Expected Return in High Frequency Trading

Ricky Cooper
Assistant Professor
Stuart School of Business
Illinois Institute of Technology
565 W. Adams, Chicago, IL 60661
312 906 6534 rcooper3@iit.edu

Ben Van Vliet
Assistant Professor
Stuart School of Business
Illinois Institute of Technology
565 W. Adams, Chicago, IL 60661
312 906 6513 bvanvliet@stuart.iit.edu

Abstract

Defining α in high frequency trading is more complicated than in low frequency since not all strategies are based on price forecasts. More components are required, as is an understanding of the interactions between them. In this paper, we develop the α attribution model for high frequency trading by explicating its components and the trading tactics used to implement high frequency strategies. The results show why high frequency traders need to be fast in order to generate positive expected returns and why they are better at providing liquidity. We provide an example implementation using a sample of high frequency equity data.

In high frequency trading (henceforth HFT), positive expected return is the key to profitability. Typically, this expectancy is called alpha (α). The consistency of α and the frequency with which it is earned are the distinguishing characteristics of algorithmic strategies relative to human traders. In the investment literature of low frequency trading, α is equal to the product of the volatility times the information coefficient (IC) times the *z*-score, a measure of confidence in the forecast signal (Grinold [1994]). Defining α in HFT however, is a bit more complicated since not all strategies are based on price forecasts. More components are required, as is an understanding of the interactions among them.

In this paper, we develop the α attribution model for HFT. We do this by explicating the components of α, as well as the trading tactics used to implement HFT strategies. The components are:

1) *opportunity*
2) *capture*
3) *effective spread*
4) *effective rebate*.

Also we provide an example implementation using a sample of high frequency equity data.

ALPHA IN HFT

The HFT industry most often defines α in terms of absolute return[1]. The average absolute return (either on a per-trade or per-unit-of-time basis) generated by a backtest or simulated trading should rightly be called the backtested α or the simulated α. We certainly use backtested and/or simulated α's as a justification for belief in the future α (i.e. once the strategy is operational). Breaking these α's into their components allows for improvement of a trading strategy, or as is often the case, an *ex post* analysis of why a strategy deviated from its expected performance.

This should be possible if one begins with the perspective that high frequency strategies, like their low frequency counterparts, are primarily seeking to profit by removing inefficiencies from the marketplace. In doing so they must be aware of the same basic ideas that affect all investment strategies: how much opportunity is available to capture; how much can be captured; and, what is the cost of capturing it? To this end we define the components necessary for a systematic study of α in HFT.

Opportunity (O)

The starting point of any discussion of α is the amount of price movement or opportunity (O) available for capture. Given a particular holding period, the amount of price movement over that period represents the available profit. A common measure of this movement is the standard deviation of the changes in the bid-ask mid-point price[2]. While standard deviation is certainly the appropriate measure for portfolio strategies that require continuous exposure to the market, for opportunistic HFT strategies, which only enter positions under certain conditions, different measures of opportunity may be appropriate (such as, say, the 90^{th} percentile movement, or even a fixed number of cents or handles, in the case of futures trading). However, in the absence some other measure, we suggest using standard deviation as a proxy for opportunity.

Capture (C)

We define the capture (C) as the percentage of the opportunity that not just a forecast signal, but more generally, *any* strategy can capture. In the case of portfolio strategies, capture is IC × z-score (see Grinold [1994]) and is most often measured as the correlation of forecasted returns with the actual, realized returns. Because IC is based upon price forecasts, any negative value for IC is bad. But, in HFT a negative value for C may very well be acceptable, because measures other than correlation may be more appropriate. For statistical arbitrage strategies that have a fixed payoff, something like a hit rate might be better. The idea is that strategies that are based on a forecast should have some positive C closely related (if not exactly) to IC, whereas strategies that are based on liquidity provision may have some other C, which could even be less than zero. Whatever the case, given the two components we have so far and before taking into account the trading tactics, α is simply $C \times O$, the captured opportunity[3].

