FE670 Algorithmic Trading Strategies

Alpha Factor Trading Strategies

Steve Yang

Stevens Institute of Technology

steve.yang@stevens.edu

09/28/2023

Overview

CAPM and Fama-French Factor Models

Econometric Considerations for Cross-Sectional Factor Models

Cross-Sectional Methods for Evaluation of Factor Premiums

Factor Based Trading Strategies

Python Time Series Factor Analysis Package

- Risk factors have been a key ingredient to quantitative models since the capital asset pricing model (CAPM) explained returns of all N assets $r_i, i = 1, ..., N$ using their respective exposure β_i to a single factor, the expected excess return of the overall market over the risk-free rate r_f. The CAPM model takes the following linear form:

$$E[r_i] = \alpha_i + \beta_i(E[r_m] - r_f) \tag{1}$$

where r_m is the market return.
- This differs from the classic fundamental analysis, Dodd and Graham, where returns depend on firm characteristics. The rationale is that in the aggregate, investors cannot eliminate this so-called systemic risk through diversification.
- In equilibrium, they require compensation for holding an asset commensurate with this systematic risk. The model implies that, given efficient markets where prices immediately reflect all public information, there should be no superior risk-adjusted returns.

Risk Fama-French Factors Theories

- Joseph Stiglitz earned the 2001 Nobel Price in economics in part for showing that markets are generally not perfectly eficient: if markets are efficient, there is no value in collecting data because this information is already reflected in prices.
- Stephen Ross proposed asset pricing theory (APT) in 1976 as

  an alternative that allows for several risk factors while eschewing market efficiency. In contrast to the CAPM, it assumes that opportunities for superior returns due to mispricing may exist but will quickly be arbitraged away.
- Kenneth French and Eugene Fama (who won the 2013 Nobel Price) identified additional risk factors that depend on firm characteristics and are widely used today. In 1993, the Fama-French three-factor model added the relative size and value of firms to the single CAPM source of risk. In 2015, the five-factor model further expanded the set to include firm profitability and level of investment.

Fundamental Factors

- Arguably the mostly widely used factors today are fundamental factors. Fundamental factors capture stock characteristics such as industry membership, country membership, valuation ratios, and technical indicators, to name a few.
- The most popular factors today - Value, Growth, Size, Momentum - have been studied for decades as part of the academic asset pricing literature and the practitioner risk factor modeling research.
- Fama and French (1992, 1993) put forward a model explaining US equity market returns with three factors: the "market" (based on the traditional CAPM model), the size factor (large vs. small capitalization stocks) and the value factor (low vs. high book to market). The "Fama-French" model, which today includes Carhart's (1997) momentum factor, has become a canon within the finance literature.

Risk Fama-French Factors Theories

- The Fama-French risk factors are computed as the return difference on diversified portfolios (PFs) with high or low values, according to metrics that reflect a given risk factor.
- These returns are obtained by sorting stocks according to these metrics and then going long stocks above a certain percentile, while shorting stocks below a certain percentile.
- The metrics associated with the risk factors are defined as follows:
  1. Market (Rm-Rf): Value-weighted return of all firms incorporated in and listed on major US exchanges, minus the one-month
  2. Size (SMB): Nine small stock PF minus mine large stock PF.
  3. Value (HML): Two value PF minus two growth (with low BE/ME value) PF.
  4. Profitability (RMW) : Two robust OP PF minus two weak OP PF (with Operating Profitability).
  5. Investment (CMA): Two conservative investment portfolio, minus two aggressive investment portfolio.

Risk Fama-French Factors Sorting Strategy

- The Fama-French risk factors are computed as the return difference on diversified portfolios with high or low values, according to metrics that reflect a given risk factor.
- These returns are obtained by sorting stocks according to these metrics and then going long stocks above a certain percentile, while shorting stocks below a certain percentile.
- The Fama-French 5-Factor Model is widely used in empirical finance and asset pricing research to provide a more nuanced explanation of asset returns compared to the simpler Capital Asset Pricing Model (CAPM) or the Fama-French 3-Factor Model.
- Factor investing has become a widely discussed part of today's investment canon. It acknowledges that various factors beyond market risk and size and value effects can influence asset returns.

