Quantitative Investment Handbook

Xinhe Liu

2018-2-28

Contents

Chapter 1

Basic Financial Concepts

This chapter summarizes some basic financial concepts you should know about.

1.1 Asset Management

Asset Owners : From conservative to active: Banks, Insurance Company, Defined Contribution Plan, Pension, Endowment/Foundations

Investment Purpose Statement(IPS): Consider Return, Risk, Time Horizon, Tax, Liquidity, Legal and Unique.

Chapter 2

Economics and Econometrics

2.1 Corporate Finance and Fundamental Analysis

Chapter 3

Asset Valuation Theory

1. NPV and IRR
2. Discounted Cash Flow Model
    - Free Cashflow
    - Required Rate of Return = Cost of Capital = Risk-adjusted discount rate : Usually from the CAPM
3. Valuation using Multiples - P/E, P/B

Chapter 4

Portfolio Optimization Study

1. Efficient Market Hypothesis (Weak, Semi-strong, Strong forms)
2. Markowitz Portfolio Optimization
    - Minimum Variance Portfolio / Tangency Portfolio
    - Jensen's alpha
3. Capital Asset Pricing Model(CAPM) Model

$$r = \beta(r_m - r_f) + r_f$$

$$\beta = \frac{cov(r, r_M)}{var(r_M)}$$

4. APT(Arbitrage-Free-Pricing) Model
5. No-arbitrage(weak, strong) and Law-of-one-price
6. Metrics
    - Sharpe Ratio
    - Jensen's alpha
    - Required Rate of Return = Cost of Capital = Risk-adjusted discount rate : Usually from the CAPM
7. Valuation using Multiples - P/E, P/B

Mainframe Portfolio Optimization Models Include

1. portfolio optimization - mean-variance short-comings and 1. Monte-Carlo simulation 2. reverse optimization and black-litterman(less sensitive to inputs) 3. resampled mean-variance optimization 4. add additional constriants (other than budget) - concentrated position 5. resampled mean-variance 6. non-normal optimization approach 7. allocating to less liquid asset class - (no investable benchmark) 8. risk budgeting vs factor based asset allocation(risk parity)

1. liability-relative 1. surplus optimization 2. hedging/risk-seeking portfolio 3. integrated asset-liability (non-linear correlation) 2. goal-based approach 1. describing goals, constructing sub-portfolios (selecting a module) One-fund theorem, two fund theorem

convex reformulation of maximizing Sharpe ratio problem(normalizing)

Multi-factor Risk Model

$$r = \mathbf{B}f + u \; V = \mathbf{B}cov(f)\mathbf{B}^T + cov(u)$$

Mixed Integer Optimization

Stochastic and Dynamic Optmization

Issue

Active Holding Optimization optimize alpha holdings and tracking error

Improvements on MVO estimation error Shrinkage estimators Jason-Stein Estimator Black-Litterman Model Robust Estimators

optimization issue add constraints resampled efficiency use other diversification approach risk-parity maximize diversification

Mixed Integer Optimization

Heuristic forward backward selection clustering

Stochastic Optimization scenario optmization approach Var cVar

Kellys Crirterion

Multi-period optmization room

Excel solver CVX matlab CVXOPT python optmization toolbox matlab Gorubi

4.1 Estimating Return and Co-variance Matrix

4.2 Transaction Cost

4.3 Tax

4.4 Dynamic Portfolio Choice

4.5 Covariance Matrix

Table 5.4
Covariance estimators

Equally weighted	Exponentially weighted
$\sigma_{12}^2 = \frac{1}{T}\sum_{t=1}^{T}(r_{1t}-\bar{r}_1)(r_{1t}-\bar{r}_2)$	$\sigma_{12}^2 = (1\text{-}\lambda)\sum_{j=1}^{T}\lambda^{j-1}(r_{1t}-\bar{r}_1)(r_{1t}-\bar{r}_2)$

4.6 Easy

1. Two coins, one is double tail, you see one
   - three pancakes: golden-golden, golden-brunt, brunt-brunt, see one golden, probability that the other side is golden two
   -
2. Two coins - one toss n+1, one toss n, probability that n+1 gets more tail than n man?