Effective Spread (S_E)

In low frequency trading, the bid-ask spread is generally ignored as a component of α because the opportunities sought are far larger. In HFT, however, where the holding periods are very short, the bid-ask spread has a large impact on α. The bid-ask spread (S) is simply the difference between bid (i.e. the price received by someone who needs to sell immediately) and the ask (i.e. the price paid by to someone who needs to buy immediately). It is traditionally thought of as the premium paid to market makers for assuming the risk of adverse selection when transacting against informed trades as in Stoll [1978]. Whether an opportunistic trading strategy earns or pays S depends upon the tactics used to implement it.

Trading tactics are how a trading strategy uses marketable orders and limit orders to enter into and exit from positions in financial instruments. A limit order is a request to transact at a given price at or below (above) the top-of-book bid (ask) price. Such orders supply liquidity and *make* one side of the market, either bid or ask. Limit orders are passive, remaining in the exchange limit order book until they are matched against an incoming marketable sell (buy) order. A marketable order is any request to transact immediately at the best available bid (ask) price. Such orders demand liquidity and *take* the market price. Marketable orders can be either market orders or limit orders that have limit prices through the top-of-book ask (bid) price[4].

Combinations of make or take orders to create round trip trades define the three types[5] of trading tactics. Take-take tactics use two marketable orders to both enter into and exit from positions in the market. Make-take tactics use a limit order to enter into positions and a marketable order to exit from positions. Make-make tactics use a limit order to both enter into and exit from positions. The different tactics incur different trading costs with respect to the bid-ask spread S. Take-take tactics incur a transaction cost of one times S for each round trip trade. Make-take tactics incur a cost of zero times S for each round trip trade, and make-make tactics earn S for each round trip trade.

As an example, consider a simple market as in Exhibit 1. The inside market, or top-of-book, is 99 bid and 100 asked, and the bid-ask spread is simply one. (For simplicity, we ignore the quantities at these levels.) A strategy employing take-take tactics that enters a position by taking the market price of 100 to buy and then immediately sells by taking the market price of 99 loses one point simply by virtue of incurring the cost of the bid-ask spread S.

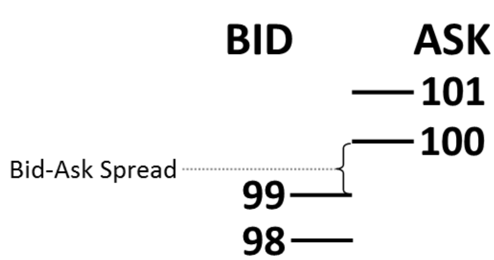

Exhibit 1: Simplified Market with Bid Ask Spread

A trading strategy that uses make-take tactics to enter a position by way of a limit order to buy at 99 and then immediately exits the position taking the market price to sell at 99 incurs no cost with respect to the bid-ask spread. Finally, a trading strategy that use make-make tactics enters a position by way of a limit order to buy at 99 and then immediately enters and at some later time gets filled on a limit order to sell at 100 earns the bid-ask spread S. These simple scenarios leads to the values for the effective spread (S_E) in equation (1).

$$S_E = \begin{cases} +S & \text{for take-take trades} \\ 0 & \text{for make-take trades} \\ -S & \text{for make-make trades} \end{cases} \tag{1}$$

Effective Rebate (R_E)

In equity markets, exchanges often pay a fee, called a rebate (R), to trading firms that supply liquidity by placing limit orders in the limit order book. Incentivizing liquidity suppliers is thought to be good for the exchange. Having deeper, more liquid markets ought to attract more and larger institutional liquidity takers, which increases trading volume and fees for the exchange. When limit orders are executed, or matched, the trading firm earns R. Thus, rebates can be an important component of α. The trading tactics also affect the effective rebate (R_E) as in equation (2). Since take-take tactics do not use limit orders, strategies that use them earn no rebate. A make-take tactics earn one rebate per round trip trade, and make-make tactics earn two times R per round trip trade.

