**Technical Details**

Return comes from: a. mispricing(tend to decay). b. affording the risk(can persist).

Strategy type: trending(behavior), value(mispricing), carry, vol trade(risk taken).

What is the risk the strategy take?

**Strategy**

* **Effective volume**

Mechanism:

1. Market driven by big money, big players use order fragmentation, this reconstruct big money.
2. 25% volume explain 75% price change. Tell you whether significant enough to push market. Market price change are driven by big players. Research on how big player play order and then know how to tract them.
3. Volatility adjust between price and volume.
4. Indicator should as close as transaction level. And separate large from small volume(top 50%). And tract the change of this volume
5. Tradition A/D problems: position of close, manipulation of open price(sells intend to push high open to attract buyers),
6. Assumption: volume equal distributed in a bar.

Market: Market driven by big player, small players are noise. To win is to track them.

Solution: How to define it, by effective volume which measures effective purchasing, and track large purchasing power.

The distinction to MFI is that.

Const: need special market.

* **Heikin Ashi candle bar**

Trend is noise, candle is noise.

Smooth candle trend, chart is more readable and easier to analysis trend.

Const: Delay problem,

* **Alternative bar**

Time bar not equal information, series correlation, normality.

Contribute equal information, equal weight smooth method.

Const: not usual market convention, a lot of action match time stamp.

* **HP Filter**

Traditional MA lack of sequence information.

L2, consider adjacent 2 sample relation. And aim to reduce the change between sample.

Const:

* **EMD Filter**

Good for nonlinear and nonstationary signal.

Breakdown signal into serval intrinsic mode(equal peak and zero crossing pint), and formulate a trend and mean reverting part, by comparing the strength of these two part, we could know whether the trend is strong or weak and choose to do trending or mean reverting.

* **Hilbert Filter**

Breakdown series into trend part(traditional moving average) and periodical part(sin wave), to tell what is the current market status.

Const: require stationary signal.

* **Kalman Filter**

Recursive filter, it consider the path of the price series compare to ma,

Prediction stage use current status and transition matrix to predict.

Updating stage use the latest observation to update the transition matrix.

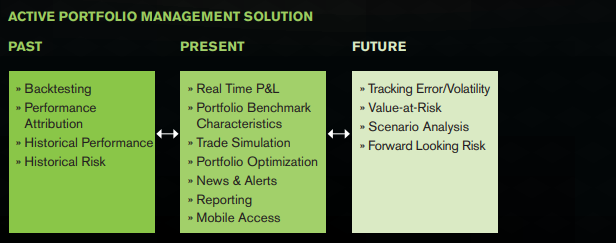
* **High low band**
* **Pivot-point**
* **Correlation/differentiation**

