QML Functions

QML Alpha LLC, March 1st, 2019

1. Factors Tab

FN residual of x with respect to factors

F_f: $x_a = \sum_{f=0}^{K} c_f F_{f,a} + y_a$, for all *d*, replace x with y.

Buckets create quintiles for all input factors and construct groups as their intersection.

b1	0	1	2	3
b2	4	5	6	7
b3	8	9	10	11
	a1	a2	a3	a4

Clusters *k*-mean clustering starts with previous day clusters (assure consistent naming).

Interaction create new alphas for each alpha**X**factor pair (component-wise product).

2. Groups Tab

Market DeMean extract mean of the alpha $y_a = x_a - Mean(x)$, for all *d.*

Market Normalize divide positive (negative) values by sum of all positive (negative) ones.

$$y_a = x_a / Sum(x_b; same\ sign\ as\ x_a)$$

This transformation makes the sum of all positive to be equal to 1, and the sum of all negatives to be equal to -1.

Group DeMean similar to **Market DeMean,** applied to each group.

Group Normalize similar to **Market Normalize,** applied to each group.

Scale amplify the alpha values such that the sum of all positive is less or equal than 1, the sum of all negative is greater or equal than -1, and at least one equality holds.

Project all values that correspond to the groups selected are preserved and everything else is changed to **nan.**

Interaction the selected alphas are decomposed on the selected groups.

3. Operations Tab

Sign $y = sign(x)$, for all *d*, all *a*

Rank $y_a = Rank(x_a)/Count$, for all *d*, produces numbers between 0 and 1.

MT $y_d = Sign\left(\sum_{k\geq 0,\ \ having\ same\ month\ as\ d}^{d} x_k\right)$ for all *a*,

Flip $y = -x$, for all *d*, all *a*

FlipIF if PnL for the history is positive then no change otherwise apply flip

Grandfathering if no value then fill with the last know value (recursively)

Log $y = log_{10}(x + shift)$, for all *d*, all *a*

abs $y = abs(x)$, for all *d*, all *a*

Tails % tails of the distribution (50% max = no change)

NaN2Zero replace all nan values by zero

Zero2NaN replace all zero values by nan

Quantiles split the alpha into its quintiles

Split split the alpha below and above the threshold

Delta $y_d = x_d - x_{d-h}$, for all *a*

MA $y_d = \sum_{k=d-h}^{d} x_k$, for all *a*

Stoch $y_d = \frac{\max(x_k, k=d-h,\ldots,d) - x_d,}{\max(x_k, k=d-h,\ldots,d) - \min(x_k, k=d-h,\ldots,d)}$, for all *a*

Momentum $y_d = \frac{x_d}{x_{d-h}} - 1$, for all *a*

Stdev $y_d = stdev(x_k,\ k = d-h, \ldots, d)$, for all *a*

CCI $y_d = \frac{x_d - average(x_k, k=d-h,\ldots,d)}{stdev(x_k, k=d-h,\ldots,d)}$, for all *a*

ADI $y_d = \frac{2x_d - \max(x_k, k=d-h,\ldots,d) - \min(x_k, k=d-h,\ldots,d)}{\max(x_k, k=d-h,\ldots,d) - \min(x_k, k=d-h,\ldots,d)}$, for all *a*

RSI $y_d = \frac{\mathrm{count}(x_k > 0, k=d-h,\ldots,d)}{h}$, for all *a*

SR $y_d = \frac{\text{average}(x_k * ret_{k+1+delay}, k=d-h-1-delay,\ldots,d-1-delay)}{\text{stdev}(x_k * ret_{k+1+delay}, k=d-h-1-delay,\ldots,d-1-delay)}$, for all *a*

CORR $y_d = correlation(x_k, ret_k.\ k = d-h, \ldots, d)$, for all *a*

MACD $y_d = MA(x_k, k = d - h_{short}, d) - MA(x_k, k = d - h_{long}, d)$, for all *a*

Convergence $y_d = max(x_k, k = d - h_{long}, d) - max(x_k, k = d - h_{short}, d) -$
$min(x_k, k = d - h_{short}, d) + min(x_k, k = d - h_{long}, d)$, for all *a*

