

方法论

- **量化投资的哲学基础0：不懂哲学的程序员不是好投资经理**
- <https://zhuanlan.zhihu.com/p/37333831>
- 投资有三个境界：**交易，策略和投资哲学**。
 - 交易就是在什么时间买卖哪只股票。
 - 好交易能够帮你赚钱。
 - 策略就是一套交易的逻辑和规则。简单讲，就是怎么在几千只股票中选到茅台和乐视。
 - 好策略指导你做赚钱的交易。
 - 采用价值投资策略的老师，关心市盈率市净率，关心公司盈利有没有护城河；
 - 采用技术分析策略的老师，关心技术指标发出了什么信号，关心技术形态是否好看；
 - 采用机器学习策略的老师，关心算法是否高逼格，特征是否好又多。
 - 投资哲学就是在回答什么是好策略。有了自己的投资哲学，才知道自己应该采用什么策略。
 - 正确的投资哲学帮你找到好策略。
- 知识可分为两类：
 - **逻辑体系型的**
 - 需要思考，把各种概念连起来，或归纳，或演绎，最后，在大脑中形成一种**逻辑网状结构**。比如可计算理论，分布式系统，解释器等。
 - 基本概念和之间的联系
 - 多因子
 - **使用操作型的**
 - 基本不需要思考，只需要看一下目录，要用的时候参考一下手册即可。比如linux的各种命令，go,python的一些奇怪语法,elastic-search dsl的使用。
 - **比如各种常用策略和使用场景，需要先记录下来**
-
- **量化投资的哲学基础1：量化投资不是“量化”投资**
- <https://zhuanlan.zhihu.com/p/29375085>
- 量化投资不是“量化”投资
- 量化投资，字面上解释，就是用数量化的方法进行投资。关键词在“量化”。
- 在外人看来，量化投资就是一群具有数理教育背景的人，用电脑程序分析大量的数据，建立数学模型，并根据数据进行投资。量化投资人还特别喜欢在展示业绩的时候，用到大量的风险收益指标：Sharpe Ratio和Max Draw Down Duration等等。
- 然而，我只需要以的三个投资策略为例，就能证明量化投资的本质并非“量化”：
- 例1：数量化的量化策略：买入过去一个月涨跌幅最小的20只股票并持有一个月。
- 例2：非数量化的量化策略：当公司发出回购公司股票的公告时，买入并持有这只股票2天。
- 例3：数量化的非量化策略：判断螺纹钢期货走出第3浪时，买入螺纹钢期货。
- 通过上面三个例子，我们能够发现，量化策略似乎不必然用到数量，比如例2。而一些用到数量的策略通常不被认为是量化投资，比如例3。

-
- 量化投资的模型做出的解释和预测是**可以检验的**。
-
- **科学哲学**是关于科学的基础，方法和含义哲学分支。这门学问的核心问题包括科学的标准，科学理论的可靠性和科学的终极目的。
-
- 量化投资同样依赖这些**哲学原则**：
 - 本体论：世界不依赖心智。量化投资不影响市场。历史测试就是假设你在历史上交易并统计交易结果。如果量化投资会影响市场，那么历史测试就是不准确的。这也就是量化投资容易受到资金规模限制的原因。因为资金规模一大，就不可避免的影响市场，导致历史测试和实际交易存在差异。
 - 本体论：不存在超自然的世界。当然不是神在操纵市场。
 - 本体论：世界按照**规律**运行。市场按照规律运行。如果假设市场的运行都是巧合。那么经过历史测试的模型将不能外推到未来。
 - 本体论：原因优先于结果。显而易见，量化投资模型假设模型的条件先发生，市场的波动后发生。否则，市场运行优先于条件，那么再依据条件去交易就迟了。
 - 本体论：无不能生有。市场的运行都是有原因的。不会凭空出现某个市场现象或者影响市场的因素。这个市场现象和影响因素也不会凭空消失。否则，历史测试又不能外推到未来了。
 - 本体论：精神不能直接作用于物质。量化投资者的想法不会直接影响市场。这个是显而易见的。
 - 认识论：可观察之物可知。可以通过观察市场获得知识。
 - 认识论：不可观察之物可知。通过历史测试的模型可以，不是必然可以，外推到未来。
 - 方法论：奥卡姆剃刀。量化模型追求简洁。**如无必要，勿增实体**。
 - 方法论：可错论。量化模型是可以是错误的。因此量化模型才是可以比较，修正和否定的。
-
- **量化投资的哲学基础2：笛卡尔说，股票多因子模型是个好框架**
- <https://zhuanlan.zhihu.com/p/30132661>
- **提出问题**
- 首先定义交易的对象。如果对象是螺纹钢，通常这是一个CTA策略。如果对象是黄金和白银期货的价差，那么这是一个套利策略。如果是对象是股票和股指期货的价差，那么这是一个Alpha策略。如果是上证50ETF的波动率，那么这个一个期权策略。
- 其次定义交易的时间。典型的高频策略的持仓时间很短，甚至短于一秒。典型的日内CTA策略和股票T+0策略的持仓时间在几分钟到几个小时。典型的股票Alpha策略的持仓时间为几周到一个月。
- 怎样去构建模型？怎样去寻找模型输入？怎样去构建模型输入？应该采取什么样的研究顺序？应该注意哪些事项？这些问题才是难点。
- **提出解决框架**
- 笛卡尔和《谈谈方法论》
- 获得知识的方法四条箴言广为流传：
 - 绝不接受任何我没有确认为真的事物。
 - 把每一个难题拆分为尽量多的部分，直到可能充分解决难题为止。
 - 执行思考时，从最简单和最容易了解的对象开始，然后一点点上升到更复杂的知识，即便不是按照对象的时间先后顺序。
 - 列举要完整，检验要普遍，直到我确信没有遗漏。

