Trading with an Edge

Nastja Bethke[1], Ed Tricker[2]

Abstract
Systematic as well as discretionary trading strategies attempt to forecast future returns to then position themselves in line with anticipated market moves. Intuitively, successful prediction should lead to a profitable trading strategy. By most standard measures, however, it appears that many well-known trading strategies ought not to be successful at all, as their success rate in predicting market moves is relatively low. Or to paraphrase loosely, most strategies, when viewed from a certain angle, are not much better than a random coin toss. This note illustrates why a small edge over a random positioning is all you need.

[1] *Senior Quantitative Research Analyst*
[2] *CIO - Quantitative Strategies*

1. Every Little Bit Helps

Over the course of the last decade, the overall performance of the S&P 500 has been impressive, with an average annual return of around 10%, see the historical price chart in Figure 1. Yet on closer inspection, we find that the index had a positive return on 'only' 55% of the days. While we certainly would not expect a much larger percentage of positive-versus-negative return days for an aggregate of individual stock prices, it bears noting that this small deviation from an even 50-50-split has over the long run added up to a substantial cumulative return.

In the following discussion we investigate this phenomenon more closely, specifically looking at whether 'every little bit helps' also carries over to systematic trading strategies.

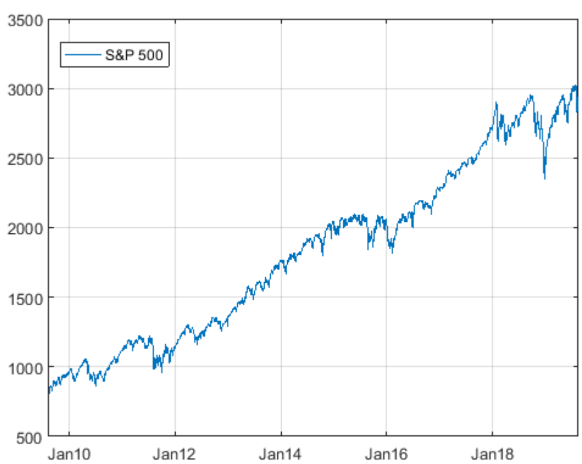

Figure 1. The level of the S&P 500 over the last decade.

2. Predicting Returns

Any systematic trading strategy has at its core a model to predict future returns. Trading decisions, i.e. a decision to go long or short a market, are then made according to the sign and potentially the magnitude of the predicted future return. Typical models used for prediction are built on a variety of techniques from statistics, signal processing and machine learning. There are various measures of the *goodness of fit* for these models and trading strategies in general, two of which we will look at more closely here. Looking at these measures in isolation, they both tend to suggest that such models are fairly bad. Yet we find that when placed in context, what is 'good' or 'bad', is relative.

For illustration, let us consider a prediction model built on linear regression. Choosing a predictor, e.g. yesterday's market return, one can then regress the given market's return on the following day on the predictor. Using a lookback of a year, such a regression model would then be built with 250 sample points (reflecting the number of business days).

Regression models are ubiquitous across the biological and social sciences, specifically in economics. Denoting the predictor, or *independent* variable, as x, and the variable to be predicted, the *dependent* variable as y, linear regression finds a linear function $f(x) = y$ that fits the N observed data points best.[1] A typical measure of the goodness of fit of a linear regression model is the *coefficient of determination* or R^2, which is defined as

$$R^2 \equiv 1 - \frac{S_{\text{res}}}{S_{\text{tot}}}, \tag{1}$$

where S_{res} measures the squared difference between the fit and actual observation, and S_{tot} represents the analogous discrepancy from the sample mean $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$:

$$S_{\text{res}} = \sum_{i=1}^{N}(y_i - f_i)^2 \text{ and } S_{\text{tot}} = \sum_{i=1}^{N}(y_i - \bar{y})^2. \tag{2}$$

Intuitively, the R^2 is best understood as the proportion of the variance in the dependent variable (here, the future returns) that is explained by the independent variable (here, the past returns), with a perfect model having an R^2 of 1 and an entirely unsuitable model having an R^2 of 0.[2]

[1] It does this by minimizing the discrepancy between the model fit and the actual observation.

[2] The R^2 also equals the square of the correlation coefficient when the regression includes an intercept.