NormDist $y_a = InvNormalDistribution(Rank(x_a))$, for all *d*

BETA mask $y_a = \frac{covariance(x_k, mask_k, k=d-h, d)}{stdev(mask_a, k=d-h, d)}$, for all *d*

Delay $y_d = x_{d-h}$, for all *a*

DecayLin $y_d = \frac{2}{h*(h+1)} \sum_{k=d-h}^{d} (k-h+1) x_k$, for all *a*

DecayHL $y_d = \lambda x_d + (1-\lambda) y_{d-1}$, for all *a*

DecayTarget apply DecayHL with $\lambda = 0.5$ until the target is achieved (10 times max)

Decay % $y_{d,a} = \lambda_{d,a} x_{d,a} + (1 - \lambda d,_a) y_{d-1,a}$, $\lambda_{d,a}$ is computed to generate asset-wise bounded turnover

Max Concentration $y = \text{assert}(x\ belong\ [-thr, thr],)$ for all *d,* all *a*

Distribute similar to Max Concentration, but will redistribute the excess weights

Max Traded Volume limit volume traded in dollar relative to budget ($Mil)

Min Traded Volume do not trade small amounts

Max Pos Volume limit positions as fraction of traded volume

Implied Alpha $Y = QX$, multiply daily position by covariance matrix (as defined by the risk model)

Implied Position $Y = Q^{-1}X$, multiply daily position by inverse covariance matrix

4. Aggregate Tab

Average weighted average

MVO weights are computed to optimize reward to risk ratio:

$$\begin{cases} \max W^T R \\ subject\ to\ W^T Q W = 1 \end{cases}$$

Stable Set simple average of alphas whose pairwise correlation is below threshold

DD weights are generated from historical drawdowns: $w_{d,p} = \frac{1}{1+100*DD(p,hist)}$

SR weights are generated from historical Sharpe Ratios

Omega weights are generated from historical Omega Ratios
(positive mass/total mass of PnL)

LogPerformance weights are defined as: $w_{d,p} = log_e(\ 1 + Rank\big(SR(p)\big)\big)$

Combine change from one alpha to another: $y_d = \begin{cases} x_d^1,\ if\ d < D \\ x_d^2\ if\ d \geq D \end{cases}$, for all a

SPerf initially, weights are equal, after a while the weight is scaled up if alpha realized 2% of its weight and it is scaled down if it enters a drawdown of 2% or more

OrthogonalPerf sequentially add best performing alphas with equal weights.

After each addition, the remaining alphas are neutralized against the selected ones and their performance is reevaluated

5. Machine Learning Tab

Factor Learning construct many alphas by interacting the original alpha with factors and apply MVO aggregation method

Group Learning construct many alphas by interacting the original alpha with groups and apply one of the available aggregation or Machine Learning methods: SR, MVO, Fisher, CART.

Regression linear regression with the independent variables being the selected alphas and the dependent variable defined in the regression page

Patterns rule-based induction learning. Positive/negative observations are defined based on the value of the threshold around the mean. Positive Rules are detected such that their Positive Homogeneity (i.e., frequency of the positive observations) is above the specified value, and their Positive Prevalence (i.e., the fraction of observations that trigger the rule) is larger that prescribed. Similar requirements are imposed to Negative Rules. The value of the alpha is generated as the weighted vote between Positive and Negative Rules.

Data is digitized using a grid in each attribute. The complexity of the rules is bounded to maximum “Degree” number of attributes.

Fisher Fisher Linear Discriminant Analysis

LDA Linear Discriminant Analysis based on the centers of positive and negative sets

NN Discretized version of Nearest Neighborhood for multiple subsets of alphas

CART Classification And Regression Trees based on branching with linear discriminants

PLS Partial Least Square regression

Logit Logistic regression

6. Statistical Factors Tab

PCA Factors	Generate Principal Components of the selected alpha.
PCA Remove	The alpha is changed so that exposure to its Principal Component are removed
Orthogonal	Alphas are orthogonalized
Corr	Construct the factor-factor covariance matrix
Match Corr	Align factors by matching correlation
Hungarian	Align factors by Hungarian Algorithm