

知乎



首发于
川流不息



科学回测中的大学问



石川

量化交易 话题的优秀回答者

已关注

53 人赞同了该文章

1 引言

回测 (backtesting) 是任何一个量化投资策略上线之前必须要经过的一道流程；它是量化投资策略和其他类型投资（如主动投资）最大的区别之一。回测将投资策略应用到历史数据中并考察它的表现。具体的，回测有如下作用：

- 验证投资策略是否有效：**我们很容易从学术文献、投资书籍、券商报告、以及互联网得到很多交易策略的想法。回测可以快速验证一个想法是否有效（净值曲线至少得是向上的）。
- 对策略进行参数调优：**当一个交易策略有效的时候，回测让我们优化该策略的参数。比如，一个均线策略，我们使用 20 天的均线还是 30 天的均线？当然，在优化的过程中，我们需要非常小心、避免过拟合的问题。
- 给出策略在未来表现的预期：**任何一个靠谱的量化投资策略都是基于对市场行为的某种假设；该假设背后反映的是某种特定的业务逻辑。量化策略假设市场的这种行为在未来会重复。因此策略在回测中的表现将对其未来上线后的表现（如收益率、最大回撤、交易费用等）起到参照作用。

知乎

首发于
川流不息

损。今天我们就来和各位小伙伴聊聊我们是怎么理解回测的。本文将从下面几个角度论述：

1. 超参数；
2. 训练集、测试集之纠结；
3. 各种偏差；
4. 不可忽视的交易费用；
5. 回测并非无所不能。

2 超参数

超参数 (hyperparameters) 是参数的参数，是源自机器学习中的一个概念。在机器学习中，超参数是在模型开始学习之前需要人为设置值的参数，而不是通过训练得到的模型参数。

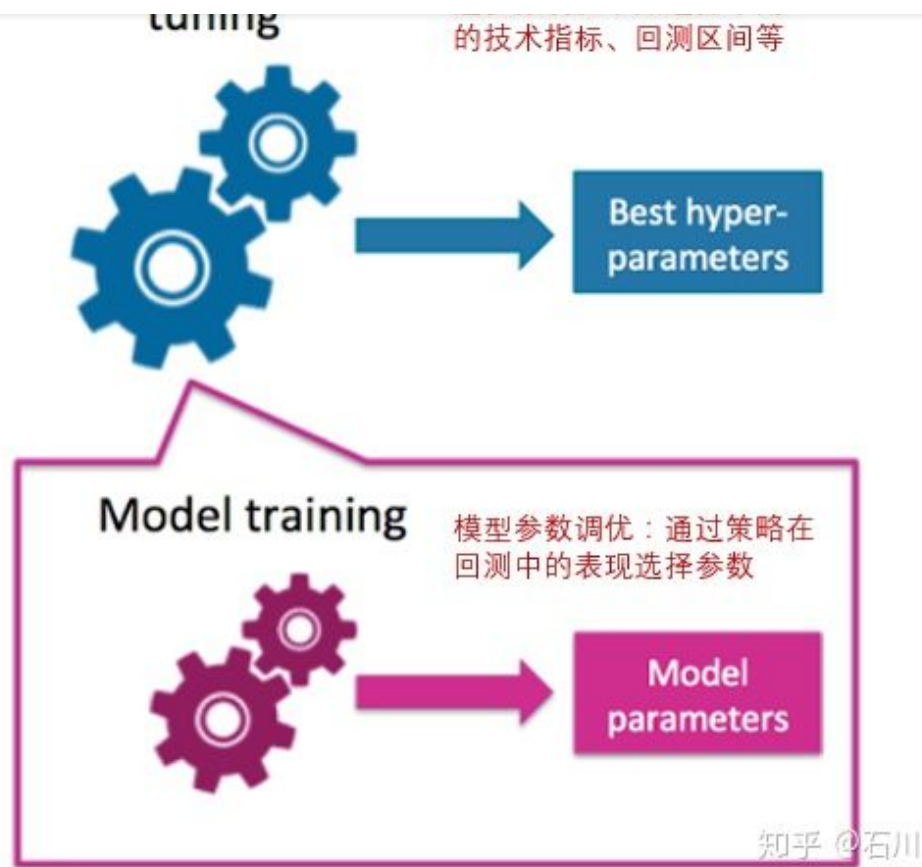
在构建量化投资策略时，虽然不一定会用到机器学习算法，但是超参数的概念依然非常重要。比如在技术分析策略中，我们是使用通道突破还是均线交叉？是使用 ADX 指标还是 SO 指标来测量动量的大小？这些需要人为决定的参数就属于超参数。一旦确认具体使用的指标后，至于计算指标的参数（比如计算通道和均线需要的时间窗口）就可以通过回测来优化。

