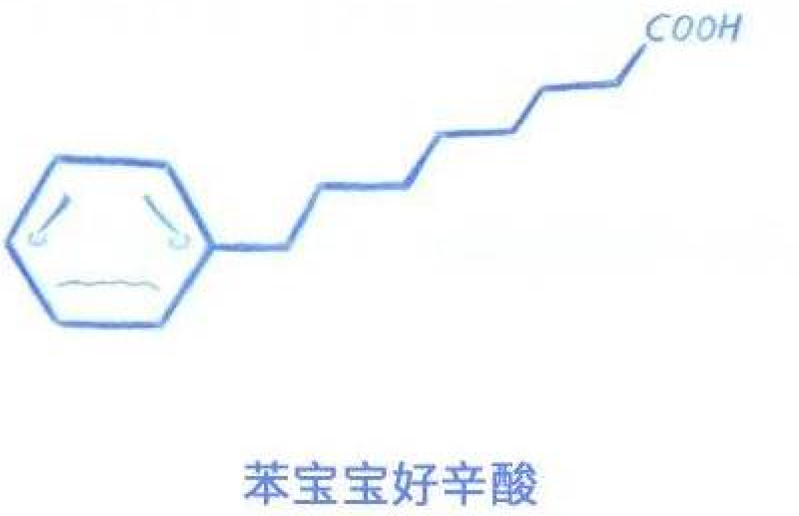
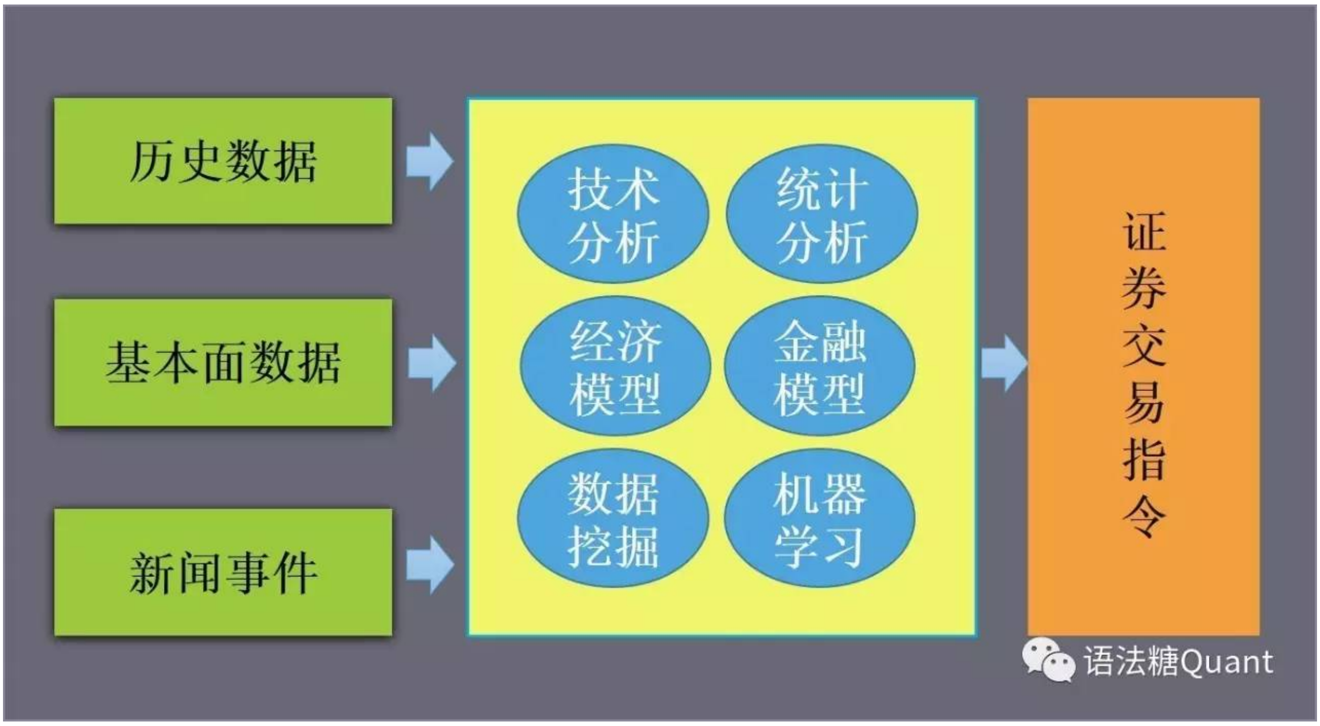
算法交易 Frame