- 股票多因子模型-其实目前的评级模型也是多因子模型
-
- **量化投资的哲学基础3：休谟说，量化投资模型都是过度优化**
- <https://zhuanlan.zhihu.com/p/30543252>
- 知识从哪里来？最传统的方法就是**演绎（Deduction）和归纳（Induction）**
- **演绎推理**，从陈述（前提）到逻辑上确定的结论的推理过程。
- **套利策略是最典型的通过演绎得到的量化策略**：关于衍生品的期现套利和跨期套利、从Put Call Parity推导出来的看涨期权和看跌期权套利、ETF套利和分级基金套利等等。
- 这类策略的特点有几点：1. 策略都是教科书式的，因为众所周知。2. 策略的收益在建仓时就已经锁定了，也就是演绎推理的“**确定**”的结论。3. 因此，这类策略的获利能力取决于市场犯了多少的错误。随着市场更加理性，套利策略逐渐演变成IT的军备竞赛，策略收益很难持续。
- **归纳**是从特例到总体的不受限制的概括推理。归纳推理中，前提并不保证结论。
- 在量化投资领域，绝大多数的模型都是通过归纳得出的。
- 技术分析相关的因子完全依赖于归纳，而基本面分析中归纳也是最关键步骤。
- **FF模型**发现小市值股票的预期收益率高于大市值的股票，低估值的股票的预期收益率高于高估值的股票，这两个因素无法用市场因子来解释。因此股票价格受到市场，市值和估值三个因素的影响。FF模型没有为后续因子的发现给出任何限定。FF模型纯粹是“归纳”出的模型。
- 对演绎的依赖最小和对归纳依赖最大的FF模型无论在学术界还是业界都获得了成功，成为量化投资和股票多因子模型最重要的源头。
- 归纳方法存在一个致命问题：过度优化
- 休谟问题。哪怕数据足够充分，哪怕模型很好地适应了全部数据，归纳出的知识是无法保证可以外推的。
-
- **量化投资的哲学基础4：科学哲学家告诉你怎样避免过度优化**
- <https://zhuanlan.zhihu.com/p/30940020>
- 在量化投资领域中，任何人都想追求**高收益，低风险和大资金容量**的策略。但是任何一个策略只可能满足其中两条：高频策略通常资金容量小；Alpha策略通常收益率低；CTA策略通常风险大。在给定的资金容量下，一个“聪明”的策略可能可以在一个“愚蠢”的市场中获得较高的风险收益比。但是，总会有力量能够把一个策略的风险收益比拉回平均水平。交易的“圣杯”是不存在的。
- 孔德的思想启示我们，**放弃追求过高的风险收益比，就能够帮助我们避免过度优化。**
- 波普尔的思想启示我们，多多开发策略，淘汰掉不赚钱的，使用还能赚钱的，然后继续多多开发策略。
-
- 那么Abduction直接引出了下一个问题，既然多个猜想都能解释现象，那么什么样的解释是最有可能的？或者说，什么样的理论最有可能是正确的？
- 一个好理论通常满足以下四个条件：
 - 有**内在机制**支撑的理论是更好理论。
 - 在**很多领域**都有效的理论是更好的理论。
 - 能够**定量**解释的理论比定性解释的理论更好。技术指标是典型的定性分析理论。通常给出多空两个状态，或者多、空和空仓三个状态。而多因子alpha模型是典型的定量分析理论，能够计算出来每只股票的alpha值。尽管很难有因子20个buckets的状态下仍然保持线性关系，但是仍然比技术指标的3个状态进步很多了。

- 一个**简单**的理论比复杂的理论更好。以**多元线性回归**为基础的多因子模型其实就是一个非常简单的模型。举几个例子，首先，线性回归模型假设自变量和因变量之间是**线性**关系。线性关系是拟合问题中最简单的模型的，稍微复杂一点的模型有抛物线和三角函数等。其次，一个传统的多因子模型会通过多种办法将成百上千的因子缩减到几个因子，即假设几个因子可以解释因变量。最后，线性回归模型假设**因子之间没有相关性**。否则，一旦要考虑因子间的相关性，模型的复杂度就会朝着几何倍数发展。
-
- **量化投资的哲学基础5：根据Abduction理论挑选机器学习模型**
- <https://zhuanlan.zhihu.com/p/37867708>
- 本文会比较以下三个最最简单的算法：**KNN, ID3, OLS**。
- 选用以上三个算法，主要是因为简单的算法更方便读者看清楚本质。
- KNN, K Nearest Neighbors, K近邻法是最简单的机器学习算法。
- 决策树是和KNN并列的最简单的机器学习算法。ID3又是决策树中最简单的一种。
- OLS，最小二乘法，是被研究地最透彻的统计方法。OLS是最传统的统计方法，通常被视作机器学习算法的对立面。
-
- 我推荐一个相对完美的算法——**GBDT**（Gradient Boosting Decision Tree）。
- 一个极简的GBDT的算法的思路如下：首先选出对标签最有解释力的特征，做回归，然后求得残差。然而选出对残差最有解释力的特征，做回归，然后求得残差。然后继续迭代。具体算法，读者可以自行百度。