4.7 Medium

1. Use coins to create probabilities : fair coin to create 1/3 probability - toss twice, one result is retoss unfair coin to create 1/3 probability - combine two toss as one

4.8 Hard

Chapter 5

Statistical Arbitrage

5.1 mean-reversion

intraday mean-reversion

Volatility Clustering and Leverage Effect

Chapter 6

Factor Investing

6.1 Smart beta and Smart Alpha

long-short STS strategy short extension(120/20) Strategy alpha and beta separation portable alpha

long short equities (factor) convertible arbitrage

6.2 Key things to Notice in Index Making

6.3 factor analysis and risk premia

value, momementum, quality, volatility, betaRM-RF The return spread between the capitalization weighted stock market and cash.

QUality Minus Junk (quanlity)

SMB The return spread of small minus large stocks (i.e., the size effect).

HML The return spread of cheap minus expensive stocks (i.e., the value effect).

RMW The return spread of the most profitable firms minus the least

profitable.

CMA The return spread of firms that invest conservatively minus aggressively.

UMD (momentum/trend) UMD is long winners and short losers and also from Ken Frenchs website)

Carry Vol Carry

Special: liquidity premia

Market Ineffciency Analysis: funding constraint of financial institutions, grand move of large funds constraints

6.4 Market Anomalies

6.4.1 Behavioral Finance

Behavioral Finance Theories Includes Prospect Theory(People suffice rather than optimize, the utility curve is concave at the gain part and convex at the loss part(loss aversion). Bounded Rationality and Behavioral Market Anomalies/bias

1. loss aversion(herding)
2. illusion of control(TAA)
3. Mental accounting(goal),
4. availability bias(familiarity, home-bias)
5. recency bias(tactical shifts)
6. framing(risk-return presented in a different way)

other: liquidity risk premia

1. Risk Management and Hedging

2. Leverage
3. Correlation
4. Strategy replacements, leverage rebalancing and rebalancing frequency, leverage reset

other: liquidity risk premia

1. Risk Management and Hedging
2. Leverage
3. Correlation
4. Strategy replacements, leverage rebalancing and rebalancing frequency, leverage reset

Chapter 7

Single (Asset Class) Trading Strategies

7.1 Equity

7.2 Volatility and Dispersion Trades

7.2.1 Volatility Models and Time-Series Models

7.3 fixed income + macro

7.4 fixed income derivatives

MBS Convexity Trade

1. real estate-direct vs indirect environment 2. benchmarks of real estate
3. other alternative, commodity to hedge