Econometric Considerations for Cross-Sectional Factor Models

In cross-sectional regression where the dependent variable is a stock's return and the independent variables are factors, inference problems may arise that are the result of violations of classical linear regression theory. The most common problems:

Measurement Problems Some factors are not explicitly given, but need to be estimated. These factors are estimated with an error. The estimation errors in the factors can have an impact on the inference from a factor model. This problem is commonly referred to as the "errors in variables problem".

Common Variation in Residuals The residuals from a regression often contain a source of common variation. Sources of common variation in the residuals have heteroskedasticity and serial correlation.

Econometric Considerations for Cross-Sectional Factor Models

Common Variation in Residuals

- Heteroskedasticity occurs when the variance of the residual differs across observations and affects the statistical inference in a linear regression. In particular, the estimated standard errors will be underestimated and the t-statistics will therefore be inflated. Ignoring heteroskedasticity may lead the researcher to find significant relationships where none actually exist. Several procedures have been developed to calculate standard errors that are robust to heteroskedasticity.
- Serial correlation occurs when consecutive residual terms in a linear regression are correlated, violating the assumptions of regression theory. If the serial correlation is positive then the standard errors are underestimated and the t-statistics will be inflated.

Econometric Considerations for Cross-Sectional Factor Models

- **Multicollinearity** Multicollinearity occurs when two or more independent variables are highly correlated. We may encounter several problems when this happens.
  1) First, it is difficult to determine which factors influence the dependent variable.
  2) Second, the individual p values can be misleading – a p value can be high even if the variable is important.
  3) Third, the confidence intervals for the regression coefficients will be wide.
- There is no formal solution based on theory to correct for multicollinearity. The best way is by removing one or more of the correlated independent variables. It can be reduced by increasing the sample size.

Fama-MacBeth Regression

- To address the inference problem caused by the correlation of the residuals, Fama and MacBeth proposed the following methodology for estimating cross-sectional regressions of returns on factors. For notational simplicity, we describe the procedure for one factor. The multifactor generalization is straightforward:
- First, for each point in time t we perform a cross-sectional regression

$$r_{i,t} = \beta_{i,t} f_t + \epsilon_{i,t}, i = 1, 2, ..., N \tag{2}$$

- In the academic literature, the regressions are typically performed using monthly or quarterly data, but the procedure could be used at any frequency.
- The mean and standard errors of the time series of slopes and residuals are evaluated to determine the significance of the cross-sectional regression.

Fama-MacBeth Regression

- We estimate f and ϵ_i as the average of their cross-sectional estimates:

$$\hat{f} = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t, \hat{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{i,t} \tag{3}$$

- The variations in the estimates determine the standard error and capture the effects of residual correlation without actually estimating the correlation.
- We used the standard deviations of the cross-sectional regression estimates to calculate the sampling errors for these estimates,

$$\hat{\sigma}_{\hat{f}} = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{f}_t - \hat{f})^2, \sigma^2_{\hat{\epsilon}_i} = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\epsilon}_{i,t} - \hat{\epsilon}_i)^2 \tag{4}$$

Fama-MacBeth Regression

- Given data on risk factors and portfolio returns, it is useful to estimate the portfolio's exposure to these returns to learn how much they drive the portfolio's returns.
- It is also of interest to understand the premium that the market pays for the exposure to a given factor, that is, how much taking this risk is worth.
- More formally, we will have $i = 1, .., N$ asset or portfolio returns over $t = 1, ..., T$ periods, and each asset's excess period return will be denoted. The goal is to test whether the $j = 1, ..., M$ factors explain the excess returns and the risk premium associated with each factor.
- To address the inference problem caused by the correlation of the residuals, Fama and MacBeth proposed a two-step methodology for a cross-sectional regression of returns on factors.

