon_{p,t}) = \boldsymbol{w}'\boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}'\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{D}\boldsymbol{w}$$

2-11. 능동적(Active) 포트폴리오와 고정적(Static) 포트폴리오

능동적 포트폴리오는 액티브한 자산배분 결정에 의해 시간에 따라서 변화하는 가중치를 가진다.

고정적 포트폴리오는 시간이 지나면서 고정된 가중치를 가진다(예: 동일 비중 포트폴리오).

요인 모델은 능동적과 고정적 포트폴리오 둘 다의 리스크를 분석하는 데 사용될 수 있다.

$\boldsymbol{R} = (R_1, \dots, R_N)$을 N개 자산의 수익률, $\boldsymbol{w} = (w_1, \dots, w_N)$를 N개 자산의 포트폴리오 비중이라고 할 때, 포트폴리오 수익률은 다음과 같다.

$$R_p = \boldsymbol{w}'\boldsymbol{R}$$

포트폴리오 분산은 다음과 같이 계산된다.

$$var(R_p) = E\left[(R_p - \overline{R_p})(R_p - \overline{R_p})'\right]$$

$$= E\left[(\boldsymbol{w}'\boldsymbol{R} - \overline{\boldsymbol{w}'\boldsymbol{R}})(\boldsymbol{w}'\boldsymbol{R} - \overline{\boldsymbol{w}'\boldsymbol{R}})'\right] = E[(\boldsymbol{w}'(\boldsymbol{R} - \overline{\boldsymbol{R}}))(\boldsymbol{w}'(\boldsymbol{R} - \overline{\boldsymbol{R}}))']$$

$$= E[(\boldsymbol{w}'(\boldsymbol{R} - \overline{\boldsymbol{R}}))((\boldsymbol{R} - \overline{\boldsymbol{R}})'\boldsymbol{w})] = \boldsymbol{w}'E[(\boldsymbol{R} - \overline{\boldsymbol{R}})(\boldsymbol{R} - \overline{\boldsymbol{R}})']\boldsymbol{w}$$

$$= \boldsymbol{w}'E[(\boldsymbol{R} - \overline{\boldsymbol{R}})(\boldsymbol{R} - \overline{\boldsymbol{R}})']\boldsymbol{w} = \boldsymbol{w}'\, cov(\boldsymbol{R})\boldsymbol{w}$$

여기서, 2-9절 식 (4)를 보면, $cov(\boldsymbol{R}) = \boldsymbol{\Omega}_{FM} = \boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}' + \boldsymbol{D}$이다. 따라서 다음의 결과가 나온다.

$$var(R_p) = \boldsymbol{w}'\, cov(\boldsymbol{R})\boldsymbol{w} = \boldsymbol{w}'\left(\boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}' + \boldsymbol{D}\right)\boldsymbol{w} = \boldsymbol{w}'\boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}'\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{D}\boldsymbol{w}$$

Contents

III. 거시경제 요인 모델

3-1. 거시경제 요인 모델 (Macroeconomic Factor Models)

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

$\boldsymbol{f}_t$ = 관측된 경제/금융 시계열

계량경제적 문제:

- 요인의 선택
- 시계열 회귀분석 기술을 사용하여, 요인 베타 $\boldsymbol{\beta}_i$와 잔차 분산 σ_i^2을 추정하기
- 요인의 관측된 역사적 데이터로부터, 요인 공분산 행렬 $\boldsymbol{\Omega}_f$을 추정하기

거시경제 요인 모델은, 자산(혹은 주식)의 수익률을 결정하는 요인이 외부의 거시경제 변수라고 보는 모델이다. 예를 들어, 삼성전자 주식의 수익률을 GDP, 원달러 환율 등의 거시경제 요인들의 선형 결합으로 본 앞의 예가 이에 해당한다.

거시경제 요인 모델 예제: $R_{삼성전자} = \beta_{삼성전자,GDP} \cdot f_{GDP} + \beta_{삼성전자,원달러} \cdot f_{원달러} + \varepsilon_{삼성전자}$

거시경제 요인 모델 상의 요인에는, 실제로 관측 가능한 경제 및 금융 변수 데이터가 사용된다. GDP 성장률, 원달러 환율 변화율, GDP 성장률의 서프라이즈 지표 등의 눈에 보이는 수치를 쓴다.

거시경제 요인 모델에서 결정해야 할 주요 프로세스는 첫째로는 어떤 경제 변수를 요인으로 사용할 것이냐는 점이다. 두 번째 프로세스는 시계열 회귀분석을 통해서 요인 베타를 구하는 일이다.

삼성전자의 원달러 환율 민감도가 얼마인가, 자동차 업종의 원달러 환율 민감도가 얼마인가 등의 답에 해당하는 거시경제 요인 베타는 시계열 회귀분석을 통해서 얻어진다. 예를 들어 삼성전자의 과거 5년간 주가수익률, 원달러 환율의 과거 5년간 변화율 데이터를 놓고, 이들에 대한 시계열 회귀분석을 통해서 삼성전자의 원달러 변수에 대한 요인 베타를 산출하는 방식이다.

3-2. Sharpe의 단일 요인 모델

Sharpe의 단일 요인 모델은, 하나의 '시장' 요인을 가지는 거시경제 요인 모델이다:

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it}, \quad i = 1, \dots, N; t = 1, \dots, T \tag{5}$$

여기서, R_{Mt}는 t시점에서 시장 지수(보통 S&P500 지수와 같은 시가총액 가중 지수)의 절대수익률 혹은 초과수익률(무위험 이자율 대비 상대수익률)을 나타낸다.

