

会议主题：方正金工Q培训第三期

会议内容：概率学家诠释技术分析稳定又可靠

主讲人：郑伟安 长江学者、千人计划、华东师范大学终身教授

时间：2017年6月9日

概率学家诠释技术分析稳定又可靠

(平稳性过程在金融市场中的应用)

郑伟安

郑伟安：今天我讲的主要是平稳性过程在金融市场中的应用。首先讲一下什么是平稳性过程。假设我们有 m 组数据，如果它们的统计规律与从中任何一个时间重新开始得到的数据相同，我们称这 m 列数据为平稳的。实际上这个只在模型里面存在，真的给你 m 列数据是没有办法验证的，因为给你一个数据就是一个数据，你如何知道他的分布呢？所以统计里有一个比较弱的平稳过程。他只要求前面的一列数据和后面的一列数据在统计特征上不变，称为弱平稳。弱平稳是什么意思？要求这两列的均值、协方差不变，那么他们就平稳了。当然真正让学生做的时候还有更简单的办法，如果你们会用统计软件，上面有一个方法更简便一点——ADF检验。ADF检验只能检验1维数据，现在是 m 维的，如何验证呢？我们可以采用各维的线性组合。实际上我们没有办法证明他是平稳的，只能用假设检验看他们是不是平稳的。如果假设检验证明他不是平稳的，那么我们需要的数值就是平稳的。

平稳的数据序列

- 假定

$\{x_{1,1}, x_{1,2}, x_{1,3}, \dots\}$

.....

$\{x_{m,1}, x_{m,2}, x_{m,3}, \dots\}$

是 m 列数据，如果它们的统计规律（概率分布）与从中任何一个时间重新开始得到的数据

$\{x_{a+1}, x_{a+1}, x_{a+3}, \dots\}$

.....

$\{x_{m,a+1}, x_{m,a+2}, x_{m,a+3}, \dots\}$

相同，我们称这 m 列数据为平稳的。如果只要求统计特征（均值，协方差）不变，则称弱平稳的。

平稳的数据有什么用呢？如果一组数据在第一种概率分布意义下是平稳的，那么对于任何一个“好的”函数，这列数据的函数的均值是收敛的，学过概率统计的都知道这个叫大数定律。所以如果我们讨论随机变量的时候，可以假定我们在研究平稳过程，如果给你一系列同分布的随机变量，从当中看它们都是一样的，分布

律没有改变。

平稳的数据有个特性。对任何一个“好的”函数 f ，
 $(f(x_{1,1}, \dots, x_{m,1}) + \dots + f(x_{1,n}, \dots, x_{m,n}))/n$
收敛。这被数学里被称为**遍历定理**。我们看一个例子
假定有两列数据：

3 2 2 3 0 2 3 0 3 3 2 2 0 3 2 2 3 0 0 3 3 2 3 0 3 2 3 0 2
2 0 0 2 0 1 0 1 1 0 2 0 0 0 1 0 2 2 2 0 2 0 2 0 2 0 0 1 1
假定我们用统计的方法检验，无法否定它们是平稳的。

• 我们记函数 $f(x_{1,1}, x_{2,1})=1$ （如果 $x_{1,1}=3, x_{2,1}=1$ ）；否则为0；
则

$(f(x_{1,1}, x_{2,1}) + \dots + f(x_{1,n}, x_{2,n}))/n$ 收敛，假定是0.31。这是
“3”出现的概率。

下面我讲讲技术分析的问题。华尔街认为技术分析是伪科学，其中有一本非常出名的书用统计的办法否定了技术分析，他把常用的技术分析用到了美国几千只股票里，结果证明所有的技术分析都没有用，不管是什么时候买进还是什么时候卖出，都没有盈利。但是里面有一个很大的问题，就是他把几千只股票都假定是不同的样本。但是我们在考虑统计的时候要假定他们之间是相互独立的，不独立怎么能做统计呢？所以这本书是错的。

我和王肇东先生合写了一本书，《高频交易与概率论》，书里面介绍了如何用统计来做交易，具体形式是什么样子呢？举个例子，大家都知道加利福尼亚轮盘赌吧，里面有18个黑的格子，18个红的格子，还有2个绿的格子，你和他赌的话，压在黑色上，赢的概率是18/38，如果你赌红的话概率也是18/38，但是，赌场赢的概率是多少？20/38！赌场是凭什么赢呢？大数定律！因为他知道你每次去转，都是独立同分布的随机变量，所以你平均赢的次数只有18/38，也就是9/19。而赌场平均赢的次数是10/19，赌场赢就赢在这里。同样地，我们也可以在市场中利用真实的交易数据去开设“赌场”。

