目录

[第一章 引言 1](#_Toc469261762)

[1.1 一级标题 1](#_Toc469261763)

[1.1.1 二级标题 1](#_Toc469261764)

# 影响高频交易策略的因素汇总

正文每页三十八行每行三十四字，正文段落和标题一律采用固定行间距二十PT, 正文每页三十八行每行三十四字，正文段落和标题一律采用固定行间距二十PT, 正文每页三十八行每行三十四字，正文段落和标题一律采用固定行间距二十PT, 正文每页三十八行每行三十四字，正文段落和标题一律采用固定行间距二十PT

## 高频交易策略的评价方法

### 策略年化收益率

策略年化收益率（Annualized Returns）表示投资期限为一年的预期收益率，是把当前收益率（日收益率、周收益率、月收益率）换算成年收益率来计算，是一种理论收益率而非真正的已取得的收益率，例如日收益率是0.01%，则年化收益率为365\*0.01%=3.65%，因为年化收益率是变动的，所以年收益率与年化收益率之间会有一定的差距。年化收益率指标的设立目的是为投资者提供比较直观的数据，供投资者将不同的投资项目进行比较与参考。

### 最大回撤率

最大回撤率（Max Drawdown）是在选定周期内任一时间节点开始计算，产品净值的最低点的收益回撤幅度的最大值。最大回撤比率用来描述策略可能出现的最糟糕的情况。回撤用来描述任一投资者可能面临的最大亏损，关注策略的最大回撤率可以帮助投资者了解该策略的风险控制能力可知道自己可能面临的最大亏损幅度，

目前对最大回撤率的主流认识有两点：

a)最大回撤越小越好；

b)回撤和风险成正比，回撤率越大，风险越大，回撤率越小，风险越小；

一般而言某个策略的最大回撤率越大，意味着该策略的风险较高。但并非最大回撤率越小越好，往往高收益的产品会伴随着高风险匹配，而风险偏好的投资者会对这些产品青睐有加。

最大回撤率的计算公式可以用如下公式描述：设D为某一天的净值，i为某一天，j为从i开始算起的之后一天，Di为第i天的策略净值，而Dj为第j天的策略净值，则公式如下：

Drawdown = max (Di-Dj)/Di

即对每一天的净值进行回撤率计算，然后求其最大值。在高频交易中，交易数据可能是按照“时、分、秒、每笔交易”来计算，则此处的回撤率也可以使用“时、分、秒、每笔交易”来计算。

### 夏普比率

夏普比率（Sharpe Ratio），又称为夏普指数。在投资活动中有一个显著的特点：投资标的的预期报酬率越高，投资人所能承受的波动风险越高；反之预期报酬率越低，投资人所能承受的波动风险越低。因此说理性的投资人选择投资标的与投资组合的主要目的是：在可以承受的风险范围内，追求最大的收益；或者在固定的收益内追求最低的风险。

夏普理论告诉我们，需要用最小的风险来换取最大的回报，因此投资者要避免一些不值得冒的风险，同时如果投资者如果缺乏投资经验与研究时间，可以让专业人士来帮助自己建立一些合适的投资组合，这些投资组合可以通过夏普比率来衡量风险和回报的比率。

夏普比率的计算公式如下：

SharpeRatio =( E(Rp)-Rf)/ σp

其中：E(Rp)为投资组合预期报酬率

Rf为无风险利率，一般取一年期国债利率

σp为投资组合的标准差

夏普比率的计算尽管非常简单，但是在具体运用中需要注意其适用性：夏普比率没有基准点，因此其本身大小没有意义，只有在与其他组合的比较中才有意义；使用标准差作为风险指标被人们认为是不是特别合适的；夏普比率是线性的，但是风险与收益之间的变换并非线性，因此夏普比率在衡量标准差较大的产品时会存在偏误；夏普比率同其他很多指标一样，衡量的是基金的历史表现，而历史表现并不能说明未来会与过去完全相同；此外夏普指数存在一个稳定性问题：夏普指数的计算结果与时间跨度和收益计算的时间间隔的选取有关。