위험 조정 수익률과 비정상 수익률은 다음과 같다:

$$E[R_{it}] = \alpha_i + \beta_i E[R_{Mt}]$$

$$\alpha_i = E[R_{it}] - \beta_i E[R_{Mt}]$$

3-3. Sharpe의 단일 요인 모델: 자산의 공분산 행렬

$$\boldsymbol{\Omega}_{FM} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{D} \tag{6}$$

여기서,

$$\sigma_M^2 = var(R_{Mt})$$

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_N)'$$

$$\boldsymbol{D} = diag(\sigma_1^2, \dots, \sigma_N^2)$$

$$\sigma_i^2 = var(\varepsilon_{it})$$

2-9절 내용을 보면, 자산 수익률들의 $(N \times N)$ 공분산 행렬은 다음과 같다.

$$cov(\boldsymbol{R}_t) = \boldsymbol{\Omega}_{FM} = \boldsymbol{B}\boldsymbol{\Omega}_f \boldsymbol{B}' + \boldsymbol{D} \tag{4}$$

Sharpe의 단일 요인 모델에서는, 요인이 하나이므로 $\boldsymbol{f}_t$의 공분산 행렬이 스칼라 값이 된다. 따라서 아래와 같이 바꿔 쓸 수 있다.

$$\boldsymbol{\Omega}_f = \sigma_M^2 = var(R_{Mt})$$

$$cov(\boldsymbol{R}_t) = \boldsymbol{\Omega}_{FM} = \boldsymbol{B}\boldsymbol{\Omega}_f \boldsymbol{B}' + \boldsymbol{D} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{D}$$

3-4. Sharpe의 단일 요인 모델: 추정

R_{Mt}가 관측 가능하기 때문에, 각 자산의 단일 요인 모델 (5)의 파라미터 β_i와 σ_i^2는 OLS(보통최소제곱법, ordinary least squares) 방식으로 시계열 회귀분석을 통해 추정될 수 있다

$$\boldsymbol{R}_i = \hat{\alpha}_i \mathbf{1}_T + \boldsymbol{R}_M \hat{\beta}_i + \hat{\boldsymbol{\varepsilon}}_i, \qquad i = 1, \dots, N$$

$$\hat{\beta}_i = \frac{\widehat{cov}(R_{it}, R_{Mt})}{\widehat{var}(R_{Mt})} = \frac{\hat{\sigma}_{iM}}{\hat{\sigma}_M^2}$$

$$\hat{\alpha}_i = \bar{R}_i - \hat{\beta}_i \overline{R_M}$$

$$\hat{\sigma}_i^{\,2} = \frac{1}{T-2} \hat{\boldsymbol{\varepsilon}}_i{}' \hat{\boldsymbol{\varepsilon}}_i$$

추정된 단일 요인 모델의 공분산 행렬은 다음과 같다.

$$\widehat{\boldsymbol{\Omega}}_{FM} = \hat{\sigma}_M^2 \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\beta}}' + \widehat{\boldsymbol{D}}$$

보통최소제곱법(최소자승법, ordinary least squares)은 잔차의 제곱의 합을 최소화하는 기법으로, 일반적인 회귀식을 추정하는 방법이다.

계량경제학에서 보통 $\hat{\alpha}, \hat{\beta}$ 등의 hat 변수는 모델 추정에 의해 나온 변수의 추정량을 뜻한다.

$$\boldsymbol{R}_i = \hat{\alpha}_i \mathbf{1}_T + \boldsymbol{R}_M \hat{\beta}_i + \hat{\boldsymbol{\varepsilon}}_i, \qquad i = 1, \dots, N$$

위 식은, 하나의 자산 i에 대해 전기간(t=1,…,T) 데이터를 회귀분석식의 행렬식 형태로 표시한 것이다.

$$\hat{\beta}_i = \frac{\widehat{cov}(R_{it}, R_{Mt})}{\widehat{var}(R_{Mt})} = \frac{\hat{\sigma}_{iM}}{\hat{\sigma}_M^2}$$

위 식에서 $\hat{\beta}_i$는 β_i(자산 i의 마켓 베타)의 추정량을 말하며, 위 공식은 일반적인 단순회귀분석에서의 베타(기울기) 공식과 완전히 동일하다.

3-5. Sharpe의 단일 요인 모델: 부연설명 1)

1. 다변량 회귀분석을 사용함으로써, 계산 효율을 얻을 수 있다. 계수 α_i와 β_i와 잔차 공분산 σ_i^2는 다변량 회귀분석 모델에서 한 단계로 계산될 수 있다.

$$\boldsymbol{R} = \boldsymbol{X\Gamma}' + \boldsymbol{E}$$

$\boldsymbol{\Gamma}'$의 다변량 OLS 추정량은 다음과 같다.

$$\widehat{\boldsymbol{\Gamma}}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}$$

잔차 공분산 행렬의 추정치는 다음과 같다.

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T-2}\widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}}$$

여기서,

$\widehat{\boldsymbol{E}} = \boldsymbol{R} - \boldsymbol{X}\widehat{\boldsymbol{\Gamma}}'$는 다변량 최소 제곱 잔차의 행렬이다.

일반적인 일변량 다중회귀분석의 행렬 표현식은 다음과 같다(2-4절의 해설 참고).

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underset{\left(n\times(k+1)\right)}{\boldsymbol{X}}\ \underset{\left((k+1)\times 1\right)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{u}} \tag{B}$$

이 식에서 $\boldsymbol{\beta}$의 OLS 추정량 공식은 다음과 같다. 이에 대한 증명은 다음 페이지에 정리되어 있다.

$$\underset{((k+1)\times 1)}{\widehat{\boldsymbol{\beta}}} = \underset{((k+1)\times(k+1))}{(\boldsymbol{X}'\boldsymbol{X})^{-1}}\ \underset{((k+1)\times n)}{\boldsymbol{X}'}\ \underset{(n\times 1)}{\boldsymbol{y}} \tag{C}$$

2-6절에서 시계열 다변량 회귀분석 수식을 보인 적이 있다($\boldsymbol{R} = \boldsymbol{X\Gamma}' + \boldsymbol{E}$).

일변량 회귀분석의 OLS 추정처럼, 다변량 회귀분석의 OLS 추정도 동일한 방식으로 할 수 있다.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \quad \Rightarrow \quad \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\boldsymbol{R} = \boldsymbol{X\Gamma}' + \boldsymbol{E} \quad \Rightarrow \quad \widehat{\boldsymbol{\Gamma}}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}$$

[회귀분석의 OLS 추정법 증명]

일반적인 다중회귀분석의 행렬 표현식은 다음과 같다(2-4절의 해설 참고).