量化交易通常也被称之为算法交易，指的是严格按照计算机算法程序给出的买卖决策进行的证券交易。如果把量化交易系统当做一个黑箱子，那么它的输入通常是证券历史行情数据和基本面数据，有些量化交易系统可以把新闻事件转化为输入。这个黑箱子的输出为证券买卖指令。如果你有兴趣打开这个黑箱子的话，你会看到一些千奇百怪的东西，包括但不限于以下类别的内容：技术分析，统计分析，经济模型，金融模型，数据挖掘，机器学习，人工智能。此外，人类所有关于预测未来的聪明才智都可能被放进这个黑箱子里，例如阴阳五行，周易八卦……

所以，算法交易真的是一门艰难的学问。人生已经如此地艰难，有些箱子就不要拆开！但苯宝宝禁不住自己的好奇心，还是决定拆开来看看！拆开之后，苯宝宝的心情是这样的！



本文主要探讨基于统计分析理论的量化交易系统。



一，寻找切实可行的策略

通常，没有人能够无中生有凭空想出一个绝妙的策略。创新往往不是像上帝那样灵光一闪创造一个全新的物种，而是像大自然普遍的生物进化法则那样在既有的物种基础上通过遗传、变异和选择机制迭代出'不死的小强'!

寻找策略通常分为三步，首先是从金融投资相关书籍，学术论文，网络获得有价值的策略。第二步对它们做适当的整合和改进。最后通过历史数据回测筛选出经得住市场考验的策略。正如同生物进化需要不断地循环迭代，策略进化也需要在获得策略、改进策略、筛选策略这三个环节上不停地循环迭代。

1，一些有参考价值的获得策略的途径（后续持续更新）：

(1)，金融投资书籍：

《量化交易——如何建立自己的算法交易事业》 欧内斯特·陈 著

《量化投资策略——如何实现超额收益alpa》理查德·托托里罗　著

《量化投资——策略与技术》 丁鹏　著

(2)，学术论文：

《The Cross-Section of Expected Stock Returns》 EF Fama，KR French

……

(3)，网络：

量化投资平台论坛。如：米筐，聚宽，优矿

量化投资类公众号。如：广发金融工程研究

量化投资类QQ群。如：量化投资研究

2，选择策略的标准

重要的是策略的杠杆收益率，而不是名义收益率。

越高的夏普比率事实上是的你最终获利越多，因为高夏普比率让你可以运用更高的杠杆进行交易。

（1），如果一项策略的年交易次数有限，夏普比率很可能不会太高，但这并不影响你把它作为多元策略的组成部分，但不足以把它当作主要盈利来源的策略。

（2），如果一项策略的挫跌很大 （如超过10 %），或者挫跌时间很长 （如4个月或者更长），也不大可能有很高的夏普比率。

（3），根据经验规则：任何夏普比率低于1的策略都不适合单独使用。几乎每月都实现盈利的策略，其年化夏普比率通常大于2；几乎每天都盈利的策略，其夏普比率通常大于3。

有些投资顾问散布一种错误的观点：如果你的目标是最大化长期资本增长，最好的策略是买入并长期持有。这一说法在数学上早已被证明是错误的。实际上，'只要你能获得足够高的杠杆'，最大化长期资本增长可以通过最大化夏普比率来实现。因此，即使你的目标是长期增长，若不考虑税收以及保证金借贷限制，在一个持有期限短，年度收益率较低，夏普比率很高的策略和一个持有期很长，年度收益率较高，夏普比率较低的策略之间，你依然应该选择短期策略。

3，改进策略

有一些常用的方法可以改进初次回测业绩不佳的策略。如何在策略改进时不引入数据迁就偏差，并保持少参数的简单模型，更像是艺术而非科学。一个与参数优化相同的指导原则是：任何策略改进要同时提高训练集和测试集的业绩。

经常有一些非常简单的策略，为交易员所共知，收益在不断下降，但仍然盈利。股票配对交易就是一个例子。收益下降的原因在于众多交易员会利用这种套利机会，从而消除盈利。然而，通常可以对基本策略进行微小调整，来提升收益。

这些微小调整常常不如基础策略那样为人熟知，因此交易员也很少采用。例如有时会排除某只或某组特定股票。例如，交易员会倾向于从技术交易程序中排除价格易受消息影响的医药股，或面临并购的股票。还可以改变进出市场的时间或交易频率，甚至简单到选择一组不同的股票。