另一个超参数就是回测期的时间窗口到底选多久。这是任何量化投资策略都必须面对的问题。比如我们测试一个量化选股因子，那回测期是从 2000 年开始，还是 2005 年开始，还是 2010 年开始等，这是事先必须确定的。

随着机器学习算法在量化投资领域的应用愈发普及，我们在使用它们时也势必面临着同样的问题。比如在集成学习中（如随机森林），决策树的数量和深度是多少；在 K-means 算法中簇数的个数；在 SVM 算法中核函数应该选哪个。这些都是超参数的范畴。



知乎

首发于
川流不息

对于典型的机器学习问题，超参数也是需要优化的，以此提高机器学习算法在样本外的性能和效果。对于量化投资策略来说，超参数的调优除了从数学角度（如策略的 Profit/Loss 曲线）考量外，还需要考虑超参数本身是否具有清晰的业务含义以及是否适合策略的使用者。这里举两个例子。

- **例子一：**假设使用 SVM 来预测股票收益率，输入的特征可以是各种量价数据和基本面数据，输出为涨跌的预测。如何选择 SVM 的核函数就是一个超参数问题。如果选择了非线性的核函数，它会把特征向量进行非线性变换、映射到更高维的空间。这当然有助于提升预测的准确性，但是人们往往难以从金融投资角度解释特征向量的非线性变换，因此这样的策略就不折不扣的成为了黑箱。
- **例子二：**趋势追踪策略既可以使用均线交叉也可以使用通道突破。在它们之间的选择就是超参数的问题。然而，这两类策略的风险收益比显著不同：均线类策略交易次数多、胜率低、平均盈利和平均亏损都较低；反观通道突破策略，它交易次数少、胜率高、平均盈利和平均亏损都较高。不同的策略背后折射的是完全不同的“投资哲学”，投资者需要根据自己的风险偏好选择。如果仅从数学角度考虑选取了一个并非让自己十分舒服的交易策略，那么在实盘中注定会失败。

除了上述这些明确的超参数外，有一些因素需在构建策略之前确定，但是在回测中我们会根据策略的表现对它们进行反馈。这类因素介于超参数和策略参数之间。比如在资产配置策略中选择哪些资产放入资产池；又如在量化多因子选股策略中，使用哪些风格因子进行选股。

在资产配置中，我们必须事先决定一个资产池（asset pool），然后构建策略。但是随着回测的深入，如果发现某类资产的加入对于策略的表现没有正贡献，我们便会从资产池中剔除掉该资产。

