

交易策略研究之回归策略 . 第13讲

 渔阳 2018-03-15 17:37

字数 5,878 阅读需 15分钟

“
欢迎来到量化小学”

▲ 加入[“量化小学”校友圈儿](#)提问交流

详细内容请观看PPT视频

上一篇:

期货策略研究基础:
趋势跟踪 第12讲

0

下一篇: ★

CTA交易策略的绩效
分析 第14讲



内容梗概

大家好, 欢迎来到量化小学, 我是渔阳。

今天我们来谈一谈**回归策略 (Mean-Reversion)**, 主要是以下几个内容:

先做简单的**复习**, 然后讲几个比较**基础的回归策略**, 最后特别要花一点时间讲一讲**量化研究当中的常见问题**。

首先, 我还是要再一次强调: 量化研究没有那么容易, **好的策略是需要结合经验和市场情况, 自己慢慢做出来的**, 千万不要有一种不劳而获的想法, 想着简单地寻找一个交易圣杯, 然后就可以躺着赚钱, 实际上这是不可能发生的。

我们前面讲过交易策略会随着时间推移、随着市场情况变化慢慢失效的, 重要的还是要掌握正确的方法, 这样才可以不停更新自己的交易策略。

复习: 交易策略的两个方向

讲具体的回归策略之前, 我们先来复习一下前面讲过的交易策略的两个主要方向——**趋势 (trend following)** 和**回归 (mean-reversion)**, 它们有什么相同与不同。右面列了回归策略的一些性质, 首先配对交易、反转交易, 这都是常见的回归策略。

来自特辑



量化小学

解放你的投资动手能力

最近更新

【学业总结】量化学习的脉络梳理, 以及
继续学习提高的路径

2019-04-12更新

进阶研究: 集成学习和深度学习 第31讲

2019-03-28更新



复习 - 交易策略的两个方向

| | 趋势 (trend following) | 回归 (mean-reversion) |
|--------------------|---|--|
| 策略举例 | CTA (趋势跟踪) , 做多动量因子 (momentum factor) , 做多期权 (long gamma) | 配对交易, 反转交易, 做多价值因子 (value factor) , 做空期权 (short gamma) |
| 波动率 (volatility) | Long | Short |
| 流动性 (liquidity) | Taker (拿走流动性) | Provider (提供流动性) |
| 交易量 (turnover) | 通常较低 | 通常较高 |
| 盈亏特点 | 胜率低, “经常输小钱, 偶尔赢大钱” | 胜率高, “经常赢小钱, 偶尔输大钱” |
| 杠杆使用 | 不适合高杠杆 (凯利公式的原理) | 适合加杠杆, 但需要特别注意尾部风险 |

W | PREMIUM

01

从波动率的角度来说, 我们希望开仓之后波动率向下走, 这样市场不那么动荡, 很多就会发生回归. 我们这样做是在为市场提供流动性, 交易量通常会比较高, 也意味着在回撤当中要对交易成本特别加以注意。

从盈亏的特点来说, 回归交易通常来说胜率比较高, “经常赢小钱、偶尔输大钱”, 所以它比较适合于加杠杆, 比较适合成为主力的策略, 但是也要特别注意尾部风险。

回归策略

回归策略举例——卷螺差

我们来看一看策略的例子：第一个卷螺差，螺纹钢和热卷板的**跨品种套利交易策略** (cross-product spread) 。



回归策略举例 – 卷螺差

螺纹钢、热轧卷板跨品种套利策略 (cross-product spread)

- 基本面：螺纹钢和热轧卷板的生产流程大部份重合，共同因素多
- 价差波动原因：基建 vs. 汽车/家电，钢厂生产线搭配
- 交易思路：当卷螺差拉开时，开仓赌价差回归

具体策略：

- ① 利用分钟级数据计算价差
- ② 计算价差的移动平均值和标准差
- ③ 当价差偏离移动平均值 \times 倍标准差时，开仓赌回归
- ④ 当价差回到移动平均值时，平仓

