Factor Selection and Factor Strength
A study base on U.S. stock Market

Research Plan Presentation

Zhiyuan Jiang
28710967

Supervisors: Dr. Natalia Bailey
Dr. David Frazier

May 17, 2020

Introduction and Motivation

Capital Asset Pricing Model has shaped the way people evaluate the risk and return of assets.

$$r_{it} = c + \beta_0 x_m + \beta_{\mathbf{j}}{}' \mathbf{x_j} + \varepsilon_{it} \tag{1}$$

By adding new factors, we can improve the model's ability of pricing the risk and predict the return. Scholars are always trying to find new factors to add into the $\mathbf{x_j}$ part.
Now we have over 500 factors inside our factor zoo, and it is still growing. (Harvey & Liu, 2019)

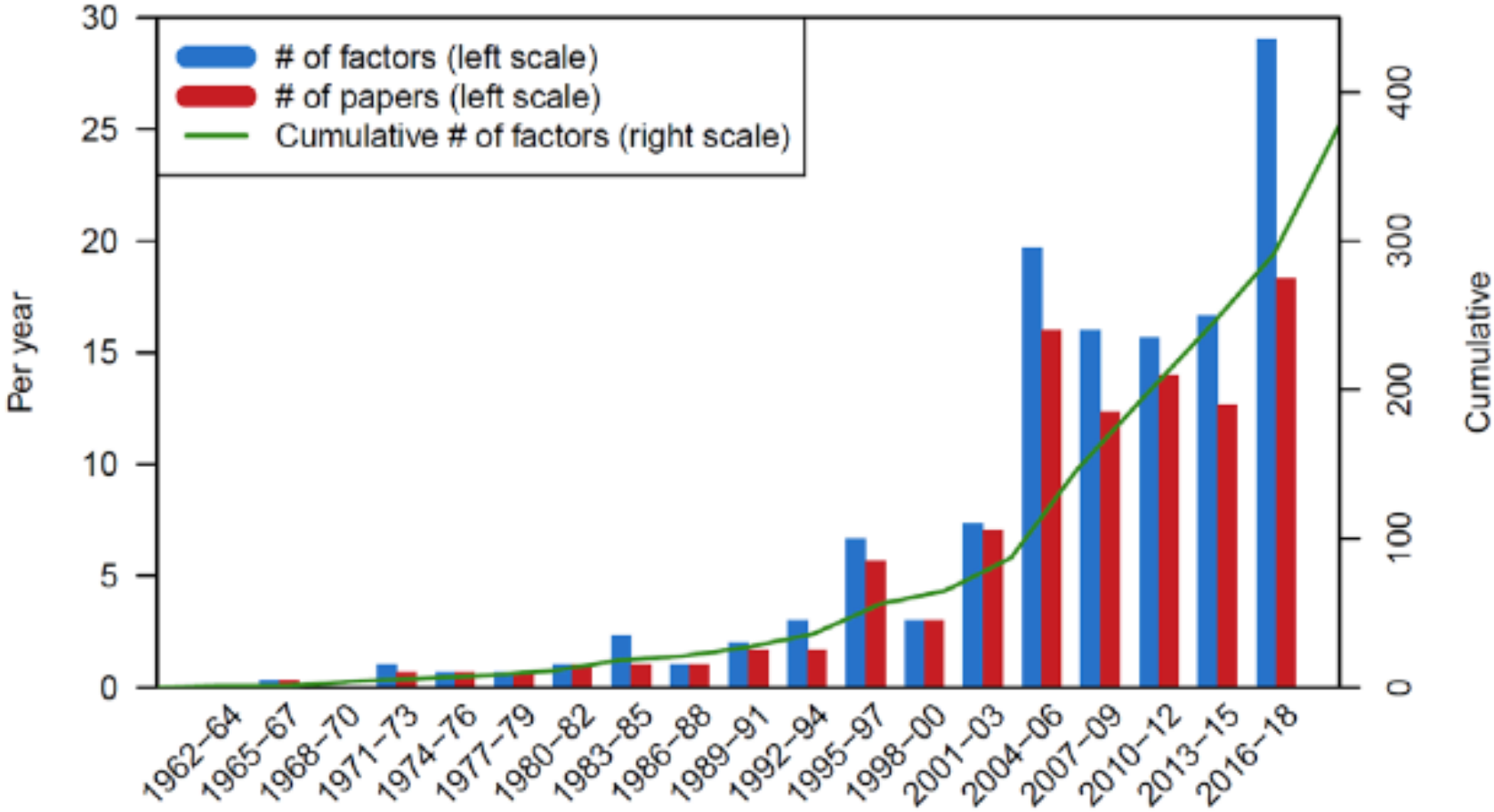

Figure: Factor amount growing through the year.

(Harvey & Liu, 2017)

Introduction and Motivation

But some problem exist inside the factor zoo.