尽管夏普比率存在如此多的限制和问题，但是由于它在计算上的简便和没有过多的假设条件而在实践中获得了较大范围的使用。

### 詹森指数

詹森指数是测定证券组合经营绩效的一种指标，是证券组合的实际期望收益率与位于证券市场线上的证券组合的期望收益率之差，它通过比较考察期基金收益率与由资本资产定价模型CAPM得出的预期收益率之差，即基金所获得的实际收益超出它承受风险对应的预期收益的部分来评价基金，此差额就是基金经理的工作所带来的额外收益。因此当詹森指数大于零，表明基金业绩表现优于市场基准组合，詹森指数越大代表基金业绩越好，如果詹森指数小于零，则说明基金的表现低于市场基准组合，绩效较差。

詹森指数的计算公式如下：

詹森指数=Ri,t—[Rf,t+βi(Rm,t-Rft)]

其中Rm,t为市场[投资组合](http://baike.baidu.com/view/194522.htm" \t "_blank)在t时期的[收益率](http://baike.baidu.com/view/969420.htm" \t "_blank)；Ri,t为i基金在t时期的收益率；Rf,t为t时期的[无风险收益率](http://baike.baidu.com/view/2258897.htm" \t "_blank)，βi为基金投资组合所承担的[系统风险](http://baike.baidu.com/view/165085.htm)。

### 市场容量

在高频交易的实际操作中会遇到一个常见的问题：某个策略在资金量小时收益稳定运行正常，但资金量增加时就会发生亏损。这是因为策略可容纳的资金量是有限的，当资金量较小时，策略对市场的影响可以忽略不计，策略可以正常运行，如果资金量增大时，策略本身就会对市场造成扭曲，导致策略运行的市场环境产生一定的变化，策略就会不可避免的产生亏损。市场容量是一个偏实际应用的问题，一般策略的最大资金容量很难说出准确的数字是多少，但毫无疑问的，市场容量是客观存在的，股指期货高频交易策略通常容量几千万，资金再多时收益率就会开始下降。

因此对交易策略的又一评估标准为该交易策略能容纳多少资金。

影响策略可容纳资金容量的因素有如下几点：策略本身的逻辑、策略交易频率、策略交易品种的日内总成交量和总持仓量、策略交易品种的瞬时挂单量、资金的风险偏好等。

策略本身的逻辑对容量的影响是巨大的，事件驱动模型和量化对冲策略可容纳的资金容量在10亿以上，而期现套利模型的资金容量可达30亿，统计套利交易模型的容量在1亿以上，单一证券的量化交易策略依投资标的而定[16]。

策略交易频率对容量的影响也比较大，一般而言交易频率越高，市场上可供交易的交易对手就越少，能成交的资金也越少，因此短线交易不能容纳很大的资金容量。

策略交易品种的日内总成交量和总持仓量对资金容量的影响是正向的，交易品种的日内总成交量和总持仓量越大，则能容纳的资金容量越大。

由于市场容量限制的存在，高频交易中必须针对每一个策略设置资金限制，避免因市场容量不足而导致的策略亏损。

## 共性影响因素

## 市场大小对交易容量的影响

市场大小是指市场中某商品的成交额与成交量的大小，衡量商品成交量大小的指标有很多，例如：单日总成交额、单日总成交量、年总成交额、年总成交量。一般而言，单日总成交额和单日总成交笔数越大，称之为市场比较大。

很显然市场大小是影响交易容量大小的决定性因素：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 今开盘 | 收盘价 | 成交量 | 持仓手 |
| 沪铜1703 2017-01-26 | 47460.00 | 48040.00 | 188630 | 159580 |
| 沪铝1703 2017-01-26 | 13965.00 | 13835.00 | 280242 | 233714 |
| 沪锌1703 2017-01-26 | 22710.00 | 23205.00 | 377530 | 159562 |
| 沪铅1703 2017-01-26 | 18750.00 | 19015.00 | 63892 | 34952 |
| 沪金1706 2017-01-26 | 269.60 | 267.60 | 142822 | 318662 |
| 沪银1706 2017-01-26 | 4073.00 | 4025.00 | 396474 | 641770 |
| 沪螺纹钢1705 2017-01-26 | 3292.00 | 3369.00 | 1896002 | 2254872 |
| 郑强麦1705 2017-01-26 | 2994.00 | 3006.00 | 2098 | -- |
| 郑棉花1705 2017-01-26 | 15620.00 | 15665.00 | 95640 | -- |
| 郑玻璃1705 2017-01-26 | 1315.00 | 1323.00 | 192204 | -- |