知乎

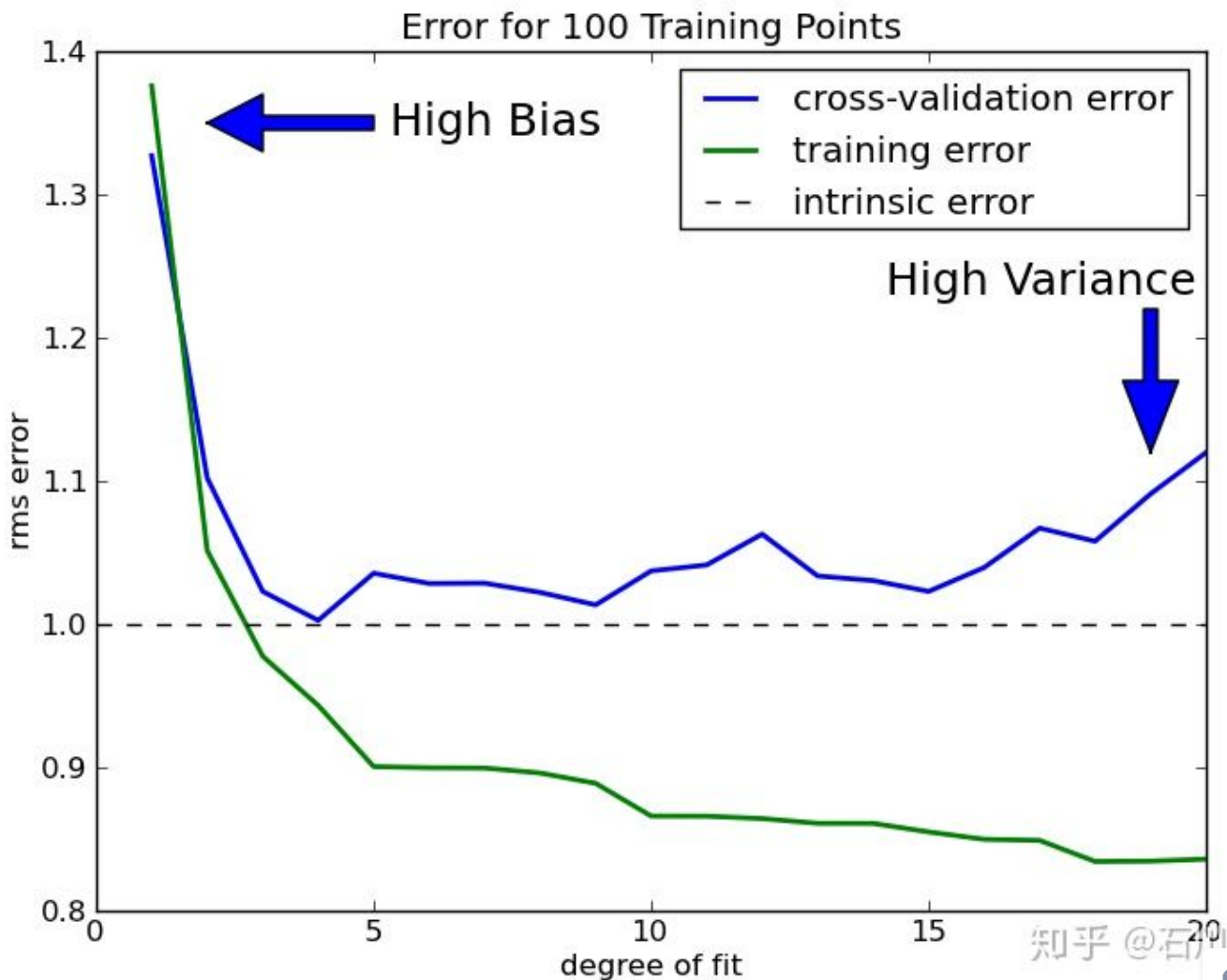
首发于
川流不息

剔除掉这些因子。像资产池、因子库这类“参数”，都需要在构建策略前确定，但是策略的表现又会指导我们进一步优化它们。

3 训练集、测试集之纠结

任何一个介绍量化策略或交易系统的文献都会强调在构建策略时，一定要把历史数据分成训练集和测试集，在训练集上优化策略的参数，在测试集上考察策略的表现。（我曾经看过一本英文的关于交易系统构建的书，洋洋洒洒 300 来页，基本上唯一的核心思想就是要有测试集……）这个论调固然没有任何问题，但是交易数据中的一些特点使得是否区分训练集和测试集有时变得很纠结。

区分训练集和测试集的初衷是避免参数在样本内过拟合，从而导致策略在样本外（上线后）的表现和其在样本内的表现大相径庭。在一个典型的通过训练集和测试集来对参数调优的过程中，通常的做法是调整算法迭代的次数或者惩罚系数之类的。随着参数的调优，训练集上的误差单调递减，等测试集上的误差开始提升，就停止调优（下图）。



一个经科学回测过的策略应该能够捕捉训练集和测试集数据表现出来的某种市场共性。任何交易数据都包含一定的噪声成分，使用测试集是为了避免模型（策略）过度错误的考虑训练集内样本。



我们说交易数据严重匮乏是就数据的**独立性**以及数据的**可交易特征**而言的。从这两点而言，在构建策略时是否一定要划分训练集和测试集、以及如何划分就需要非常谨慎。

由于市场流动性环境和宏观经济的变化，以及波动率聚类等特点，投资品价格数据的短期相关性比较高，而在不同的历史时期往往反应出不同的特点。**如果训练集和测试集的数据代表了不同的市场状态，那么它们之间的交易数据就没有多少共性可言。**相对于训练集来说，测试集中的数据满足训练集中假设的样本太少，以此来优化策略势必忽视了测试集数据中传达出的新的市场环境。对于一个策略来说，我们都不希望忽略市场的任何特征（除了噪音）。