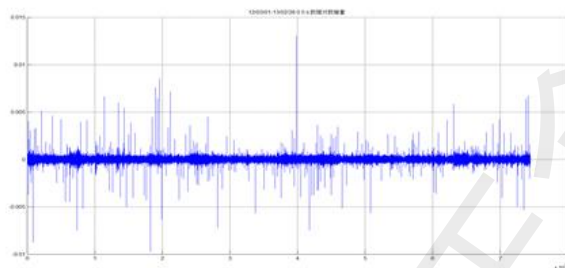
下面再给大家看一个有市场依据的东西。如果我们用 P_t 表示股指期货的价格， $p(t)$ 是 $P(t)$ 的对数，即 $p(t)=\log P(t)$ 。那么 $p(t+\Delta t)-p(t)$ 是平稳过程。 $p(t+\Delta)-p(t)$ 在金融中称为“对数收益率”，其中 Δ 可以取一秒，半秒，或者1/4秒，就代表时间 t 前面一个时刻。我们选取的是一年的数据，从验证结果看这是个平稳过程。数据结果从头开始看，和从当中开始看，其实是一样的，均值收敛，所以遍历定理成立。这个在金融上是站得住脚的。

股指期货的对数增量平稳性

以 $P(t)$ 记股指期货的价格

$$p(t) = \log P(t)$$

则 $p(t+\Delta) - p(t)$ 是平稳过程。它在金融中称为“对数收益率”



今天的第一部分我们就采用平稳性这一个假设，然后想办法找到套利的方法。我前面提到的那本：批判技术分析的书，这本书可以到亚马逊上买到。他是用统计里面的抽样理论来做的，但是做的方法我觉得是不对的，而我们用的是另外一个理念，就是平稳过程。

下面给大家看一下如何利用平稳过程赚钱。我们先找到买入点，卖出点，然后每次一个来回有一个收益，我的目的是要使我的收益构成平稳过程，这样的话他的均值就会收敛，平均收益就存在。我的目标是平均收益的均值是正的，如果是负的就赔钱了。总之，平均收益可以收敛，而且是正的，那就可以赚钱了。

把收益造成平稳过程

- 前面对股指期货的结论对一般股票数据也对。对数收益率是平稳过程。
- 我们的方法是在这平稳过程中找到买入与卖出点，让收益 $X(1), X(2), X(3), \dots$ 也构成平稳过程，于是其均值
- $(X(1) + X(2) + \dots + X(n)) / n$
- 收敛。如果这个均值是正的，那么我们可以反复进行，形成统计套利。

接下来给大家看个在中国市场上股指期货的例子。大家知道，股指期货现在的交易实际上是没有效的，股灾之后基本上交易规则就没办法通过股指期货赚钱了。

因此我们的数据是股指期货在股灾一年前的数据，那么我如何做呢？目标其实很简单，我要买进卖出，使我的收益构成平稳过程，那么均值就收敛了，并且我要他根据历史数据做出来的均值是正的。我知道这个收益是平稳的，而且他的均值已经收敛了，而且是正的，我继续玩下去不是能持续挣钱吗？

那么要怎么建立套利模型呢？我的思想很简单，我只相信模型越简单越好，很复杂的模型往往是不行的。8年前，我在瑞士听过一位教授的报告，他把过去的股票价格用一个数学模型表示出来，他在模型上加了70项，一个个指数加上去，我觉得这样做也许是没有用处的，因为这个项数加的越多，看着和过去的曲线越来越接近，但是却和将来的模型越来越远。所以我要求学生做模型的时候，一定要能几句话就能讲清楚模型，而且一般模拟2、3次就要出来，这时候我才会相信你。如果你根据一个模型拟合出一条曲线，看看不像了再修改程序，几次操作下来，看看好像很对了，但是一点用处都没有，因为你很难预测将来，所以我不相信现在部分大数据在金融领域的研究，因为大数据研究的前提要保证过去发生的模式在将来也会发生。下围棋是可以的，361格，你是可以算的出来的。但是金融最难的就是预测将来，所以模型必须越简单越好，否则很容易过度拟合。

