FE670 Algorithmic Trading Strategies

Alpha Factor Trading Strategies

Steve Yang

Stevens Institute of Technology

steve.yang@stevens.edu

09/23/2021

Overview

CAPM and Fama-French Factor Models

Factor Based Trading

Risks to Trading Strategies

Desirable Properties of Factors

Building Factors from Company Characteristics

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 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 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 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. Size (SMB): Nine small stock PF minus mine large stock PF.
  2. Value (HML): Two value PF minus two growth (with low BE/ME value) PF.
  3. Profitability (RMW) : Two robust OP PF minus two weak OP PF.
  4. Investment (CMA): Two conservative investment portfolio, minus two aggressive investment portfolio.
  5. Market (Rm-Rf): Value-weighted return of all firms incorporated in and listed on major US exchanges, minus the one-month Treasury bill rate.

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:

$$\mathbf{r}_i = \mathbf{F}\beta_i + \epsilon_i \tag{2}$$

  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:

$$\mathbf{r}_t = \hat{\beta}_i \lambda_t \tag{3}$$

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:

$$t = \frac{\lambda_j}{\sigma(\lambda_j)/\sqrt{T}} \tag{4}$$

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

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 metrics associated with the risk factors are defined as follows:
  1. Size: Market equity (ME)
  2. Value: Book value of equity (BE)
  3. Operating profitability (OP): Revenue minus cost of goods sold/assets
  4. Investment: Investment/assets
- Factor investing has become a widely discussed part of today's investment canon.

Factor Based Trading

- Broadly we can classify investment strategies into the following categories:
  1. Factor-based trading strategies (also called stock selection or alpha models).
  2. Statistical arbitrage.
  3. High-frequency strategies.
  4. Event strategies.

* Most academics and practitioners agree that the efficient market hypothesis does not hold all the time and that it is possible to beat the market.

** Industry survey shows factors and factor-based models form the core of a major part of today's quantitative trading strategies.

- *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:

Defining a trading idea or investment strategy
Developing factors
Acquiring and processing data
Analyzing the factors

Building the strategy
Evaluating the strategy
Backtesting the strategy
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.

Risks to Trading Strategies

In investment management, risk is a primary concern. The majority of trading strategies are not risk free but rather subject to various risks. Here we describe some common risks to factor trading strategies as well as other trading strategies.

- *Fundamental risk* is the risk of suffering adverse fundamental news. For example, a good company with high earnings to price ratios may suddenly face a class-action litigation. We can minimizing the exposure to fundamental risk by diversifying across many companies. But sometime, fundamental risk could be systemic. In this case, portfolio managers that were sector or market neutral in general may do well.

- *Noise risk* is the risk that a mispricing may worsen in the short run. The idea here is that the premium or value takes too long to be realized, resulting in a realized lower than a targeted rate of return.
- *Model risk*, also referred to as misspecification risk, refers to the risk associated with making wrong modeling assumptions and decisions. This includes the choice of variables, methodology, and context the model operates in. We reviewed several remedies based on information theory, Bayesian methods, shrinkage, and random coefficient models.
- *Liquidity risk* is a concern for investors. Liquidity is defined as the ability to trade quickly without significant price changes, and the ability to trade large volume without significant price changes. Liquidity could be dried up under a stressed market circumstance.

Desirable Properties of Factors

Factors should be founded on sound economic intuition, market insight, or an anomaly. In addition to the underlying economic reasoning, factors should have other properties that make them effective for forecasting:

- It is an advantage if factors are **intuitive** to investors. Many investors will only invest in particular funds if they understand and agree with the basic ideas behind the trading strategies. Factors give portfolio managers a tool in communicating to investors what themes they are investing in.
- The search for the **economic meaningful** factors should avoid strictly relying on pure historical analysis. Factors used in a model should not emerge from a sequential process of evaluating successful factors while removing less favorable ones.

- A group of factors should be **parsimonious** in its description of the trading strategy. This will require careful evaluation of the interaction between the different factors. For example, highly correlated factors will cause the interferences made in a multivariate approach to be less reliable.

- The success or failure of factors selected should not depend on a few outliers. It is desirable to construct factors that are reasonably **robust to outliers**.

