Predictive Relationship: Difference Relationship

Contents

1 Relationship Description

In a difference relationship, recent change in the value of the signal directly affects future changes in the price:

$$\Delta p_{t+1} \propto k\left(\frac{r_t}{p_t} - 1\right) + c \tag{1}$$

where Δp_{t+1} represents the future price change, r_t represents the current value of the research signal, p_t represents the current price of market, k is a scalar, and c is a constant. The bracketed term calculates the percentage difference between p_t and r_t. Intuitively, this means that when the signal moves above the price, it is good to buy the market, as the price will tend to drift higher. When the signal moves below the price, it is good to sell the market as the price will tend to drift lower.

After finding the relationship which maximises risk adjusted returns, InferTrade runs tests for statistical significance to verify that the relationship gives a predictive edge. A predictive difference relationship can be used to when the current percentage gap between the price and the signal is high versus the historical average, and when current percentage gap between the price and the signal is high versus the historical average. If the scalar k is negative, inverting the signal will make this true.

Examples of time series that might show this kind of relationship are price forecasts or fair value models:

analyst or model thinks the security is cheap -> increases price target -> larger percentage difference -> price goes up as investors see value

We would typically not expect a signal that is a sentiment index to show a difference relationship, as the index is likely dimensionally inconsistent with the price and of a different scale. However if the share price is very stable versus the index changes then a partial (inefficient) relationship may occur as the level of difference will correlate closely with the absolute level of the signal.

Similarly a technical positioning indicator will be unlikely to show a Difference Relationship as the signal and price series are likely dimensionally inconsistent, such that the difference has limited predictive value except where stable price means the change in the gap correlates with the change in the signal level.

2 Trading Strategy Description

A predictive difference relationship can be reflected in many kinds of rules. InferTrade uses a 120 period (6 months for daily data) rolling regression of the percentage gap between the current signal and price against next day's price change as a benchmark. This trading rule recommends taking a position sized as a percentage of the capital you have allocated to trading this particular market.

The rolling regression difference relationship trading rule recommends taking a position sized as a percentage of the capital you have allocated to trading this particular market and will show higher returns than usual if a significant difference relationship is present between the price and signal series. The equation given below is used to estimate the future percentage price change Δp_{t+1} to occur from today to tomorrow's price change using difference between today's research level r_t and price level p_t:

$$z_t = k_1\left(\frac{r_t}{p_t} - 1\right) + k_2 + \epsilon_{t+1} \tag{2}$$

with the error symmetric such that

$$E_t^{model}[\Delta p_{t+1}] = k_1 \left(\frac{r_t}{p_t} - 1 \right) + k_2 \tag{3}$$

where $E_t^{model}[\Delta p_{t+1}]$ is the predicted next day percentage price change, r_t is the research value at time t, and k_1 and k_2 are the difference and static coefficients. ϵ_t represents the in sample regression error between the model predicted and observed price change for time t, such that $\epsilon_t = E_{t-1}^{model}[\Delta p_t] - \Delta p_t$.

To give a suggested portfolio allocation z_t for the market, e.g. "invest 5% of strategy allocation into the security", the regression forecast's expected next day % price move is divided by the square of the Standard Root Mean Squared Percentage Prediction Error $\varepsilon_{rms_t}^2$ in our above regression:

$$z_t = \begin{cases} \frac{\Delta p_{t+1}}{\varepsilon_{rms_t}^2} & \text{if } t > L, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where z_t is the suggested allocation as a percentage of the capital which you have allocated to trading this market at time t.

Worked example for liquid market: current share price is \$50, regression coefficients are 0.00005 $(5 * 10^-5)$ for level and 0.0 for the constant, with \$1,000 allocated to this single-stock strategy. The current research series - a analyst or model perception - is +5.0, with a RMS prediction error from the regression of 5%. In this situation the strategy recipe recommends an investment of $0.00005 * 5.0/(0.05)^2 = 10\%$ be allocated. So \$1,000 * 10% = \$100 would be allocated, so we would buy $\frac{\$100}{\$100} = 2$ shares. We see from this example the expected gain over the next day is 2.5 basis points (minus trading costs), versus 500 basis points of daily risk, so in practice this kind of strategy is most useful where the signal both has a predictive edge and is stable (auto-correlated) in recommendation over time, so that the 'edge' from multiple time steps can compound linearly without incurring additional bid-offer spreads whilst the 'risk' aggregates scales with the square-root of time (for random walk).

(Note that if the magnitude of the error term is smaller than the magnitude of the price change prediction, i.e. $|\varepsilon_{rms_t}| < |\Delta p_{t+1}|$, dividing Δp_{t+1} by $\varepsilon_{rms_t}^2$ would result in $z_t > 1$. Limits can then be applied to cap the allocation. E.g. between 0% and 100% if short selling is prohibited. In practice it is generally unlikely that in liquid markets a genuine statistical trading signal this strong will be found.)

3 Rule Parameters

Below is a table summarizing the parameters specific to this trading rule.

Parameter Name	Default Value	Description	Symbol
Regression Length	120	This is the number of previous days used to estimate the regression coefficients.	L

Note that rule parameters are fixed to their default values for this and any other relationships optimised on simplified mode. For estimating the regression coefficients, a rolling window is used, so that at time t, the regression data uses research data from time t-121 to t-1 to calculate 120 research changes as the regression's independent variable. Price data from t-120 to t is used to calculate 120 next day price changes as the regression's dependent variable.

4 Glossary

- **Bullish:** Positive outlook on the market. Expectation of positive returns.
- **Bearish:** Negative outlook on the market. Expectation of negative returns.
- **Allocation:** The allocation is the fractional amount of the portfolios value used to determine the size of the trading position.
- **Parameter:** Value used by the trading rule in the calculation for trading position
- **Trading Rule:** Strategy to determine when to buy, hold or sell a position.

Further Links

1. InferTrade: `https://www.infertrade.com`
2. Privacy Policy/Legal notice: `https://www.infertrade.com/privacy-policy`
3. InferStat Ltd: `https://www.inferstat.com`

Appendix

Calculating Regression Error

This PDF omits any explanation of how the regression coefficients k_1 and k_2 are calculated as we are using a standard regression for calculation of both the coefficients and standard error. For InferTrade we used the open source *sklearn* Python library. The same can be achieved in Microsoft Excel by using the *FORECAST.LINEAR* function.

As a worked illustration of the error calculation, assume that InferTrade predicted that a security would increase in price by 0.1% between yesterday and today, and the price actually decreased by 4.9%. In this example, $\Delta p_t = +0.1\%$, and $\Delta o_t = -4.9\%$, giving $\Delta P_t - \Delta p_t = 0.1\% - 4.9\% = -5\% = -0.05$. This value is then squared to give $(\Delta p_t - \Delta o_t)^2 = (-0.05)^2 = 0.0025$. To obtain $\varepsilon^2_{rms_t}$ for day t, this calculation is repeated for every day from day $i = t - L$ to day $i = t$, with values summed before dividing by the regression length L.