Finding and Backtesting Strategies

Tinghao Li

Jun 2019

Agenda

Introduction

- Why care ? Plans to consistently win more, not one time luck
- More importantly, understand the other side of the trades.
  - Why do they do that ?
  - Will they continue doing these ?
  - Do we have an edge over them ?
  - Examples, farmers use agriculture futures to hedge prices fluctuations or oil producer BP own a large oil futures trading team.
- The author summarized the profit sources into **Fig.1**

Introduction

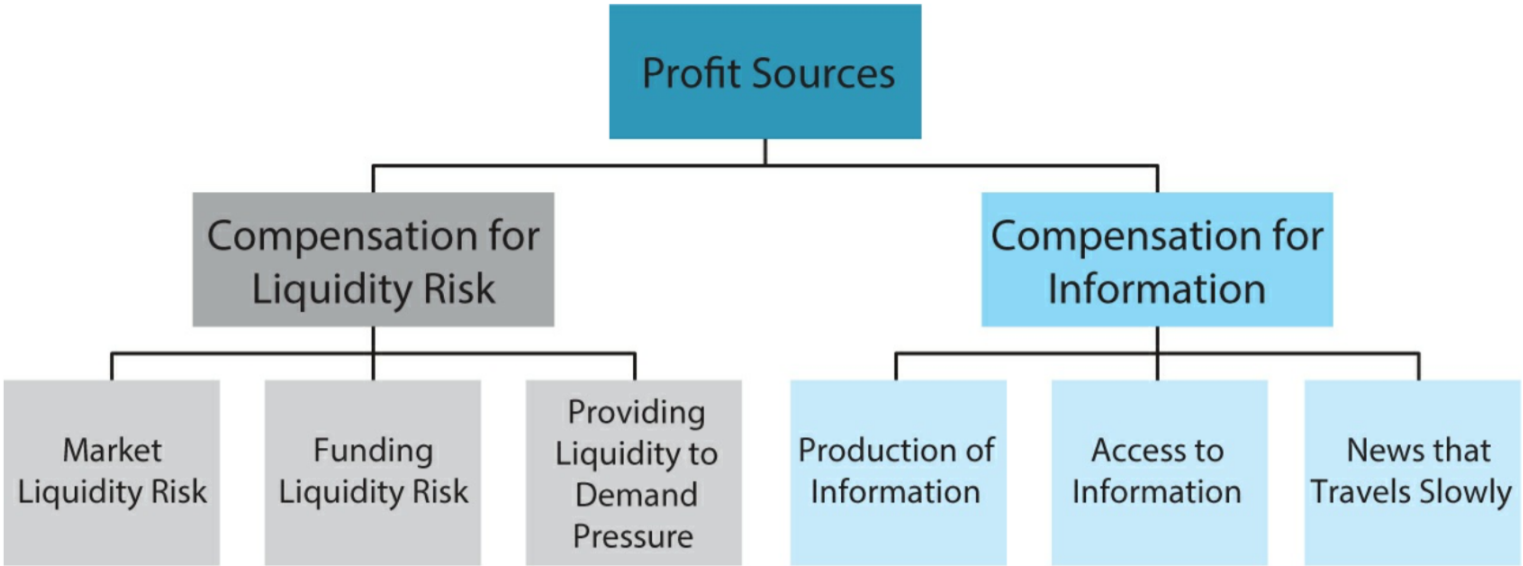


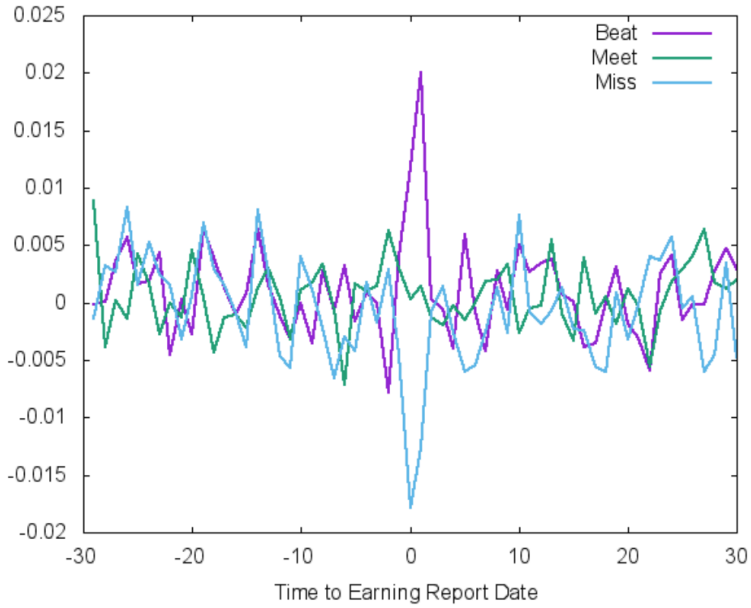
Figure 3.1. The main sources of profit for hedge fund strategies.

PnL Source 1 - Information

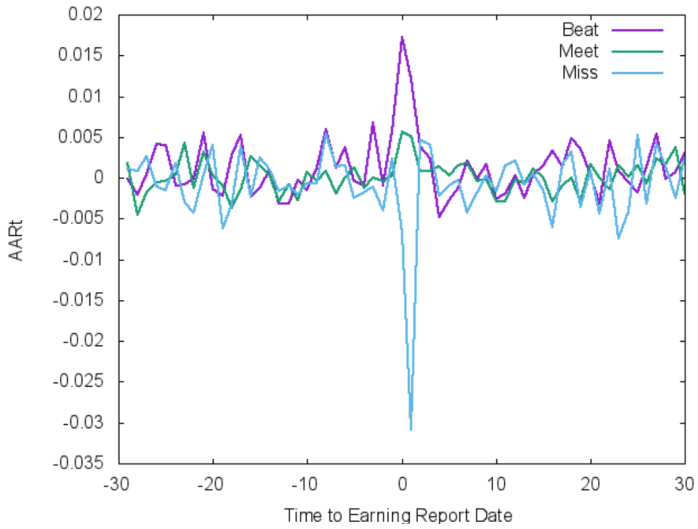
Why information can be a repeatable source of **profits**?
Because markets cannot be perfectly efficient and always reflect the full information.

- **Production of information**
- **Access of information** credit card transaction data, and other alternative data sources/superior information
- **Processing of information** hard to interpret the news and time to adjust the news (post-earnings-announcement drift ?)

Is that **Slow** ?



(a) Consumer Stocks from Russell 2000 in 2015



(b) Financial Stocks from Russell 2000 in 2015

- y-axis is the daily returns - market return.
- It might take weeks to digest but market responses are efficient

PnL Source 1 - Information con't

The pitfalls of price discovery from information/arbitrage