7.5 credit

7.6 commodities

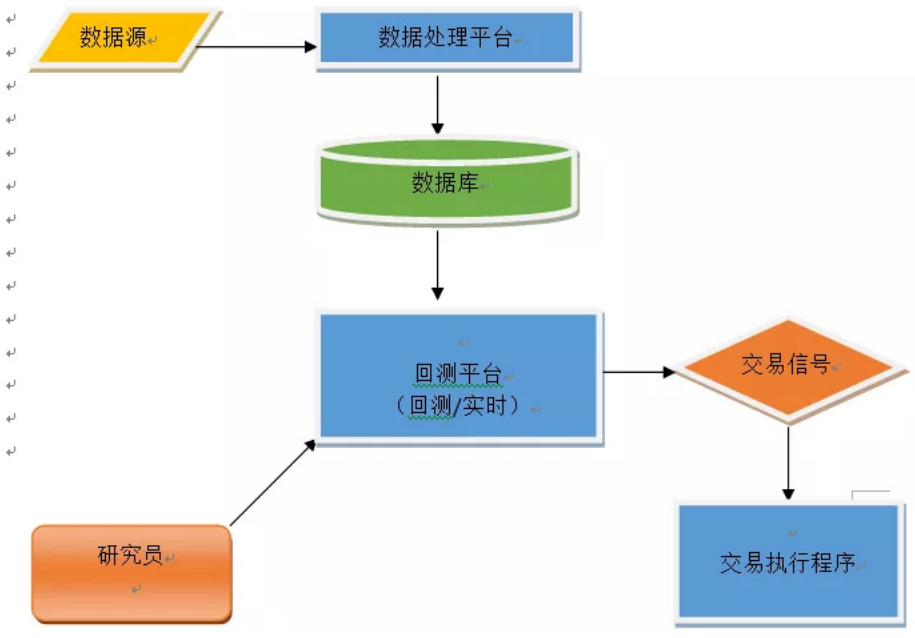
Chapter 8

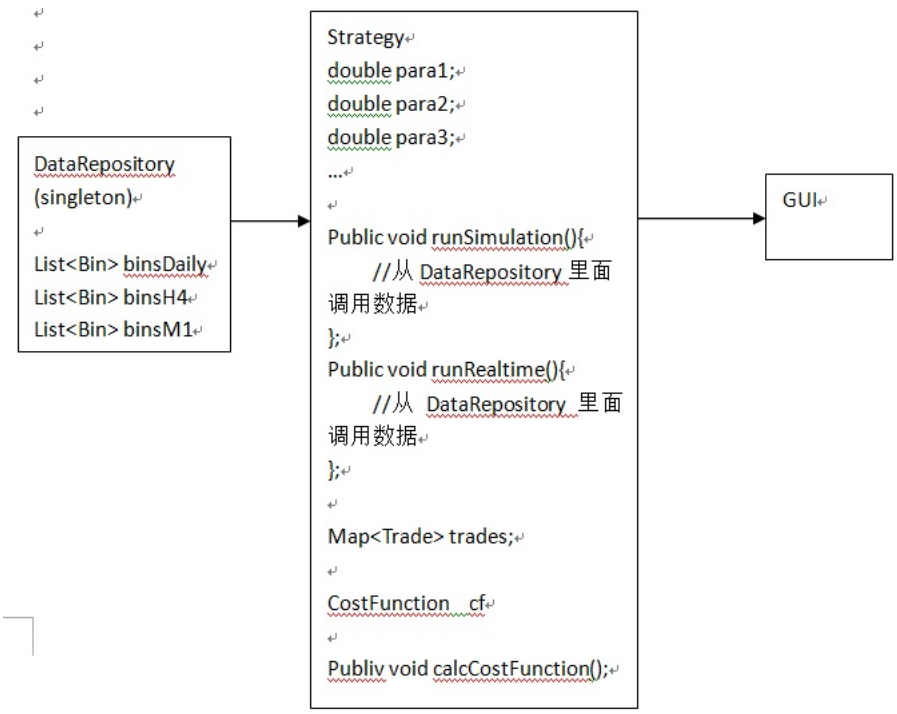
Strategies Implementation/Platform Building