因此我的要求就是利用平稳性赚钱，接下来给大家看看怎么赚。这个差叫对数收益率。我们可以证明，对数收益率是平稳的过程，所以我第一个念头就是是不是可以把技术指标化为对数收益率的函数，因为平稳过程的函数还是平稳过程。这有什么好处呢？技术指标化为平稳过程之后，就可以算出这件事发生了多少次，有多少次是成功的。统计的应用主要是靠大数定律，所以我只相信大数定律。接下来给大家看看这句话，把收益构造成平稳过程，我们把这个叫做统计套利。我们知道，真正的套利在金融市场上，只会赚不会赔，但是统计套利的意思是我允许你赔，但是平均下来你必须赚，就像开赌场一样，我开赌场有可能输，因为可能会有个赌徒运气很好，一直赢钱，但是我也必须让他赌下去，虽然我有倒霉的可能，但是倒霉的概率很小，就像开赌场一样。所以我们要争取做到统计套利。接下来我来看一看，如何利用常用的技术指标找到买入卖出点？首先，我们知道布林带不是平稳过程，那么我要把他化成平稳过程。移动平均MA大家都知道，指数移动平均EMA大家也都知道，他们都不是平稳过程，但是我们可以证明，如果把短期平均除以长期平均以后， $MA[P, n](t)/P(t)$ 与 $EMA[P, n](t)/P(t)$ 都是平稳过程。这么一来我们就可以判断什么时候短期平均会超越长期平均。如果商等于1就是他们相等；如果短期平均从下面穿过长期平均，也就是分子本来是小于分母的，小于1就变为大于1了，这就是黄金穿越。我让我的学生去研究，这个平稳过程中有几次穿越1，并且我让他们在找到穿越1的时候买进，过一秒卖出，然后看看一年内的收益有多少。很多学生计算下来，收益是收敛的，但是没有赚钱，并且扣掉手续费后还是收益收敛至负值，所以不行，但收益还是平稳的，只不过是平稳地输钱。那么我们要想办法平稳地赢钱。

技术指标

- 我们发现，常用的技术指标都可转换为平稳过程。所以可以统计它们的有效性。下面举例。
- 对时间序列 $Y(t)$ ($t=1, 2, \dots$) 定义
- $MA[Y, n](t) = [Y(t) + \dots + Y(t-n+1)]/n$
- 为移动平均
- $EMA[Y, n](t) = 2[nY(t) + (n-1)Y(t-1) + \dots + Y(t-n+1)]/n(n+1)$
- 为指数移动平均。
- 因为 $P(t-i)/P(t) = \exp\{p(t-i) - p(t)\}$ 平稳，所以它们的线性组合 $MA[P, n](t)/P(t)$ 与 $EMA[P, n](t)/P(t)$ 也平稳。
-

接下来是布林带与平稳过程，布林带如何化为平稳过程呢？我们知道，布林带本身并不是平稳过程，但我们构造这么一个分数，把这整个分数的绝对值控制在2.5以下，那么我们可以证明这个商是一个平稳过程。再给大家看看MACD，这也是经过三步来定义的，我们histogram除以股票价格，计算之后也可以证明他是平稳过程。

MACD也与平稳过程有关

- 对 $q < m < n$ 定义
- $MACD = EMA[\text{stockPrices}, m] - EMA[\text{stockPrices}, n]$
- $\text{signal} = EMA[MACD, k]$
- $\text{histogram} = MACD - \text{signal}$
- 引进平稳过程

$$\frac{\text{histogram}}{P(t - n - k)}$$

布林带与平稳过程

- 记 $\sigma(t)$ 为 $P(t), p(t-1), \dots, p(t-n+1)$ 的子样标准差, 则布林带定义为 $EMA[P,n](t) \pm 2.5\sigma$. 或者
- $2.5 > | \{P(t) - EMA[P,n](t) / \sigma(t) \}|$
- 但绝对值里面的商是平稳过程

$$\frac{e^{p(t)} - 2(ne^{p(t)} + (n-1)e^{p(t-1)} + \dots + e^{p(t-n+1)})/n(n+1)}{\sqrt{\frac{1}{n-1}(e^{p(t)} - MA(t))^2 + \dots + \frac{1}{n}(e^{p(t-n+1)} - MA(t))^2}}$$

下面就是我的学生要做的事了。我跟他说明平稳过程有了, MACD也有了, 我让他找一条线买进, 再找一条线卖出, 看看是不是赚钱。他确定了一个指标范围, 分别买入和卖出, 一共交易了2000次, 收益表现不错。但在2014年10月份后, 收益曲线开始平了, 赚不到钱了, 为什么呢? 我当时开始在外面做报告了, 金融消息很灵通, 一做报告以后, 本来有用的技术指标, 就变得没有用了。我们再看看ROC和RSI, 很明显, 他们都是平稳过程。