$$R_E = \begin{cases} 0 & \text{for take-take trades} \\ R & \text{for make-take trades} \\ 2 \times R & \text{for make-make trades} \end{cases} \tag{2}$$

Expected Return (α)

Given the four components, the α of an HFT strategy can now be fully defined as:

$$\alpha = C \times O - S_E + R_E \tag{3}$$

In equation (3), α equals the captured opportunity less the net costs of implementing the trade. It ignores commissions and margins, which in HFT are often fixed. For example, a broker-dealer doesn't worry about commissions and high frequency traders with direct market access often pay a fixed fee per stock. If these are important variables to a particular firm deciding among various strategies, they can easily be appended to equation (3).

TACTICS MATTER

The complicating fact in equation (3) is that the values of the components are dependent upon each other. There are hidden interactions. The captured opportunity is not independent of the effective spread if we consider that:

1) Capturing an opportunity is a function of entering a position quickly and exiting that position as close to the optimal time as possible.
2) The effective spread is a function of the trading tactics employed. One can execute immediately and pay the spread, or earn the spread by waiting for the market to execute a passive limit order.

Thus, earning the effective spread entails sacrificing some captured opportunity. Or alternatively, capturing more opportunity means paying the effective spread. The tactics matter because the capture percentage C declines with execution speed. We can see the impact of tactics on α if we consider a trading strategy implemented in the three ways. Let's assume the trading strategy has the following characteristics:

- The average holding period is 60 seconds.
- The average bid-ask spread S is .08, or 8 cents.
- The opportunity over the 60 second holding period in standard deviation O_{60} is .09, or 9 cents.
- The R is 0.001, or a tenth of a penny.

Example 1: Take-Take

If the strategy uses take-take tactics, then the effective spread S_E is .08 and R_E is 0. If C is .25, then the α for this strategy is -.0575. Take-take tactics result in immediate execution and capture the full C × O, but incur –S. Therefore, C × O must be larger than S in order to have a profitable strategy.

$$\alpha = .25 \cdot .09 - .08 + 0 = -.0575$$

Example 2: Make-Take

If the strategy uses make-take tactics, then the effective spread S_E is 0 and R_E is 0.001. If C is reduced to .10, then the α for this strategy is .01. Make-take tactics do not incur –S, but do incur an unknown delay before the position opening trade occurs. Because of the delay in execution and the adverse selection, the value of C has declined. Thus, it behooves traders who use make-take tactics in their strategies to minimize the time spent waiting in the limit order queue[6].

$$\alpha = .10 \cdot .09 + 0 + .001 = .01$$

Example 3: Make-Make

If the strategy uses make-make tactics, then the effective spread S_E is -.08 and R_E is 0.002. If C is -.05, then the α for this strategy is .0775. The value of C has declined further still due to the waiting time to execute both sides of the trade and adverse selection on both sides as well. In this case, even though the C is negative, the spread and the rebate turns the expectancy positive. Make-make tactics are compensated by the amount of S and 2 × R for the waiting time, so that even though C is negative, the strategy still has a positive α.

$$\alpha = -.05 \cdot .09 + .08 + .002 = .0775$$

This scenario paints a rosy picture of strategies that provide liquidity. It does not take into account that such strategies occasionally incur extreme left tail returns when adverse selection events happen, and this especially true if the technology is slow. (We will discuss this point in more detail later.) This scenario leads to new trading strategies, ones where the holding period is very short and where the value of C is kept close to 0, which both work to reduce the possibility of adverse selection, so that α is $-S + R_E$. Example 3 show why HFT strategies are better at supplying liquidity than low frequency traders. Low frequency traders need O to be large, and a negative value of C is bad. HFT strategies can more consistently earn $-S + R_E$, and because their O is smaller, they can avoid adverse selection.