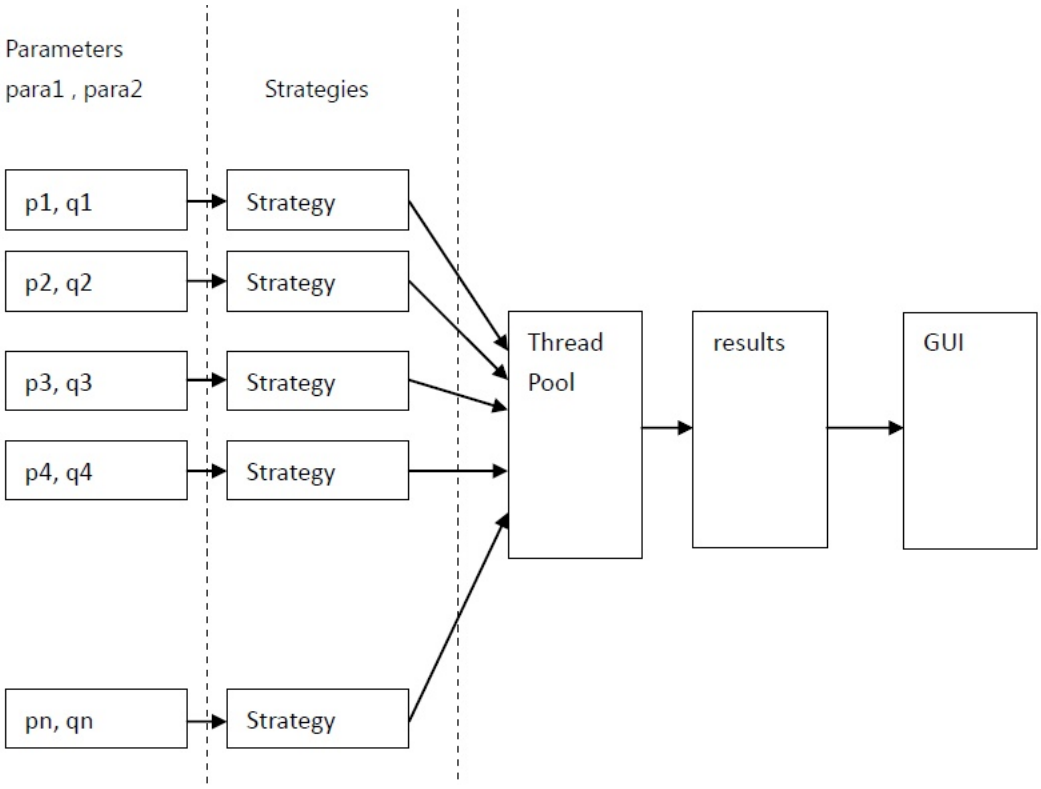
8.1 Trading System

8.2 Strategy Development Pipeline(Single/Linear Strategy)

8.2.1 Basic Architecture/System




DataRepository
(singleton)
List<Bin> binsDaily
List<Bin> binsH4
List<Bin> binsM1
Strategy
double para1;
double para2;
double para3;
...
Public void runSimulation(){
//从 DataRepository 里面调用数据
};
Public void runRealtime(){
//从 DataRepository 里面调用数据
};
Map<Trade> trades;
CostFunction cf
Publiv void calcCostFunction();
GUI




```java
public class Bin {

    public Calendar openTime;
    public Calendar closeTime;

    public double openBid;
    public double highBid;
    public double lowBid;
    public double closeBid;

    public double openAsk;
    public double highAsk;
    public double lowAsk;
    public double closeAsk;

    public double volume;

    public double pnl;

    public int length;
}
```

```java
import java.util.Calendar;
import java.util.UUID;

public class Trade {

    private UUID uuid;

    private String instrument;

    private double entryPrice;
    private double exitPrice;

    private double stoplossPrice;
    private double profitTargetPrice;

    private double commissionrate;
    private double commission;

    private double profit;

    private int quantity;

    private Calendar entryTime;
    private Calendar exitTime;

    private DirectionType direction;


}
```

```
for binsDaily {
    for (Bin bin:binsM1) {
        if(达到入场价){
            //入场一单
            Trade trade = new Trade()
            Trades.put(trade);
        }
        For (trades){
            If(打止赢/打止损){
                Trade.close()
            }
        }
    }
}
```

Historical Data is a singleton. All Market Data (eg. a candle stick) should be organized to feed the researcher to program strategies on the backtester (eg. like quantopian). All back-testing should be parallized (ideally on GPU) to display parameter-profit relationship. (heatmap, stock charts, etc) Ideally the optimization process could be visualized (like Tensorflow)

All like a research facility feedback cycle.

Key Details: API Design, Module Separation etc.

8.2.2 Data

Key is a real-time listener. Technical Considerations: KDB, Hadoop and HDFI(?), SQL Like, Mongo Db to store Archive Data. Market Data Providers consideration buying from Wind, BBG, Reuters, Etc.
teams: platform operation engineer, analytics builder, strategy control/management and risk management, data team, execution team, researcher team (3 x tech)
data licensing and data quality insurance
data base, text file archive, big data issue
cheap data: brokerage: interative brokers.