W | PREMIUM

5.02

首先，从基本面角度来说，螺纹钢和热卷板这两个品种关系是很紧密的，因为它们生产流程大部分是重合的，共同因素很多。

那价差为什么会波动呢？因为这两个品种用途毕竟还有所不同，螺纹钢主要是用在基建上；那么热卷板主要是汽车，包括家电制造。因此当这些需求发生变化的时候，钢厂生产线一旦配到螺纹钢或热卷板上，是需要时间才能够调到另外一个品种上面的。因此当需求发生变化的时候，钢和板的价差就会发生上下波动。

从交易思路的角度来说，既然是一些短期因素，那么当卷螺差拉开的时候，如果我们觉得基本面原因也好、技术面原因也好，都是短期倾向于回归的，那么我们就可以在合适的价格开仓赌价差的回归。

实操视频：具体策略实现



那么接下来我们用quant<OS>来看一看怎样实现这个策略。

从步骤上来说就是以下几步：第一，这次我们会用分钟级的数据来做回测，上次用的是日级别的数据。那么涉及到回归交易，它交易比较频繁，所以我们尝试着用一些更高频的数据，有了分钟级数据之后，就会计算价差的移动平均值和标准差。

然后当实际价差偏离移动平均值到一定的距离的时候，我们就考虑开仓来赌回归，当这个价差回到移动平均值，我们就平仓，很直白的一个交易策略。

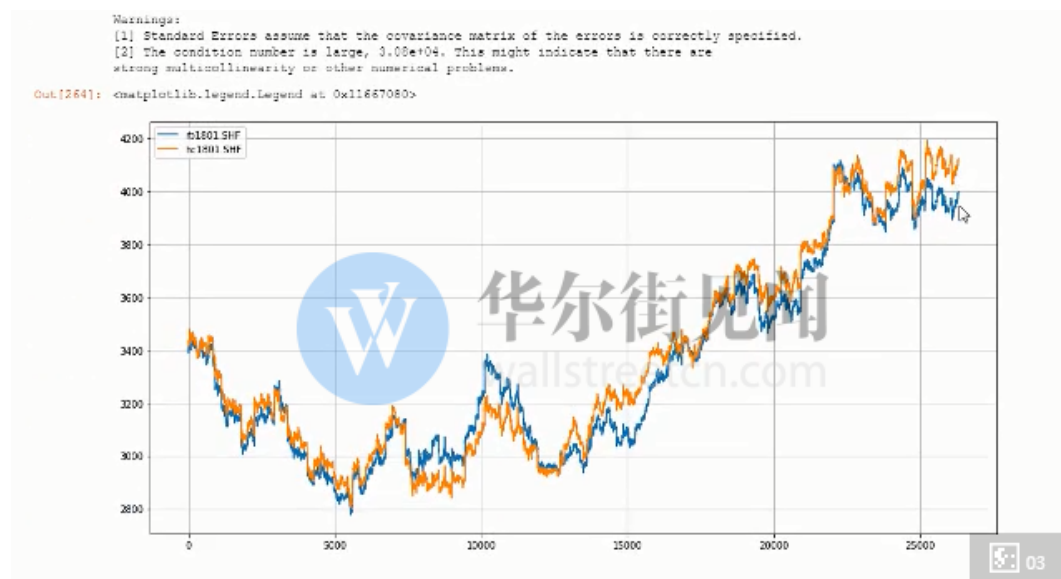
策略实现操作视频

现在我们切换到quant<OS>终端，有些代码每次都有，像登录、设置等等。那么接下来这些程序是设置品种，然后提取数据,在这儿我研究的是rb1710合约和HC1710合约。



上次我们也讲过，螺纹钢和热卷板的主力合约都是1月、5月和10月，研究区间是从去年3月15号一直到8月31号，把数据取出来，在这我设置参数取的是分钟级别的数据。

然后可以先画出一些图，来看一看他们是不是有什么关系（如下图），橙色的是热卷板，蓝色的是螺纹钢，我们看到它们的走势基本是一样的。



从宏观来讲，可以想象它们价差不会变动特别大。当然也可以画一个数据拟合线，发现它们都是比较接近的。那么下面这张图是把它们的价差转换成the score跟标准差挂钩的打分，发现它是在0上下震动的，也就是说它还是具有比较明显的回归特性。