表X：上期所与郑期所部分商品主力合约交易情况，数据来源《期货日报》

从成交量分析，商品的成交量排名前三的商品如下：

沪螺纹钢(1896002手)>沪银(396474手)>沪锌(377530手)

成交量代表着市场上的交易活跃程度，成交量大则意味着两点：一是买方和卖方数量众多；二是此时的商品价格与双方的心理预期一致。由于这两个因素，投资者在一个成交量大的商品中寻找交易对手方显然要容易的多，并且在成交量大的商品中交易意味着投资者面对的是一个“大致达成一致的价格”。

成交额也是不得不考虑的一个方面。由于上期所和郑期所并未正式公布单日成交额，因此我们采用如下公式来计算成交额：

成交额=交易单位\*合约报价单位\*成交量\*收盘价\*最低保证金比例

对上表进行调整后计算结果如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 交易单位 | 合约报价单位 | 成交量 | 收盘价 | 最低保证金比例 | 成交额(万元) |
| 沪铜1703 2017-01-26 | 5吨/手 | 元(人民币)/吨 | 188630 | 48040.00 | 11% | 498,391 |
| 沪铝1703 2017-01-26 | 5吨/手 | 元(人民币)/吨 | 280242 | 13835.00 | 11% | 213,243 |
| 沪锌1703 2017-01-26 | 5吨/手 | 元(人民币)/吨 | 377530 | 23205.00 | 11% | 481,832 |
| 沪铅1703 2017-01-26 | 25吨/手 | 元(人民币)/吨 | 63892 | 19015.00 | 11% | 334,099 |
| 沪金1706 2017-01-26 | 1000克/手 | 元(人民币)/克 | 142822 | 267.60 | 10% | 382,191 |
| 沪银1706 2017-01-26 | 15千克/手 | 元(人民币)/千克 | 396474 | 4025.00 | 11% | 263,308 |
| 沪螺纹钢1705 2017-01-26 | 10吨/手 | 元(人民币/吨) | 1896002 | 3369.00 | 11% | 702,639 |
| 郑麦1705 2017-01-26 | 20吨/手 | 元(人民币/吨) | 2,098 | 3006.00 | 10% | 1261 |
| 郑棉1705 2017-01-26 | 5吨/手 | 元(人民币/吨) | 95,640 | 15665.00 | 10% | 74,910 |
| 郑玻璃1705 2017-01-26 | 20吨/手 | 元(人民币/吨) | 192,204 | 1323.00 | 10% | 50,857 |

表X：上期所与郑期所部分商品主力合约成交额，数据来源作者计算

从成交额分析，商品的成交量排名前三的商品如下：

沪螺纹钢（702639万元）>沪铜(498391万元)>沪锌(481832万元)

成交额可以直观的体现出该商品可以容纳的资金容量。如果我们假设对某个商品投入的资金量对大盘的扰动是下图的假设关系：



图X:对单个商品投入的资金占该商品当日成交额的比率与市场扰动关系

我们简单假设存在一个临界点P，如果单个商品投入的资金占该商品当日成交额的比率在此比率之下，则市场不会收到扰动，则既定的交易策略可以不受阻碍的实现；如果单个商品投入的资金占该商品当日成交额的比率超过了P点则市场会受到扰动，使原本符合既定交易策略的市场环境不复存在，进而造成亏损。

关于此P点的值是一个偏实践性的数值，并没有精确的方法可以去计算出一个标准的值供大家参考。并且对此P点的见解有很大的争论，部分投资者的投资策略是力争对市场无扰动，使得投资策略可以在市场稳定的情况下获得应有收益；而另一部分投资者会考虑如何突破P点，对市场进行充分的扰动，然后在扰动中进行获利。这也可以从另一方面解释在中国股票市场或者其他市场尚未发展完全的时代，为什么价值投资性的股票偏向与大盘股，而游资性的资金喜欢小盘股。