在可交易特征方面，满足一个策略假设的样本其实很少。举个极端的例子，比如 A 股中追踪大牛市的趋势类策略。在过去 20 年中，也仅有三波牛市，而且它们表现出来的市场特征均不相同（比如以 2007 年大牛市构建的趋势追踪策略在 2015 年大牛市中的逃顶效果并不好）。在这种情况下，如果还把数据分成训练集和测试集会怎样呢？**我们一定会把策略在测试集中的体现出来的新市场环境反馈到训练过程中，这已经违背了分训练集和测试集调参的本意；这等价于我们在整个历史数据中对策略的参数调优了。**比如我们使用训练集构建一个策略，但是在测试集上遇到股灾了，为了对付它，就强行加上一个下跌 XX% 就止损的条件。这其实不是在对原始策略的参数调优，而是针对测试集的样本去改变策略的初衷！如此来“优化”策略，又还有什么训练集、测试集之分？当然，这么做的出发点是好的——让策略尽可能的适应不同的市场环境，在真正的样本外（上线后）的适应性更强。

对于是否一定要分训练集和测试集，我的看法是**这种防止样本内过拟合的思想是绝对可取的，但如何执行应该试策略而定。**比如对于用机器学习算法来预测短期收益率的策略，使用训练集和测试集就是恰当的。对于量化因子选股策略，这似乎又是个伪命题。众所周知，小市值因子在股灾 2.0 之前一直有效，但从 2016 年初开始，随着价值投资的回归，小市值因子彻底失效。又比如，价值因子（或者说就是“上证 50”因子、“沪深 300”因子）在最近两年有效，但是随着它们估值的提升，谁也不能保证它在未来持续有效。**一个经科学回测过的策略必须能够捕捉市场的某种特点。只要这个特点有明确的业务支持而非数据挖掘的产物，那么针对这个特点构建及优化策略参数时，是分别使用训练、测试集，还是使用所有数据来整体优化，这并不重要。**

4 各种偏差

本节介绍回测中的一些常见的偏差，它们包括**优化偏差（过拟合）、前视偏差、以及幸存者偏差。**常见的介绍回测的文章的重点就是这些。



知乎

首发于
川流不息

知乎 @石川

优化偏差 (optimization bias) 指的是通过不断地增加参数个数，直到策略的收益曲线在回测中表现的非常好。参数越多，越容易在样本内“处处精准”。这意味着对样本内数据的噪音过度建模，及过拟合。**防止优化偏差的方法是减少参数的个数 —— 比如能够使用一个技术指标就不要使用两个 —— 以及增加回测数据。**在这方面，可以从两个维度来增加。一个是使用更长的历史数据，比如一个策略在过去 10 年有效就比在过去 3 年有效更有说服力。但这么做时需要考虑市场状态变换（有可能太久远的历史数据并不满足你策略的假设），因此回测数据也并不是越长越好。第二个维度是考虑不同的投资品。如果策略和一组给定的参数在多个投资品上都有效，它无疑更有说服力。

此外，**通过考察参数平原 (parameter surface) 也可以检查过拟合。**将策略表现的指标（比如收益率或者最大回撤）画成参数的函数就是参数平原。**对于一个好的策略，在微调参数取值时，它的表现应该比较稳定，参数平原应该十分光滑。**比如如果最优的均值参数是 20 个交易日，则当我们使用 19 或 21 个交易日计算均值时，策略的表现和 20 日均线的策略不应该差很多。如果策略的参数平原非常不规则，则意味着参数背后没有合理的业务逻辑，而更多的是数据挖掘的结果，这时就要小心。