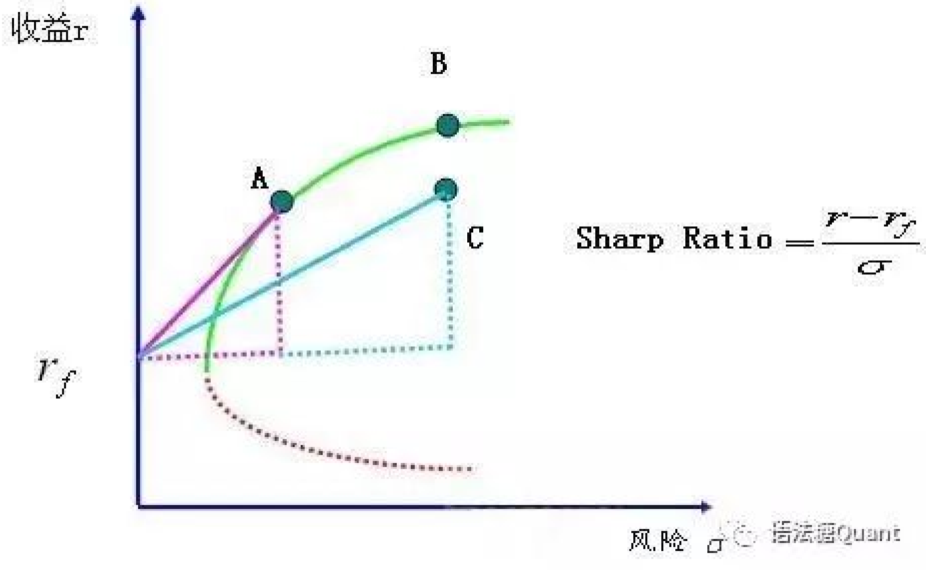
策略的改进，最好基于经济学基本原理，或者透彻研究过的市场现象，而不是依据一些主观的试错法则。否则，就有可能产生数据迁就偏差。

二，回测

1，业绩度量标准：夏普比率和挫跌

夏普比率和挫跌是在策略之间，交易员之间进行横向比较的最好指标。

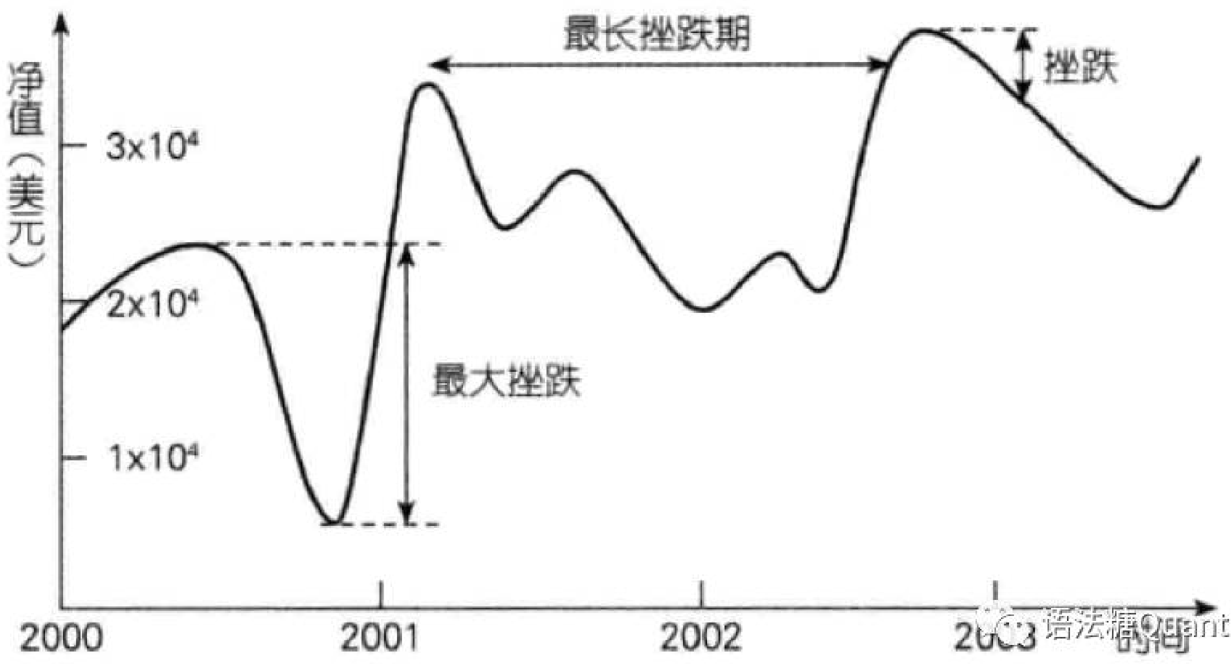
在资本资产定价模型中(CAPM)，资产收益率被分解成无风险收益率和风险补偿之和。而夏普比率反映的是投资组合承受单位风险所获得的超额收益率补偿。夏普比率越高，通过使用杠杆，将能够获得越高的杠杆收益率。



$$\text{Sharp Ratio} = \frac{r - r_f}{\sigma}$$

挫跌是直观反映投资组合风险的一个指标，挫跌包括包括最大挫跌和最长挫跌期。最大挫跌指的是在考察期内从任何时候开始向后计算，净值的最大亏损比例。而最长挫跌期指的是在考察期内从任何时候开始向后计算，净值处于亏损状态所经历的最长时间。由于收益和亏损之间具有一种不对称性，例如100元亏损10%，变成90元，后期必须盈利11.11%才能回复本金；如果100元亏损20%，变成80元，后期必须盈利25%才能回复本金；如果100元亏损50%，变成50元，后期必须盈利100%才能回复本金。 在亏损比例较高时，需要后期高的多的盈利比例才能回复本金，所以控制挫跌是非常重要的。