ROC与RSI

- ROC是指 $[P(t) - P(t-n)]/P(t-n)$ 。从形式可以看出是
- $p(t) - p(t-n)$ 的函数, 所以平稳。
- 记 $RS = \text{上涨时间段数} / \text{下跌时间段数}$, 这也是 $p(t)$ 增量的函数,
- 则RSI是指函数
- $RSI = 100 - 100/(1 + RS)$
- 所以也平稳。

最后我们再看一下遍历定理与高频交易。如果我们用 $I(t)$ 作为持有一手期货的示性函数。持有期货时等于1, 否则就为0, 那么我们就可以表示出 $(k/n, (k+1)/n)$ 时间段对冲该期货的对数收益率。我前面把技术指标划为平稳过程了, 如果这个对数收益是平稳的, 那么由遍历定理, 我们可以证明这个对数收益率也是平稳过

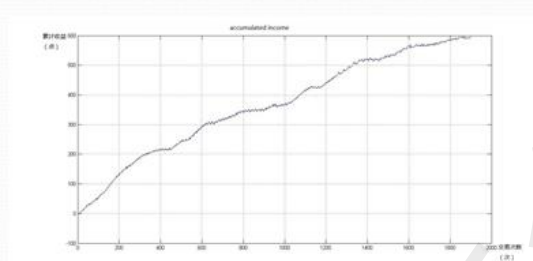
程。另外，高频交易里面交易费用很重要。关于交易费用，这里面有两个难题，一个是机器有多快，那时候我到一家大的证券公司，我的几个朋友告诉我，他们的速度内部时间可以做到56微秒了，结果我说不行，你看我们的书上的要求，需要控制在10个微秒。他们说光风控就要占30个微秒，10个微秒是不可能的。我让他把风控部分拿到外面来，结果他们照做了，过了两周告诉我，他们的速度大大加快了。当然风控还是很重要，15年股灾之后这部分不能放到外面来了，所以现在高频交易要赚钱已经不是很容易了。

遍历定理与高频交易

- 如果我们用 $I(t)$ 作为持有一手期货的示性函数，也即 $I(t)=1$ (持有期货时) 否则为0. 则
- $$I(k/n)[p((k+1)/n)-p(k/n)] \quad (*)$$
- 是 $(k/n, (k+1)/n)$ 时间段对冲该期货的对数收益率。
- 如果 $I(t)$ 是依据前面技术指标的平稳过程而操作的，则它是期货对数收益率的函数，所以关于 k 是平稳的。于是，**(*)构成平稳过程**。这样，各态历经定理就告诉我们平均对数收益率收敛
- $$\sum_{k=0, \dots, N-1} I(k/n)[p((k+1)/n)-p(k/n)] / N$$

- 记 $K(T)$ 为到时间 T 为止的交易次数，则从上式需扣除交易费用
- $$\text{Log} (1-0.0025\%) K(T)$$
- 对前面两式取和得到**累积对数收益是平稳过程的和**
- $$\sum_{(k+1)/n < T} I(k/n)[p((k+1)/n)-p(k/n)] + \text{Log} (1-0.0025\%) K(T)$$
- 从遍历性定理知道，
- $$\{\sum_{(k+1)/n < T} I(k/n)[p((k+1)/n)-p(k/n)] + \text{Log} (1-0.0025\%) K(T)\} / T$$
- 收敛。到时间 N 为止的盈利是
- $$P(0)\{\text{Exp}\{\sum_{(k+1)/n < T} I(k/n)[p((k+1)/n)-p(k/n)] + \text{Log} (1-0.0025\%) K(T)\}-1\}$$

扣除交易费后的累积收益



互动环节

提问1：我刚刚看了您的PPT，有两个疑问， $p(t+\Delta)-p(t)$ 中的 Δ 取的是多少？

回答：其实我们取了从1/2秒一直到5分钟，收益曲线都差不多，都是这个形状。

提问2：还有刚刚说到了交易成本， $\text{Log}(1-0.0025\%)*K(T)$ ，这个是卖出的时候吗？

回答：对的，买入的时候也有交易费用。

提问3：郑教授您好，请问这个图是不是一个理想状态下的收益？

回答：对的，这个并不是实战图。

(本会议纪要为精华内容)