以上的数据来源于期货市场，但是得出的结论也适用于其他的市场，如股票市场、外汇市场、期权市场等。

## 交易成本对于高频交易的影响

交易成本（Transaction Cost）又称为交易费用，是指在完成一笔交易时，交易双方在买卖前后所产生的各种与此交易相关的成本。此成本并非仅仅指与货币有关的成本，而且也包含了时间成本、机会成本等。在1975年Williamson指出交易成本可以区分为如下几项内容：搜寻成本、信息成本、议价成本、决策成本、监督成本。1985年Williamson进一步将交易成本加以整理区分为事前时候两个大类，事前交易成本含签约、谈判、保障契约等成本，事后成本含契约不能适用而导致的成本。

高频交易的交易成本种类比较多，我们可以总结为如下几类：准入资格成本、交易费用成本、软硬件设备成本、附属运营成本。

准入资格成本是指获得高频交易准入资格的成本。高频交易有多种形式，初级的可以采用第三方系统的方式，例如交易开拓者、金字塔、文华赢顺等软件，较高级的方式可以直连上期所CTP接口，当前最新的模式是采用优矿等互联网量化交易平台。仅从费用方面考量：

|  |  |
| --- | --- |
|  | 准入费用 |
| 交易开拓者 | 期货账户入金10万，交易费用按交易所费用的25%比例收费 |
| 文华盈智 | 按年收费，1800元/年 |
| 金字塔决策交易系统 | 按年与模块收费，标准版2800元/年，机构版28000元/年 |
| 上期所CTP | 不同期货公司的价格不同，笔者所在的期货公司要求账户入金100万 |

表X：不同接入方式的准入费用，数据来源相关公司询价

在费用方面，不同的接入方式费用差距很大，这是在进入高频交易领域时不得不考虑的一点。可以从投资者的多个方面来综合考量：初始资金是否充足、收费方式倾向于按年收费或者按交易收费、是否有独立模块需求、是否有自行研发交易系统的能力、学习并掌握此系统所需的时间成本等。

交易费用成本是指每进行一次交易所需要付出的成本，此成本与多个因素相关：商品种类、期货公司或证券公司优惠、交易系统成本等方面构成。

我们假设所有投资者的议价能力是一致的，可以从期货公司以及软件提供商处得到一致的最低的优惠（实际操作中上一般是交易费用的25%左右，即实际付出成本=期货公司收取佣金+软件供应商佣金+交易费用=1.5\*交易费用，此处的参数1.5会随着期货公司不同而不同），那么交易成本主要体现在交易费用上：

|  |  |
| --- | --- |
|  | 交易所交易手续费 |
| 沪铜1703 2017-01-26 | 0.5%% |
| 沪铝1703 2017-01-26 | 3元/手 |
| 沪锌1703 2017-01-26 | 3元/手 |
| 沪铅1703 2017-01-26 | 0.4%%，平今仓免费 |
| 沪金1706 2017-01-26 | 10元/手 |
| 沪银1706 2017-01-26 | 0.5%% |
| 沪螺纹钢1705 2017-01-26 | 1%% |
| 郑强麦1705 2017-01-26 | 2.5元/手 |
| 郑棉花1705 2017-01-26 | 6元/手 |
| 郑玻璃1705 2017-01-26 | 3元/手 |

