Understanding of Factor Strength

Zhiyuan Jiang
I.D:28710967

May 3, 2020

Contents

1 Introduction and Motivation

Capital Asset Pricing Model (CAPM), created by Sharpe (1964) and Lintner (1965) is one of the most famous, and wildly used model to explain the relationship between financial asset's risk and return, especially for the securities. The original model only contains one explanatory variable nowadays called market factor. Since then, many scholars are trying to find and add new variable to the CAPM model to enhance it's ability of capturing the dynamics between stock return and return volatility. Two examples of new factors are the size factor (SMB) and book-to-market factor(HML) found by Fama and French (1992)

A recent study by Harvey and Liu (2019), after examining top economics and financial economics journals, revealed the fact that, after 2004, new factors and paper illustrate how those new factors are helping explain the relationship between risk and return were in abundance. In their 2015 papers, Harvey, Liu, and Zhu coined a term ”factor zoo” which precisely capture the situation that the field of financial economics has too many factors. And some, if not most,of them can provide seemly significant results purely because of luck. ? (?) provides some insight and explanation about this phenomena.

Among all those factors, the ability of each factor to explain the risk and return is different. In their recent paper, (Bailey, Kapetanios, & Pesaran, 2020) introduce a new framework to measure a factor's ability or strength of pricing the risk of assets. The idea is that for a factor, the more different asset's relationship between risk and return it can capture, the stronger it is. We will review this concept in the following section 2.1

Some researches (see Kleibergen, 2009, and Gospodinov, Kan, & Robotti, 2017) argues that for a factor with small coefficient or even no coefficient, the statistic inference for model containing such factors will be unreliable. Kan and Zhang (1999) warned that when including a factor has no correlation with the asset return in a two-pass method of testing the pricing model, the model will falsely identified that useless model as significant more usually than it should. Therefore, appropriately identified the strength of factors and eliminate the factors that has limited explaining power before contains it into the CAPM becomes crucial.

This project, adopt the framework provided by Bailey et al. (2020) to exam the strength of factors from ( )'s factor zoo. In addition, we implied several machine learning algorithm to help exam ( )

2 Methodology

2.1 Factor Strength: Definition and Estimation

Factor strength has been elaborated by Bailey et al. (2020), the following section will re-iterate their definition of factor strength as well as the method to estimate it.
To start with, we define a single factor CAPM model:

$$y_{it} = \beta_i + \theta_i x_t + \varepsilon_{it} \tag{1}$$

Assume we have n different assets (for instance, n = 500 if using data from S&P 500 index). Collecting and calculating those assets returns from T different observations. y_{it} on the left hand side of equation (1) is the excess return of asset i at time t, The excess return equals to the asset return minus the risk free return. x_t in the right hand side is the factor with interest at time t. Therefore, θ_i is the loading of factor x_t. β_i is the constant term, represent the asset's ability to generate abnormal return from the market. ε_{it} as the idiosyncratic error term has been assumed to follow independent, identical distribution, with zero mean and time invariant variance σ_i^2.

After settle down, we run OLS for this model and obtain the results:

$$y_{it} = \hat{\beta}_i + \hat{\theta}_i x_t + \hat{\varepsilon}_{it}, \quad t = 1, 2, 3, \ldots T$$

Both $\hat{\beta}_i$ and $\hat{\theta}_i$ are the OLS estimation results of equation (1). Because we want to investigate the differences between estimated factor loading $\hat{\theta}_i$ and zero, we can construct a t-test with $t_i = \frac{\hat{\theta}_i - 0}{\hat{\varsigma}_i}$ where ς_i is the standard error of $\hat{\theta}_i$. Then we defined π_{nT} as the proportion of significant factor's amount to the total observations amount:

$$\hat{\pi}_{nT} = \frac{\sum_{i=1}^{n} \hat{\ell}_{i,nT}}{n} \tag{2}$$

$\ell_{i,nT}$ is an indicator function as: $\ell_{i,nT} := \mathbf{1}[|t_i| > c(n)]$. If the t-statistic t_i 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 $\hat{\theta}_i$ is significant. $c_p(n)$ represent the critical value of a test with test size p. The critical value is calculated by:

$$c_p(n) = \Phi^{-1}(1 - \frac{p}{2n^{\delta}}) \tag{3}$$

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 has not fixed the multiple testing problem. One of the most commonly used adjustment for multiple testing problem is Bonferroni correction. When n as sample size goes to infinity, however, the Bonferroni correction can not yield satisfying results since the $\frac{p}{2n^{\delta}} \to 0$ when $n \to \infty$. Therefore, Bailey, Kapetanios, and Pesaran (2016) provides another adjustment with additional exponent δ to constrain the behaviour of n.