一个挫跌太深，挫跌期限太长的策略，实盘运行起来，这酸爽，足以让你怀疑人生！所以，应当选择挫跌参数比较美丽的策略，让生活更加容易一些！



2，避免常见的回测陷阱

（1）数据迁就偏差

如果你构建一个有100个参数的策略，完全可以通过优化参数，使得历史业绩看起来非常棒。同样可能的是，该策略的未来业绩和回测结果截然不同，非常糟糕。实际上，即使只有一两个参数，也很难避免所谓数据迁就偏差。一般而言，模型的参数越多，就越有可能遭遇数据迁就偏差。能经得住时间考验的往往是简单的模型。

下面的一些方法可以帮助避免数据迁就偏差。

合适的样本含量——根据经验规则，通常假定优化参数所需的数据点个数，是模型中自由参数个数的252倍。例如回测3个参数的日交易模型，至少要用三年的日价格数据。

样本外测试——将历史数据根据时间先后分为两段，后一段数据用于样本外测试。构建模型时，参数优化和定性选择使用前一段数据(称为训练集)，所得模型的测试使用后一段数据(称为测试集)。在理想情况下，基于训练集的最优参数和决策，对于测试集也是最优的，不过实际上很难做到这一点。但测试集上的业绩起码要合理。否则，模型就存在数据迁就偏差，需要进一步简化并减少参数。

动态参数优化——通过使参数不断适应变化的历史数据，来消除数据迁就偏差。假如你问：'模型中有固定的盈利上限吗？'交易员会诚实地回答：'没有，盈利上限并不是一个输入参数，它是由模型本身所决定的。'例如ATR指标就是一个很好的设置动态止盈止损参数的方法。

敏感性分析——在完成模型的参数和各种特征的优化，通过测试集的检验之后，可以通过改变这些参数或改变模型的定性决策来观察模型的业绩的变化。如果业绩变化很大，在参数取其它任何值时都很糟糕，模型很可能存在数据迁就偏差。各种简化模型，减少参数的方法都应该尝试。如果参数微小变化对模型业绩影响不大，应考虑将资金分配到不同的参数集和条件集。

（2）数据存活偏差

股票价格的历史数据库往往不包括那些由于破产，退市，兼并或收购而消失的股票，因此存在所谓的存活偏差，因为数据库中只有幸存者。使用有存活偏差的数据进行回测是很危险的，因为会夸大策略的历史业绩。这在策略有 '价值' 偏好时 （倾向于买便宜的股票） 尤为突出，因为有些股票便宜的原因是公司即将破产。

存活偏差影响早年业绩的原因在于，回测回溯的越早，消失的股票也就越多。一个纯多头策略在回测的早期会比当年实际的盈亏看起来好。在判断一项策略的适用性时，要重点关注其近几年的业绩，而不要被包括早年光鲜数字的总体业绩欺骗。另外，金融市场的 '状态转换'，也意味着早年的金融数据不能简单地应用于今天的相同模型。证券市场监管的变化或其他宏观经济事件，都可能导致重要的制度转换。很多具备统计学思维的读者很难接受这一点。他们中的很多人认为，数据越多，回测在统计上就越可靠。但这只对由平稳过程产生的金融时间序列来说才是正确的，不幸的是，由于前面提到的原因，金融时间序列显然是非平稳的。

以下方法可以帮助避免数据存活偏差。

使用无存活偏差的数据回测——设法获得没有存活偏差的数据进行模型回测。

使用近期数据进行回测——如果模型在十几年的数据回测绩效显著优于近几年的数据回测绩效，那么模型存在数据存活偏差，只有近几年的数据回测结果才具有较强的参考价值。

（3）交易成本偏差

交易成本不仅包括交易佣金，还得考虑流动性成本，以及自身指令所引起的市场价格的变动效应即 '市场冲击'。用市价指令买卖证券要支付流动性成本，若用限价指令买卖证券则要承担机会成本。此外，还要考虑滑价：由于互联网和各种软件的滞后，触发指令的价格和执行价格之间存在一定的差别。滑价有正有负，但平均来说，对交易员是成本而不是收益。如果你觉得滑价是收益，我建议你修改程序，在发送指令的时候，故意延迟几秒钟。

下面的方法可以帮助你避免交易成本偏差。

选择大盘股回测——通常大盘股的流动性比小盘股更好，受到的市场冲击更小。

适当调高佣金比例——由于有市场冲击成本，流动性成本，机会成本，滑价等因素的影响，回测时适当调高佣金比例是必要的。

3，策略例子

（1）配对交易

黄金的现货价格GLD和一揽子采金企业股票的价格GDX之间高度相关。利用协整分析，可以发现GLD多头和GDX空头所形成的差价呈现均值回归。通过训练集上的回归分析可以得出GLD和GDX之间的对冲比率，并设定配对交易策略进出市场的阈值。阈值在训练集上的优化会改变策略在测试集上的业绩。