前视偏差 (look-ahead bias) 指的是错误的使用了未来数据。在构建策略时，尽管我们会非常小心，但仍然会时不时的步入前视偏差的陷阱，这往往是编程时的疏忽造成的。举两个我自己的例子。有一次我使用卡尔曼滤波（一种状态空间建模方法）构建了一个预测模型，我很小心的使用截止到每个交易日当日的滚动窗口以避免使用未来数据，结果回测的效果非常好，“好到一看那净值曲线就知道一定是哪里出了问题”。后来经过检查发现，在使用该方法的第一步有一个初始化的过程，在那里传入了所有历史数据，而非截止到每个交易日的数据，由此引入了未来数据。另一次我测试一个技术分析策略，该策略的收益曲线也是非常吸引人。后来发现，程序中使用了最高价和最低价数据，并假设当它们出现时触发交易。这显然是有问题的，因为只有当一个交易周期结束后，才能知道这段时间内的最高价和最低价是多少，而到了那个时候，早已无法按该价格成交。所以，通常在使用最高、最低价时，必须至少使用滞后一期的数据。同样的问题也会发生在使用基本面数据或宏观经济数据的策略中，我们必须对这些数据何时发布非常小心，以免在回测中提前使用这些数据，造成前视偏差。



知乎

首发于
川流不息

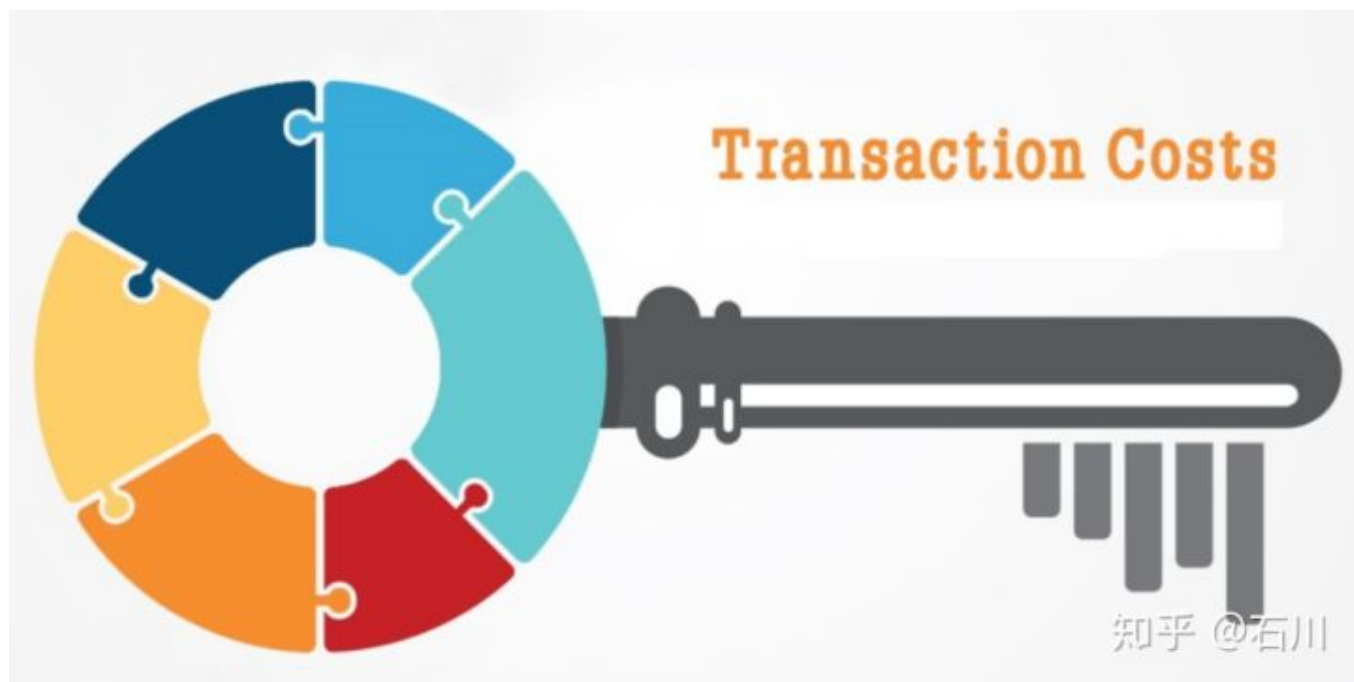
而忽略掉在回测期内退市的股票。这么做会提高提高策略在回测期内的表现，因为那些差的、退市的股票根本就没有被考虑在内。但在实际交易中，我们无法提前预知哪些股票会退市。因此，忽略退市的股票是不恰当的。这也是一种“前视偏差”。

量化投资较主动投资的优势是，它可以摒弃人的情绪，排除人在交易时由各种认知偏差带来的额外的风险（见让你投资亏钱的15个“偏差”）。在回测中，也只有尽量摒弃各种偏差，才能使它发挥出最大的作用。