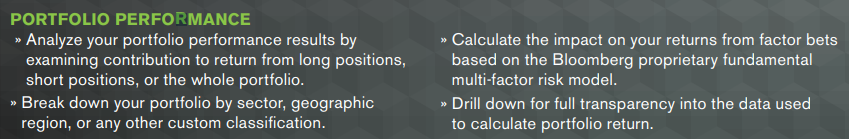
On price return

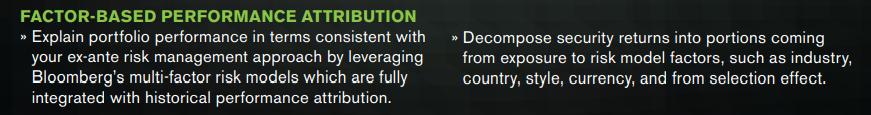
On accumulated effective volume

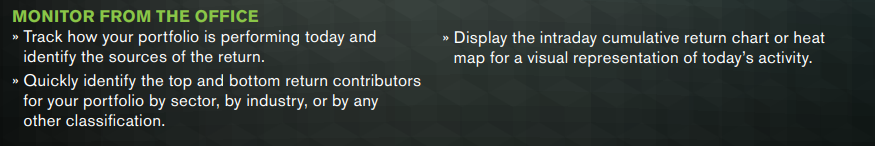
On volume trigger

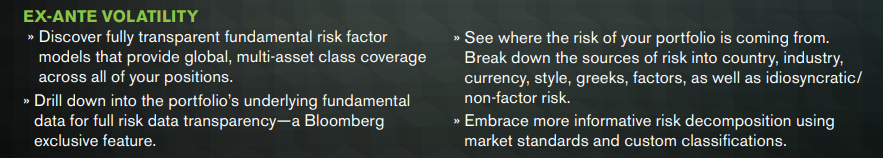
**Portfolio**

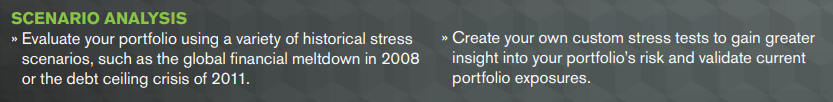


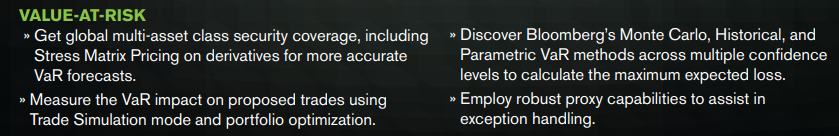


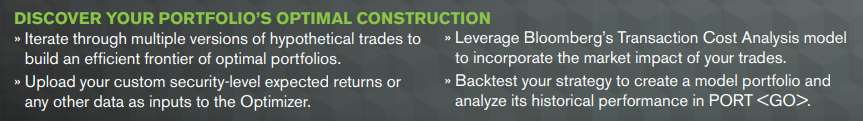




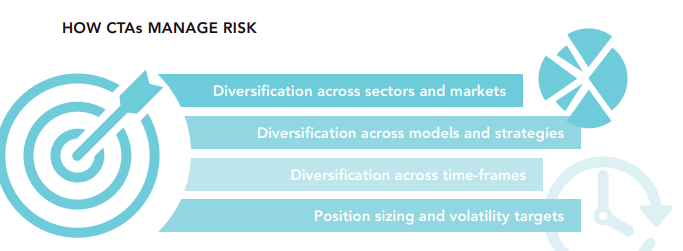


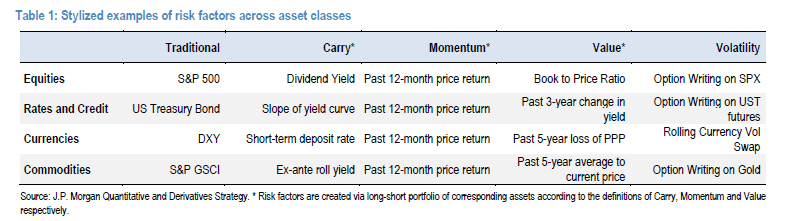












* **Risk analysis**

Risk style analysis:

value, momentum, carry, volatility.

Should use factors with positive risk premia, independent(no correlation).

Risk factor attribution:

Stock: country, sector, momentum, market cap, pb, yield, impv.

Forex: dxy, yield curve, momentum, carry rate, impv,

Commodity: benchmark index, roll-yield, momentum, current price level, gold impv,

Rates and fixed income: US bond, yield curve, momentum.

Risk factor exposure selection, limits, allocation:

Optimization.

stress test:

performance in historical event, incorporate correlation between factors.

Consider extra factors like liquidity, spread, volatility,

scenario analysis:

change market variable to gauge potential impact(rate, currency, credit spread, inflation shocks), no historical data or calculation and not bounded by historical correlation, distribution should towards to tail.

value at risk:

calculation:

historical:

monte carlo: assume distribution of return.

parametric method: regression on risk factor, calculate weight, mix volatility.

Covariance matrix:

Direct: nonstable(small change may lead to large effect), singular if dimention exceed sample number.

Equal Correlation:

Market/Factor model: get the coefficient and then know covariance. Only one factor to capture stock movement is not adequate.

VaR contribution, attribution, how to reduce var.

* **Performance measurement**

Understand how return being achieved.

Breakdown active return into risk factors:

Stock(sector allocation, stock selection, market timing, currency movement)

Fixed-income(interest rate, spread effect, duration)

Benchmark: sp500, Barclays indexes, SG CTA.

Own breakdown:

Return break: factor contribution

Compare to benchmark CTA index.

Breakdown strategy performance into market selection, allocation,

intraday execution performance(to vwap),

strategy long to buyhold,

strategy timing attribution and allocation attribution.

Investment style analysis: factor analysis,

Performance measure:

regression on benchmark or risk factors, check the value and the interception and whether it is significant.

Regression on style, check the predicted and real return difference.