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{(B)}$$

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underset{(n\times(k+1))}{\boldsymbol{X}} \underset{((k+1)\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{u}}$$

$\boldsymbol{\beta}$의 추정에 사용하는, 보통최소제곱법(최소자승법, ordinary least squares)은 잔차의 제곱의 합을 최소화하는 방식이다.

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

잔차 제곱 합은 다음과 같이 쓸 수 있다.

$$\sum \hat{u}_i^2 = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

$$= (y' - (\boldsymbol{X}\hat{\boldsymbol{\beta}})')(y - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = (y' - \hat{\boldsymbol{\beta}}'\boldsymbol{X}')(y - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

$$= \mathbf{y}'\mathbf{y} - \mathbf{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

여기서, $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$는 스칼라이므로, $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$와 $\mathbf{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})'$는 동일하게 바꿔 쓸 수 있다.

$$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

잔차 제곱 합이 최소가 되기 위해서는 $\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$를 $\hat{\boldsymbol{\beta}}$로 미분한 값이 0이 되어야 한다.

$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$의 미분값을 보자.

$$\frac{\partial(\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}})}{\partial\hat{\boldsymbol{\beta}}} = \frac{\partial(\mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\partial\hat{\boldsymbol{\beta}}}$$

여기서, 임의의 행렬 $\boldsymbol{A}$, 열벡터 $\boldsymbol{x}$가 있을 때, $\frac{\partial(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}$ 이다(행렬의 정의에 따라 유도 가능). 그리고, $\boldsymbol{A}$가 대칭행렬($\boldsymbol{A} = \boldsymbol{A}'$)일 경우에는, $\frac{\partial(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$ 로 쓸 수 있다.

위의 미분식에서 $\boldsymbol{X}'\boldsymbol{X}$은 대칭행렬이므로, $\frac{\partial(\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}})}{\partial\hat{\boldsymbol{\beta}}} = 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$ 로 쓸 수 있다. 따라서 위의 미분식은 다음과 같다.

$$\frac{\partial(\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}})}{\partial\hat{\boldsymbol{\beta}}} = 0 - 2\boldsymbol{X}'\boldsymbol{y} - 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

위 방정식을 0이라고 놓으면,

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

역행렬이 존재한다는 전제 하에, $\boldsymbol{\beta}$의 OLS 추정량은 다음과 같다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \quad \text{(C)}$$

3-6. Sharpe의 단일 요인 모델: 부연설명 2)

2. 시계열 회귀분석에서 R^2는 "시장" 위험의, $1 - R^2$는 자산 개별 위험의 비중을 나타내는 수치다. 또한, $\hat{\sigma}_i$는 자산 개별 위험의 일반적인 크기에 대한 척도다. 분산의 분해는 다음과 같다.

$$var(R_{it}) = \beta_i^2 var(R_{Mt}) + var(\varepsilon_{it}) = \beta_i^2 \sigma_M^2 + \sigma_i^2$$

R^2는 다음과 같이 측정된다.

$$R^2 = \frac{\hat{\beta}_i^2 \hat{\sigma}_M^2}{\widehat{var}(R_{it})}$$

3. β_i와 σ_i^2를 추정하기 위해서 robust[4] 회귀분석 기술이 사용될 수 있다. 또한 σ_M^2의 robust 추정치도 계산 가능하다.

4. 단일 요인 공분산 행렬 식 (6)(3-3절 참조)은 시간에 대해서 일정하다. 이 가정은 좋은 가정이 아닐 수 있다. 시간에 따라서 식 (6)이 변화하도록 허용하는 여러 가지 방법이 있다. 일반적으로 β_i, σ_i^2, σ_M^2는 시간에 따라 변동할 수 있다.

$$\beta_i = \beta_{it}, \quad \sigma_i^2 = \sigma_{it}^2, \quad \sigma_M^2 = \sigma_{Mt}^2$$

시간 변동 베타를 포착하기 위해, 롤링 회귀분석이나 칼만 필터 기술이 사용될 수 있다. 조건부 이분산성을 포착하기 위해, GARCH 모델이 σ_{it}^2와 σ_{Mt}^2를 구하는 데 사용될 수 있다. 또한 β_i, σ_{it}^2, σ_{Mt}^2의 추정치를 계산하는 데 지수 가중치를 사용할 수 있다. 시간 변동 요인 모델 공분산 행렬은 다음과 같다.

$$\widehat{\boldsymbol{\Omega}}_{FM,t} = \hat{\sigma}_{Mt}^2 \widehat{\boldsymbol{\beta}}_t \widehat{\boldsymbol{\beta}}_t' + \widehat{\boldsymbol{D}}_t$$

[4] robust 회귀분석은 이상치의 영향을 적게 받도록 고안된 회귀분석을 말함

3-7. 일반적인 거시경제 다요인 모델

모델은 K개의 관측 가능한 경제변수를 사용한다.

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{f}_t + \varepsilon_{it}$$

- Chen, Roll and Ross (1986)는 주식을 위한 공통적으로 사용되는 거시경제 요인에 대한 설명을 제시했다. Lo (2008)는 헤지펀드에 대해 논의했다.
- 때때로, 거시경제 요인은 평균이 0이고 동일한 척도를 가지기 위해 표준화된다.
- 요인은 안정적이어야 한다(추세가 있으면 안됨).
- 때때로, 요인들은 직교하도록 만들어진다.