8.2.3 Backtester/Simulator

Key Components

*Send Singal to Quoting/Trading/Exection Tool(Real Time) *Market

Data Objects (eg. loop for every time bins) *stop loss/risk control system integration *parameter-backtest profit/statistics result: optimization and loss function set function to tune the parameters *multi-thread: Java backter (Java thread pool*) *human selection of parameters: parameter table and visualization

8.2.4 Trade Record and Money Management

record every trade, summarize execution shortfall, statistical trends and information (shortcomings of strategy executions) and market information (learning material) build statistics and storage

More: order book and trade book level data handling

8.2.5 Analytics

Strategy Management

, Sharpe Anslysis, Holding Period, Slippage visualization to better assistant strategic allocation

Execution Analysis and Cost

quantitative trading/systematic trading strategies: * equity long/short

8.2.6 Research Team

Key problems: * Optimization and Combination of Sub-Strategies (Eg. factors) * Market Regime Change Detection(problem not solved): Distinguish between trend and oscillation market * market supply/demand imbalance analysis (risk-premia) * volatility trading, dispersion trading - 2nd and 3rd degree trading, (vol model, vol clustering effect, vol leverage effect) * hedging/overlay strategy research: hedging cost and hedging risk management, how to adjust hedge according to market condition. * common ideas: market imbalance, mean-reversion, autocorrelation

patterns etc (find patterns and trade) - based on statistics. Risk factors, implied arbitrage - based on math.

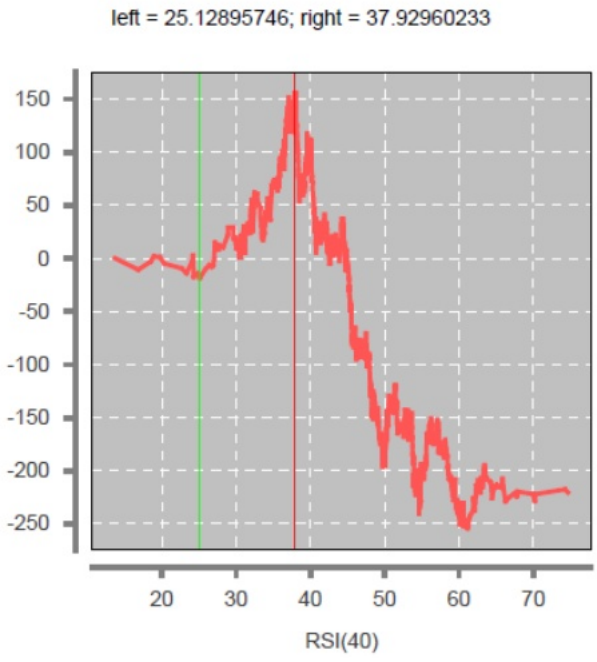
Parameter Optmization and Control

Rely on GUI - parameter distribution and selection optmization methdologies from machine learning (see optimization chapter) robustness anslysis and out-of sample test ** (random cut the universe of rolling window on selection period)