Goal of portfolio manager:

Above average return within a given risk class.

Portfolio diversification to eliminate all unsystematic risk.

Statistics:

Jensen’s Alpha: the excess return of an asset over the required rate of return base on CAPM.

Information ratio: excess alpha divided by per unit of risk.

Portfolio attribution:

Compare benchmark, get return by differ allocation and asset selection, timing skill.

* **Portfolio Optimization**

Market condition: whether to achieved our desired return.

Diversification: time scale, instrument, model.

Efficient frontier: control(risk exposure, turnover, transaction cost). Get return/vol.

Tangent portfolio: generate the most return for per risk taken.

Black Litterman: with view, introduce return term with view, confidence and expected return.

* **Portfolio sizing**

Reward/drawdown rule.

* **Risk management**

Risk indicator control: VIX, riskline, market breath, global correlation,

Concentration limit: position concentration, strategy concentration.

Monitor: VaR determine magin.

* **CTA Risk Analysis**

Factors: liquidity, volatility, correlation, capacity.

Cta style factors: market size, equity bias, trading speed.

Constraints: liquidity, trade cost, inner market correlation, position&risk limits, capacity.

Own: liquidity, volatility, correlation, paramer risk,

* **Risk parity**

Return distribution is negative tail, while CTA normally exhibit positive tail.

PCA tail risk parity: equal risk contribution to portfolio, using expected shortfall.

Under constraint: market, strategy time frame,

* **Sizing, Kelly/drawdown rule**

Concern: normally increase position when CTA drawdown and would put many money when just launch a new strategy, investor need to gain confidence along strategy trading and thus we increase position gradually as we make more and more money.

Rule: portfolio allocation give us the final target size, we use gain&drawdown rule, if we achieve certain percentage return(3 month to full, 1.5% per month, 5 steps to 12), first drawdown should not exceed 50% of all gains.

Kelly: determine the leverage to maximize your growth rate while would not loss all.

* Backtesting
* Alpha decay and fail

Exceed maximum drawdown time or drawdown.

Exceed downside of strategy which could get from cross-sectional test results.

* CAPM/APT

CAPM: the main driven for investment is expected return and volatility, return is normally distributed, excess return is from reward of carry risk.

APT:

**Macro Econometrics**

* **Glossary**

Average duration:

bond price sensitivity to interest rate. Express as Number of Years.

Effective duration:

Average weighted maturity:

Number of years to maturity.

Information ratio:

Measure performance against a benchmark.

Sharpe ratio:

Measure performance against the risk-free rate.

Yield to maturity:

Consider current market price, par value, coupon rate, and expire time. Reinvested.

Yield to worst:

Compare market yield to coupon, decide prepayment.

* **Fixed income risk**

Interest rate risk: duration

Yield curve risk: interest rate structure, not change for all maturity. Key duration

Volatility risk: interest rate volatility.

Credit risk: default risk, downgrade risk, credit spread risk(premium on similar risk free bond),

Currency risk: not in domestic currency.

Liquidity risk: bid price.

Hedging:

Duration hedging: hedge ratio is opposite to duration.

Bond price volatility inverse to maturity.

Convexity hedge: 2 more assets.

PCA hedging:

* **Fixed-income valuation**

Sum of the discounted value of the future cash flow.

Duration: modified duration with unchanged cash flow. Effective duration opposite.

Convexity: for big interest rate change.

Term structure: factor explain the change

Parallel movement:

Slope oscillation: stepness, short-term and long-term spread.

Curvature: concave or convex

Theory:

Expectation theory: expectation about short-term rate. Don’t consider reinvestment

Risk premium theory: risk premium to hold long term bond.

Market segmentation: different investor, by supply and demand.

* ESG

Environmental, Social, Governance.

Affect sovereign bond spread and measure of a country’s long-term economic sustainability. Could identify potential upside and early transformation.

* ESG

**Statistical Analysis**

****

* **Dataset check, pre-processing**

Function:

Ensure data quality before model building.

Know the property of you dataset, then know potential problem when modelling.

Typical problems: Missing values, duplicate values, abnormal extreme values,

Procedure: data cleaning, data aggregation, data formatting, data transformation.

Data Clean:

Missing:

Delete missing value accounts for 80%,

Replace with previous/statistics/interpolation/classification,

Outlier:

Delete, 3-sigma rule/clustering,

Replace, mean/log value,

Noise:

Bin formatting, use average of the bin to replace all.