好，那么我们注意到这个品种间有回归特性，就可以用回测引擎试一试能不能通过交易来赚这个钱。在这儿第一段程序是我设置的一些参数，包括用分钟级别数据做回测，一开始有一万块钱，还设置一个手续费等等。

设置策略参数

```
In [258]: props = {
    "symbol"          : ','.join(symbol_list),
    "start_date"       : start_date,
    "end_date"         : end_date,
    "bar_type"         : "1m",
    "init_balance"     : 1e4,
    "buffer_size"      : 30,
    "commission_rate"  : 0.00025
}
props.update(data_config)
props.update(trade_config)
```

I

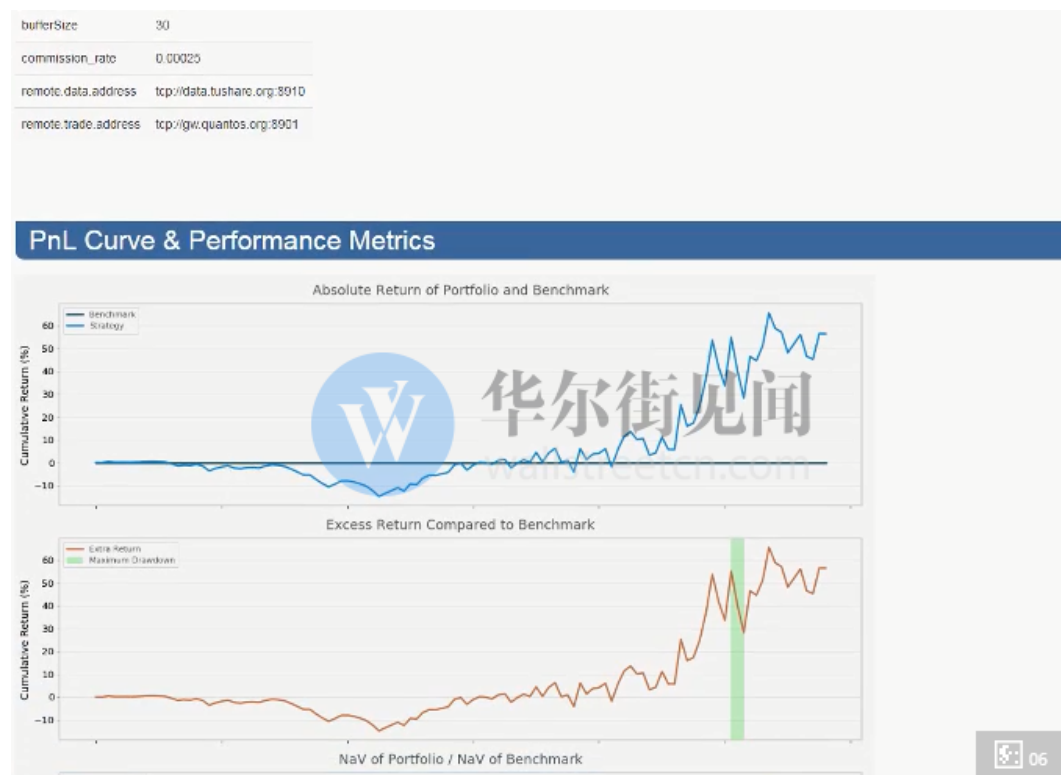
05

主要策略设置在底下一大段程序，特别是“on_bar”的函数，上次我们也讲过，这是事件驱动引擎中最重要的一个函数，就是当一个新的分钟线来了，我要做一些什么，就把需要做的事情在里面定义。



