Factor Models

Ernest Chan, Ph.D.
QTS Capital Management, LLC.

About Me

- Researcher at IBM T. J. Watson Lab in machine learning Quantitative researcher/trader for Morgan Stanley, Credit Suisse, and various hedge funds.
- Principal of QTS Capital Management which manages a hedge fund as well as client accounts.
- Author:
  - *Quantitative Trading: How to Build Your Own Algorithmic Trading Business* (Wiley 2009).
  - *Algorithmic Trading: Winning Strategies and Their Rationale* (Wiley 2013).
- Blogger: epchan.blogspot.com

Factor Models in Practice

- A factor (or factor loading) is simply any variable that can be used to predict returns.
  - Every number in the financial statement of a company or technical indicator can be a (cross-sectional) factor loading.
    - E.g. ROE, Book/Market ratio, Dividend yield, Recent return.
  - Every macroeconomic variable can be a (time-series) factor for a stock.
    - E.g. HML return, SMB return, Gold return, GDP growth
    - Each stock will have different "factor loading" i.e. regression coefficient w.r.t. common time-series factor.

What are Factors?

- Factors imply both returns and risks:
  – E.g. HML: long value vs short growth stocks generates returns over long run, but can suffer prolong drawdown during financial crises.
  – Returns is the compensation for those risks.
  – Factor returns are not easy to arbitrage away and is enduring, since not every investor want to suffer the risks.
  – If risks diminish, returns will diminish. E.g. SMB generates minimal returns in recent years.

Computing Time-Series Factors

- This is straightforward: for each stock, just take as long a returns series as we like, and regress it against the factor(s) such as HML returns.
  - Of course, we need to lag the HML return by one period in order to be predictive.

Computing Cross-Sectional Factors

- E.g. ROE, B/M, Dividend Yield are observable factor *loadings*.
- It is *not* as straightforward to compute cross-sectional factors.
- Naively, we can just take one snapshot in time t-1, and regress the 1-day return from t-1 to t against the factor loadings of all the stocks.

Computing Cross-Sectional Factors

- E.g. regress dependent variable vector

  [FutRet(AAPL) FutRet(GOOG) FutRet(MSFT) …]^T

against independent variable vector

[Earnings(AAPL) Earnings(GOOG) Earnings(MSFT) …]^T

results in one factor (regression coefficient).

- Multiple regression using matrix for earnings, dividends, etc., can accommodate multiple factors.

Computing Cross-Sectional Factors

- This suffers from insufficient data, and factors can vary greatly and unrealistically from day to day (or month-to-month, quarter-to-quarter).
- More robust method: *aggregate* data.
  - Aggregate returns over many periods in history, therefore “tying” the factors of different periods to be the same number.

Even Simpler than Regression…

- In finance, sometimes even linear regression is overfitting.
  - If we have multiple factors, linear regression will inevitably assign different weights /factor loadings/regression coefficients to them.
  - Sometimes only the sign of each coefficient is reliable, not the magnitude.
  - We might just “standardize” each factor by its mean and standard deviation, apply correct sign, and add all factors with equal weight.

Standardization of Factors

- Hypothetical Example
  - ROE of stocks in an index has mean of 0.6 and standard deviation of 0.4.
  - B/M of stocks in same index has mean of 0.1 and standard deviation of 0.5.
  - MSFT is in that index, and currently has ROE=0.3, B/M=0.2
  - Factor for MSFT = (0.3-0.6)/0.4 + (0.2-0.1)/0.5

    =-0.55
  - "+" would be "-" if B/M anti-correlates with future returns.

Equal Weight, Adding Ranks, Multi-sort

- Sometimes even this “standardization” is unnecessary: just rank the stocks according to each factor, and add up those ranks for each stock to get a summary rank*!
- Alternatively, we can sort a portfolio of stocks with one factor, pick top and bottom quintiles, then re-sort with different (less predictive) factor, again pick top and bottom quintiles within the previous quintiles, and so on. (*i.e.* Multi-sort*.)

Simpler the Better

- The equal weight/rank method is found to outperform* many regression-based method in many different areas of social science including finance. (*Daniel Kahneman, “Thinking, Fast and Slow”).

Some Exotic Factors for Stocks

- “Variance Risk Premium”: Difference Between Implied Volatility and Historical Volatility.
  - High VRP predicts negative returns.
- “Implied Skew”: Skew of returns implied by difference between OTM call and put option prices.
  - High Implied Skew predicts positive returns.

Some Exotic Factors for Stocks

- "Implied Kurtosis": Kurtosis of returns implied by difference between OTM call + put option prices and ATM call + put option prices.
  - High Implied Kurtosis predicts positive returns.
  - Unintuitive given VRP results!
- Short interest (sign?)
  - Depends on how exactly* you measure short interest.
- Liquidity
  - Low liquidity (volume) predicts positive returns.

Some Exotic Factors for Stocks

- "News sentiment"
  - Natural language processing algorithms used to parse and analyze all news feed automatically.
  - "Sentiment score" assigned to each story indicating possible price impact.
  - Aggregation of sentiment score from fixed period is predictive of future returns.
  - See www.ravenpack.com/research/shorttermstockselectionpaperform.htm

Nuances

- If we are ranking stocks based on a single factor and not in a multi-factor model, beware that the sign of the regression coefficient may change between large cap stocks and small cap stocks: better segregate them!
  - Same problem can occur with other factors, as mentioned previously.
- Similarly, some factor models do not work on all industry groups. (E.g. Joel Greenblatt’s model). Need to exclude some groups!

Thank you for joining us!

- Please check out my online workshop: Artificial Intelligence for Traders, Jul 16, 23.

  See epchan.com/workshops

- Keep in touch!
  - Twitter @chanep
  - Blog epchan.blogspot.com