Data Aggregation: from multi source

Clear identifier,

Unique, not easy to derive from other data,

Contradiction, remove duplication,

Data Formatting: Get the best subset to do modelling

Feature elimination, by SVM/regression, single feature importance, person correlation/kalfang test/lasso.

Dimension reducing, PCA/FA/SVD/clutering/kmeans

Data Transformation:

Normalization/Standardization, zsocre/log/minmax,

Discretization, equal frequency/bin,

Rarefaction, dummy variable,

Data Normalization/Standardization:

When require distance calculation or gradient decent computing, like regression, adaboost, svm, knn, kmeans. Not for tree model.

* **Statistical analysis method**

Regression: linear, logistic, lasso, ridge, multi-variable, polynomial, stepwise, elasticnet,

Clustering: k-means, clustering,

Discriminate analysis: LDA, SVM,

Decision tree: random forest, bagging, boosting, adaboost,

Time series analysis: arma, arch, garch,

Distribution analysis:

Hypo-testing:

Reliability analysis:

Frequency analysis:

Correlation analysis:

Variance analysis:

Bayes:

Principle analysis:

Factor analysis:

RNN:

Compare regression:

Regression: explainationary, while cannot get non-linear relation.

Random forest: could do non-linear relation, not sensitive to outlier, while may be overfitting.

Compare classification:

Logistic: probability explaination, while weak in decision boundary.

SVM: could use non-linear kernel, while hard to do parameter adjust and choose kernel.

Compare clustering:

K-means: fast, flexible, effective while not clear class number and require ball shape distribution.

GMM:

* **Distribution analysis**

Mean, median, variance,

Normality test:

Autocorrelation:

Correlation:

* **Regression, linear/logistic/lasso/ridge**

Linear regression:

Assumption: residual should be independent normal distribution with mean 0 and equal variance. Independent variables should not correlate with each other and with the residual, not random variable.

Correlation: both are random variable,

T-stats: test parameter=0,

F-stats: model significance, parameters are all not equal to zeros. Construct F.

Heteroscedasticity:

Result: T stats is t-distribution, cannot use it as a measure of significance and cannot get confident interval.

Detect: BP-test, White-test, regression between variance and variables(1st order, 2nd order, cross). F-stats.

Solve: weighted LS when we know variance relation with variable. If not, find relation first, then use WLS.

Autocorrelation:

Result: 1st or 2nd autocorrelation for residual.

Detect: DW test,

Solve: take differential of the variable.

Multicollinearity:

Result: parameter estimation,

Detect: correlation matrix of variable, VIF(variance inflation method) x gression on others.

Solve: stepwise, lasso, ridge regression. PCA on feature.

Logistic:

Require large sample, cause of maximum likehood desire this.

Should not multicollinearity.

Stepwise:

Use stats like information, bias, to determine.

Lasso:

Reduce overfiting by introduce parameter penality part. L1, use absolute penality, tend to set parameter to be 0.

Ridge:

When encounter multicollinearity, L2, tend to set parameter to very small of no need.

ElasticNet: Lasso+Ridge.

* Time series, Garch/Arch, autocorrelation/cointegration
* **PCA/LDA**

Unsupersied/supervised, maximize variance along vector/minimize inner sample and maximize outer sample.

Get linear uncorrelated factors,

PCA const:

PCA remove linear relationship, while some may be non-linear and could be removed.

PCA assume sample distribution is normal.

PCA assume sample is one class and find the most explain one while may not be efficient, LDA could do better.

* SVM
* K-means
* Random-forest

Sub-decision tree: random sampling, random feature selecting, voting for final.

* Beyes
* Boosting

Sample weight update, bagging classifier.

* HMM

**Database:**

* Tick data

hehe

* KDB
* Postgres SQL

**Market:**

* Contract roll
* Forex spot/forward/ndf/forward curve

Ndf: non-deliverable forward, no exchange of notional, just settle up the forex rate difference. Mainly exist for currency with strict government regulation.

Forward rate: multiply spot rate with interest rate differential, reverse order of spot.

Forward point: forward rate – spot rate

Forward curve: linear interpolation.

Swap rate curve: steepen,

* Yield curve

Interest rate vs. maturity.

Normally it is upward sloping. Could also be downward and flat.

Economy expansion indicates upwards.