3-8. 일반적인 거시경제 다요인 모델: 추정

요인 실현값은 관측 가능하기 때문에, 파라미터 행렬 $\boldsymbol{B}$와 $\boldsymbol{D}$는 시계열 회귀분석을 사용하여 추정될 수 있다.

$$\boldsymbol{R}_i = \hat{\alpha}_i \boldsymbol{1}_T + \boldsymbol{F}\widehat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\varepsilon}}_i = \boldsymbol{X}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\varepsilon}}_i, \quad i = 1, \dots, N$$

$$\text{여기서, } \boldsymbol{X} = [\boldsymbol{1}_T \vdots \boldsymbol{F}], \ \ \hat{\boldsymbol{\gamma}} = \left(\hat{\alpha}_i, \widehat{\boldsymbol{\beta}}_i{}'\right)' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}_i$$

$$\hat{\sigma}_i^2 = \frac{1}{T-K-1}\hat{\boldsymbol{\varepsilon}}_i{}'\hat{\boldsymbol{\varepsilon}}_i$$

요인 실현값의 공분산 행렬은, 시계열 샘플 공분산 행렬을 사용하여 추정될 수 있다.

$$\widehat{\boldsymbol{\Omega}}_f = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{f}_t - \bar{\boldsymbol{f}})(\boldsymbol{f}_t - \bar{\boldsymbol{f}})', \quad \bar{\boldsymbol{f}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_t$$

추정된 다요인 모델 공분산 행렬은 다음과 같다.

$$\widehat{\boldsymbol{\Omega}}_{FM} = \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Omega}}_f\widehat{\boldsymbol{B}}' + \widehat{\boldsymbol{D}} \tag{7}$$

$\boldsymbol{R}_i = \boldsymbol{X}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\varepsilon}}_i$는 자산 i에 대한 t=1,…,T 구간의 시계열 회귀분석식을 의미한다(2-5절).

따라서, 추정량 $\hat{\boldsymbol{\gamma}}$은 일반적인 회귀분석 공식을 사용해 구할 수 있다. (3-5절의 해설 “회귀분석의 OLS 추정법 증명”을 참고)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \quad \Rightarrow \quad \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\boldsymbol{R}_i = \boldsymbol{X}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\varepsilon}}_i \quad \Rightarrow \quad \hat{\boldsymbol{\gamma}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}_i$$

3-9. 일반적인 거시경제 다요인 모델: 부연설명

1. 단일 요인 모델에서와 같이, $\boldsymbol{\beta}_i$, σ_i^2를 계산하기 위해 robust 회귀분석이 사용될 수 있다. $\boldsymbol{\Omega}_f$를 계산하고 추정하는 데 robust 공분산 행렬 추정량이 사용될 수 있다.

2. $\boldsymbol{\beta}_i$, $\boldsymbol{\Omega}_f$, σ_i^2 (i=1,...N)이 시간에 따라 변동하도록 허용함으로써, $\boldsymbol{\Omega}_{FM}$이 시간 변동할 수 있다.

예제: Berndt (1991)의 투자 자료를 이용한 R에서의 단일 지수 모델의 추정

Contents

IV. 펀더멘털 요인 모델

4-1. 펀더멘털 요인 모델 (Fundamental Factor Models)

펀더멘털 요인 모델은 공통 리스크 요인을 결정하기 위해, 산업 분류, 시가총액, 스타일 분류(가치, 성장) 등과 같은 관찰 가능한 자산 고유의 특성을 사용한다.

- BARRA 접근법에서, 요인 베타는 관측 가능한 자산의 특성으로 구성된다(즉, $\boldsymbol{B}$가 알려짐).
- BARRA 접근법에서, 요인 실현값 $\boldsymbol{f_t}$는 주어진 $\boldsymbol{B}$와 각 t시점에서 추정되거나 만들어 진다.
- 실무적으로, 펀더멘털 요인 모델은 2가지 방식으로 추정된다.

4-2. BARRA 접근법

- 이 접근법은 BARRA의 설립자인 Bar Rosenberg가 개척했으며, Grinold and Kahn (2000), Conner 외 (2010), Carino 외 (2010)에서 오랫동안 논의되었다.
- 이 접근법에서, 관찰 가능한 자산 고유의 펀더멘털(또는 그의 일부 변형)은 요인 베타, $\boldsymbol{\beta_i}$로 취급된다.
- 시간 t에서의 요인 실현값 $\boldsymbol{f_t}$는 관측 불가능하다. 그러면 계량경제적 문제는, 요인 베타가 주어졌을 때 시간 t에서의 요인 실현값을 추정하는 것이다. 이는 T개의 횡단면 회귀분석을 실행하여 수행된다.

MSCI Barra equity model 등은 BARRA 접근법의 요인 모델이다. BARRA 접근법에서는 자산(주식) 고유의 펀더멘털이 요인 베타로 사용된다. 예를 들자면, 개별 주식의 'PBR 수치', 혹은 '해당 PBR 수치가 전종목 분포 내에서 차지하는 z-score 수치' 등이 해당 주식의 Value 요인 베타(익스포저)가 되는 식이다. 이에 따라, 매일 혹은 매월의 각 주식들의 각 요인에 대한 요인 베타(익스포저)가 먼저 정해진다. 그 다음에 각 요인들의 매일 혹은 매월의 수익률이, 회귀분석을 통해서 산출된다.

특정 t시점에서 전종목의 수익률 $\boldsymbol{R_t}$는 눈으로 관찰 가능하므로 사전에 정해진다. t시점에서의 전종목의 요인 베타 $\boldsymbol{\beta}$도 관찰되는 값이므로 사전에 정해진다. BARRA 접근법 상에서 Value, Momentum, Liquidity와 같은 팩터들의 수익률 $\boldsymbol{f_t}$는 눈에 보이지 않는다. 주어진 $\boldsymbol{R_t}$가 종속변수이며, $\boldsymbol{\beta}$가 독립변수인 상황에서, 횡단면 회귀분석을 통해 t시점의 요인 수익률 $\boldsymbol{f_t}$가 정해진다.

예를 통해서 보자.

A, B, C의 3개 주식만을 가진 주식시장 하에서 Growth와 Dividend 2개 팩터만을 가진 BARRA 모델을 가정하면, 주식별 수익률 분해는 다음의 순서로 진행된다(4월 2일의 일일 데이터 생성 예제).

1. 주식 A, B, C의 4월 2일 수익률을 측정해 각각 4%, 1%, -4%가 나왔다.

2. 각 주식들의 당시 팩터 익스포저(베타)를 계산한다.

a) 먼저 Growth 팩터를 측정하는 기준(descriptor)으로 5년 EPS 성장률을 사용한다. 각 주식의 5년 EPS 성장률을 구한 다음, 전종목을 대상으로 한 z-score 값이 주식별 팩터 익스포저가 된다. 그 결과로 주식 A, B, C의 Growth 팩터 익스포저는 각각 0.7, -0.2, -0.5가 나왔다.

b) Dividend 팩터를 측정하는 기준(descriptor)로 현재 배당수익률을 사용한다. 각 주식의 배당수익률 수치를 구한 다음, 전종목을 대상으로 한 z-score 값을 주식별 팩터 익스포저로 쓴다. 최종적으로 주식 A, B, C의 Dividend 팩터 익스포저는 각각 0.1, -0.5, 0.4가 나왔다.

