1 Factor Strength

The concept of factor strength in this project was introduced by Pesaran and Smith (2019). In general, the factor strength α represents the pervasiveness of factor's loading.

1.1 Definition

Consider the following multi-factor model for n different assets and T observations with k various risk factors.

$$r_{it} - r_{ft} = a_{it} + \beta_{im}(r_{mt} - r_{ft}) + \sum_{j=1}^{k} \beta_{ij} f_{jt} + \varepsilon_{it} \tag{1}$$

In the left-hand side, we have r_{it} denotes the return of security i at time t, where $i = 1, 2, 3, \cdots, n$ and $t = 1, 2, 3, \cdots T$. r_{ft} denotes the risk free rate at time t. In the other hand, a_{it} is the constant term. r_{mt} is the market average return and therefore, $(r_{mt} - r_{ft})$ is the excess return of the market, or market factor. Corresponding β_{im} is the loading of market excess return. f_{jt} of $j = 1, 2, 3 \cdots k$ is potential risk factor. ε_{it} is the stochastic error term.

The factor strengthα_j of factor f_{jt} as is dependent on how many significant factor loadings it can generate. The more statistically non-zero loading a factor can generate, the stronger the factor is. When we have n assets, there are $[n^{\alpha}]$ are not zero, $[\cdot]$ denotes the integer number operator, which will take the integer part of number inside. For factor f_j with loading β_j, we assume it should have:

$$|\beta_j| > c_p(n) \quad i = 1, 2, \ldots, [n^{\alpha_j}]$$

$$|\beta_j| = 0 \quad i = [n^{\alpha_j}] + 1, [n^{\alpha_j}] + 2, \ldots, n$$

The factor loadings of first $[n^{\alpha_j}]$ terms are all bigger than critical value $c_p(n)$, this indicates that those factors are all significantly different from zero. Then the rest $n - [n^{\alpha_j}]$ term are equal to zero, which means the factors can not pricing the risk for those $n - [n_j^{\alpha}]$ assets. For a factor has strength $\alpha = 1$, factor loading will be significant for every assets at every time. And if we have factor strength $\alpha = 0$, it means that the factor cannot generate any loading different from zero, in other words the factor can not pricing risk of any assets.

1.2 Estimation

To estimate the α, Bailey, Kapetanios, and Pesaran (2020) provides a consistent method.

Here we consider a simplified model (1), the CAPM model with only one factor named f with different value f_t at different time. For simplify, we use $x_{it} := r_{it} - r_{ft}$ to denoted the left hand side of the model (1), the excess return of assets, β_{it} is the factor loading of assets i at time t. v_{it} is the stochastic error term.

$$x_{it} = a_{it} + \beta_{it} f_t + v_{it} \tag{2}$$

Assume we have n different assets and T observations for each assets: $i = 1, 2, 3, \cdots, n$ and $t = 1, 2, 3, \cdots T$. Running the OLS regression for each $i = 1, 2, 3 \cdots, n$, to obtain:

$$x_{it} = \hat{a}_{iT} + \hat{\beta}_{iT} f_t + \hat{v}_{it}$$

For every factor loading $\hat{\beta}_{iT}$, we can exam their significance by constructing a t-test. The t-test statistic will be $t_{iT} = \frac{\hat{\beta}_{iT} - 0}{\hat{\sigma}_{iT}}$. Then the test statistic for the corresponding $\hat{\beta}_i$ will be:

$$t_{iT} = \frac{(\mathbf{f}'\mathbf{M}_\tau \mathbf{f})^{1/2} \hat{\gamma}_{iT}}{\hat{\sigma}_{iT}} = \frac{(\mathbf{f}'\mathbf{M}_\tau \mathbf{f})^{-1/2} (\mathbf{f}'\mathbf{M}_\tau \mathbf{x}_i)}{\hat{\sigma}_{iT}}$$

Here, the $\mathbf{M}_\tau = \mathbf{I}_T - T^{-1}\tau\tau'$, and the τ is a $T \times 1$ vector with every elements equals to 1. $\mathbf{f}$ and $\mathbf{x_i}$ are two vectors with: $\mathbf{f} = (f_1, f_2 \cdots, f_T)'$ $\mathbf{x_i} = (x_{i1}, x_{i2}, \cdots, x_{iT})$. The denominator $\hat{\sigma}_{iT} = \frac{\sum_{i=1}^{T} \hat{v}_{it}^2}{T}$.