EMPIRICAL DATA[7] AND RESULTS

In order to demonstrate the characteristics of equation (1) and the impact of the various tactics on α, we use data on Apple, Inc. (AAPL) for January 3, 2012. (We tried various samples, but the results do not qualitatively change.) The dataset contains every message about every event in the NASDAQ limit order book, including all additions, cancelations, and executions. These messages are time-stamped to the nanosecond, so that we have an exact timing and ordering of all events. Using this data we calculate opportunity O using the standard deviation of the price changes in the bid-ask mid-point price over a range of times.

Using the data just described, the average bid-ask spread S over the day was 0.088704, or about 9 cents. The standard deviations in dollars over various holding periods are shown in Exhibit 2.

1 sec	5 sec	30 sec	1 min	5 min	10 min	20 min	30 min
0.0199	0.0415	0.0572	0.0935	0.1342	0.3451	0.5521	0.6890

Exhibit 2: Standard Deviations Over Various Holding Periods

Using the standard deviations in Exhibit 2 as proxies for opportunity, we calculate the α's according to equation (3) over values for capture C ranging from -1 to 1. (The case where C = 1 is logically equivalent to the "omniscient trader" of Kearns, et al. [2010].) We assume R = 0. Exhibits 3, 4 and 5 show the α's for each of the three tactics over the various holding periods. So, for example, in Exhibit 3, if the holding period is one second, C = -1.00, O = .0199, S = .088704, and R = 0, then for take-take tactics, the value of α is -0.109 as can be seen in the upper left hand corner. In each Exhibit 3-5, the shaded cells indicate where the value of α is positive. In all other cells, α is negative or 0.

Take-Take								
C	**1 sec**	**5 sec**	**30 sec**	**1 min**	**5 min**	**10 min**	**20 min**	**30 min**
-1.00	-0.109	-0.130	-0.146	-0.182	-0.223	-0.434	-0.641	-0.778
-0.75	-0.104	-0.120	-0.132	-0.159	-0.189	-0.348	-0.503	-0.605
-0.50	-0.099	-0.109	-0.117	-0.135	-0.156	-0.261	-0.365	-0.433
-0.25	-0.094	-0.099	-0.103	-0.112	-0.122	-0.175	-0.227	-0.261
0.00	-0.089	-0.089	-0.089	-0.089	-0.089	-0.089	-0.089	-0.089
0.25	-0.084	-0.078	-0.074	-0.065	-0.055	-0.002	**0.049**	**0.084**
0.50	-0.079	-0.068	-0.060	-0.042	-0.022	**0.084**	**0.187**	**0.256**
0.75	-0.074	-0.058	-0.046	-0.019	**0.012**	**0.170**	**0.325**	**0.428**
1.00	-0.069	-0.047	-0.032	**0.005**	**0.046**	**0.256**	**0.463**	**0.600**

Exhibit 3: Alphas Given Take-Take Tactics

In Exhibit 3, we can see that for take-take tactics, α is positive only when the value of C is impossibly high (i.e. .75 or 1.00), or the holding period is fairly long, at least by HFT standards. In practice, high values of C are available for strategies that chase very fleeting opportunities. For strategies that depend on price forecasts, values of C any higher than around .25 are very difficult to discover, and 20 to 30 minute holding periods are likely outside the bounds of the definition of high frequency. This combination makes it very difficult for an HFT strategy to have positive α using take-take tactics. The cost of the bid-ask spread is too difficult to overcome with better forecasts over short time frames.

Make-Take								
C	**1 sec**	**5 sec**	**30 sec**	**1 min**	**5 min**	**10 min**	**20 min**	**30 min**
-1.00	-0.020	-0.042	-0.057	-0.093	-0.134	-0.345	-0.552	-0.689
-0.75	-0.015	-0.031	-0.043	-0.070	-0.101	-0.259	-0.414	-0.517