（2）均值回归

麻省理工学院的 Amir Khandani 和 Andrew Lo 提出了一个简单的均值回归模型。策略非常简单，买入前一交易日收益最差的股票，卖空前一交易日收益最好的股票。在不考虑交易成本的情况下，针对小盘股做回测，这一策略的表现非常出色。如果考虑交易成本，并对大盘股做回测，这个策略的业绩会变差很多。

三，资金和风险管理

1，平均复合收益率

一道有趣的智力题(风险为什么不是个好东西)

这道智力题可能会难住不少职业交易员。假设某只股票的股价服从(几何)随机游走模型。也就是说每分钟以相同的概率上升或下降1%，如果购买了这只股票，是最有可能是盈利，亏损，还是持平(长期持有并忽略融资成本)？

绝大多数交易员会脱口而出'持平！'，但实际上是错误的。正确的答案是，平均以每分钟0.005% (0.5个基点) 的速度亏损！这是因为，对几何随机游走模型来说，平均复合收益率不再是短期(或单期)收益率m(此处为零)，而是 $g = m - \sigma^2/2$. 其中g 为平均复合收益率，σ是单期收益率的波动率。这也与几何平均值通常小于算数平均值的事实相一致。(除非所有数值完全相等，否则这两个均值不会相等。)若我们假定收益率的算数均值为零，几何平均值(平均复合收益率)就必然是负的。

此处给我们的教训是，经过风险调整之后的长期复合增长率总是会小于简单平均收益率，风险总是会减少长期增长率！

平均复合收益率公式：$g = r + m - \sigma^2/2$ 的一个数学证明：

(*g 为平均复合收益率，r是无风险收益率，m是超额收益率，σ是单期收益率的波动率*)

假设股票价格波动是一个 Ito 过程（几何布朗运动）：$\frac{dS}{S} = \mu\, dt + \sigma \epsilon \sqrt{dt} = \mu\, dt + \sigma\, dz$

由伊藤 (Ito) 引理：设W[S, t] = Log[S]，则W[S, t] 也是一个伊藤过程。

$$d\,\mathrm{Log}[S] = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma \epsilon \sqrt{dt} = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dz$$

$\Rightarrow$ (*该微分方程的解为*) $\mathrm{Log}[S_T] = \mathrm{Log}[S_0] + \left(\mu - \frac{\sigma^2}{2}\right)(T - t) + \sigma\,(z[T] - z[t])$

$\Rightarrow$ (*该微分方程的解为*) $S_T = S_0\,\mathrm{Exp}\left[\left(\mu - \frac{\sigma^2}{2}\right)(T - t) + \sigma\,\Delta z\right]$

(*其中*) $\Delta z \approx \mathrm{NormalDistribution}\left[0, \sqrt{T - t}\right]$

$$\mathrm{Log}[S_T] \approx \mathrm{NormalDistribution}\left[\mathrm{Log}[S_0] + \left(\mu - \frac{\sigma^2}{2}\right)(T - t), \sigma\sqrt{T - t}\right]$$

将t和T之间的连续复利年收益率记为 g

则 $S_T = S_0\,\mathrm{Exp}[g\,(T - t)]$;

$$\Rightarrow g = \frac{1}{T - t}\left(\mathrm{Log}[S_T] - \mathrm{Log}[S_0]\right) =$$

$$\left(\mu - \frac{\sigma^2}{2}\right) + \frac{\sigma \epsilon \sqrt{T - t}}{T - t} \approx \mathrm{NormalDistribution}\left[\left(\mu - \frac{\sigma^2}{2}\right), \frac{\sigma}{\sqrt{T - t}}\right];$$

所以：$\mathrm{Expectation}[g] = \mu - \frac{\sigma^2}{2}$;

或者：$g = r + m - \frac{\sigma^2}{2}$

2，凯利公式

假设策略i (这里用i代表第i个策略) 的收益率服从正态分布，其均值m[i]和标准差σ[i]已经给定(收益扣除融资成本，也就是超额收益)。用列向量F ={ f[1], f[2], f[3], ……， f[n] } 表示分配到n个策略的净值比例，列向量 M ={ m[1],m[2],m[3],……， m[n] }表示n个策略的超额收益。

则最优净值比例 F= C^(-1) M，其中C表示协方差矩阵，矩阵的元素C[i,j]表示第i个策略和第j个策略收益率的协方差，C^(-1) 表示协方差矩阵的逆。

如果假设所有策略在统计上独立，协方差矩阵就变为对角矩阵，对角线元素等于每个策略收益率的方差，公式十分简单：

$$f_i^* = \frac{m_i}{\sigma_i^2} = \frac{SharpeRatio}{\sigma_i} \left(*SharpeRatio = \frac{Rp - Rf}{\sigma} = \frac{m}{\sigma}*\right)$$