Fama-MacBeth Regression

- The two stage Fama-MacBeth regression is designed to estimate the premium rewarded for the exposure to a particular risk factor by the market. The two stages consist of:

  1. First state: N time-series regression, one for each asset or portfolio, of its excess returns on the factors to estimate the factor loadings. In the matrix form, for each asset:

  $$\mathbf{r}_i = \mathbf{F}\beta_i + \epsilon_i \tag{5}$$

  2. Second stage: T cross-sectional regression, one for each time period, to estimate the risk premium. In matrix form, we obtain a vector of risk premia for each period:

  $$\mathbf{r}_t = \hat{\beta}_i \lambda_t \tag{6}$$

Fama-MacBeth Regression

- Now we can compute the factor risk premia as the time average and get a t-statistic to assess their individual significance, using the assumption that the risk premia estimates are independent over time:

$$t = \frac{\lambda_j}{\sigma(\lambda_j)/\sqrt{T}} \tag{7}$$

- If we had a very large and representative data sample on traded risk factors, we could use the sample mean as a risk premium estimate.
- The Fama-MacBeth methodology leverages the covariance of the factors with other assets to determine the factor premia.

Information Coefficients

- To determine the forecast ability of a model, practitioners commonly use the information coefficient (IC). The IC is a linear statistic that measures the cross-sectional correlation between a factor and its subsequent realized return

$$IC_{t,t+k} = \text{corr}(\mathbf{f}_t, \mathbf{r}_{t,t+k})$$

  where $\mathbf{f}_t$ is a vector of cross sectional factor values at time t and $\mathbf{r}_{t,t+k}$ is a vector of returns over the time period t to $t+k$.
- Just like the standard correlation coefficient, the values of the IC range from -1 to $+1$. A positive IC indicates a positive relation between the factor and return. A negative IC indicates a negative relation. ICs are usually calculated over an interval, for example, daily or monthly. We can evaluate how a factor has performed by examining the time series behavior of the ICs.

- An alternative specification of this measure is to make $\mathbf{f}_t$ the rank of a cross-sectional factor. This calculation is similar to the Spearman rank coefficient. By using the rank of the factor, we focus on the ordering of the factor instead of its value. Ranking the factor value reduces the unduly influence of outliers and reduces the influence of variables with unequal variances.
- The subsequent realized returns to a factor typically vary over different time horizons. For example, the return to a factor based on price reversal is realized over short horizons, while valuation metrics such as EBITDA/EV are realized over longer periods. It therefore makes sense to calculate multiple ICs for a set of factor forecasts whereby each calculation varies the horizon over which the returns are measured.

- Information coefficients can also be used to assess the risk of factors and trading strategies. The standard deviation of the time series (with respect to) of ICs for a particular factor can be interpreted as the strategy risk of a factor. Examining the time series behavior of $std(IC_{t,t+k})$ over different time periods may give a better understanding of how often a particular factor may fail.
- The expected tracking error can be used to understand the active risk of investment portfolios. Qian and Hua (2004) defined an expected tracking error as:

$$\sigma_{TE} = \mathrm{std}(\mathbf{IC}_{t,t+k})\sqrt{N}\sigma_{\mathrm{model}}\mathrm{dis}(\mathbf{R}_t)$$

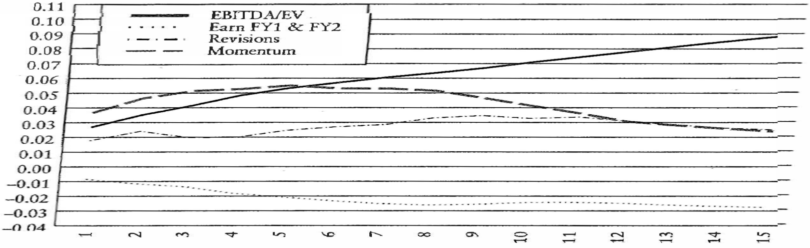
where N is the number of stocks in the universe (breath), σ_{model} is the risk model tracking error, and $\mathrm{dis}(\mathbf{R}_t)$ is dispersion of returns defined by

$$\mathrm{dis}(\mathbf{R}_t) = \mathrm{std}(r_{1,t}, r_{2,t}, ..., r_{N,t})$$

Example: Information Coefficients

Exhibit 7.4 displays the time-varying behavior of ICs for each one of the factors EBITDA/EV, growth of fiscal year 1 and 2 earnings estimates, revisions, and momentum. The graph depicts the information horizons for each factor. The EBITDA/EV factor earns higher returns. The overall pattern shows that the return realization pattern to different factors varies.