-0.50	-0.010	-0.021	-0.029	-0.047	-0.067	-0.173	-0.276	-0.344
-0.25	-0.005	-0.010	-0.014	-0.023	-0.034	-0.086	-0.138	-0.172
0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.25	**0.005**	**0.010**	**0.014**	**0.023**	**0.034**	**0.086**	**0.138**	**0.172**
0.50	**0.010**	**0.021**	**0.029**	**0.047**	**0.067**	**0.173**	**0.276**	**0.344**
0.75	**0.015**	**0.031**	**0.043**	**0.070**	**0.101**	**0.259**	**0.414**	**0.517**
1.00	**0.020**	**0.042**	**0.057**	**0.093**	**0.134**	**0.345**	**0.552**	**0.689**

Exhibit 4: Alphas Given Make-Take Tactics

In Exhibit 4, we can see that for make-take tactics, α is positive for any positive value of C. This is fairly clear, since positive capture leads to positive α and negative capture leads to negative α when S = 0. The implicit assumption, however, is that the time spent in the queue, waiting to be executed, is short. Orders often remain in the queue for several seconds, even minutes, which precludes obtaining the α over these time frames. Of course, the faster one's technology, the more forward in the queue one's order will be, and therefore, the waiting time will be shortened. Thus, being fast enables trading firms to obtain α over these shorter time frames. Another implicit assumption in Exhibit 4 is that the value of C remains constant over time, which it certainly does not as we will show.

Make-Make								
C	**1 sec**	**5 sec**	**30 sec**	**1 min**	**5 min**	**10 min**	**20 min**	**30 min**
-1.00	**0.069**	**0.047**	**0.032**	-0.005	-0.046	-0.256	-0.463	-0.600
-0.75	**0.074**	**0.058**	**0.046**	**0.019**	-0.012	-0.170	-0.325	-0.428
-0.50	**0.079**	**0.068**	**0.060**	**0.042**	**0.022**	-0.084	-0.187	-0.256
-0.25	**0.084**	**0.078**	**0.074**	**0.065**	**0.055**	**0.002**	-0.049	-0.084
0.00	**0.089**	**0.089**	**0.089**	**0.089**	**0.089**	**0.089**	**0.089**	**0.089**
0.25	**0.094**	**0.099**	**0.103**	**0.112**	**0.122**	**0.175**	**0.227**	**0.261**
0.50	**0.099**	**0.109**	**0.117**	**0.135**	**0.156**	**0.261**	**0.365**	**0.433**
0.75	**0.104**	**0.120**	**0.132**	**0.159**	**0.189**	**0.348**	**0.503**	**0.605**
1.00	**0.109**	**0.130**	**0.146**	**0.182**	**0.223**	**0.434**	**0.641**	**0.778**

Exhibit 5: Alphas Given Make-Make Tactics

In Exhibit 5, we can see that for make-make tactics, α is positive for almost all values of C. Even where C is severely negative, the value of the spread earned overcomes essentially any strategy, no matter how dumb, as long as the technology is fast. As with the previous example, the ability to obtain the positive α's associated with the shorter holding periods depends upon getting the limit orders executed quickly. This can only happen consistently if the waiting period is short, meaning that you are consistently at the front of the queue. Being toward the back of the queue means waiting a long time for execution, and waiting a long time increases the possibility of adverse selection[8].

IMPACT OF SPEED

Technological speed has a profound impact on realized captured opportunity. First, as Exhibit 6 shows, the correlation of forecasts to actual price movements decays as a function of time. This decay is a function of the length of the forecast. Exhibit 6 shows the decay for 1 and 5 second forecasts given delays in tenths of a second. Thus, any delay in execution will negatively impact capture. Thus, many trading strategies will be ineligible either because using make-make tactics is too expensive in terms of the fixed technology costs required to be fast enough, or because using take-take tactics is too expensive terms of the spread cost.