下面是对一般的凯利公式的一个证明：

凯利公式的证明

用正态随机变量 x_i 描述策略 i 的单期预期收益，策略i的超额收益率 $m_i = \text{Expectation}[x_i] - r$；

策略组合 $X = \sum_{i=1}^{n} f_i x_i$ 的预期超额收益率为 $m = \sum_{i=1}^{n} f_i m_i = F^T M$，

策略组合的方差 $\sigma^2 = \text{Variance}[X] = F^T C F$

年化复合增长率 $g = r + m - \frac{\sigma^2}{2} = r + F^T M - \frac{F^T C F}{2}$；对g求极值条件：$\frac{\partial g}{\partial F} = 0$

$\Rightarrow M - \frac{C F}{2} - \frac{(F^T C)^T}{2} = M - C F = 0$；

$\Rightarrow F = C^{-1} M$

3，用凯利公式计算最优杠杆

凯利公式给出了特定交易策略的最优杠杆。即若某个单一策略的超额收益率为m，收益率的波动率为σ，那么最优杠杆为：

$$f^* = \frac{m}{\sigma^2} = \frac{SharpeRatio}{\sigma};$$

采用凯利杠杆，将使净值的长期复合增长率最大。那么最大复合增长率为多少呢？简单运算可以得到答案，应为：

$$g = r + \frac{SharpeRatio^2}{2};$$

可以看到，最大杠杆收益率只和无风险收益率以及夏普比率有关。而夏普比率和杠杆倍数无关。所以夏普比率是比名义年化收益率衡量投资策略业绩更好的标准。

通常，由于参数估计存在误差，加上收益率不一定完全服从正态分布，诸多的黑天鹅事件造成了所谓的'厚尾效应'，交易员出于对安全的考虑，所使用的杠杆只有最优杠杆的一半。这就是'半凯利'投机。

例子：假设投资组合只包含跟踪标准普尔500指数的ETF基金SPY的多头头寸。设SPY的平均年收益率为11.23%，年标准差为16.91%，无风险利率为4%。因此，投资组合的年平均超额收益率为7.23%。可以算出夏普比率为0.4257。

根据凯利公式，最优杠杆 $f = 0.07231/0.1691^2=2.528$。注意一个有趣的地方，凯利公式的f与时间长短无关，所以并不涉及是否年化收益率和标准差，这和夏普比率与时间长短有关正好相反。最后，考虑了融资成本的年化复合杠杆收益率为13.14%。

这个策略的凯利杠杆2.58意味着，如果投资10000美元现金，并且相信收益率和标准差的期望价值，就可借钱买入价值252800美元的SPY，此时，这10000美元投资的预期年复合增长率为13.14%。作为比较，我们来计算一下没有杠杆时的年化复合增长率， $g=r+m-\sigma^2/2=9.8\%$. 这是仅用现金购买SPY的长期增长率，不等于11.23%的年收益率。

4，用凯利公式计算最优资产配置

凯利公式可以用来计算如何在不同的策略或资产之间进行最优化资产配置，这个方法比获得过诺贝尔经济学奖的马柯维茨的投资组合管理理论实用许多。

例子：假设有3只特定行业的ETF，来看看它们之间如何进行资本配置，从而获得投资组合的最大增长率。这三只ETF是：OIH(原油服务)、RKH(区域银行)和RTH(零售)。设它们的每日收盘价格数据分别存储在 cls1, cls2, cls3这三个数组中。Matlab代码如下：

```
clear;
%每日收盘价
cls1=[100 101 102 100 99 98 102 103]';
cls2=[20 23 22 24 20 19 20 18]';
cls3=[50 55 60 55 52 45 50 60]';
%计算每日收益率
ret1=(cls1(2:end)-cls1(1:end-1))./cls1(1:end-1);
ret2=(cls2(2:end)-cls2(1:end-1))./cls2(1:end-1);
ret3=(cls3(2:end)-cls3(1:end-1))./cls3(1:end-1);
%计算每日超额收益率，假设年化无风险利率是 r = 0.04
Ret=[ret1,ret2,ret3];
r=0.04;
excessRet=Ret-repmat(0.04/252,size(Ret));
%计算年平均超额收益率
M=252*mean(excessRet)';
%计算年协方差矩阵
C=252*cov(excessRet);
%计算凯利最优杠杆
F=C\M;
%最大年化复合增长率
g=r+F'*C*F/2;
%投资组合的夏普比率
S=sqrt(F'*C*F);
```