5 不可忽视的交易费用

一个回测中赚钱的策略和一个实盘中亏损的策略之间的差别是什么？交易费用！

交易费用包括手续费（commission）、延迟造成的滑点（slippage）以及流动性造成的冲击成本（market impact）。其中手续费的估计最简单，券商有固定的手续费比例，只需要在回测中考虑到即可。滑点和冲击成本则很难在回测中准确的反应。



滑点指的是在交易信号产生到交易实际执行之间价格的（不利）变动造成的额外费用。这对于中、高频策略额外重要。这些策略捕捉日内转瞬即逝的交易机会，如果回测中策略的表现依赖于按照交易信号产生瞬间的价格来交易，那么这个策略在实盘不会有什么前途。这是因为实盘中难以保证会交易到信号发生瞬间的价格（价格出现后才能计算交易信号，交易信号需要送到券商，然后送到交易所交易，这之间都需要时间）。为此，回测中可以考虑的是按照每个 bar 的收盘价来交易。这个 bar 的频率可以很高，比如分钟级，但一定是按照它的 close 价格来成交，而非在这个 bar 形成过程中信号触发的瞬时价格。又或者，在回测时考虑信号出现后一段时间内按照 TWAP（时间平均）成交。在这些设定下并考虑适当的滑点，如果该策略还能赚钱，那么在实盘中才有机

知乎

首发于
川流不息

冲击成本指的是投资品流动性不足带来的额外交易费用。对于流动性高的投资品，限价指令簿（limit order book，就是我们在交易软件上看到的 level 1 买卖各几档的挂单情况）上，买一和卖一之间的价差（bid-ask spread）很小，且两边的挂单量都足够。交易这类投资品不会带来太大的市场冲击。但如果投资品的流动性不足，那么为了完成交易，策略就势必会 walk the book（在一个方向上“收单”），直到完成我们的交易量需求。这个冲击成本和每笔交易时的具体流动性有关，因此回测中难以准确估计。

鉴于交易成本的重要性，在回测中必须假设合理的交易费率，宁可高估也不应低估。通常的做法是估计几档不同的交易费用，比如每笔交易单边千分之一、千分之三、千分之五甚至更高等。以此感受一下策略在不同费率下的收益和回撤情况。但真实的交易成本仍然需要一段时间的稳定实盘交易来评判。

能否有效的控制交易费用直接关系到一个策略的资金容量。资金量（交易量）越大，完成交易所需的时间越长，而交易成本（及不确定性）随交易时间呈非线性增长。由于交易量巨大，华尔街的大型金融公司都会针对市场微观结构（即 order book 上买单、卖单出现的 dynamics）来建模，以此构建交易模型（注意是交易模型，而不是投资策略模型），从而尽可能的降低交易成本。

6 回测并非无所不能

科学的回测对于量化投资策略至关重要。但是，**回测也并不是万能的。**

一个经过严格回测的策略投放到实盘时最大的敌人是一个人的心理关。它指的是交易者能否克服实盘中的心理压力从而坚持使用这个策略。回测中的净值曲线通常是非常稳健的（几年的回测期内慢慢爬升），其各种指标——收益率、最大回撤、夏普率等——也非常出色。面对回测中优秀的净值表现，我们会错误的高估自己对于回测中亏损以及回撤周期的容忍程度。在实盘中面对如此的亏损和回撤周期完全是另一个 story。

对于任何一个策略，几乎可以确定的是它在回测中的表现是其在实盘中表现的上限（比如以最大回撤而言，我们自己的经验和很多量化投资前辈的经验都指出，实盘的最大回撤差不多是回测中最大回撤的 2 倍）。此外，当一个策略上线之初，亏损先于利润出现也是十分常见的。面对真金白银的亏损，我们会比想象的更脆弱、更容易怀疑策略的开发中是否存在没有考虑到的问题、更容易自我动摇从而想要放弃该策略。在实际交易中，价格时刻在波动，充斥着噪音的各路消息以远超过我们能够接受的速度袭来，我们会快步踏入行为金融学中的各种认知偏差陷阱、丧失冷静、做出一系列干预交易系统的错误操作。这些心理压力是哪怕再精细、再严格的回测也无法呈现的。