3. 4월 2일 기준의 주식별 수익률과 팩터 익스포저가 주어져 있으므로, 횡단면 회귀분석으로 4월 2일의 팩터 리턴을 산출할 수 있다(절편이 없는 식 (2) 형태).

$$\underset{(\mathrm{N}\times 1)}{\boldsymbol{R_t}} = \underset{(\mathrm{N}\times \mathrm{K})}{\boldsymbol{B}}\underset{(\mathrm{K}\times 1)}{\boldsymbol{f_t}} + \underset{(\mathrm{N}\times 1)}{\boldsymbol{\varepsilon_t}}, \quad t = 1, \dots, T$$

$$\begin{bmatrix} 4\% \\ 1\% \\ -4\% \end{bmatrix} = \begin{bmatrix} 0.7 & 0.1 \\ -0.2 & -0.5 \\ -0.5 & 0.4 \end{bmatrix} \boldsymbol{f_t} + \boldsymbol{\varepsilon_t}$$

$$\boldsymbol{\hat{f}_t} = (\boldsymbol{B'B})^{-1}\boldsymbol{B'R_t} = \begin{bmatrix} 5.8\% \\ -3.6\% \end{bmatrix}$$

4월 2일의 주식 A, B, C 수익률의 팩터별 분해는 다음과 같다.

$$\begin{bmatrix} 4\% \\ 1\% \\ -4\% \end{bmatrix} = \begin{bmatrix} 0.7 & 0.1 \\ -0.2 & -0.5 \\ -0.5 & 0.4 \end{bmatrix} \begin{bmatrix} 5.8\% \\ -3.6\% \end{bmatrix} + \begin{bmatrix} 0.3\% \\ 0.3\% \\ 0.3\% \end{bmatrix}$$

계산 결과, 4월 2일의 Growth 팩터 수익률은 5.8%로 계산되고, Dividend 팩터 수익률은 -3.6%로 계산된다. Growth 팩터 수익률이 5.8%로 높기 때문에 Growth 익스포저가 큰 A 주식의 최종 수익률이 4%로 높았던 것으로 해석할 수 있다. 그리고, Dividend 팩터 수익률이 -3.6%로 낮았기 때문에 Dividend에 플러스 익스포저를 가진 C 주식의 최종 수익률이 -4%로 낮았다고 해석할 수 있다. (실제로는 각 주식의 익스포저와 실제 수익률을 종합하여, 이를 잘 설명하는 팩터 리턴을 회귀분석으로 찾은 것이다.)

4. 위의 방법을 통해서, 일별 팩터 수익률을 횡단면 회귀분석을 통해서 하나씩 만들어낸다. 이 정보를 가지고 팩터별 공분산 행렬도 만들어낸다.

이상으로, 전체 주식의 일자별 수익률, 팩터 익스포저, 팩터별 수익률, 팩터 공분산 행렬 정보가 만들어지게 된다. 이를 포트폴리오를 구성에 활용할 경우, 우선 팩터 예상 수익률을 historical 수치를 써서 전망하거나 경제변수와의 관계 등을 분석해 전망할 수 있다. 그러면, 팩터별 예상 수익률을 기초로, 개별 주식의 예상 수익률과 포트폴리오의 예상 수익률을 모두 계산할 수 있다. 또한 포트폴리오 리스크도 팩터 공분산 행렬을 통해서 산출할 수 있게 된다.

4-3. Fama-French 접근법

- 이 접근법은 Eugene Fama and Kenneth French (1992)에 의해 도입되었다.
- 관찰된 자산 고유 특성(예로 사이즈)이 주어졌을 때, 그들은 두 단계 프로세스를 사용하여 요인 실현값을 결정했다. 먼저, 그들은 자산 고유 특성의 값에 기초하여 자산의 횡단면을 정렬했다. 그리고, 그들은 정렬된 자산들의 최상위 5분위를 long하고 최하위 5분위를 short하는 헤지 포트폴리오를 형성했다. 이 헤지 포트폴리오의 시간 t에서 관찰된 수익률이, 해당 자산 고유 특성에 대한 관측된 요인 실현값이다. 이 프로세스가 자산 고유의 각 특성별로 반복된다.
- t=1,…,T의 관찰된 요인 실현값이 주어졌을 때, 각 자산의 팩터 베타들은 N개의 시계열 회귀분석을 사용해서 추정된다.

BARRA와 달리, Fama-French 접근법은 요인 실현값을 먼저 구하는 방식이다. SMB 팩터 수익률, HML 팩터 수익률로 불리는 것들이 Fama-French 접근법에서 사용하는 자료들이다. Fama-French 접근법은, 우선 요인의 수익률을 눈에 보이도록 계산한다. 이 때 보통, 전종목을 특정 특성에 맞추어 정렬한 뒤, n개 분위로 나누어 최상위 분위를 long하고 최하위 분위를 short하는 헤지 포트폴리오를 구성해, 이 포트폴리오의 수익률을 해당 요인 수익률로 사용한다.

예를 들어, Fama and French (1993) 논문에서, SMB 요인 수익률은 시가총액 최하위 30% 종목군을 Long하고, 최상위 30% 종목군을 Short한 포트폴리오의 수익률로 만들어진다.

이렇게 개별 자산의 전기간 수익률이 준비되고 요인들의 전기간 수익률이 미리 계산되어 준비되면, 이를 바탕으로 개별 자산의 전기간 요인 베타들이 회귀분석을 통해 산출된다.

예를 들어, Fama and French (1993)에서는, 사이즈와 BE/ME 기준으로 형성된 25개 주식 포트폴리오를 준비하고, RM-RF, SMB, HML 3개 요인 수익률을 준비했다. 이를 가지고 각 (25개) 주식 포트폴리오의 수익률을 종속변수로 놓고 3개 요인 수익률을 독립변수로 놓은 상황에서, 각 주식 포트폴리오의 요인 베타를 계산하는 회귀분석을 25번 실시했다.