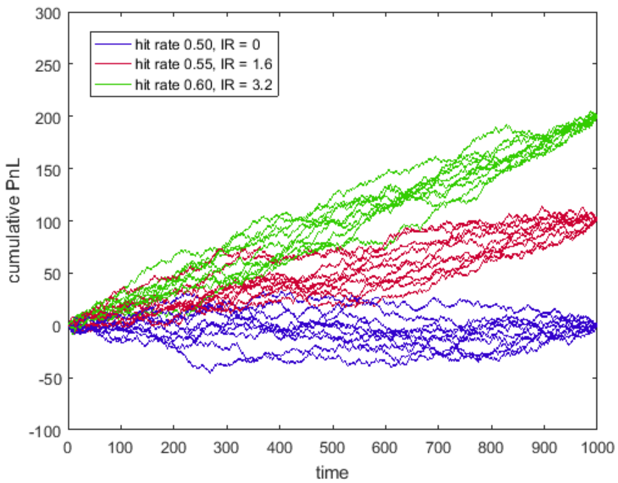

Figure 2. Simulation of PnL paths for three hit rates: h = 50%, 55% and 60%. PnL on any day can either be 1 or -1; each path comprises 1000 days, representing four years of trading.

While 'typical' R^2 values encountered in many regression models[3] in, e.g., biology or economics can be at least as high as 0.4, when it comes to predicting returns such values are unlikely to be attained. This is a reflection of the low signal-to-noise ratio present in financial data.[4]

The question then becomes: how can successful trading strategies exist, if the models at their core have such a 'bad fit', especially when measured against the yardstick of other disciplines? In this note we try and answer this question.

3. Successful Trading Strategies

The success of a trading strategy hinges on the relative number of profitable trades, as well as the relative magnitude of the associated profit, when compared to unprofitable trades. Generally, a higher frequency of profitable trades compared to that of unprofitable trades is desirable. For profits and losses of equal magnitude, and ignoring transaction costs (which we do throughout this note), such a frequency split would lead to positive cumulative PnL over time. We can of course imagine scenarios where just a few instances of extremely profitable trades lead to an overall gain, but, in the extreme, this type of strategy is not robust, as it is possible that those few instances are merely due to luck, and can equally likely occur in the other direction, causing substantial losses.

We can define the *hit rate*, the second of our measures of how good our model is, as the fraction of the number of profitable days out of all days in which a position was taken. To get a feeling for realistic and attainable hit rate values, we conduct a toy experiment: we fix the hit rate h and draw 1000 random numbers X which can only take the values 1 (a profit) and -1 (a loss), such that the number of 1s is equal to $h \times 1000$.[5] Each draw represents a trading day. Summing up the resulting draws of 1s and -1s gives us a cumulative PnL curve. For each hit rate, we plot ten such curves in Figure 2, and we test three different hit rates. For a given hit rate all cumulative PnL curves of course end at the same point, as the total number of 1s and -1s per path is identical; they also have the same IR (the ratio of annualized return over annualized volatility). As we can see, hit rates of 50% and 55% lead to IRs of 0 and 1.6, respectively. A hit rate of 60% results in an IR of 3.2. IRs between 0 and 2 can thus be viewed as bounds on what hit rates are realistic in practice for strategies of slow to medium frequency.

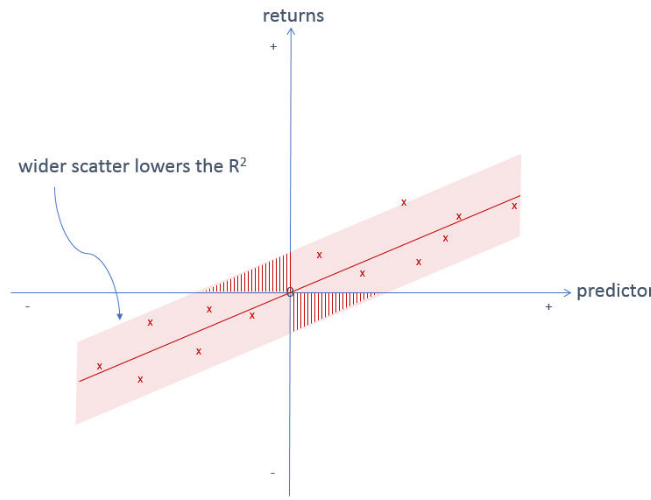

Figure 3. Illustration of a regression fitting returns to a predictor variable. The shaded red region is the extent of the scatter of samples around the model fit; the larger the scatter, the lower the R^2. Using the linear regression model to determine the position for a trade, the striped triangular regions represent unsuccessful trades, where the model predicts a negative or positive return, but the actual realization is a positive or negative return, respectively.

4. An Edge Is All You Need

Our toy experiment indicates that a moderate hit rate marginally above 50% can lead to satisfactory performance. We now want to link this hit rate to the quality of the model generating the return predictions. Earlier, we introduced the R^2 as a measure of the quality of fit of a linear regression model. To gain some intuition on whether R^2 values typically observed when trying to predict market returns can be reconciled with positive performance, we set up another stylized example.

