GROUP PROJECT REPORT #1 LITERATURE SURVEY - A COMPARISON BETWEEN DIFFERENT POPULAR FACTOR MODELS

AFM 423 Machine Learning Approach to Quantitative Investing

Prepared by Xuntian Lin, Sue Lu, Aman Halwasia

March 2, 2023

1.0 Introduction

In economic research, asset pricing has always been the central topic. Capital Asset Pricing Model (CAPM) is a single factor model where the study of the relationship between market risk and the expected stock returns has attracted much attention over the years. Fama and French first introducing the three-factor model has changed how academics associate market risks with other factors. While more and more factor models are introduced to the research field, people are interested in which model best fits into the market and gives the best predictions regarding monthly stock returns.

2.0 Analysis on Different Models

Four articles are summarized below to assess how different models with various factors would affect the overall fit of the multi linear regression models and how they predict the average returns for a portfolio of assets.

2.1 Fama-French Three-Factor Model

This model (FAMA & FRENCH, 1992) was first introduced by Eugene Fama and Kenneth French in 1992. The model identified three factors that explain the average returns of a portfolio or asset: market risk, size of firms, and book-to-market value.

Summary:

The authors first stated two empirical contradictions of the Sharpe-Lintner-Black (SLB) Model, which has long shaped the way academics think about the relation between risks and returns. However, they suggested that stock risks are multidimensional, where two dimensions of risk are proxied by size (ME) and by the ratio of the book value of common equity to its market value (BE/ME). Investigation to the relation between beta and size, and relevance between book-to-market equity, E/P, and leverage are rigorously tested respectively. The article also summarizes the data and approach to estimate beta, as well as the relation between average return and beta and between average return and size.

Strengths:

1. **Data source and reliability**: This paper clearly cited where the data from in the research, for example, stocks returns files from the Center for Research in Security Prices (CRSP) and COMPUSTAT. Data merging methods are also summarized to address the discrepancies between accounting variables and stock return data for replicability.

2. **Rigorous statistical methodology**: Fama and French used a rigorous methodology to analyze the data, including subdividing each size decile into 10 portfolios on the basis of pre-ranking betas for individual stocks, fitting Fama-MacBeth regression analysis and various statistical tests, which enhanced the reliability and validity of their findings.
3. **Data and results presentation:** Results to statistical regression are formatted and labelled in order including *t*-values, *p*-values. Authors also analyzed how those statistical results can explain their hypothesis between the returns and various factors. Appendix data and or table is also available to support the main arguments provided throughout the entire article.
4. **Timeless**: The paper has stood the test of time and remains relevant today, as the concepts and framework introduced by the authors continue to be used in empirical research and investment management.

Weakness

1. The paper relies on the accuracy of data from the Center for Research in Security Prices (CRSP), which has been found to contain errors and inconsistencies. The paper did not mention raw data processing and analysis in detail. Potential data issues while analyzing include: missing data, and unknown data qualities.
2. While the three-factor model proposed in the paper has been influential, the only three variables may not entirely explain the relationship between the risks and the expected returns. For example, additional factors beyond size, book-to-market, and market risk may be important in explaining expected returns. This will be mainly addressed in other papers sourced in this literature survey.

3. This paper published in 1992 incorporated a large amount of data, yet barely any graphs or visualizations are presented in this article, making the audience harder to understand how the original dataset looks like. Readers lacking adequate financial fundamentals may also find it difficult to understand key arguments and analysis. A list of glossaries may be helpful for readers with all ranges of knowledge while reading this article.

2.2 Carhart Four-Factor Model

Carhart, M. M. (1997). On persistence in mutual fund performance. is the original paper that introduces the Carhart Four-Factor Model. It tests the model on a sample of US mutual funds and finds that the momentum factor has significant explanatory power for mutual fund returns. This model was developed by Mark Carhart in 1997 and builds on the Fama-French model by adding a momentum factor. The momentum factor represents exposure to stocks that have recently performed well.

Summary:

Persistence in mutual fund performance does not reflect superior stock-picking skill. Rather, common factors in stock returns and persistent differences in mutual fund expenses and transaction costs explain almost all of the predictability in mutual fund returns. Only the strong, persistent underperformance by the worst-return mutual funds remains anomalous. Individual mutual funds that appear to follow the one-year momentum strategy earn significantly lower abnormal returns after expenses. Thus, Mark made three conclusions for investors who want to maximize their mutual fund returns:

1. Avoid funds with persistently poor performance;

2. funds with high returns last year (JEGADEESH & TITMAN, 1993) have higher-than-average expected returns next year which can be referred to as hot hands effect (HENDRICKS et al., 1993), but not in years thereafter; and
3. The investment costs of expense ratios, transaction costs, and load fees all have a direct, negative impact on performance.

Strengths:

1. **Introduces a new factor**: The paper introduces a new factor, the momentum factor, that captures the effect of past performance on mutual fund returns. This is an important contribution to the literature as it shows that momentum is an important factor in explaining mutual fund performance. Prior to this paper, the Fama-French Three-Factor Model was the dominant model used to explain mutual fund performance.
2. **Large sample size and long time period**: The paper uses a sample of 1,892 US equity mutual funds and a long time period from 1962 to 1993 to test the model. This increases the reliability of the results and helps to ensure that the findings are not driven by extreme outliers or short-term fluctuations.
3. **Robust statistical methods**: The use of robust methods helps to ensure that the results are not sensitive to small changes in the model specification or estimation technique.
4. **Improvement over the Fama-French Three-Factor Model**: The paper shows that the Carhart Four-Factor Model provides a better explanation of mutual fund returns than the Fama-French Three-Factor Model. This demonstrates the importance of including the momentum factor in asset pricing models.