  Source of Factors: The sources are widespread with no one dominating clearly. Search through a variety of sources seems to provide the best opportunity to uncover factors that will be valuable for new models. Example sources include economic foundations, inefficiency in processing information, financial reports, discussions with portfolio managers or traders, sell-side reports or equity research reports, and academic literature in finance, accounting, and economics, etc.

Building Factors from Company Characteristics

- We desire our factors to relate the financial data provided by a company to metrics that investors use when making decisions about the attractiveness of a stock such as valuation ratios, operating efficiency ratios, profitability ratios, and solvency ratios.
- Factors should also relate to the market data such as forecasts, prices and returns, and trading volume. We distinguish three categories of financial data: time series, cross-sectional, and panel data.
- *Time series* data consist of information and variables collected over multiple time periods. *Cross-sectional* data consist of data collected at one point in time for many different companies. *A panel* data set consists of cross-sectional data collected at different points in time.

Data Integrity

Quality data maintain several attributes such as providing a consistent view of history, maintaining good data availability, containing no survivorship, and avoid look-ahead bias. It is important for the quantitative researchers to be able to recognize the limitations and adjust the data accordingly.

1. *Backfilling* of data happens when a company is first entered into a database at the current period and its historical data are also added.
2. *Restatement* of data are prevalent in distorting consistency of data. Many database companies may overwrite the number initially recorded.
3. *Survivorship bias* occurs when companies are removed from the database when they no longer exist.
4. *Lookahead bias* occurs when data are used in a study that would not have been available during the actual period analyzed.

Methods to Adjust Factors

- A factor may need to be adjusted using analytical or statistical techniques to be more useful for modeling. The following three adjustment are common:
- **Standardization:** It rescales a variable while preserving its order. Typically, we choose the standardized variable to have a mean of zero and a standard deviation of one by using the transformation

$$x_i^{new} = \frac{x_i - \bar{x}_i}{\sigma_x}$$

- **Orthogonalization:** Orthogonalizing a factor for other specified factor(s) removes this relationship. To orthogonalize the factor using averages according to industries or sectors, we can first calculate industry scores

$$s_k = \frac{\sum_{i=1}^{n} x_i \cdot \text{ind}_{i,k}}{\sum_{i=1}^{n} \text{ind}_{i,k}}$$

where x_i is a factor and $\text{ind}_{i,k}$ represent the weight of stock i in industry k. Next we subtract the industry average of the industry scores, s_k, from each stock. We compute

$$x_i^{new} = x_i - \sum_{k \in \text{Industries}} \text{ind}_{i,k} \cdot s_k$$

where x_i^{new} is the new industry neutral factor.

We can also use linear regression to orthogonalize a factor. We first determine the coefficients in the equation

$$x_i = a + b \cdot f_i + \epsilon_i$$

where f_i is the factor to orthogonalize the factor x_i by, b is the contribution of f_i to x_i, and ϵ_i is the component of the factor x_i not related to $f_i \cdot \epsilon_i$ is orthogonal to f_i (that is, ϵ_i is independent of f_i) and represents the neutralized factor $x^{\text{new}} = \epsilon_i$

In the same fashion, we can orthogonalize our variable relative to a set of factors by using the multivariate linear regression

$$x_i = a + \sum_j b_j \cdot f_j + \epsilon_i$$

and then setting $x_j^{new} = \epsilon_i$.

The interaction between factors in a risk model and an alpha model often concerns portfolio managers. One possible approach to address this concern is to orthogonalize the factors or final scores from the alpha model against the factors used in the risk model.

- **Transformation:** It is a common practice to apply transformations to data used in statistical and econometric models. In particular, factors are often transformed such that the resulting series is symmetric or close to being normally distributed. Frequently used transformations include natural logarithms, exponentials, and square roots.

- **Outliers Detection and Management:** Outliers are observations that seem to be inconsistent with the other values in a data set. Financial data contain outliers for a number of reasons including data errors, measurement errors, or unusual events.

- Outliers can be detected by several methods. Graphs such as boxplots, scatter plots, or histograms can be useful to visually identify them. Alternatively there are number of numerical techniques available. One common method is to compute the inter-quantile-range and then identify outliers as a measure of dispersion and is calculated as the difference between the third and first quartiles of a sample.

- Winsorization is the process of transforming extreme values in the data. First, we calculate percentiles of the data. Next we define outliers by referencing a certain percentile ranking. It is important to fully investigate the practical consequences of using either one of these procedures.

Python Time Series Factor Analysis Package

- 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.