按照凯利公式进行最优化资产配置以后，策略组合的年化复合增长率和夏普比率分别为：

$$g = r + \frac{F^T C F}{2};$$

$$SharpeRatio = \sqrt{F^T C F};$$

5，用凯利公式进行风险管理

凯利公式不仅可用于资本的最优配置以及确定最优杠杆，还可用于进行风险管理。按照凯利公式，随着净值变化，需要对资本配置不断调整，以保证其最优。

例如在用凯利公式计算最优杠杆，投资SPY的例子中，假如你按照凯利公式买入价值252800美元的投资组合。第二天在SPY上亏损10%，则投资组合的价值变为227520美元，你的净值只有74720美元。这时候该怎么办？凯利准则要求立刻将投资组合价值减少到188892美元，因为这是最优杠杆2.528与净值74720美元的乘积。

应至少在每个交易日结束时做一次资本配置的调整。除了更新资本配置，还应该周期性地跟踪计算最新的均值和标准差，从而更新F。那么回溯期应为多久？更新凯利公式中这些输入变量的频率又是怎样的呢？这取决于策略的平均持有期。如果仅持有一天左右，根据经验规则，建议回溯期为6个月。使用较短回溯期的一个好处是，可以逐步减少正在失效策略上的风险暴露。至于更新频率，如果编写一个程序，F每天更新一次应该不是什么难事。

最后一点，有些策略每天生成的交易信号是一个变数，从而导致每天的头寸以及总资本也是变数。在这种情况下，该怎样使用凯利公式来决定资本呢？可以用凯利公式来计算最大头寸数量以及允许的最大资本。使用低于凯利公式计算出的杠杆永远是安全的。

在面临交易亏损时，凯利公式会建议你减少投资组合规模。无论风险管理方案是否基于凯利公式做出，亏损减仓都是风险管理的惯常做法。无论何时发生亏损，风险管理总会要求建仓。相反，当策略盈利时，最优杠杆会建议增仓。

随着交易亏损的增大，除了逐渐降低模型的杠杆直到零，没有什么更多的办法可以降低模型风险。根据最新的历史收益率均值和标准差，使用凯利公式不断调整杠杆，可以系统地做到这一点。随着回溯期内的历史超额收益率均值降为零，凯利杠杆也将调整为零。这比突然弃用一个发生了很多挫跌的模型要好。

四，统计套利专题

1，均值回归策略和惯性策略

只有当证券价格是均值回归的或趋势的，交易策略才能盈利。否则，价格是随机漫步的，交易将无利可图。

这样的假设通常是安全的：除非公司的预期盈利发生了变化，股票价格会均值回归。当公司的预期盈利发生变化时，股票会显示惯性。

惯性可能产生于信息的缓慢扩散——当越来越多的人开始注意到某条新闻，越来越多的人决定买入或者卖出某只股票，就使得价格朝某一方向运动。根据这种现象，我们可以构建一个叫做'后盈利公告漂移'的惯性策略。这一策略建议你在盈利超出预期时买入股票，低于预期时卖出股票。更一般地，很多信息公告都有改变股票未来的预期盈利的潜力，从而又激发一个趋势时段。至于何种消息会触发趋势时段，以及趋势会持续多久，需要靠交易员自己去发现。

除了信息的缓慢扩散，大额指令因流动性需求分拆执行或大投资者的私募投资决策，也会导致惯性。随着大型经纪商越来越多地使用日益复杂的执行算法，要判断一个观察到的趋势背后是否存在一个大额指令变得越来越难。

惯性也可能由投资者的羊群行为所引发的：投资者把他人的买卖决策作为自己交易决策的唯一判断标准。

使用相同策略的交易员之间日益激烈的竞争会对策略本身造成怎样的影响？对于均值回归策略而言，典型的结果就是套利机会的逐步消失，从而使得收益率逐渐降低至零。当套利机会消失殆尽时，均值回归策略就危险了，因为越来越多的交易信号来自股票估值的基本面变化，而这并不会均值回归。对于惯性策略而言，竞争的后果是减少该趋势持续下去的时间周期。当消息以更快的速度扩散，从而更多交易员们更早地利用这一趋势，均衡价格就会很快实现。所有在均衡价格实现之后建仓的交易员都将无利可图。

2，平稳性和协整性分析

均值回归的数学基础是时间序列的平稳性和协整性分析。如果一个时间序列不会越来越大地偏离初始值，这个时间序列就是'平稳的'。显然，如果证券的价格序列是平稳的，它就很可能适用于均值回归策略。不过，大多数股票的价格序列都不是平稳的，通常表现为几何随机游走，不断地离初始点价值越来越远。尽管如此，你能找到像买入一只股票，卖出另一只股票这样的股票配对，配对的市场价值是平稳的。这种情况下，两个独立的时间序列被称为'协整'。通常，协整配对中的两只股票来自同一行业。交易员早已很熟悉这种配对交易策略。他们在配对的差价低的时候买入配对组合，在差价高的时候卖出配对——这就是经典的均值回归策略。