在今天这个策略当中，我们先算了一些平均值、标准差等等，然后我们做一些判断:如果价差超过平均值2.5倍标准差，我们就开空头去赌回归。反过来说，如果价差低于平均值再减掉2.5倍标准差，我们就开一个多头。什么时候平仓呢？当回到平均值的时候，我们就做平仓操作，挺简单的一个逻辑。

都定义好之后，我们就可以开始跑回测，跟上次一样，run strategy 把回测跑出来之后，再跑一遍分析函数，它就会生成一个研究报告，复制这个连接，然后在浏览器里面打开就可以。



我们就看到在这组回测参数下还是能赚一些钱的，大概三四个月的时间，居然赚了50%，当然上下是有些起伏的。底下列出了一些参数，如果年化的话，收益率可以高达160几，



因为我三四个月赚了50%，但是波动率也是相当惊人的，另外最大回测也是蛮大的（有17%）。

当然这是这一组和约、这一组时间点，如果换一组合约、换一些时间点就可能会有不一样的情况，我也鼓励大家课后要去试一试。

刚才我们讲过回归策略通常会做很多交易，那么来看看是不是这样，在quant<OS>做回测的时候，其实除了回测报告之外还会有一些其他数据，在这儿我们也看一看怎么来看这些更细致的数据。

在金融终端的左下角有一个加号，如果你点开会看到几个选项可以更新JAQS，也可以直接管理一些Python的包，最上面选项打开工作目录，会帮你切到计算机文件系统，看到这里面有output文件夹，里边有calendar spread，也就是今天跑的这段程序输出在这里面。

我们看到这些输出除了一些图、报告之外，底下还有一个excel文件，它实际上是把所有交易都列出来了，你在哪天买的，什么样的点位，到底做了些什么，都在这个文件里面。

那么我们看到还是做了很多笔交易的，一共大概是2800笔交易，在开仓实际上算两手交易的，因为同时交易螺纹钢和热卷板，因此2800笔除以4，从价差角度来说大概做了700个来回。

这个也还是挺正常，因为我们用的是分钟级数据，所以每天都有200多个数据点。回测周期大概有五个月，因此还是挺长的时间周期做了700个来回，所以刚才我们看了一个比较简单的回归策略的例子。

回归策略举例：跨期套利



接下来另一大类回归策略就是直接做跨期套利，比如说螺纹钢**月间套利策略（calendar spread）**，这是同一个品种，那么为什么价差还会有波动呢？这就涉及到几个常见的原因：

第一个是现货供需可能会发生变化，如果现货特别紧俏的话，就会导致近月期货价格比较高，远月相对会低一些。

第二个对于未来经济情况的预判，如果大家认为下半年经济情况不如上半年的话，基建开工情况不如上半年，远月就会比较便宜。

那么最后还有季节性因素，因为做基建、盖房子是有时间的，像北方冬天就不能施工，一般来说在春天特别还有在秋天，需求会比较大。

那么这些原因会导致不同月份之间价差有波动，我们就要去判断波动是真实还是长期的，还是有一些短期因素造成的。

从交易思路来说，如果我们认为1805和1810合约价差的上下波动受到一些短期因素的驱动，可能有投资人过激反应包含在里面，那么我们就更有理由相信它会逐步回归正常，因此可以考虑在一些合适的点来开仓赌价差回归，大家也可以用刚才那段程序把其中标的换一换测试一下策略。

那么涉及到跨期套利，这里面有几个一般性规律，我在这儿简单提一下，前面我们讲过期现套利，简单来说，**如果远月期货价格过高，那么可以先把现货买进来，付仓储成本，到远月期货到期的时候再把现货交出去进行套利。**



因此，如果期现套利是可行的，也就意味着远月升水不容易拉开很大，因为太大的话，期现套利盘就进来了，这意味着什么？我们在做回归交易的时候，**如果远月升水到一定水平，就可以先冲进去做价差回归**，因为向上的风险被限制住了。

从另外一个角度来说，如果远月发生贴水幅度就可以很大，因为你没有办法通过期现套利把贴水进行强制回归，需要先借现货把它卖掉，操作起来就很困难。比如说在2017年，黑色系很多品种远月都曾经出现过大幅贴水，而且维持挺长时间的。