表X：不同商品的交易所交易手续费，此手续费会由各交易所随时调整

可以看到不同种类的商品的交易所交易手续费是不同的，我们可以根据投资者的实际计划投资金额来计算，如果可投资金额为40万，计划投资25%到市场中，即购买10万元保证金的商品，则需要付出的手续费如下：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 交易所交易手续费 | 收盘价 | 交易单位 | 保证金 | 每手金额 | 可购买手数 | 手续费 |
| 沪铜1703 2017-01-26 | 0.5%% | 48040 | 5 | 11% | 26422 | 3 | 36.03 |
| 沪铝1703 2017-01-26 | 3元/手 | 13835 | 5 | 11% | 7609.25 | 13 | 39 |
| 沪锌1703 2017-01-26 | 3元/手 | 23205 | 5 | 11% | 12762.75 | 7 | 21 |
| 沪铅1703 2017-01-26 | 0.4%% | 19015 | 25 | 11% | 52291.25 | 1 | 19.02 |
| 沪金1706 2017-01-26 | 10元/手 | 267.6 | 1000 | 10% | 26760 | 3 | 30 |
| 沪银1706 2017-01-26 | 0.5%% | 4025 | 15 | 11% | 6641.25 | 15 | 45.28 |
| 沪螺纹钢1705 2017-01-26 | 1%% | 3369 | 10 | 11% | 3705.9 | 26 | 87.59 |
| 郑强麦1705 2017-01-26 | 2.5元/手 | 3006 | 20 | 10% | 6012 | 16 | 40 |
| 郑棉花1705 2017-01-26 | 6元/手 | 15665 | 5 | 10% | 7832.5 | 12 | 72 |
| 郑玻璃1705 2017-01-26 | 3元/手 | 1323 | 20 | 10% | 2646 | 37 | 111 |

表X：投资金额为100000元，最终付出的手续费

从表中可以看到，使用同等量的资金购买不同商品的手续费千差万别，甚至会有5倍的差距（郑玻璃与沪铅），在高频交易中，交易费用由于高频率的交易单而成为一个天文数字。因此投资者在做交易时必须考虑手续费对交易利润的影响，目前业内对手续费有如下处理方式：a)认可交易手续费为必须成本；b)投资者通过与期货公司商议返佣等形式减免交易手续费；

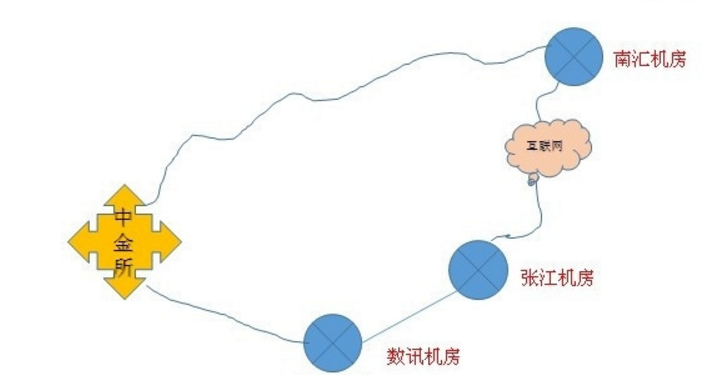
## 交易延时对于高频交易的影响

国内期货市场目前的数据传送情况如下：

|  |  |  |
| --- | --- | --- |
|  | 系统名称 | 行情数据发布频率 |
| 上海期货交易所 | 上期所CTP | 2次/秒 |
| 郑州商品交易所 | 郑州易盛 | 2次/秒 |
| 大连商品交易所 | 大连飞创 | 基本行情2次/秒 Level-2行情4次/秒 延时行情2次/秒 |
| 中国金融期货交易所 | 中金所飞马系统 | 4次/秒 |

表X：国内期货市场行情数据发布频率，数据来自各交易所官网网址

上期所CTP接口由于启动时间较早、功能完善，且同时支持四大期货交易所，所以运用范围比较广泛。中金所的飞马系统由中金所技术公司研发，仅能支持中金所的股指期货与国债期货，但由于其系统部署在数讯机房，物理位置较优（上期所在张江机房，中金所在数讯机房，物理上数讯机房更加靠近中金所），目前各期货公司做实地测试单程耗时仅2-3ms，在速度上有较大优势。



图X：上海各机房物理部署，来源于中金所飞马系统官网

对于高频交易来说，无论是做市或者套利，速度都是十分重要的因素。如果使用做市商策略，则交易目标是在尽可能短的时间内完成一个回合的交易，这样可以使市场风险降到最低，否则价格有可能会朝着不利于做市商的方向发展。投资者尽可能的完成一个回合的交易，这自然带来的要求就是提高交易的速度。此外，作为高频交易者，与其他高频交易者也存在着竞争关系：能够比对方更快发出指令的交易商，哪怕只快一毫秒，都可以抢先把订单发布出去或者成交。当使用套利策略时，速度同样十分重要，尤其是在高频交易者竞争激烈的市场上，套利机会将在几百毫秒内出现后就被其他高频交易者拿下，在这种情形下，第一时间发现这些机会，就意味着能够利用这个机会获得利润。