- Not all factors are good or useful
  - Predict power drop once published (McLean & Pontiff, 2016)
  - Can not been replicated (Hou, Xue, & Zhang, 2018)
  - Significant because of luck (Harvey & Liu, 2014)
- Including weak or useless factors is terrible
  - Useless factor in FM first-regression(Fama & MacBeth, 1973) will yield misleading second regression result (Kan & Zhang, 1999)
  - If the loading is small, regression will report incorrect result Kleibergen (2009)

Literature

- **Identify factors from the zoo**
  Harvey, Liu, and Zhu (2015), McLean and Pontiff (2016), Harvey and Liu (2017), Barillas and Shanken (2018),Pukthuanthong, Roll, and Subrahmanyam (2019)
- **Using machine learning method**
  Rapach, Strauss, and Zhou (2013), Feng, Giglio, and Xiu (2019),Gu, Kelly, and Xiu (2020), Lettau and Pelger (2020), Freyberger, Neuhierl, and Weber (2020), Kozak, Nagel, and Santosh (2020)

Methodology

This project facing two challenges:

1. A Big Zoo: Too many candidates
   Curse of Dimensionality
2. Correlation among factors
   Traditional factor selection algorithm can not handle this

Methodology

This project facing two challenges:

1. A Big Zoo: Too many candidates
   Curse of Dimensionality
2. Correlation among factors
   Traditional factor selection algorithm can not handle this

Divided this project into two parts:

1. Calculate the factor strength, only focus on factor with certain strength
2. Use Elastic Net model, overcome the correlation problem

Factor Strength

Elaborated by Pesaran and Smith (2019), Bailey, Kapetanios, and Pesaran (2020)
Factor strength represents the factor's pervasiveness.
The stronger a factor is, the more assets it can price, i.e. has loading significantly different from zero.
Recall the CAPM model introduced at the beginning, but ignore the market factor.
If a factor x_j has strength α_j, we let this factor regress with N different assets.

$$\beta_j \neq 0, j = 1, 2, 3, \cdots, [N^{\alpha_j}]$$
$$\beta_j = 0, j = [N^{\alpha_j}] + 1, [N^{\alpha_j}] + 2, [N^{\alpha_j}] + 3, \cdots, N$$

Elastic Net

Introduce by Zou and Hastie (2005), is a improved method to select factor.

Recall OLS estimation, for the factor model (1), OLS will estimates the factor loading by minimise the residual sum of square:

$$\hat{\beta}_j = argmin\{\sum_{i=1}^{N}(r_{it} - \beta_{\mathbf{j}}\mathbf{x_j})^2\}$$

For Elastic net, it add two penalty terms into the OLS regression:

$$\hat{\beta}_j = argmin\{\sum_{i=1}^{N}(r_{it} - \beta_{\mathbf{j}}\mathbf{x_j})^2 + \lambda_1 {\beta_{\mathbf{j}}}^2 + \lambda_2|\beta_{\mathbf{j}}|\} \qquad (2)$$

Advantage of Elastic Net: it can select factors from a group of correlated candidates.

Future Plan

For the next step, we will start the empirical analyses.
We will collect and exam the data for the empirical research.
Assets: **Companies from Standard & Poor (S&P) 500 index**
We will use those securities return as the left hand side part of the CAPM.
Factor: **Factors from Harvey and Liu (2019)'s factor list**
Using factor strength as the criterion to trim first, and then applied the elastic net method.

Thanks for listening
Question or Comment

Bibliography I

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

Barillas, F., & Shanken, J. (2018, 4). Comparing asset pricing models. *The Journal of Finance*, *73*, 715-754. Retrieved from http://doi.wiley.com/10.1111/jofi.12607 doi: 10.1111/jofi.12607

Fama, E. F., & MacBeth, J. D. (1973, 5). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*, 607-636. doi: 10.1086/260061

Feng, G., Giglio, S., & Xiu, D. (2019, 1). *Taming the factor zoo: A test of new factors.* Retrieved from http://www.nber.org/papers/w25481.pdf doi: 10.3386/w25481

Bibliography II

Freyberger, J., Neuhierl, A., & Weber, M. (2020, 4). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*, 2326-2377. Retrieved from https://doi.org/10.1093/rfs/hhz123 doi: 10.1093/rfs/hhz123

Gu, S., Kelly, B., & Xiu, D. (2020, 2). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*, 2223-2273. Retrieved from https://doi.org/10.1093/rfs/hhaa009 doi: 10.1093/rfs/hhaa009

Harvey, C. R., & Liu, Y. (2014, 11). Lucky factors. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2528780

Bibliography III

Harvey, C. R., & Liu, Y. (2017, 12). False (and missed) discoveries in financial economics. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3073799

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

Bibliography IV

Hou, K., Xue, C., & Zhang, L. (2018, 12). Replicating anomalies. *The Review of Financial Studies*, *33*, 2019-2133. Retrieved from https://doi.org/10.1093/rfs/hhy131 doi: 10.1093/rfs/hhy131

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

Bibliography V

Kozak, S., Nagel, S., & Santosh, S. (2020, 2). Shrinking the cross-section. *Journal of Financial Economics*, *135*, 271-292. doi: 10.1016/j.jfineco.2019.06.008

Lettau, M., & Pelger, M. (2020, 2). Estimating latent asset-pricing factors. *Journal of Econometrics*. doi: 10.1016/j.jeconom.2019.08.012

McLean, R. D., & Pontiff, J. (2016, 2). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*, 5-32. Retrieved from `http://doi.wiley.com/10.1111/jofi.12365` doi: 10.1111/jofi.12365

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*.

Bibliography VI

Pukthuanthong, K., Roll, R., & Subrahmanyam, A. (2019, 8). A protocol for factor identification. *Review of Financial Studies*, *32*, 1573-1607. Retrieved from https://doi.org/10.1093/rfs/hhy093 doi: 10.1093/rfs/hhy093

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013, 8). International stock return predictability: What is the role of the united states? *The Journal of Finance*, *68*, 1633-1662. Retrieved from http://doi.wiley.com/10.1111/jofi.12041 doi: 10.1111/jofi.12041

Bibliography VII

Zou, H., & Hastie, T. (2005, 4). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*, 301-320. Retrieved from http://doi.wiley.com/10.1111/j.1467-9868.2005.00503.x doi: 10.1111/j.1467-9868.2005.00503.x