可以用DF或ADF检验等方法来判断两个时间序列是否是协整的。可以用最小二乘法ols方法确定配对因子。

如果你认为同行业的任意两只股票都是协整的，我们可以给出一个反例：可口可乐和百事可乐。用ADF检验方法可以发现两只股票协整的可能性小于90%。如果一个价格序列(可以是一只股票，一对股票或者是一个投资组合)是平稳的，只要未来继续保持平稳(这未必能保证)，采用均值回归策略一定能盈利。反之则不然，成功的均值回归策略，并不一定要求一个平稳的价格序列。正如许多交易者所知道的，即使一个非平稳的价格序列，也可能有很多可以利用的短期回归机会。

协整性和相关性是完全不同的概念。两个价格序列的相关性指的实际上是一段时间内其收益率的相关性，如果两只股票正相关，其价格变化方向在大多数时间都是相同的，然而，有正相关关系并不反映两只股票的长期行为特征。特别地，正相关无法保证两只股票在长期内价格偏离不会越来越大，即使在大多数时间里它们的变动方向相同。类似地，两只股票是协整的，但可能并不相关。它们的变动方向可以有时相同，有时相反。但是它们的差价总会在一段时间内回到1美元左右。

3，高频交易策略

（1）什么是高频交易策略?

许多高频交易专家认为高频交易策略是那些持仓不超过几秒的策略，更广泛的定义也会把日内交易策略也归为高频交易策略。由于重组的流动性，高频交易策略最早出现在外汇市场，随后是期货市场。最近几年，随着股票市场流动性增强，逐笔历史数据普及以及计算速度大幅提高，高频交易策略也被广泛应用于股票交易。

（2）高频交易有什么优点?

高频交易能够获得很高的夏普比率。根据大数定律，交易的次数越多，收益率相对于均值的偏差就越小。而在高频交易策略下，一天可以交易成百上千次。因此，如果你的策略具有正的平均收益率，高频交易日收益率与平均收益率的偏差将会达到最小。由于高频交易策略具有高的夏普比率，与非高频策略相比，可以使用的杠杆水平更高，高杠杆进而大大提高策略的净值收益率。

此外，高频交易策略进行风险管理非常容易：亏损时可以很快地'去杠杆'，市场糟糕时可随时停止交易。最坏的结果也不过是当某个策略太过普通，收益率不断下滑而最终无利可图。至于突发大额亏损或多账户的连环亏损则是不太可能的。

（3）高频策略如何获得正的平均收益率?

高频策略有均值策略，也有惯性策略，有市场中性的策略，也有多头单向的策略。但通常来说，这些策略要么是利用市场中出现的微小的无效性而获利，要么是通过提供短暂的流动性以获得微小的报酬。与基于宏观趋势或公司基本面的策略会时常经历大的波动不同，高频交易策略依赖于每天都会出现的市场短暂无效和流动性需求，这使得它可以每天持续盈利。

（4）高频策略有什么缺点?

由于平均持有期只有几分钟甚至几秒钟，对策略进行回测不那么容易。交易成本对最终结果会产生重大影响。此外，高速执行往往决定交易的实际盈亏。专业的高频交易公司都是用C语言编写策略，并且会将服务器放置于交易所或主要网络节点附近以降低时滞。所以，虽然高频交易策略夏普比率较高并且收益率可观，但还是不大适合初涉此领域的独立交易员。当然，随着交易经验和资源的积累，我们没有理由不向这个目标前进。

4，高杠杆组合优于高贝塔组合

根据资本资产定价模型CAPM，投资组合的收益率和它的贝塔值成正比。因此我们可以提高贝塔值(选择高贝塔值股票)来提高投资组合的收益率。根据凯利公式，我们也可以通过提高杠杆水平来提高投资组合的收益率。这两种方法似乎都合乎常理。事实上，可以通过提高一个低贝塔值投资组合的杠杆使其收益率与高贝塔值投资组合的收益率相当。问题是，高杠杆低贝塔的投资组合与低杠杆高贝塔的投资组合是等价的吗?

回答是否定的。在使用凯利杠杆的条件下，投资组合的长期复合增长率和夏普比率的平方成正比，而不是与平均收益率成正比。所以，如果两个投资组合的平均收益率相同，我们应当选择风险或标准差较小的那个组合。而实证研究表明，由低贝塔值股票构成的投资组合往往风险较低，夏普比率较高。

例如，PanAgora资产管理公司的Edward Qian博士在《风险对等组合》中写到，股票和债券60:40的资产配置并不是最优的，因为它的风险资产比重过大。如果想在保持风险水平不变的情况下，实现更高的夏普比率，Qian博士建议23：77的股票和债券配置，并使用1.8倍的杠杆。

不过要注意，所有这些结论都是基于收益率服从正态分布的假设。因为真实收益率的分布具有厚尾特征，对低贝塔值的股票使用过高的杠杆并不是明智的。