最后，既然说近远月升贴水跟期间套利挺有关系的，因此我们就要注意区别期现套利什么时候可以做，什么时候不可以做。在商品期货里面涉及到仓单概念，可以把货存在仓库里面，然后做一个仓单，交割的时候不用交货，把仓单给别人就行了。

但是商品都是有一定保质期，即便像螺纹钢这种品种，也不能永远放着，食品更是如此。因此仓单的注册里有一定的学问，是定期需要更新的。

举个例子来说，豆油去年存在仓库里是合格的，但是重新注册仓单的时候，可能就不合格，因为时间长了。因此在具体交易商品期货的时候，要对交易规则、仓单注册包括交割有一定的认识，知道什么样的品种可以交割，什么样的品种不能交割的，然后去判断在什么样的点位可能发生期现套利，在什么样的点位可以考虑做价差回归。

除了仓单之外，交易有时候还会修改交割规则，比如以铁矿石为例，从2018年10月份开始，原来一些可以交割品质比较差的矿石就不能再交割。因为有这样的事件在那里，因此不同月份的期货合约也会有一个价差，这些价差存在是有原因的，未必会修复，在做价差交易的时候要特别小心这些事情。

价差回归的常见数学性质-协整 (co-integration)



好，刚才我们讲了跨期价差和跨品种价差，那么在做价差回归的时候，我们会经常用到一个数学概念叫**协整（co-integration）**。

这儿我们也简单过一下，在讲时间序列的时候，前面曾经反复提到过，首先要关心它是不是稳态的，稳态的时间序列有比较好的数学性质，方差是不随时间而改变的。

但是在实践当中，很多时间序列本身不是稳态的，特别是价格序列，从定义上来说，它就不可能是稳态的。这时候我们怎么办呢？尽管螺纹钢和热卷板本身价格序列都不是稳态的，但它们线性组合，比如说价差可能是稳态的，因为在一个区间以内上下振动，本身比价格的数学性质要好！

我们怎么来检验一个时间序列是不是稳态的呢？常用的方法包括**ADF（augumented Dickey-Fuller）**，大家可以自己去看数学上的定义。那么如果我们确定A、B这两个时间序列本身有一种协整性质的话，就知道协整的时间序列它是有长期均值，会回归的，就可以去设计相关的交易策略。

从标的组合的选择上来说，通常在高相关性基本面联系比较密切的品种中，容易出现一些协整的关系。比如前面讲的螺纹钢和热卷板，或者是焦煤、焦炭，这是上下游生产的关系。还有豆油、菜油、棕榈油，都是食用油有可替代关系。在这样的品种当中，比较容易找到协整的关系，比较容易找到能够做回归的组合。

好，所以通常的流程先去找一些可能的配对，然后做协整检验，如果OK长期会回归，就可以设一些开仓、平仓条件，把它放到刚才回测的框架中去，看看交易策略是不是有效。

■ 量化研究常见的坑



这两次我们讲到量化研究的具体策略，所以我觉得要也花一点时间过一下常见的坑有哪些，因为量化研究确实不容易，赚钱策略挺麻烦，倒是挺容易掉到一些坑里面。如果你犯了一些错误，研究结果就不能相信，我们有一些常见问题是要避免的：

第一个**看到未来数据 (Look-ahead bias)**，我在这里举了几个例子，比如如果期货的夜盘数据处理不当，你把它当作是今天的数据，就有可能发生在白天交易的时候，已经看到了晚上会发生什么，你当然是可以赚钱的，但这是错误的。

第二个常犯的错误就是**幸存者偏差(Survivorship bias)**，比如我在30多个商品期货上面都测试某个策略，其中有五个品种取得比较好的效果，这时候我就要当心幸存者偏差，因为完全可能是因为巧合，或者是试了那么多品种，总得有几个是行的。