역시 예를 통해서 보자. 포트폴리오 기준으로 분석한 Fama and French (1993)와 다르게, 논의의 단순성을 위해 이번 예제에서는 개별 종목 기준으로 분석해보았다.

RM-RF, SMB, HML의 3개 팩터를 가진 가상의 Fama-French 모델에서, 주식별 수익률 분해는 다음의 순서로 진행된다.

1. 주식 A, B, C, D의 전기간 수익률을 측정한다.

일자	종목별 수익률 (%)			
	A	B	C	D
2010-04-06	-0.7	1.6	0.2	0.4
2010-04-07	1.7	-1.2	-1.9	-1.0
2010-04-08	-0.7	8.4	3.7	2.0
2010-04-09	0.0	0.7	0.6	0.6
2010-04-12	0.3	-3.2	-1.5	-1.4
2010-04-13	1.6	-0.2	1.9	2.6
…	…	…	…	…
2019-04-05	-0.7	-1.9	2.5	-1.2

2. 팩터 RM-RF, SMB, HML의 전기간 수익률을 측정한다.

이 때, RM-RF는 시장의 초과수익률을 나타내는 팩터로서, 코스피 지수 수익률에서 3년물 국채금리를 차감한 수치를 사용할 수 있다. SMB(Small Minus Big)는 소형주가 대형주를 아웃퍼폼하는 현상을 나타내는 팩터로서, 시가총액 하위 30%의 수익률 평균에서 상위 30% 수익률 평균을 차감한 수치를 사용할 수 있다. 소형주가 대형주보다 수익률이 좋았던 시기에는 SMB 팩터가 플러스 값을, 대형주가 소형주보다 수익률이 좋았던 시기에는 마이너스 값을 가지게 된다. 마지막으로 HML(High BE/ME Minus Low BE/ME)은 고BE/ME(=저PBR) 주식이 저BE/ME(=고PBR) 주식을 아웃퍼폼하는 현상을 나타내는 팩터로서, BE/ME 상위 30%의 수익률 평균에서 하위 30% 수익률 평균을 차감한 수치를 사용할 수 있다. Value가 잘 작동했던 시기에는 HML 팩터가 플러스 값을, 잘 작동하지 않았던 시기에는 마이너스 값을 가지게 된다.

일자	팩터별 수익률 (%)		
	RM-RF	SMB	HML
2010-04-06	-0.3	0.9	0.0
2010-04-07	0.6	-1.9	3.0
2010-04-08	0.3	2.6	0.0
2010-04-09	-0.3	-1.9	-0.1
2010-04-12	-0.1	0.0	0.7
2010-04-13	0.5	2.2	-1.3
…	…	…	…
2019-04-05	-0.2	-0.8	-0.7

3. 주식별 수익률과 팩터 수익률이 주어져 있으므로, 각 주식에 대해 시계열 회귀분석으로 전기간 β값(익스포저)을 추정할 수 있다. A주식을 예로 들어보면 다음과 같다. (절편이 없는 식 (3) 형태임. 이후로 수익률에서의 % 표기는 생략함.)

$$\underset{(T\times 1)}{\boldsymbol{R}_i} = \underset{(T\times K)}{\boldsymbol{F}}\underset{(K\times 1)}{\boldsymbol{\beta}_i} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}_i}, \quad i = 1, \dots, N$$

$$\begin{bmatrix} -0.7 \\ 1.7 \\ -0.7 \\ \vdots \\ -0.7 \end{bmatrix} = \begin{bmatrix} -0.3 & 0.9 & 0.0 \\ 0.6 & -1.9 & 3.0 \\ 0.3 & 2.6 & 0.0 \\ \vdots & \vdots & \vdots \\ -0.2 & -0.8 & -0.7 \end{bmatrix} \boldsymbol{\beta_A} + \boldsymbol{\varepsilon_A}$$

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{A}} = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{R}_{\boldsymbol{A}} = \begin{bmatrix} 1.3 \\ -0.1 \\ 0.1 \end{bmatrix}$$

A주식의 전기간 수익률을 각 팩터들로 분해해보면 다음과 같다.

$$\begin{bmatrix} -0.7 \\ 1.7 \\ -0.7 \\ \vdots \\ -0.7 \end{bmatrix} = \begin{bmatrix} -0.3 & 0.9 & 0.0 \\ 0.6 & -1.9 & 3.0 \\ 0.3 & 2.6 & 0.0 \\ \vdots & \vdots & \vdots \\ -0.2 & -0.8 & -0.7 \end{bmatrix} \begin{bmatrix} 1.3 \\ -0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} -0.2 \\ 0.4 \\ -0.7 \\ \vdots \\ -0.5 \end{bmatrix}$$

4. 위의 방법으로, B, C, D에 대해서도 시계열 회귀분석을 통해 전기간 팩터 익스포저를 만들어낸다. 이를 가지고 팩터별 공분산 행렬도 만든다.

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{A}} = \begin{bmatrix} 1.3 \\ -0.1 \\ 0.1 \end{bmatrix}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{B}} = \begin{bmatrix} 0.9 \\ 2.1 \\ -0.2 \end{bmatrix}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{C}} = \begin{bmatrix} 1.3 \\ 0.7 \\ 0.2 \end{bmatrix}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{D}} = \begin{bmatrix} 0.6 \\ 0.2 \\ -0.2 \end{bmatrix}$$

각 종목의 익스포저를 보면, B종목의 SMB 익스포저(2.1)가 다른 종목들에 비해 큰 것을 알 수 있다. 이에 따라 SMB가 잘 작동할 때, B종목의 수익률이 좋을 것이라고 예상할 수 있다. 실제로 SMB 팩터 수익률이 높았던(2.6%) 2010/4/8에, B종목이 다른 종목들에 비해 높은 수익률(8.4%)을 기록했다.

이상으로, 전종목의 일자별 수익률, 팩터 익스포저, 팩터별 수익률, 팩터 공분산 행렬 정보가 만들어진다. Fama-French 접근법에서는 개별 종목과 팩터의 수익률 관측값을 이용해 각 팩터에 대한 익스포저를 추정한다.