另外，任何策略都会有失效的那一天。一个技术指标会因为使用的人越来越多而失效，一个选因子会因为市场状态的变化而失效，一个赚钱的策略会因为流动性突然枯竭（比如股指期货的突定制）而失效。**任何量化策略都需要定期进行事后评价，比较它在样本外和回测期内的表现，评价是否仍然有效，或者根据实盘的反馈进行调整。**

知乎

首发于
川流不息

然而，如果我们在这两年一直按照小市值选股却早已亏的一塌糊涂了。又比如，在资产配置策略中，随着实盘的进行，哪些投资品应该被挪出资产池、哪些回测中没有考虑的投资品又应该被加进来，这些都是回测无法回答的。

当然，这些并非回测之过。一个经过科学回测的量化投资策略仍然是投资中的利器，它刻画了市场的某种具备清晰业务逻辑支撑的特性，并假设该特性在未来会重复，从而捕捉这些机会。如果市场变了，该特性不再发生而导致策略失效，这不是回测的问题，我们只要欣然接受它就是了。科学的回测让我们更有希望接近实现交易的一致性；坚持一致性的交易才能让我们成为更优秀的交易者；只有不断追求卓越、成为更优秀的交易者，才可能在市场的沉浮中生存的更久。

(全文完)

免责声明：文章内容不可视为投资意见。市场有风险，入市需谨慎。

原创不易，请保护版权。如需转载，请联系获得授权，并注明出处，谢谢。已委托“维权骑士”
([维权骑士](#) 免费版权监测/版权保护/版权分发) 为进行维权行动。

编辑于 2019-07-02

[交易策略](#) [量化交易](#) [算法交易](#)

▲ 赞同 53 ▼ 11 条评论 分享 ★ 收藏 ...

文章被以下专栏收录



川流不息

北京量信投资管理有限公司是一家在中国基金业协会备案登记的专业私募基金管理人...

关注专栏

推荐阅读

那些年，做期货量化与高频的日子

读书的时候写了不少留学申请与就业的文章，现在就谈谈这几年做 futures quantitative and high





babyquant

张生

岑秋苑

11 条评论

切换为时间排序

写下你的评论...



江海

1 年前

请问怎么判断策略是否失效呢？

👍 1



Artilia 回复 江海

1 个月前

真正的样本外检验，最好实盘

👍 赞



Chandler Chan

1 年前

一看完本文，就迅速发现了系统中的一个前视偏差，改完bug后效果显著，收益减半，夏普由3变1😏

👍 1



石川 (作者) 回复 Chandler Chan

1 年前

我也有这样的经历.....😏

👍 赞



石川 (作者)

1 年前

另外有时回测一些日内策略，效果也特别好，后来一看就是绝大部分收益都来自bid ask spread.....

👍 赞



雪在烧

12 个月前

2种理论

1 是要选现在测试结果不好的模型

1种是选择现在测试非常优秀的模型

这2种都有令人信服的理论

科学回测中的大学问 - 知乎

知乎

首发于
川流不息

但我认为 依靠测试出来的模型， 依然问题非常大

真正有用的模型，是 哲学思想变成代码， 测试目的是看编程有没有实现这个哲学思想.....

类似案例，就是头皮马丁， 完全可以通过计算， 知道风险和利润所在， 并通过设计， 实现这些控制----这是说真正可以略长期使用的模型， 应该是通过思想设计， 而不是回测， 回测是看设计， 是不是真正变成代码而已

👍 赞 ↩ 回复 🗑 踩 🚩 举报



去特么的智障

8 个月前

不知道您了解股票T0不，我有个策略按照下一个tick对手价成交，在今天的平均单笔收益是千1.8，比印花税加手续费多万四左右。这样的策略可以么？因为我了解的T0按照下一个tick对手价成交已经算比较严格了

👍 赞



Artilia 回复 去特么的智障

1 个月前

市场冲击成本没考虑吧？

👍 赞



ERIC FU

8 个月前

没有用

👍 赞



longtaizi

8 个月前

这篇文章的一些内容在successful algorithm trading里看到过

👍 赞



zvcinq

6 个月前

没找到业务支持的规律不就是过拟合吗，上线赚也是运气。关键是就算找到业务支持这个本来就是各说各有理的东西。还是要找到规律的规律的规律，有一个能完全抽象成平稳的序列，然后根据序列反复模拟，才能知道P-value，才知道找到那个规律是不是扯淡。

👍 赞