EXHIBIT 7.4 Information Coefficients over Various Horizons for EBITDA/EV, Growth of Fiscal Year 1 and Fiscal Year 2 Earnings Estimates, Revisions, and Momentum Factors



- Certain factor-based strategies have delivered excess returns over the last 20 years, suggesting they can add long-term value to a portfolio, albeit with higher risk than their benchmark index in some cases.
- A long-term approach to factor investing involving one of the following:
  - Single factor strategies: For example, an investor's portfolio might be allocated to momentum, or to dividend growers, or to quality.
  - Multifactor strategies: For example, an investor may choose a strategy that combines momentum, quality, size and value by maximizing exposure to stocks with several of those factors, in combination.
  - A combination of the two: offering both diversification and flexibility.
- In practice, to determine the right approach for a given situation there are several issues to consider 1). the structure of the financial data. 2). the economic intuition underlying the factor. 3). validity of the underlying assumptions of each approach.

Portfolio Sorts

- The portfolios are constructed by grouping together securities with similar characteristics (factors). The goal of this process is to determine whether a factor earns a systametic premium.
- The return of each portfolio is calculated by equally weighting the individual stock returns. The portfolios provide a representation of how returns vary across the different values of a factor. By studying the return behavior of the factor portfolios, we may assess the return and risk profile of the factor.
- Overall, the return behavior of the portfolios will help us conclude whether there is a premium associated with a factor and describe its properties.

- The construction of portfolio sorts on a factor is straightforward:
  1. Choose an appropriate sorting methodology.
  2. Sort the assets according to the factor.
  3. Group the sorted assets into N portfolios (usually $N = 5$, or $N = 10$).
  4. Compute average returns (and other statistics) of the assets in each portfolio over subsequent periods.
- The standard statistical testing procedure for portfolios sorts is to use a Student's $t-$test to evaluate the significance of the mean return differential between the portfolios of stocks with the highest and lowest values of the factor.

Choosing the Sorting Methodology

The sorting methodology should be consistent with the characteristics of the distribution of the factor and the economic motivation underlying its premium. Here six ways to sort factors:

Method 1:

Sort stocks with factor values from the highest to lowest.

Method 2:

Sort stocks with factor values from the lowest to highest.

Method 3: (Example: dividend yield factor)

First allocate stocks with zero factor values into the bottom portfolio.

Sort the remaining stocks with nonzero factor values into the remaining portfolios.

Compute average returns (and other statistics) of the assets in each portfolio over subsequent periods.

Choosing the Sorting Methodology

Method 4:

1). Allocate stocks with zero factor values into the middle portfolio.

2). Sort stocks with positive factor values into the remaining higher portfolios (greater than the middle portfolio).

3). Sort stocks with negative factor values into the remaining lower portfolios (less than the middle portfolio).

Method 5: (Example: rank stocks according to earnings growth on a sector neutral basis)

1). Sort stocks into partitions.

2). Rank assets within each partition.

3). Combine assets with the same ranking from the different partitions into portfolios.

Choosing the Sorting Methodology

Method 6: (Example: share repurchase factor)

1). Separate all the stocks with negative factors values. Split the group of stocks with negative values into two portfolios using the median value as the break point.

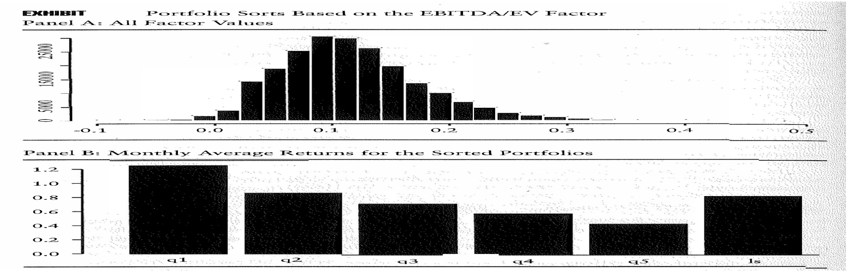
2). Allocate stocks with zero factor values into one portfolio.

3). Sort the remaining stocks with nonzero factor values into portfolios based on their factor values.