Weakness:

1. **Lack of theoretical explanation for momentum factor**: This leaves open the question of whether momentum is a genuine risk factor or simply a statistical anomaly.
2. **Limited exploration of economic significance**: The paper is unclear whether the momentum effect is strong enough to generate profitable trading strategies for investors. Further research is needed to examine the profitability of trading strategies based on the momentum factor.
3. **Survivorship bias:** The paper does not account for survivorship bias in the mutual fund data. This could potentially bias the results if poorly performing funds drop out of the sample over time. Future research could address this limitation by using survivorship bias-corrected datasets.
4. **Limited generalizability**: The paper focuses solely on US equity mutual funds and does not test the model on other asset classes or in other geographies. This limits the generalizability of the results. Future research could extend the analysis to other asset classes and geographic regions to examine the cross-sectional variation in asset pricing models.

Limitation: Measuring whether funds follow the momentum strategy is imperfect in Mark's sample.

2.3 Barra Risk Factor Model Based on Idiosyncratic Momentum

Summary: This paper introduces an idiosyncratic momentum factor based on the residuals from Barra Global Risk Models, which generate higher long-term Information Ratio in the Chinese equity market than traditional momentum. This strategy also displays consistent performance in varying economic environments and is less affected by market dynamics. Additionally, it is not

oriented toward small-cap stocks, reducing transaction costs and firm-specific risk. These results solve the momentum puzzle and enhance the effectiveness of the momentum strategy in the Chinese equity market.

Strengths:

1. The study introduces a new idiosyncratic momentum factor that generates higher long-term Information Ratio than traditional momentum strategies.
2. The paper provides evidence of the strategy's effectiveness even during multiple-year periods where traditional momentum generates negative returns.
3. The strategy is substantially less affected by market dynamics, reducing transaction costs and firm-specific risk.

Weaknesses:

1. The study is limited to the Chinese equity market, which may not generalize to other markets.
2. The paper does not provide a thorough explanation of the theoretical mechanism underlying the idiosyncratic momentum strategy.

2.4 Comparison between Fama-French three factor model and liquidity based 3 Factor Model in Predicting Portfolio Return

Summary:

This paper compares the ability of the Fama French 3 Factor Model and 2 Liquidity Based Models to predict the Malaysian stock market. The findings of the paper suggests that when considering small markets a factor from which risk is derived is the firm’s illiquidity as well as the chance of being in distress. By introducing these factors while making predictions the models

have been able to outperform Fama French as well as CAPM in predicting returns in the Malaysian stock market.

Strengths:

The paper does well in explaining the significance of choosing liquidity as a factor to compete against Fama French's traditional factors of size and Book-to-market. This paper also highlights that this holds especially true for the size of market being considered and the potential bias that could be a result of looking at a smaller market like Malaysia. This ensures that the reader is not misled by the scope of the study.

Weakness:

While this paper finds that illiquidity as a factor does provide more accurate results it is unable to statistically prove it and is unable to reject the hypothesis of there being a significant difference between the Fama French and liquidity based 2 factor models. This leads to the question of whether this should be considered for larger markets since Fama French has proven to be more robust irrespective of market size.

Limitations:

The scope of the paper doesn't allow us to judge the forecasting power of these liquidity based factor models in larger markets and see if they are still marginally more effective compared to the traditional Fama French 3 Factor Model and leaves scope for further research.

3.0 Implementation of Plan

We will use multilinear regressions with Fama French three factor models, Carhart Four-Factor Model, Barra Risk Factor Mode and liquidity based Three Factor Model using the dataset provided from the course website.

Key metrics/methods will be used to assess the overall fitness of different models mentioned in this literature survey

- Least R squared values for different models
- *t*-statistics, *p*-values, estimated betas and intercepts
- Suggested methods mentioned in the academic papers; compare the results
- Discuss why the discrepancies between models exist and propose practical solutions to improve accuracy of each model

A summary of steps in terms of how to implement this topic is listed as below:

Summary of steps:

1. Download the data from the course website and inspect the summary statistics of the data downloaded (ranges, samples means and medians)
2. Preprocess the data resolving issues including data missing (removal and/or imputation), outliers detection, etc.
3. Scaling predictors: standardization, winsorization, uniformization would be useful
4. Prepare the three different models and train the model fitting the dataset by linear regression and features selection
5. Use the key metrics mentioned above to measure the performance among the different factor models by statistical analysis
6. Visualization: Use plotting, tables to show the graphical representation of the results

References

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Hendricks, D., Patel, J., & Zeckhauser, R. (1993). Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988. *The Journal of Finance*, *48*(1), 93–130.

Lu, S., & Lu, C. (2018). Barra risk model based idiosyncratic momentum for Chinese equity market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3140113

Rahim, R. A., & Nor, A. H. S. M. (2006). A COMPARISON BETWEEN FAMA AND FRENCH MODEL AND LIQUIDITY-BASED THREE-FACTOR MODELS IN PREDICTING THE PORTFOLIO RETURNS . *Asian Academy of Management Journal of Accounting and Finance*, *2*(2), 43–60.