With those test statistic, we then defined π_{nT} as the proportion of significant factor loading amount to the total factor loadings amount:

$$\hat{\pi}_{nT} = \frac{\sum_{i=1}^{n} \hat{\ell}_{i,nT}}{n} \tag{3}$$

$\ell_{i,nT}$ is an indicator function as: $\ell_{i,nT} := \mathbf{1}[|t_{it}| > c(n)]$. If the t-statistic t_{jt} is greater than the critical value $c_p(n)$, $\hat{\ell}_{i,nT} = 1$. In other word, we will count one if the factor loading $\hat{\beta}_{ij}$ is significant. $c_p(n)$ represent the critical value of a test with test size p. The critical value is calculated by:

$$c_p(n) = \Phi^{-1}(1 - \frac{p}{2n^\delta}) \tag{4}$$

Here, $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal distribution, and δ is a non-negative value represent the critical value exponent. The traditional method to calculate critical value $\Phi^{-1}(1-\frac{p}{2})$ does not take multiple-test problem into the consideration. Here, by adding the n^{δ} term into the denominator, the new critical value has adjusted the multiple-test problem. Therefore, we will use this critical value $c_p(n)$ to compare with the test statistic to justify does the factor loading is significantly different from zero.

After obtain the $\hat{\pi}_{nT}$, we can use the following formula provided by Bailey et al. (2020) to estimate our strength indicator α_j:

$$\hat{\alpha} = \begin{cases} 1 + \frac{\ln(\hat{\pi}_{nT})}{\ln n} & \text{if } \hat{\pi}_{nT} > 0, \\ 0, & \text{if } \hat{\pi}_{nT} = 0. \end{cases}$$

Whenever we have $\hat{\pi}_{nT}$, the estimated $\hat{\alpha}$ will be equal to zero. From the estimation, we can find out that $\hat{\alpha} \in [0,1]$

2 Monte Carlo Design

In order to study the limited sample property of factor strength α, we designed several Monte Carlo simulations, to observe the behaviours of the factors strength under different settings. We consider the following data generating process (DGP): a CAPM model with market factor and risk factor

$$x_{it} = q_1(r_{mt} - r_f) + q_2(\sum_{j=1}^{k} \beta_{ij} f_{jt}) + \varepsilon_{it}$$

In the simulation, we consider a dataset has $i = 1,2,\ldots,n$ different assets, with $t = 1,2,\ldots,T$ different observations. k different risk factors represents by the subscript j and one market factors are also included.

$q_1(\cdot)$ and $q_2(\cdot)$ are two different functions represent the unknown mechanism of market factor and other risk factors in pricing asset risk. $(r_{mt} - r_f)$ is the excess market return, calculated by average market return r_{mt} minus risk free return r_{ft}. r_{it} is the stock return, f_{jt} denotes factors other than market factors and β_{ij} is the corresponding factor loading. ε_{it} is random error with structure can be defined in different designs. For each factor, we assume they follow a multinomial distribution

with mean zero and a $k \times k$ variance-covariance matrix Σ.

$$\mathbf{f_t} = \begin{pmatrix} f_{i,t} \\ f_{2,t} \\ \vdots \\ f_{k,t} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma) \quad \Sigma := \begin{pmatrix} \sigma_{f_1}^2, & \rho_{12}\sigma_{f1}\sigma_{f2} & \cdots & \rho_{1k}\sigma_{f1}\sigma_{fk} \\ \rho_{12}\sigma_{f2}\sigma_{f1}, & \sigma_{f2}^2 & \cdots & \rho_{2k}\sigma_{f2}\sigma_{fk} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1k}\sigma_{fk}\sigma_{f1}, & \rho_{k2}\sigma_{fk}\sigma_{f2} & \cdots & \sigma_{fk}^2 \end{pmatrix}$$