Signal Indicator Design

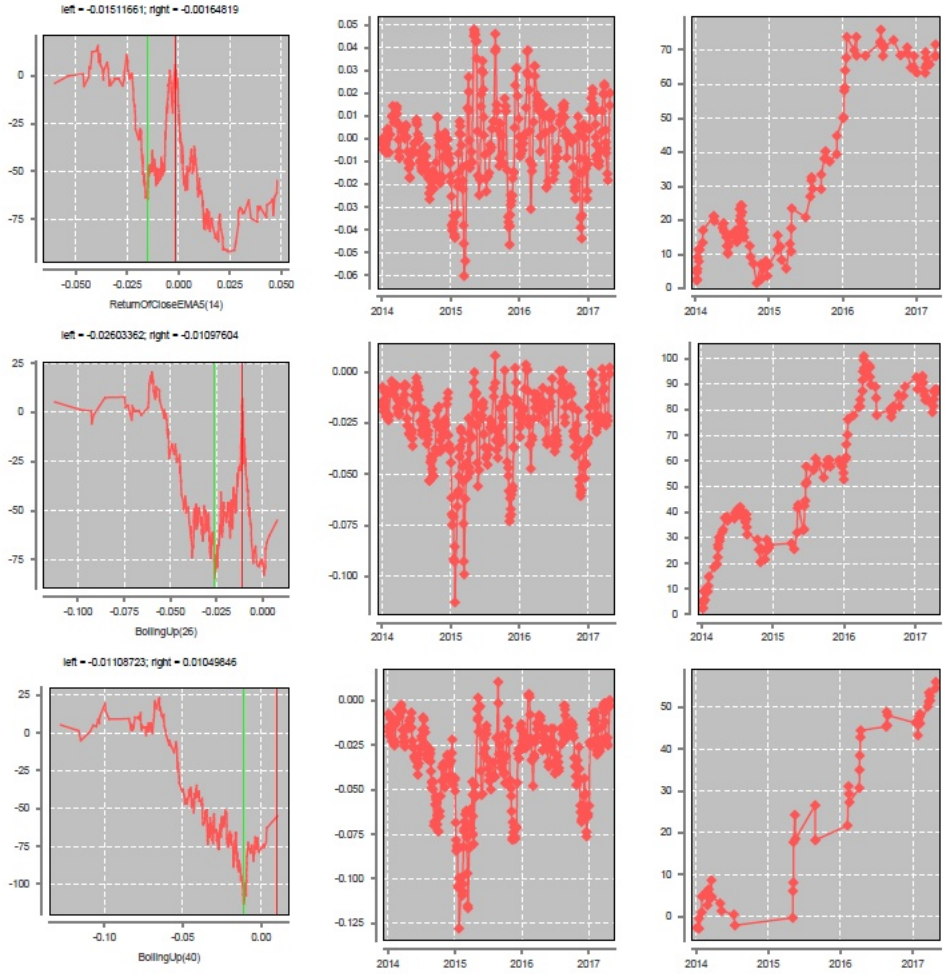
For example, based on fundemental ratio and technical indicators - design a formula. And check the level of prediction power (if any)

1.seeking stationarity: find a stationary time series use difference, integration, and normalize with volatility 2.find signal level, plot cumulative back-test return against different signal level (use own quotes, and use signal level to filter quotes)



3.Check stability of

customized indicators



4. check overfitting and type-II error in all settings, apply noise filtering if possible 5. design a interface to input indicator(math formula parser to read string) and visualize information using GUI.(HTML/XML Render)

```
<?xml version="1.0"?>
<indicators>

<section name="priceDivEMA">
    <indicator>
        <f>OpenDivEMA(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
    <indicator>
        <f>CloseDivEMA(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
    <indicator>
        <f>HighDivEMA(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
    <indicator>
        <f>LowDivEMA(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
</section>

<section name="ReturnEMA">
    <indicator>
        <f>ReturnOfOpenEMA5(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
    <indicator>
        <f>ReturnOfCloseEMA5(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
    <indicator>
        <f>ReturnOfHighEMA5(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
    <indicator>
        <f>ReturnOfLowEMA5(\i)</f>
        <i>2,3,4,5,7,10,14,20,24,26,30,36,40,42,46,48,50</i>
    </indicator>
</section>
```

```java
public static HashMap<String, List<String>> parseXML(File fXmlFile) {

    HashMap<String, List<String>> map = new HashMap<String, List<String>>();

    DocumentBuilderFactory dbFactory = DocumentBuilderFactory.newInstance();
    DocumentBuilder dBuilder = null;
    try {
        dBuilder = dbFactory.newDocumentBuilder();
    } catch (ParserConfigurationException e) {
        e.printStackTrace();
    }
    Document doc = null;
    try {
        doc = dBuilder.parse(fXmlFile);
    } catch (SAXException | IOException e) {
        e.printStackTrace();
    }

    doc.getDocumentElement().normalize();

    NodeList nList = doc.getElementsByTagName("section");

    for (int i = 0; i < nList.getLength(); i++) {
        Node nNode = nList.item(i);
        if (nNode.getNodeType() == Node.ELEMENT_NODE) {
            Element element = (Element) nNode;
            String sectionName = element.getAttribute("name");
            map.put(sectionName, new ArrayList<String>());
            NodeList childList = element.getElementsByTagName("indicator");

            for (int j = 0; j < childList.getLength(); j++) {
                Node childNode = childList.item(j);
                Element cElement = (Element) childNode;
                String indicatorNameOrigin = cElement.getElementsByTagName("f").item(0).getChildNodes().item(0)
                        .getNodeValue();
                if (cElement.getElementsByTagName("i").getLength() > 0) {
                    String indicatorParaList = cElement.getElementsByTagName("i").item(0).getChildNodes().item(0)
                            .getNodeValue();
                    String[] indicatorParas = indicatorParaList.split(",");
                    for (String indicatorPara : indicatorParas) {
                        String indicatorName = indicatorNameOrigin.replace("\\i", indicatorPara);
                        map.get(sectionName).add(indicatorName);
                    }

                } else {
                    map.get(sectionName).add(indicatorNameOrigin);
                }

            }

        }

    }
    return map;
}
```