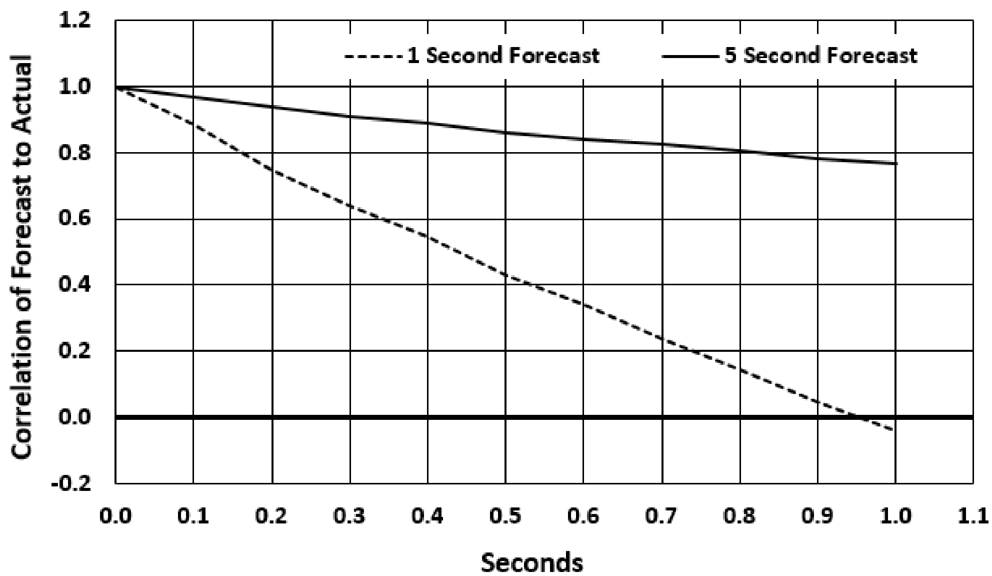

Exhibit 6: Decay in Forecast Over Time

Second, the delay in execution may affect the calculation of realized opportunity. Being slow results in being toward the back of the queue. Trades at the back of the queue tend to get executed more often against informed trades (in the wrong direction). The probability of adverse selection is higher and the realized opportunity will be worse than a simple standard deviation would indicate. This is unfortunate for a strategy that has a negative capture C. It may require a take transaction to stop out of the accumulating losses, incurring a worse effective spread than the make-make tactics assumed. Thus, the profitability of strategies that use make-make tactics in Exhibit 5 are illusory except for the very fast players.

CONCLUSION

HFT strategies face a complicated expected return equation. However, by breaking down α into its components, a trading firm can gain greater understanding of the variability in profits and losses. Of course, this variability includes not only the variability of the component, but also the correlations that must be considered. These correlations explain the need for speed. Technological speed helps prevent the components from developing large negative correlations causing a rapid downward spiral. Appreciation of the α equation can help risk managers, strategy developers, and regulators all understand the intricacies of HFT.

References

Grinold, R. C. "Alpha is volatility times IC times score." Journal of Portfolio Management, 20 (1994), pp. 9-16.

Stoll, H. R. "The supply of dealer services in securities markets." Journal of Finance, 33 (1978), pp. 1133-1151.

Kearns, M., A. Kulesza, Y. Nevmyvaka. "Empirical limitations on high frequency trading profitability." Journal of Trading, 5 (2010), pp. 50-62.

---

[1] Certain strategies could also involve residual return relative to a benchmark. Our methodology could easily be adapted in such a case.

[2] The bid-ask mid-point price is simply the bid price plus the ask price divided by two. Standard deviation is most often that of log returns, but we use it terms of dollars.

[3] For low frequency strategies, $C \times O$ would be exactly the same as in Grinold (1984).

[4] A buy limit order with a limit price at or above that the current top-of-book ask is not placed into the exchange limit order book but rather is matched immediately against a resting limit order at the market ask price.

[5] Take-make tactics are rarely, if ever, used in HFT.

[6] We assume a first-in-first-out (FIFO) queue with price and time priority.

[7] We would like to thank Xambala, Inc. for providing this data and for securing NASDAQ's permission to use it in our research.

[8] One complication we are not addressing is that adverse selection related to large market moves can lead to capitulation stop-loss trades, which are another reason that execution speed is important.