The diagonal of matrix Σ indicates the variance of each factor, and the rest represent the correlation among all k factors. In this model, we can control several parts to investigates different scenarios of the simulation:

2.1 Baseline Design

Follow the general model above, we assume both $q_1(a)$ and $q_2(a)$ are linear function:

$$q_1(r_{mt} - r_f) = a_{it} + \beta_{im}(r_{mt} - r_{ft})$$

$$q_2(\sum_{j=1}^{k} \beta_{ij} f_{jt}) = \sum_{j=1}^{k} \beta_{ij} f_{jt}$$

Therefore, if we include the market factor with other risk factors, the model with single factor can be write as:

$$x_{it} = a_{it} + \sum_{j=1}^{k+1} \beta_{ij} f_{ij} + \varepsilon_{it} \tag{5}$$

To generate factor loadings and asset's return, we follow the next procedures: First, we generate the constant term a_{it} which has a uniform distribution from -0.5 to 0.5, $a_{it} \sim U[-0.5, 0.5]$. Then, in this baseline design, we assume the error term $\varepsilon_{it} \sim N(0,1)$, this means that the error term has mean zero and variance equals to one . Next, we will set up the true factor strength α. Because here we allocate the market factor which as factor strength equals to one at the top of every factor, the factor strength vector will have k+1 elements:

$$\alpha = \begin{pmatrix} 1, & \alpha_2, & \alpha_3, & \cdots, & \alpha_{k+1} \end{pmatrix}$$

The other strengths will be adjusted to various values base on the simulation's design. After hav-

ing the factor strength, we can calculate for each factors, how many loadings will be significantly different from zero. Since we assume for any random factors j with strength α_j, it will have $[n_j^\alpha]$, so we

For each factors, we assume they follow the multinomial standard distribution with mean 0 and variance Σ, but in this baseline design, each factors are independent with each others, and each factors has same variance as one, so the Σ will be:

$$\Sigma = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

After that, we will generate the factor loadings from a uniform distribution. In order to make sure every factor loading is sufficiently larger than 0, we set the expected value of those loadings $\mu_\beta = 0.71$, $\beta_{ij} \sim IIDU(\mu_\beta - 0.2, \mu_\beta + 0.2)$. Then we randomly assign $n - [n^\alpha]$ factor loadings as zero, to reflect the fact that only $[n^\alpha]$ factor loadings are significantly different from zero. After generate constant term, factor, factor loading, and the error term, we can calculate the simulated asset's return by using the equation (5). With the return and factors, we can re-calculate the factors loading and use the estimation method discussed in section 1.2.

For this baseline design, we consider the different combinations of T and n with $T = \{120, 240, 360\}$, $n = \{100, 300, 500\}$. The market factor will have strength $\alpha_m = 1$ all the time, and the strength of the other factor will be $\alpha_x = \{0.5, 0.7, 0.9, 1\}$. For every setting, we will replicate 500 times independently, all the constant a_{it} and loading β_i will be re-generated for each replication. To exam the goodness of estimation, we calculate the bias between our true underneath factor strength α and the estimated strength $\hat{\alpha}$ as $bias = |\alpha - \hat{\alpha}|$. We also use the bias to calculate the Mean Square Error (MSE). To calculate the MSE, we will collect the bias for each replication, and then use the formula:

$$MSE = \frac{1}{n} \sum_{i=1}^{500} (bias_i)^2$$

References

Bailey, N., Kapetanios, G., & Pesaran, M. H. (2020). Measurement of factor strength: Theory and practice. *CESifo Working Paper*.

Pesaran, M. H., & Smith, R. P. (2019). The role of factor strength and pricing errors for estimation and inference in asset pricing models. *CESifo Working Paper Series*.