- It often carries fundamental or noise trades risks, e.g., an undervalued company can go further undervalued, or the divergence would be likely to increase before converging.
- The hedge funds might sometimes ride the bubble.

Overall, when looking for new signals, think :

- New information overlooked by others and thus it is not fully priced
- New ways to combine/analyze various source of information
- Get/infer information faster

PnL Source 2 - Liquidity Risk

- Market liquidity risk - "They'll let you in, but they won't let you out"
  - Liquidity-adjusted capital asset pricing model

$$E(R_i - TC^i) = R^f + \beta^i \lambda$$

, where λ is the (market) risk premium and β^i is computed as :

$$\beta^i = \frac{cov(R^i - TC^i, R^M - TC^M)}{var(R^M - TC^M)}$$

- Funding Liquidity Risks
- Providing Liquidity to Demand Pressure

Backtest considerations

- Universe
- Signals/alphas
- Trading rules
- Align the time series (time lags)

Backtest considerations

- Universe
- Signals/alphas
- Trading rules
- ~~Align the time series (time lags)~~ Risk management

Backtest considerations - trading rules

- Portfolio re-balance rule. Determine the optimal weights of each securities according to specific rules.
- Enter-exit trading rule. When to get in/out and the positions along the way

Data mining and biases

Backtests typically look a lot better than the real world trading performance, otherwise, it is less likely you'd adopt them in production - **selection bias** and other reasons :

- The world is changing
- You are slower than others

But more fundamentally, they suffer from **data mining**.