像这样的话，根据幸存者偏差选出来的这些品种，因为只是一种巧合，你要把策略放到示范当中，可能就会变成随机行走。在这样的情况下，我就要分析这五个品种为什么会赚钱？究竟是因为幸存者偏差，还是因为这五个品种存在某些特殊规律。

好，另外一个常见的例子就是在研究股票的时候，如果用现在这一时刻的成份股里面的成长股，比如说倒着推，很可能发现它们的表现都是比较好的，这是因为只有表现好的股票才会被提拔到指数里面成为成份股，表现不好就已经被踢出去了。

怎样做会没有这个偏差呢？回到历史上面有任何一天在那一时刻指数的成份股来看，这些成份股未来表现怎么样？如果你这样做的话，第一是没有看到未来，第二也就没有幸存者偏差。

第三个常犯的错误就是**过拟合 (Over fitting)**，在做量化研究的过程中非常容易犯的错误。如果一组参数不行再试一组，然后1.5不行，就试1.6，最后总有行的。这样实际上在



凑结果，尽管对历史拟合的很好，但是对于未来实际上是没有什

这里也有一些常见的处理的方法：第一**不要搞那么多参数，也不要进行过度优化**。第二要测试策略的**鲁棒性（robustness）**，把参数稍微改一改，通常情况下都还是可以赚钱的，这样的策略它对参数不是那么敏感，可靠程度也会比较高。

第四个常见的错误是**数据挖掘（data-mining）**，我还是觉得在做量化模型的时候，最好能够在金融上能够讲清楚这个道理。比如刚才我们讲的螺纹和热卷的套利，因为它基本面很接近，但如果不是这样，仅仅通过数据挖掘，又有可能找到那种巧合而已。

在这我们也再一次提下深度学习，当然也是数据挖掘的一种，可以尝试但也特别要注意学习出来的究竟是真正规律还是一种巧合。那么深度学习的模型要想做得好，是需要有大量高质量数据，我觉得在量化研究上只适用于一些比较偏高频的、有大量数据的场景。

第五个常见的错误就是因为数据处理的不精细，导致**垃圾进、垃圾出的效果（garbage in, garbage out）**。比如在原始数据中的起止点，你没有经过分辨，没有经过处理，让这些错误的数据点影响你的模型研究、参数的选择，挺常见的一个问题。

最后一个常犯的错误，就是对交易环节有不切实际的假设，初学者比较容易犯这样的错误。

比如说回收当中允许在涨停板买入、跌停板卖出，这些是不可能的，实际上如果你用quant<OS>的话，它也不会让你这样做，就没有这种可能性。第二对于交易成本没有充分考虑，比如刚才我们讲的回归策略，因为交易比较频繁，所以对于成本还是挺敏感的。



如果你不考虑成本、或者是成本收得过低，可能会得到一个貌似赚钱的策略，但是放到实盘中付了成本之后就不行，所以也要注意。


好，我们在这简单的罗列了一些量化研究常见的坑，那么更详细的说明，可以参考课后资料里的一篇券商研报。

学习资料

所以马上我们就讲到学习资料，第一，大家还是要把课上讲过的样例仔细看一看，在金融终端里自己跑一下，然后尝试着改一改参数，看看结果有什么不一样。

学习资料

- 策略实现样例：quant<OS>金融终端教学频道
<https://www.quantos.org/courses/index.html>
- 渤海证券量化研报：《商品期货跨品种择时套利策略》，2016年9月28日
- 德意志银行量化研报：*Seven Sins of Quantitative Investing*，2015年1月
- 协整（co-integration）的数学概念和检验方法：任何经典统计教材或网络资源

 | PREMIUM

07

第二份资料是**渤海证券的量化研报**，关于商品期货的跨品种套利策略，建议大家看一下，前面我们讲过对于研报的使用方法是吸取一些想法，然后吸取一些研究大致流程，对于它的结论经常还是要打一个问号，抱一点怀疑的态度。