因此高频交易的发展目标始终是“要力争比竞争对手更快”，使信息滞后度达到最低。其中最重要的策略就是“主机托管”，此处我们展示某高频交易软件厂商的主机托管说明。

图X：某高频交易软件厂商的主机托管说明

目前国内高频交易软件厂商众多，但毫无疑问的是每家公司都尽可能提供了“主机托管服务”，使高频交易者的计算机与交易所相应的计算机尽可能的靠近，这样可以减少电子信号进出交易所所耗费的时间。由于电子信号以光速传播，主机托管的优势对于普通交易者来说并不显著，但是信号沿光纤传播的速度为每公里5.5微妙，高频交易者的服务器与交易所的距离每减少100公里，单向的信号传输时间就会减少550微妙，双向速度将减少1ms，对于高频交易来说，1ms已经足够甩开竞争对手。尤其我国国土面积辽阔大纵深的情况，如果有一名高频交易者在西安开展他的股指期货高频交易事业，就必须面对西安到上海的距离问题。西安到上海的直线距离是1200公里，意味着这位西安的高频交易者每笔交易都会有24ms的交易延时，这在高频交易中将会带来惨重的损失。如果该高频交易者租用了托管机房，则可以极快的获取最新的数据信息，比其他的交易者更快的下单与成交。

## 个性影响因素

## 影响做市商策略的因素分析

在上一章节描述到，影响做市商策略的因素如下：商品价格的波动性特征、合约的交易活动性。

商品价格的波动性特征有很多方面来描述，据相关研究，我国商品期货收益普遍呈现出较为显著的左偏和尖峰胖尾特征；其价格波动具有明显的聚集性和长记忆性特征,但与证券市场不同的是,其中大多数品种的价格波动不存在非对称杠杆效应(Asymmetric leverage effect)。国内外大豆、豆粕、糖、铜和橡胶期货间均存在双向的波动溢出效应,且在各农产品期货品种中,国外产品对国内产品的波动溢出效应更强；锌期货间则仅存在国内对国外的单向波动溢出效应。整体而言,我国主要商品期货与国外相似期货品种之间存在密切联系,且在2008年金融危机后,国内期货对国外相关品种的波动溢出强度和都有不同程度的增强[17]。

商品价格的波动性特征主要体现在如下方面：偏向性、聚集性、长记忆性、跳跃性、周期性。

偏向性是指商品价格总体上是否有不断增长的历史趋势。长期而言，部分商品总体上有不断增长的历史趋势，其原因可以归结为资源的不可再生和通货膨胀的影响；

聚集性是指商品的波动是否密集，相关实证分析表明中国期货市场的价格波动具有明显的聚集性；

时间序列的长记忆性最早是由水文学家赫斯特(Hurst)于1951年提出来的。许多时间序列呈现出相距较远的观测值之间仍然存在不可忽略的相关性的特征,一般将这种特征称为时间序列的长记忆性或长期相关性。据相关实证结果表明中国股市波动性过程具有明显的长期记忆特征[18]，而中国期货市场的商品价格波动长期记忆性也十分明显[17]。

