Factor Strength and Factor Selection

An Application to U.S. Stock Market

Zhiyuan Jiang
28710967

Supervisors: Dr Natalia Bailey
Dr David Frazier

October 8, 2020

Motivation

Capital Asset Pricing Model (CAPM) is the benchmark of risk pricing.

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

- r_{it}: asset's return
- r_{ft}: risk free return
- a_i: constant/intercept
- β_{im}: market factor loading
- r_{mt}: market return
- β_{ij}: risk factor loading
- f_{jt}: risk factor
- ε_{it}: stochastic error

- **Add factors to enhance risk pricing.**
- **New factors are booming**

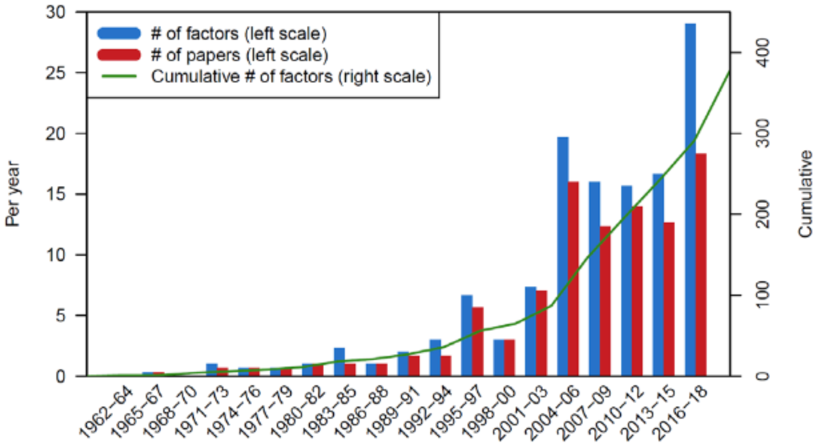

Figure: Factor amount growing through the year.

(Harvey & Liu, 2019)

'We have a lot of questions to answer:
Firstly, which characteristics really provide ***independent*** *information about average returns? Which are subsumed by others ?'* ?, ?

Factor Strength

The research interest is pricing risk, so factor strength matter.
Consistency of risk pricing is dependent on the strength of factor (Pesaran & Smith, 2019)
Strong factor ⇒ price more asset's risk ⇒ generate more significantly loadings.
Factor strength is defined in terms of factor loading (Bailey, Kapetanios, & Pesaran, 2020) as follow.
Assume we have N different assets.

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

Introduction and Motivation

But some problems exist among all those factors.

- Including Factor without correlation with return in FM first-regression(Fama & MacBeth, 1973) will yield misleading second regression result (Kan & Zhang, 1999)
- If the factor loading is small, estimated risk premia will be spurious Kleibergen (2009)

Reference to this problem is made in the literature:
Kan and Zhang (1999), Kleibergen (2009), Kleibergen and Zhan (2015), Gospodinov, Kan, and Robotti (2017), Anatolyev and Mikusheva (2018)

Literature

- **Identify factors**
  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)

Main Problem

This project faces two challenges:

1. High dimensions of data group
   How to identify the significant one. $\Rightarrow$ use factor strength as criteria.
2. Correlation among factors
   Traditional variable selection algorithm (Lasso) can not handle this.$\Rightarrow$ Will use elastic net techniques

Elastic Net

Introduce by Zou and Hastie (2005), is a improved method to select factor.
Considering the following loss function:

$$\hat{\beta}_{ij} = \arg\min_{\beta_{ij}} \{ \sum_{i=1}^{n} [(r_{it} - r_{ft}) - \beta_{ij} f_{jt}]^2 + \lambda_2 \sum_{i=1}^{n} \beta_{ij}^2 + \lambda_1 \sum_{i=1}^{n} |\beta_{ij}| \}$$

The L_1 norm $\sum_{i=1}^{n} |\beta_{ij}|$ helps select the factor, reduce redundancy.
The L_2 norm $\sum_{i=1}^{n} \beta_{ij}^2$ helps handle the correlation.

Preliminary Result

Use Monte Carlo simulation to study the property of estimated factor strength.

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

- Overestimates occurs when strength is low
  $\alpha = 0.5, \hat{\alpha} \approx 0.7$
- But the precision improved with strength increase
  $\alpha = 0.7, \hat{\alpha} = 0.8$
- When we have the strong factor, we have the unbiased estimator $\alpha = \hat{\alpha} = 1$

Data

The data set contains two part:

- **Assets**: Stocks from Standard & Poor (S&P) 500 index companies
  Three year U.S. t-bill Average return of U.S. stock market
- **Factor**: 145 risk factors plus one market factor.

Thirty year time period: Jan 1987 - Dec 2007.
Data set into three subsamples: 10/20/30 years.

Thanks for listening

Bibliography I

Anatolyev, S., & Mikusheva, A. (2018, 7). Factor models with many assets: strong factors, weak factors, and the two-pass procedure. *CESifo Working Paper Series*. Retrieved from http://arxiv.org/abs/1807.04094

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

Bibliography II

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

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

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

Bibliography III

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. (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

Bibliography IV

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

Bibliography V

Kleibergen, F., & Zhan, Z. (2015, 11). Unexplained factors and their effects on second pass r-squared's. *Journal of Econometrics*, *189*, 101-116. doi: 10.1016/j.jeconom.2014.11.006

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

Bibliography VI

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*.

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

Bibliography VII

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

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