因为券商研报总是告诉你可以赚很多钱，实际上至少根据我们的经验，很多时候也是不能重复的。但这并不妨碍通过券商研报来获取一些基本信息，获得一些策略开发上的提示。

第三份资料是**德意志银行的量化研报——量化研究的七宗罪**（Seven Sins of Quantitative Investing），主要讲到哪些常见的坑，我建议大家可以去看一看，这个报告挺好的。其实量化研究，首先你别犯一些常见的错误。

最后关于协整（co-integration）数学概念以及检验方法，任何一本经典的统计教材，或者比较靠谱的网络资源上都有，建议大家去看一看。


作业

好，最后是两个简单的作业：


作业


- 在quant<OS>终端上：
 - 理解今天讲到的跨期价差套利程序
 - 用同样方法研究rb1801, hc1801的套利交易，回测是否还赚钱？
 - 用Dickey-Fuller 方法测试协整性
- 小明利用2016年以来的黑色系商品期货数据测试了一个趋势跟踪策略，取得了很好的效果。这个结果可能有什么问题？应如何进一步测试？

答案：在金融终端《量化小学》频道，稍后公布



华尔街见闻

 | PREMIUM

 08



第一个就是在金融终端上，用今天我们讲过跨期价差套利的程序，尝试一下其他一些组合。

比如换个月份，rb1801、hc1801，尝试着用它来做跨期套利，你就会发现简单的策略，对于时间的选择、对于品种的选择还是挺敏感的。所以在实际交易当中，你要知道什么时候能用这样的策略，或者说要花时间把它改得更精细一些。

第三，我们去尝试用ADF（augumented Dickey-Fuller）的方法测试一下协整性，在Python里面都可以直接调包来完成。

好，第二个也是一个思考题：如果某个分析师利用2016年以来的黑色系商品期货数据来做测试趋势跟踪策略，觉得取得很好的效果。

那么大家可以思考一下，这里面是不是有什么问题？你应该采取什么样的方法来进行应对？答案我们都会在金融终端的《量化小学》频道来公布。

今天我们大概就讲这些内容，下一次我们来谈一谈**怎样分析交易绩效**。谢谢大家。

-END-

加入“量化小学”的见识圈，关注动态



感谢您订阅本特辑，扫描下方二维码或[点击圈子链接](#)，即可加入专属见识圈子提问交流





量化小学



渔生

小学而大不遗，量化师生联谊会

感谢大家订阅《量化小学》，这里是学校见识社群，你可以随时提问、随时互动，我们一起投资，一起分享！



风险提示及免责条款

市场有风险，投资需谨慎。本文不构成个人投资建议，也未考虑到个别用户特殊的投资目标、财务状况或需要。用户应考虑本文中的任何意见、观点或结论是否符合其特定状况。据此投资，责任自负。

写评论

请发表您的评论



表情



图片

发布评论

华尔街见闻

关于我们

广告投放

版权与商务合作

联系方式

意见反馈

声明

未经许可，任何人不得复制、转载、或以其他方式使用本网站的内容。

评论前请阅读网站[“跟帖评论自律管理承诺书”](#)

法律信息

版权声明

用户协议

付费内容订阅协议

隐私政策

违法和不良信息举报

举报电话: 021-60675200 (周一到周五9:30-11:30, 13:00-18:30)

举报邮箱: contact@wallstreetcn.com

网站举报: [点击这里](#)



华尔街见闻APP



华尔街见闻公众号



微博@华尔街见闻



中央网信办
违法和不良信息举报中心

上海市互联网
违法和不良信息举报信息

[违法和不良信息举报受理和处置管理办法](#)

[清朗·财经违规内容专项整治公告](#)



举报中心

友情链接

[腾讯财经](#) | [财经网](#) | [澎湃新闻](#) | [界面新闻](#) | [全景财经](#) | [陆家嘴金融网](#) | [富途牛牛](#) | [网易财经](#) | [凤凰网财经](#) | [虎嗅](#)

© 2010 - 2022 上海阿牛信息科技有限公司 版权所有 沪ICP备13019121号



沪公网安备 31010102002334 号 增值电信业务经营许可证沪B2-20180399