4-4. BARRA 형태 단일 요인 모델

단일 요인 모델을 시간 t에서의 횡단면 회귀분석의 형태로 고려해보자.

$$\underset{(N\times 1)}{\boldsymbol{R}_t} = \underset{(N\times 1)}{\boldsymbol{\beta}}\underset{(1\times 1)}{f_t} + \underset{(N\times 1)}{\boldsymbol{\varepsilon}_t}, \quad t = 1,\dots,T$$

- $\boldsymbol{\beta}$는 자산 고유 속성의 관측값의 N × 1 벡터다(예로 시가총액, 산업 분류, 스타일 분류)
- f_t는 미관측된 요인 실현값이다
- $var(f_t) = \sigma_f^2;\ cov(f_t, \varepsilon_{it}) = 0\ for\ all\ i,t;\ var(\varepsilon_{it}) = \sigma_i^2, \quad i = 1,\dots,N$

4-5. BARRA 형태 단일 요인 모델: 추정

t=1,…,T의 각 시점에서, 요인 베타 $\boldsymbol{\beta}$는 데이터로 취급되고, 요인 실현값 f_t는 추정해야 하는 파라미터다. 오차항 $\boldsymbol{\varepsilon_t}$가 이분산성을 가지므로, f_t의 효율적인 추정은 가중 최소 제곱법(WLS)[5]에 의해 수행된다(자산 고유 분산 σ_i^2가 알려져 있다고 가정할 때)

$$\hat{f}_{t,wls} = (\boldsymbol{\beta}'\boldsymbol{D}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{D}^{-1}\boldsymbol{R}_t, \quad t = 1,\dots,T \tag{8}$$

$$\boldsymbol{D} = diag(\sigma_1^2,\dots,\sigma_N^2)$$

주1: σ_i^2는 일치성[6]을 가지며 추정될 수 있고, feasible WLS 추정도 계산될 수 있다.

$$\hat{f}_{t,fwls} = (\boldsymbol{\beta}'\widehat{\boldsymbol{D}}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\widehat{\boldsymbol{D}}^{-1}\boldsymbol{R}_t, \quad t = 1,\dots,T$$

$$\widehat{\boldsymbol{D}} = diag(\hat{\sigma}_1^2,\dots,\hat{\sigma}_N^2)$$

주2: σ_i^2 이외의 다른 가중치가 사용될 수도 있다.

이분산성을 가정하면, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{D},\ \ \boldsymbol{D} = diag(\sigma_1^2,\dots,\sigma_N^2)$이라고 쓸 수 있다.

WLS가 아닌 간단한 OLS의 추정해는 다음과 같다. (3-5절의 해설 "회귀분석의 OLS 추정법 증명"을 참고)

$$\hat{f}_t = (\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{R}_t$$

이분산 하에서 WLS 방식의 추정해는 $\hat{f}_{t,wls} = (\boldsymbol{\beta}'\boldsymbol{D}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{D}^{-1}\boldsymbol{R}_t$로 나타난다.

5 WLS와 feasible WLS는 오차 분산에 이분산성이 있을 때 사용하는 회귀분석 기법임
6 통계에서 consistency(일치성)는, 표본의 크기가 늘어날 때 추정량이 모수 값으로 계속 수렴하는 특성을 말함

4-6. BARRA 형태 업종 요인 모델

K개의 상호 배타적인 업종들과 함께 스타일화된 BARRA 형태 업종 요인 모델을 고려할 수 있다. 각 자산에서 식 (1)의 요인 민감도 β_{ik}는 시간 불변이며 다음의 형태다.

$$\beta_{ik} = 1\ if\ asset\ i\ is\ in\ industry\ k$$
$$= 0, otherwise$$

그리고 f_{kt}는 시점에서 k번째 업종의 요인 실현값을 나타낸다.

- 요인 베타는 주어진 자산이 특정한 업종에 포함되어 있는지 여부를 나타내는 더미 변수다.
- f_{kt}의 추정값은 k번째 업종에서 운영되는 기업의 t시점에서의 가중 평균 초과수익률과 같다.

4-7. 업종 요인 모델 회귀분석

K개 업종을 가진 업종 요인 모델은 다음과 같이 요약된다.

$$R_{it} = \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \varepsilon_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T$$

$$var(\varepsilon_{it}) = \sigma_i^2, \quad i = 1, \ldots, N$$

$$cov(\varepsilon_{it}, f_{jt}) = 0, \quad j = 1, \ldots, K; i = 1, \ldots, N$$

$$cov(f_{it}, f_{jt}) = \sigma_{ij}^f, \quad i, j = 1, \ldots, K$$

여기서

$$\beta_{ik} = 1\ if\ asset\ i\ is\ in\ industry\ k\ (k = 1, \ldots, K)$$
$$= 0, otherwise$$

k번째 업종에 N_k개 기업이 있을 때, $\sum_{k=1}^{K} N_k = N$이라고 가정한다.

4-8. 업종 요인 모델 상의 요인에 대한 추정

t시점의 횡단면 회귀분석을 고려하자

$$\boldsymbol{R}_t = \boldsymbol{\beta}_1 \boldsymbol{f}_{1t} + \cdots + \boldsymbol{\beta}_K \boldsymbol{f}_{Kt} + \boldsymbol{\varepsilon}_t$$

$$= \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$$

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t] = \boldsymbol{D}, \quad cov(\boldsymbol{f}_t) = \boldsymbol{\Omega}_f$$

업종들은 상호 배타적이기 때문에, 이는 다음을 따른다.

$$\boldsymbol{\beta}_j{}' \boldsymbol{\beta}_k = N_k \; for \; j = k, \quad 0 \; otherwise$$

요인 실현값 $\boldsymbol{f}_t$의, 불편성[7]을 가지지만 효율성[8]은 없는 추정치는 OLS를 통해서 얻을 수 있다.

$$\hat{\boldsymbol{f}}_{t,OLS} = (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{R}_t = \begin{pmatrix} \hat{f}_{1t,OLS} \\ \vdots \\ \hat{f}_{Kt,OLS} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{N_1}\sum_{i=1}^{N_1} R_{it}^1 \\ \vdots \\ \dfrac{1}{N_K}\sum_{i=1}^{N_K} R_{it}^K \end{pmatrix}$$

$\boldsymbol{R}_t = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$는 일반적인 회귀분석식의 모양을 따른다.