Understand the avoidable biases

- **Survival bias** some stocks disappear in the future which we do not know when we trade your "new" trading ideas
- **Shorting** Sometimes it is impossible to short the underlying
- **Data leaking bias** One way to overcome is to split the data for in-sample and out-of-sample, only using the latter one when everything is ready ; the other way is via rolling window but hard to reduce the long-range serial dependence
- **Over-parameterization** Less complicated models and less parameters - sensitivity analysis
- **Look-ahead bias** Most importantly, never used the future data

Adjusting backtests for transaction costs (TC)

TC reduces the return of the trading strategies, especially for high turnover portfolios. So the following steps are typically considered for TC in backtesting.

- The availability of data, e.g., frequency and levels of data, for estimating the features (liquidity) of the underlying assets
- Estimate/compute the fixed and variable costs, such as commissions, exchange fees, spreads, and the cost of consuming liquidity

Predictive signals & regression coefficients

A successful trading strategy is ultimately based on a signal that can predict returns, thus running a **predictive** regression is a useful tool to examine the signals. The author proposed the following theorems :

- A time series regression corresponds to a market timing strategy
- A cross-sectional regression corresponds to a security strategy
- A Univariate regression corresponds to sorting securities by one signal while a multivariate regression for sorting by multiple signals

Time series regression

- Considering to predict the future returns with

$$R^e_{t+1} = a + bF_t + \epsilon_{t+1}$$

, where F is the factor/signal.

- According to the OLS estimate of coefficient

$$b = \frac{\sum_t (F_t - \overline{F}) R_{t+1}}{\sum_t (F_t - \overline{F})^2} = \sum_t x_t R_{t+1}$$

, where trading position $x_t = k(F_t - \overline{F}_t)$ and $k = \frac{1}{\sum_t (F_t - \overline{F})^2}$.

  - We long the security when the signal is above its average and vice versa.
  - This strategy is profitable when b is positive

Cross-sectional regression

- We can run the following cross-sectional regression with each security i at time t

$$R^i_{t+1} = a + b_t F^i_t + \epsilon^i_{t+1}$$

, where F is the factor/signal.

- Similarly,

$$b_t = \frac{\sum_i (F^i_t - \overline{F}_t) R^i_{t+1}}{\sum_i (F^i_t - \overline{F}_t)^2} = \sum_i x^i_t R^i_{t+1}$$

, where trading position $x^i_t = k_t(F^i_t - \overline{F}_t)$.

- Thus the strategy selects long positions for securities with signals larger than the average and otherwise short.

Cross-sectional regression con't

- Furthermore, since b_t is the accumulative returns of stocks regarding signal F at time t, we could compute this selection strategy's returns and volatility over time

$$\hat{b} = \frac{1}{T} \sum_{t=1,2...T} b_t$$

, and

$$\hat{\sigma} = \sqrt{\frac{1}{T-1} \sum_{t=1,2...T} (b_t - \bar{b})^2}$$

- Therefore, the sharp ratio is $SR = \frac{\hat{b}}{\hat{\sigma}}$ while $t - stat = \sqrt{T}\frac{\hat{b}}{\hat{\sigma}}$

Univariate/multivariate regression

- We can also regress returns on several factors

$$R^i_{t+1} = a + b^F F^i_t + b^G G^i_t + \epsilon^i_{t+1}$$

, where F and G are the signals.

  - We could use this to asses the benefits of adding new factors
  - Also, it could be used to analyze the PnL attributions among factors as well.

Summary & Discussions

- When researching trading strategies, understanding your profit sources.
- The connections between signals/factors predictive power and the OLS regression analysis.
- When backtesting trading strategies, understanding the common pitfalls/biases and find ways to overcome.
- **Backtesting** while **researching** is like drinking and driving. Do not research under the influence of a backtest - Marcos' second law of backtesting.

Thanks !

Questions ?