* The portfolio sort methodology has several advantages. The approach is easy to implement and can easily handle stocks that drop out or enter into the sample. The resulting portfolios diversify away idiosyncratic risk of individual assets and provide a way of assessing how average returns differ across different magnitude of a factor.

Example: Portfolio Sorts Based on the EBITDA/EV Factor

Exhibit 7.1 contains the cross-sectional distribution of the EBITDA/EV factor. This distribution is approximately normally distributed around a mean of 0.1, with a slight right skew. We use method 1 to sort the variables into five portfolios. Therefore, a strategy that goes long on portfolio 1 and short 5 appears to produce abnormal returns.



Information Ratios for Portfolio Sorts

- The information ratio (IR) is a statistic for summarizing the risk-adjusted performance of an investment strategy. It is defined as the ratio of the average excess return to the standard deviation of return.
- For actively managed equity long portfolios, the IR measures the risk-adjusted value a portfolio manager is adding relative to a benchmark.
- IR can also be used to capture the risk-adjusted performance of long-short portfolios from a portfolio sorts.
- When comparing portfolios built using different factors, the IR is an effective measure for differentiating the performance between the strategies.

New Research on Portfolio Sorts

- The standard statistical testing procedure uses a Student's t-test to evaluate the mean return differential between the two portfolios containing stocks with the highest and lowest values of the sorting factor.

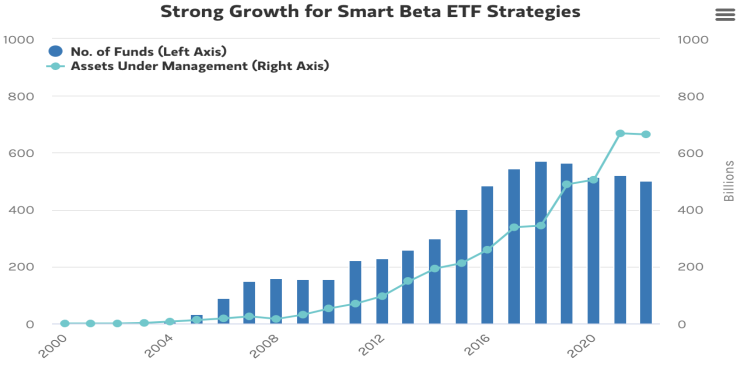
  However, this approach ignores important information about the overall pattern of returns among the remaining portfolios.

- The Monotonic Relation(MR) test can reveal whether the null hypothesis of no systemic relationship can be rejected in favor of a monotonic relationship predicted by economic theory.

  By MR it is meant that the expected returns of a factor should rise or decline monotonically in one direction as one goes from one portfolio to another.

Factor Based Trading Strategies

- Think of factor investing as a middle ground between passive and active investing. As of December 2022, there were \$664 billion in total assets under management with 503 U.S.-listed strategies.



Source: Morningstar Direct, Morgan Stanley Wealth Management ETF Research as of Dec. 31, 2022

- *Security Analysis* by Benjamin Graham and David Dodd (1934) was considered the first contribution to factor-based strategies. Today's quantitative managers use factors as fundamental building blocks for trading strategies. Within a trading strategy, factors determine when to buy and when to sell securities.

  We define a factor as a common characteristic among a group of assets. For example, the credit rating on a bond, or a particular financial ratio (P/E) or the book-to-price ratios, etc.

- We further expand: 1). Factors frequently are intended to capture some economic intuition. 2). We should recognize that assets with similar factors tend to behave in similar ways. 3). We'd like our factor to be able to differentiate across different markets and samples. 4). We want our factor to be robust across different time periods.

Factor Types

- Factors fall into three categories – macroeconomic influences, cross-sectional characteristics, and statistical factors.
  1. **Macroeconomic influences** are time series that measure observable economic activities. Examples include interest rate levels, gross domestic production, and industrial production.
  2. **Cross-sectional characteristics** are observable asset specifics or firm characteristics. Examples include, dividend yield, book value, and volatility.
  3. **Statistical factors** are unobservable or latent factors common across a group of assets. These factors make no explicit assumptions about the asset characteristics that drive commonality in returns. Statistical factors are not determined using exogenous data but are extracted from other variables such as returns.

