What's alpha?

What we are doing is statistical arbitrage from stock market inefficiency. The portfolio, which is balanced daily, contains thousands of stocks using long/short strategy.

There is a number for each stock, which is called alpha and is used to describe stocks future returns. We allocate more moneys for stocks with larger alpha values.

Logic of simulation:

Set a window=250 days which is days look back
Set size = 10M which is the money in the long and short side
Sliding window
Read 250 days worth of data prior to simulation
For day=startdate to enddate Do:
  Shift data;
  Read data for a day
  Normalize all data in the window for splits and dividends
  Alpha Calculation: alpha[ii]
  Neutralization (Market, Sector, Industry): alpha[ii] -= group_mean(alpha[ii])
  Filter alpha values: very large value...
  Scale: scale = 2 * size / sum(fabs(alpha [ii]))
  Calculate Position: today_position = alpha[ii] * scale / price[ii].
  Trading volume: trd_vol = today_position – yesterday_position.
  Calculate daily pnl(profit and loss)
Calculate simulation performance (IR, turnover, return, drawdown)

How to use websim?

Read the WebsimFaq documents

What is good alpha?

1. Models with good performances:

- Returns: annualized returns
- IR: mean(pnl)/std(pnl). This is the most important measure we are looking, where daily pnl is daily profit and lose series. Information ratio in websim is annualized IR, which is sqrt(252)*IR.
- Turnover: shares traded divided by shares hold
- Drawdown: Maximal pnl lose from peak to valley
- Fitness: sqrt(returns/turnover)*IR

2. Models with good idea

- Keep the alpha simple, the idea makes sense and can be explained.
- Be creative and low correlated with other existing models

3. Models consistently work for different regions and universe

4. Tips

- Avoid over fitting: don't include too many parameters and if-then rules, alpha expression should not be very long
- Works for liquid stocks: not only works for top3000 universe, but also for some liquid universe such as top2000, top100.
- Not only delay 0, but also delay 1. (delay 0 means today we can use today's data, and delay1 today we only use yesterday's data)
- Normalize the signal: alpha=volume is not good, alpha=volume/sharesout is better
- Avoid extremely big/small values
- Don't spend too much time for simple reversion and volume ratio, we already have a lot
- Keep the signals' purity, don't add up different signals in an alpha if we have not enough reasons
- For turnover, if it < 0.4, we don't need to care it much. if turnover > 0.4, we need higher IR, or use more heavy decays
- Parameters: Don't spend too much time to fit parameters, for example there is no differences between:std(returns, 5) / std(returns, 100) and std(returns, 5) / std(returns, 60). Performance should not be sensitive to the parameter selections.

Workflow:

1. Report best 3 models every day to WQ_Equities_Advisors@worldquant.com. The report should contain alpha expressions, idea explanations and performances. We will give comments and suggestions if any.
2. Develop alpha based on US TOP3000 Delay 1, then test the models performances for other regions/universes
3. Book recommendation:
   http://books.google.com/books?id=a1yB8LTQnOEC&pg=PP1&dq=active+portfolio+management
4. You can click logs button for status of your worthy alphas

More tries you will have more chances to get the great discovery.