따라서, $\boldsymbol{R}_t = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$의 $\boldsymbol{f}_t$ OLS 추정량 수식은 다음과 같다. (3-5절의 해설 “회귀분석의 OLS 추정법 증명”을 참고)

$$\hat{\boldsymbol{f}}_{t,OLS} = (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{R}_t$$

7 통계에서 불편성(unbiasedness)은 추정량의 기댓값이 모수와 동일한 것을 의미함
8 통계에서 효율성(efficiency)은 해당 추정량이 최소분산과 불편성을 가지는 것을 의미함

4-9. BARRA 형태 업종 요인 모델: 요인 실현값 공분산 행렬의 추정

주어진 $(\hat{\boldsymbol{f}}_{1,OLS}, \dots, \hat{\boldsymbol{f}}_{T,OLS})$에서, 업종 요인의 공분산 행렬은 시계열 샘플 공분산으로 계산될 수 있다.

$$\widehat{\boldsymbol{\Omega}}_{OLS}^{F} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{\boldsymbol{f}}_{t,OLS} - \bar{\boldsymbol{f}}_{OLS}\right)\left(\hat{\boldsymbol{f}}_{t,OLS} - \bar{\boldsymbol{f}}_{OLS}\right)'$$

여기서, $\bar{\boldsymbol{f}}_{OLS} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{f}}_{t,OLS}$

4-10. BARRA 형태 업종 요인 모델: 잔차 분산의 추정

잔차 분산, $var(\varepsilon_{it}) = \sigma_i^2$은 다음과 같이 T개의 횡단면 회귀분석에서의 잔차의 시계열로부터 추정할 수 있다. $\hat{\boldsymbol{\varepsilon}}_{t,OLS}$, $t = 1, \dots, T$가 OLS 잔차의 $(N \times 1)$ 벡터를 나타내고, $\hat{\varepsilon}_{it,OLS}$가 $\hat{\boldsymbol{\varepsilon}}_{t,OLS}$의 i번째 행을 나타낸다고 하자. 그러면, σ_i^2는 다음과 같이 추정된다.

$$\hat{\sigma}_{i,OLS}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{\varepsilon}_{it,OLS} - \bar{\varepsilon}_{i,OLS})^2, \quad i = 1, \dots, N$$

$$\bar{\varepsilon}_{i,OLS} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it,OLS}$$

4-11. 업종 요인 모델의 자산 수익률 공분산 행렬의 추정

N개 자산의 공분산 행렬은 다음을 사용하여 추정된다.

$$\widehat{\boldsymbol{\Omega}}_{OLS} = \boldsymbol{B}\widehat{\boldsymbol{\Omega}}^{F}_{OLS}\boldsymbol{B}' + \widehat{\boldsymbol{D}}_{OLS}$$

여기서 $\widehat{\boldsymbol{D}}_{OLS}$는 대각선에 $\hat{\sigma}^2_{i,OLS}$을 가지는 대각 행렬이다.

4-12. 가중 최소제곱 추정법

- 요인 실현값 $\boldsymbol{f}_t$의 OLS 추정치는 자산 수익률에서의 횡단면 이분산성으로 인해 비효율적이다.
- 잔차 분산의 추정치는 가중최소제곱(feasible GLS)의 가중치로 사용될 수 있다.

$$\hat{\boldsymbol{f}}_{t,GLS} = (\boldsymbol{B}'\widehat{\boldsymbol{D}}^{-1}_{OLS}\boldsymbol{B})^{-1}\boldsymbol{B}'\widehat{\boldsymbol{D}}^{-1}_{OLS}\boldsymbol{R}_t, \quad t = 1, \dots, T$$

$$\widehat{\boldsymbol{\Omega}}^{F}_{GLS} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{\boldsymbol{f}}_{t,GLS} - \bar{\boldsymbol{f}}_{GLS}\right)\left(\hat{\boldsymbol{f}}_{t,GLS} - \bar{\boldsymbol{f}}_{GLS}\right)'$$

$$\hat{\sigma}^2_{i,GLS} = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{\varepsilon}_{it,GLS} - \bar{\varepsilon}_{i,GLS})^2, \quad i = 1, \dots, N$$

$$\widehat{\boldsymbol{\Omega}}_{GLS} = \boldsymbol{B}\widehat{\boldsymbol{\Omega}}^{F}_{GLS}\boldsymbol{B}' + \widehat{\boldsymbol{D}}_{GLS}$$

위 4-11절의 공분산 행렬 내용은, 2-9절의 식 (4)와 동일한 구조다.

$$cov(\boldsymbol{R}_t) = \boldsymbol{\Omega}_{FM} = \boldsymbol{B}\boldsymbol{\Omega}_f\boldsymbol{B}' + \boldsymbol{D} \tag{4}$$

위 4-12절의 $\hat{\boldsymbol{f}}_{t,GLS}$ 수식은 4-5절의 식 (8)과 동일한 구조다.

$$\hat{f}_{t,wls} = (\boldsymbol{\beta}'\boldsymbol{D}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{D}^{-1}\boldsymbol{R}_t, \quad t = 1, \dots, T \tag{8}$$

$$\boldsymbol{D} = diag(\sigma^2_1, \dots, \sigma^2_N)$$

Compliance notice

- 본 보고서는 철저히 계량적 분석에 근거한 의견을 제시합니다. 따라서 당사의 대표 투자의견과 다를 수 있습니다.
- 본 조사분석자료의 애널리스트는 4월 12일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 4월 12일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자가 등 제 3자에게 사전 제공된 사실이 없습니다.

삼성증권 SAMSUNG

삼성증권주식회사
06620 서울특별시 서초구 서초대로 74길 11 10층 리서치센터
02 2020 8000

지점 대표번호
1588 2323 / 1544 1544

고객 불편사항 접수
080 911 0900

samsungPOP.com

신뢰에 가치로 답하다