跳跃性是指商品价格运动具有较短的交易时间完成较大的运动距离的特征，其特点是运动具有突发性和急速性。

周期性是指商品价格的运行具有相对稳定的重复性的特征，该特征的周期具有变异性，受市场情绪影响。它并不是简单的重复，而是会收到投资者心理和情绪的影响而变动。

交易活跃性有多种衡量标准，例如：成交金额、换手率、成交量。在期货市场中多用成交金额与成交量来衡量商品的交易活跃性。

成交金额是当日每笔成交订单的交易金额的总和；

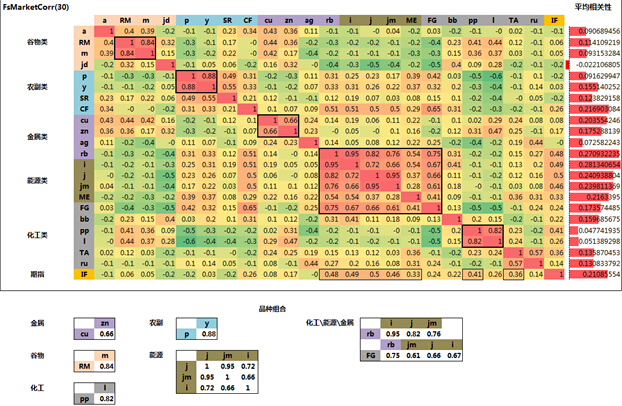
成交量是当日成交订单的总手数；

在期货市场上做市商策略受波动性特征和交易活跃性的影响主要体现在如下方面：

由于偏向性的周期较长（因通货膨胀的影响，往往会与经济周期相同），因此在期货市场上一般可以近似认为商品短期内无偏向性；商品价格波动的聚集性比较明显，意味着做市商策略必须主动适应较剧烈的长期的波动；长记忆性意味着做市商策略必须从历史数据中分析当前市场的变动（这也是技术流派的用武之地）；跳跃性意味着做市商策略必须谨慎处理价格的突发变动；周期性意味着做市商策略必须对商品的运行周期进行掌握与分析，并熟知国内外商品的产量与环境变化等影响。

## 影响统计套利策略的因素分析

在上一章节描述到，影响统计套利策略的主要因素如下：商品间价格的相关性、过度拟合现象、套利周期的影响。

商品期货之间的相关性研究用相关系数表示，是用以反映变量之间相关关系密切程度的统计指标。相关系数是按积差方法计算，同样以两变量与各自平均值的离差为基础，通过两个离差相乘来反映两变量之间相关程度；着重研究线性的单相关系数。

图X：各个期货商品之间的相关性，图片来源和讯金融实验室

相关系数的取值范围为[-1,+1]之间，当 取值为1 时，代表两组变量是完全正相关的；当 取值为-1 时，代表两组变量是完全负相关的；r越靠近0轴，两组变量间相关性越弱。一般来说，取值的绝对值在0.66以上属于高度相关。取值的绝对值介于0.33至0.66间属于弱相关。取值的绝对值介于0至0.33间属于相关性不显著。

图中可以看到，高度相关的商品有如下组合：

焦炭J与焦煤JM的相关性系数为0.95

铁矿石i与螺纹钢的相关性系数为0.95

棕榈P和豆油Y的相关性系数为0.88

菜粕RM和豆粕M的相关系数为0.84

聚丙烯PP和塑料L的相关系数为0.82

铜CU和锌ZN的相关系数为0.66

套利的基本形式有两种：价差套利和比价套利。所谓价差套利是做出商品期货品种间同一月份的价格之间的价差，并且画出价差的时间序列图，分析价差，寻找合理的价差范围，超出合理的价差变动范围时则卖出价格较高的商品并买入价格较低的商品，待日后价格恢复正常水平后获利了结。而比价套利与差价套利基本类似，通过根据历史数据计算出相关的商品期货品种间同一月份的价格之间的比值即比价，并画出比价的时间序列图，分析比价的规律，找出合理的范围，当比价超过合理的比价范围时, 则卖出价格较高的商品并买入价格较低的商品，待日后价格恢复正常水平后获利了结。

选择一个相关系数较高的产品组合开展统计套利交易，毫无疑问商品的价格会更快的恢复正常水平，使得价格向不利方向发展的风险降到最低。与此同时，我们还需要注意到另一个方面，高相关性的商品价格出现不一致的概率较低，并且不一致的幅度较小，该情况会导致风险最低的时候利润也会最低，这也是在做商品程序化时不得不考虑的内容。

过度拟合是指为了得到一致假设而使假设变得过度严格的情况，即对于样本数据，描述的准确度很高，而对于样本外数据，描述的准确度却很差。在期货市场中重点表现为如下形式：高频交易者往往会设计一个交易时机筛选条件，在符合这个筛选条件的时机下单会带来较大利润。这个筛选条件的表现形式是一个N维的函数，返回值为该时机所对应的操作如买入开仓、卖出开仓、平仓等。为了使这个筛选条件更加符合历史数据，高频交易者往往会通过计算机程序对此筛选条件的N个参数进行优化，使这个筛选条件在面对历史数据时可以表现到最好，可以把每一个交易时机都筛选出来。如果高频交易者在最后的优化步骤时优化次数过于严格，就会导致筛选条件对历史数据是高度拟合的，但是对于未来的数据却无法捕捉到交易时机。