Consider a setting where linear regression is used for predicting future returns, the 'noise' not captured by the linear regression is due to unknown and/or unobserved variables that affect the returns, and the underlying market dynamics are invariant over time [6]. Figure 3 illustrates this setup. The regression scatter -

[3] Here we have in mind models with one independent variable; adding more explanatory variables can only increase the R^2, often up to values larger than 0.5.

[4] Practitioners are also wary of overfitting, which can easily be achieved with non-linear models. Overfitting is effectively the removal of most of the discrepancy between the model fit and the actual observation, without the resulting model then generalizing to unseen data.

[5] We can also look at this from a theoretical point of view: in our experiment the probability of a profit is $P(1) = h$ and that of a loss is $P(-1) = 1 - h$. The expectation and variance of the binary random variable X are $\mathrm{E}(X) = 2h - 1$ and $\mathrm{Var}(X) = 4h(1 - h)$, respectively.

[6] This is arguably the strongest assumption we will make.

indicated in the graph by the shaded region - that drives the R^2 away from one represents sample returns that the linear model does not and will not fit well in the future.

It is quite clear then which scenarios will lead to a profitable day and an increase in hit rate, and which will not: knowing the value of the predictor variable at time t, we read off the predicted future return as given by the linear regression. As long as the realized return on $t+1$ has the same sign as the predicted return (and the according position taken), we will make a profit on $t+1$. The lined triangular areas thus indicate scenarios where the linear regression model fails to capture the correct sign of the future return.

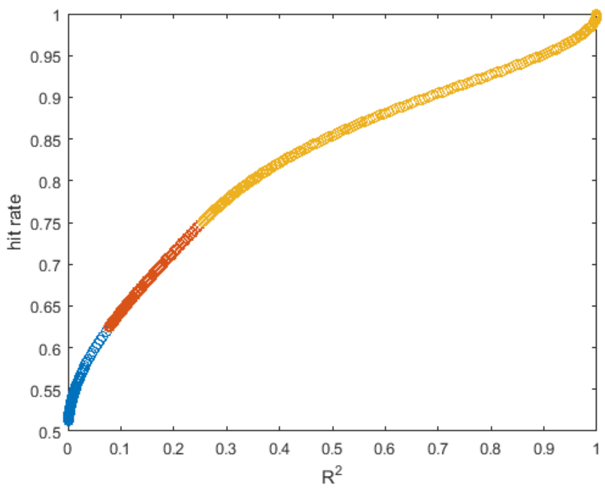

Figure 4. Relationship between R^2 and hit rate in a stylized setting, where linear regression is used to predict future returns and thus to determine the position taken.

By varying the width of the scatter and thus the value of R^2, we vary the size of the lined triangular areas relative to the entire shaded region. Using geometric arguments, specifically that a successful 'hit' falls into the correct quadrant of the coordinate system, we can calculate the associated hit rate for a given R^2. This gives us the relationship between R^2 and hit rate depicted in Figure 4. One parameter that we have not mentioned so far is the slope of the regression line (which represents the relationship between observed and predicted returns in our case). Running the same toy experiment with a modified slope turns out to only affect the range of R^2 observed, but traces the exact same relationship. Figure 4 in fact shows three curves collapsed onto one, with an increase in slope resulting in a higher R^2 (while keeping the width of the scatter the same).

We can see that the relationship between R^2 and hit rate is monotonic, but more importantly we find that R^2 values in the low single digits can result in a hit rate consistent with positive performance. Conversely, high R^2 as maybe are typical in other fields, would result in huge hit rates. When building predictive models in systematic trading we therefore need not expect the levels of goodness of fit that are found in other disciplines. It turns out that a very moderate edge, or small predictive power, suffices to build a functioning trading strategy.

5. Conclusion

This note illustrates that a small edge in predictive power can lead to a successful trading strategy in the long run. While this small edge is not trivial to come by, it is nevertheless reassuring that we do not need perfect foresight to build a profitable strategy.

Furthermore, we can often benefit from diversification, see Tricker and Mitchell (2017) for an illustration of diversification in the context of trading a range of different markets. Beyond diversification via the trading of different markets, we can achieve diversification by trading a variety of different strategies. Combining different markets for which individually low hit rates are obtained, and then combining different strategies, can yield a portfolio with an overall higher percentage of positive return days than the underlying individual hit rates.

References

E. Tricker and M. Mitchell. Market Diversification. Research Note, Graham Capital Management, August 2017.

Legal Disclaimer