6. aggregate all indicators(eg. macd, ead). Aggregate all strategies using optimization framework or selection framework to gain statistical alpha 7. indicator effectiveness test 1. test correlation - the correlation between indicator and profit vs. the correlation between correlation and white noise(hypothesis test) * use spearman correlation rather than pearson correlation* 2. Use Monte Carlo Simulation to do permutation test of effectiveness of indicator 3. Very very hard - detect sensentivity to market regime change(osicallation and trend) and identify market regime change.

Integration of single indicators and portfolio theory

Form indicator as factors: standardization to mean-0, normal/t-distributed scores. Select powerful ones (ones that passed the permutation test). Optimize to maximize holdings exposure to factor with risk penalty. The key is still feature engineering.

$$h'f - ah'Vh$$

h:列向量，portfolio 里各个品种的持仓
f:列向量, expected return
a 常数: risk aversion
V：covariance matrix
‘ 意思是矩阵转秩

For Covariance, See section ”covariance matrix”.

Strategy Risk Management and Money management

small stop loss, big stop gain level on reversion strategies.bigger stop loss, smaller stop gains on volatile markets - based on experience, market analysis.

Choose symmeteric/non-symmetric risk control based on market belief

Hedging and Market Exposture Management - Volatility Control and Automatic de-leveraging.

together with cost consdieration.

8.3 Backtesting

1. Surviorship-bias
2. Look ahead-bias
3. In-sample bias

Chapter 9

Portfolio Risk Management

9.1 Derivatives and Hedging Strategies

2. risk management 1. Var - historical, analytical, MC good and bad 2. credit risk exposure (pv only swap has) 3. derivatives 1. futures, hedge, synthetic equity/cash, pre-investing 2. options 1. spread-bull bear, butterfly 2. straddle, collar, box spread(bull, bear spread- risk free rate) 3. interest rate swap - leveraged floating-rate notes, inverse floater; currency swap 4. swaption - payer, receiver - use receiver to add/remove callable bond features

1. fixed income 1. duration matching 1. requirements 2. vs cashflow matching(tenor offer), contingent immunization, horizon matching 2. index and challenges 1. index vs mutual fund,ETF, synthetic strategies(total return swap, less cash but counterparyt risk) 3. yield curve strategies 1. laddered, bullet, barbell vs level slope curvature 2. barbell vs bullet, condor and butterfly long short at level change, slope change, curvature change, yield volatility change performance and strategy (wing and body) 4. high yield and credit stprad 1. IGB HYB : credit risk, credit migration risk, interest rate risk, liquidity risk 2. access liquidity risk and tail risk 3. emerging market difference

9.2 FX risk management

1. currency management 1. forward price (long/short base currency) 2. options- risk reversal, put spread, seagull spread 2. index and benchmark 1. capitalization-weighted, price-weighted, equal-weighted index, fundamental-weighted indexes

Chapter 10

Monitoring and Performance Evaluation

10.1 Monitoring

1. rebalancing corridor width 2. CPPI/ swaption etc

10.2 Performance Evaluation

10.2.1 Equity

Factor Model (Axima, Barra)

10.2.2 Fixed Income

1. fixed income - no contribution-risk-free-asset categories-benchmark level(style)-manger-allocation effects(residual)

Chapter 11

Alternative Data and General Machine Learning