出现过度拟合的原因大致有如下原因：建模样本抽取错误、样本里的噪音数据干扰过大、建模时的“逻辑假设”到了模型应用时已经不能成立、参数太多模型复杂度高等。而解决过度拟合的原因也有如下几种方式：

1. 清洗噪音数据，通过增加时间周期将数据平稳化；
2. 避免使用过老的数据
3. 使用简单的模型；
4. 将数据分为训练数据集与测试数据集，使用在测试数据集上产生最小误差的优化参数；

## 影响动量策略与反转策略的因素分析

在上一章节描述到，影响做动量策略与反转策略的因素如下：知情交易者的比例、信息传播速度、信息吸收程度。

动量交易策略是在行为金融理论背景下提出来的，指如果某一个股票在过去一段时间内收益率较高，那么将来的一段时间内其期望收益率也较高；相反，如果某一个股票的收益率较低，那么将来相同时期内它的期望收益率也会处于一个较低的水平。因此基于动量效应，投资者就可以买进收益率较高的股票，卖出收益率最低的股票，由此构建一个零风险投资组合进行套利，这便是动量交易策略。

De Long,Shleifer,Summers和Waldmann(1990)[20]提出了DSSW模型，该模型利用正反馈交易行为说明了反应过度现象。市场上存在这样一类投资者，他们仅关注技术面，倾向于买入近期上涨的股票，在市场上表现为“追涨杀跌”。正反馈交易者的“追涨杀跌”行为将进一步刺激股价上涨(或者下跌)，该股票短期表现为过度反应，然而当股价被一再推高到某个极限值(或者下跌到某个极限值)之后，股价的泡沫必然破灭(或者超跌反弹)，从而开始反转。此外Berberis,Shleifer和Vishny(1998)[21]认为投资者有存在两种普遍的决策偏差，一是代表性偏差(representative bias)，投资者会更关注近期数据的变化，他们会有一种规律情结，喜欢根据市场的近期的运动态势总结出某种规律或者模式，当这种模式偶然出现时，投资者容易忽略偶然性，高估自己总结的模式的普遍性，并将其应用到未来的投资中去，从而导致过度反应。比如前一个季度的盈利股票，在本季度继续保持盈利，股价不断上涨，投资者认为高盈利的规律能继续保持，从而继续买入该股票，导致股价被高估，股价表现出过度反应。当这种高盈利趋势不能保持时，股价就会表现出反转。另一种是保守性偏差（conservatism），即投资者未能根据变化了的情况及时修正自己的预测，保守性偏差会导致反应不足（underreaction）。比如公司公布了意外的盈余信息，而投资者对新信息的反应速度较慢，新信息逐渐反应到股价中去，在市场上从而表现出动量现象。Deniel，Hirshleifer和Subrahmanyam(1998) [22]提出了DHS模型，该模型认为投资者在决策时存在两种偏差：一种是过度自信, 指投资者高估自己的能力，高估拥有的私有信息价值；另一种是有偏的自我归因，指如果公开信息证实了投资者的私有信息，将极大的增强投资者的自信心；但如果公开信息和投资者私有信息不一致，投资者信心会降低一定幅度，但降低的幅度与同等条件的信心增强的幅度要小，即投资者会下意识的抑制不利于评价自己能力的信息。在过度自信和有偏的自我归因共同作用下，股票价格在短期内存在动量，而长期则表现为反转。

通常情况下市场上有影响商品价格的信息出现时，知情交易者会在价格未完全反映该商品信息前进行交易，随着交易的逐渐传播、越来越多的非知情交易者成为了知情交易者，使得证券价格延续之前的走势，并逐渐反映价值，随着价格回归价值，知情交易者退出市场，知情交易概率达到最低水平，由于通常价格变动对于信息有过度反应现象，从而价格逐渐转变原有趋势而出现逆转。

参考文献

参考文献内容

致谢

致谢词

附录

附录内容