Basic Framework and Building Blocks

- We focus on using factors to build forecasting models, also referred to as *alpha* or *stock selection models*. We begin by designing a framework that is flexible enough so that the components can be easily modified, yet structured enough that we remain focused on our end goal of designing a profitable trading strategy. The typical steps in the development of a trading strategy are:

1. Defining a trading idea or investment strategy
2. Developing factors
3. Acquiring/processing data
4. Analyzing the factors
5. Building the strategy
6. Evaluating the strategy
7. Backtesting the strategy
8. Implementing the strategy

Basic Framework and Building Blocks

- **Defining a Trading Idea or Investing Strategy:**

  A successful trading strategy often starts as an idea based on sound economic intuition, market insight, or the discovery of an anomaly. Background research can be helpful in order to understand what others have tried or implemented in the past.

  A trading idea has a more short-term horizon often associated with an event or mis-pricing. **A trading strategy** has a longer horizon and is frequently based on the exploration of a premium associated with an anomaly or a characteristic.

- **Developing Factors:**

  Factors provide building blocks of the model used to build an investment strategy. After having established the trading strategy, we move from the economic concepts to the construction of factors that may be available to capture our intuition.

Basic Framework and Building Blocks

- **Acquiring and Processing Data:**

  A trading strategy relies on accurate and clean data to build factors. There are a number of third-party solutions and databases available for this purpose such as Thomson Routers, Bloomberg, Market IQ, Factset Research Systems, and Compustats.

- **Analyzing the Factors:**

  A variety of statistical and econometric techniques must be performed on the data to evaluate the empirical properties of factors. This empirical research is used to understand the risk and return potential of a factor. The analysis is the starting point for building a model of a trading strategy.

Basic Framework and Building Blocks

- **Building Strategy:**

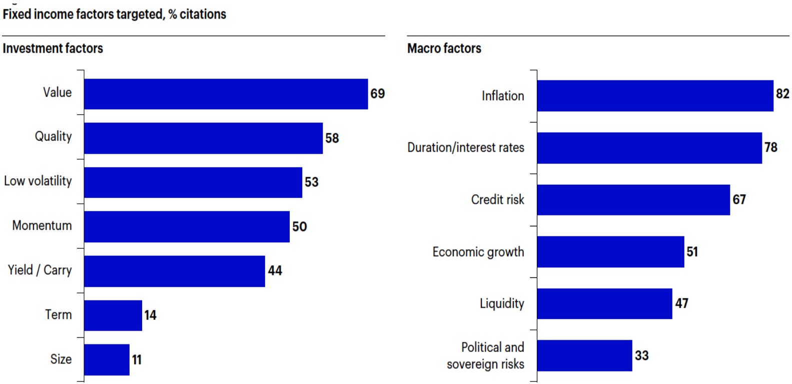
  The model represents a mathematical specification of the trading strategy. There are two important considerations in this specification: the selection of which factors and how these factors are combined. Both considerations need to be motivated by the economic intuition behind the trading strategy.

- **Evaluating, Backtesting, and Implementing the Strategy:**

  The final step involves assessing the estimation, specification, and forecasting quality of the model. This analysis includes examining the goodness of fit (often done in sample), forecasting ability (often done out of sample), and sensitivity and risk characteristics of the model.

Example of Factor Investing Trend (Invesco 2022 Survey)

- Fixed income returns are closely tied to fundamental macroeconomic variables. Investors applying a systematic approach to their fixed income portfolios often initially prioritize traditional macro drivers of return.



Which investments factors are you targeting in your fixed income portfolio? Which macro fixed income factors are you using in your factor investing strategy? Sample size: 45.

Python Time Series Factor Analysis Package

- Algorithmic trading strategies use linear factor models to quantify the relationship between the return of an asset and the sources of risk that represent the main drivers of these returns. Each factor risk carries a premium, and the total asset return can be expected to correspond to a weighted average of these risk premia.
- There are several practical applications of factor models across the portfolio management process from construction and asset selection to risk management and performance evaluation. The importance of factor models continues to grow as common risk factors are now traceable.
- This example illustrates the steps for estimating a factor model using as an example the data and process which led to results reported in Fama-French factor website.