After obtain the $\hat{\pi}_{nT}$, we can use the following formula to estimate our strength indicator α:

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

From the estimation, we can find out that $\hat{\alpha} \in [0,1]$

$\hat{\alpha}$ represent the pervasiveness of a factor. Here we denote $[n^{\alpha}]$, $[\cdot]$ will take the integer part of number inside. For factor θ_i:

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

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

For a factor has strength $\alpha = 1$, factor will be significant for every assets at every time. The more observation the factor can significantly influence, the stronger the factor is, and vice versa.

3 Monte Carlo Simulation

3.1 General Design

In this Monte Carlo simulation, we consider a dataset has $i = 1,2,\ldots,n$ different assets, with $t = 1,2,\ldots,T$ different observations. $j = 1,2,\ldots,k$ different factors and one market factors are included in the simulation. All returns of this study are generated from the following model with stochastic error.

$$r_{it} = f_1(\bar{r}_t - r_f) + f_2(\beta_{ij}\theta jt) + \varepsilon_{it}$$

$f_1(\cdot)$ and $f_2(\cdot)$ are two different functions represent the unknown mechanism of market factor and other factors pricing asset risk. $(\bar{r}_t - r_f)$ is the market return, calculated from market or index return$\bar{r}_t$ minus risk free returnr_f. We obtain the market factor from the existing data. r_{it} is the stock return, θ_{jt}denotes factors other than market factors and $\beta_{ij}$is the corresponding factor loading. ε_{it} is error. Notice that the β_{ij} will be influenced by each factor's strength α_j, where we have α as defined in section 2.1. And for each factors, we assume they follow a multinomial distributions with mean zero and a $k \times k$ variance-covariance matrix Σ. The matrix Σ indicates the correlation among all k factors.

In this model, we can control several parts to investigates different scenarios of the simulation:

1. The function $f_1(\cdot)$and $f_2(\cdot)$.

2. The strength α of each factor.

3. The correlation among each factors.

4. The stochastic error term ε_{it}

3.2 Baseline Design

As beginning point, we consider a baseline design with $k = 30$, $n = 500$, and $T = 120$. Both $f_1(a)$ and $f_2(a)$ are linear function:

$$f_1(a) = c_i + \beta_i a$$

$$f_2(a) = a$$

Here the c_i as constant are generated from a uniform distribution $U[-0.5, 0.5]$. To generate the factors loading, we employed a two steps strategy. First we generate a whole factor loadings matrix with t columns represent t different time observations, and $k+1$ rows denote k factors plus one extra market factors. All element of the matrix follow a independent identical uniform distributions $IIDU(\mu_\beta - 0.2, \mu_\beta + 0.2)$. The μ_β has been equals to 0.71 in this case to set all values apart from zero. After generate the matrix, we randomly selected $[n^{\alpha_q}]$, $\{q = 1, 2, 3 \ldots k+1\}$ elements from each column to keep their value and set the other elements to zero. This step ensures the loading reflects the true strength of each factors.

References

Bailey, N., Kapetanios, G., & Pesaran, M. H. (2016, 9). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, *31*, 929-960. Retrieved from `http://doi.wiley.com/10.1002/jae.2476` doi: 10.1002/jae.2476

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

Fama, E. F., & French, K. R. (1992, 6). The cross-section of expected stock returns. *The Journal of Finance*, *47*, 427-465. Retrieved from `http://doi.wiley.com/10.1111/j.1540-6261.1992.tb04398.x` doi: 10.1111/j.1540-6261.1992.tb04398.x

Gospodinov, N., Kan, R., & Robotti, C. (2017, 9). Spurious inference in reduced-rank asset-pricing models. *Econometrica*, *85*, 1613-1628. doi: 10.3982/ecta13750

Harvey, C. R., & Liu, Y. (2019, 3). A census of the factor zoo. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3341728

Harvey, C. R., Liu, Y., & Zhu, H. (2015, 10). … and the cross-section of expected returns. *The Review of Financial Studies*, *29*, 5-68. Retrieved from `https://doi.org/10.1093/rfs/hhv059` doi: 10.1093/rfs/hhv059

Kan, R., & Zhang, C. (1999, 2). Two-pass tests of asset pricing models with useless factors. *The Journal of Finance*, *54*, 203-235. Retrieved from `http://doi.wiley.com/10.1111/0022-1082.00102` doi: 10.1111/0022-1082.00102

Kleibergen, F. (2009, 4). Tests of risk premia in linear factor models. *Journal of Econometrics*, *149*, 149-173. doi: 10.1016/j.jeconom.2009.01.013

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*, 13-37. doi: 10.2307/1924119

Sharpe, W. F. (1964, 9). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*, 425-442. Retrieved from `http://doi.wiley.com/10.1111/j.1540-6261.1964.tb02865.x` doi: 10.1111/j.1540-6261.1964.tb02865.x