

PBO算法在量化交易中的应用

——以黄金和比特币的交易策略为例

陆羿辰¹, 张焯辰¹, 王天霞¹, 陈 蕾¹, 胡江胜¹, 孙丹丹^{2*}

¹江苏理工学院数理学院, 江苏 常州

²江苏理工学院经济学院, 江苏 常州

收稿日期: 2023年3月24日; 录用日期: 2023年4月18日; 发布日期: 2023年4月26日

摘 要

本文以黄金和比特币交易策略为例, 讨论了PBO算法在量化交易中的应用。本文提出了一种基于PBO算法的动态优化算法用以优化交易模型参数。为了测试算法的性能与稳定性, 本文介绍了五种基于传统技术分析理论的量化交易预测模型, 并展示算法对这些模型参数优化后的结果。结果表明, 与全局优化的结果相比, 本文所提出的算法在两种资产及五种模型上都具有显著提升, 说明算法具有良好的性能与稳定性。

关键词

量化交易、PBO算法、CSCV框架

The Application of the PBO Algorithm in Quantitative Trading

—Taking Gold and Bitcoin Trading Strategies as Examples

Yichen Lu¹, Ye Chen Zhang¹, Tianxia Wang¹, Lei Chen¹, Jiangsheng Hu¹, Dandan Sun^{2*}

¹College of Mathematics and Physics, Jiangsu University of Technology, Changzhou Jiangsu

²School of Economics, Jiangsu University of Technology, Changzhou Jiangsu

Received: Mar. 24th, 2023; accepted: Apr. 18th, 2023; published: Apr. 26th, 2023

Abstract

This paper discusses the application of the PBO algorithm in quantitative trading, taking the ex-

*通讯作者。

文章引用: 陆羿辰, 张焯辰, 王天霞, 陈蕾, 胡江胜, 孙丹丹. PBO 算法在量化交易中的应用[J]. 应用数学进展, 2023, 12(4): 1690-1697. DOI: 10.12677/aam.2023.124175

amples of trading strategies for gold and Bitcoin. A dynamic optimization algorithm based on the PBO algorithm is proposed in this paper to optimize trading model parameters. To test the performance and stability of the algorithm, the paper introduces five quantitative trading prediction models based on traditional technical analysis theory, and the results of optimizing the parameters of these models using the algorithm are presented. The results show that compared with the results of global optimization, the algorithm proposed in this paper has significant improvements in both assets and all five models, indicating that the algorithm has good performance and stability.

Keywords

Quantitative Trading, PBO Algorithm, Combinatorially-Symmetric Cross-Validation

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

量化交易是根据投资经验或者数学和统计的方法建立量化模型，将公开数据输入模型得到对金融资产未来收益的预测，从而指导投资实现盈利的投资方法[1]。回测是量化交易研究中的一项重要工作，也是量化交易与经典投资的最大区别之一。回测的本质是模拟和再现一种投资策略，以便其能够在历史上准确地描述它。基于回测，投资者可以推断策略在未来的表现，从而选择更有效的投资策略投入实盘中使用[2]。例如，Lo 等人[3]提出了一种名为“Bootstrap 方法”的技术，用于评估策略在历史回测中的表现是否具有统计显著性。经过在标准普尔 500 指数上的 1000 次随机模拟回测，他们得出结论：移动平均 (Moving Average, MA) 交叉策略的平均收益显著高于随机策略。因此他们认为：如果交易策略在随机路径上的平均回测表现显著优于随机策略，那么就有理由相信策略在未来市场中也可能有效。然而，在研究低频交易的投资策略时，针对有限的数据集，若采用复杂模型或是多重假设检验进行回测，将有较高的概率出现过拟合现象。

控制过拟合是量化交易中最关键的概念之一。模型在金融数据上过拟合的主要原因有两点，分别是市场规律发生了变化与模型对回测期间数据噪声进行了过度学习[4]。因此，在模型学习样本内数据的过程中，有必要进行过拟合检验，即时识别模型的过度拟合并抑制，使模型在样本外有更稳定的表现。

PBO (Probability of Backtest Overfitting) 是一种定量衡量小样本回测过拟合风险的算法，由 Bailey 等人于 2017 年首次提出[5]，其计算方式基于组合对称交叉验证框架 (Combinatorially-Symmetric Cross-Validation, CSCV)。PBO 算法将夏普比率 (Sharpe Ratio, SR) 作为策略的评估指标，其中过拟合概率，即 PBO 值的定义如下

$$PBO = P[SR_{\bar{n}} < ME(SR)],$$

其中， SR 为每个策略在“测试集”中的夏普比率， $SR_{\bar{n}}$ 为最优策略在“测试集”中的夏普比率，而最优策略为“训练集”中夏普比率最高的策略。 $ME(SR)$ 为所有 SR 的中位数。需要注意的是，此处的“测试集”与“训练集”并不完全等同于机器学习中的测试集、训练集概念，但具有异曲同工之处。

上述 PBO 定义的含义是：“训练集”夏普比率最高的策略 \bar{n} ，在“测试集”的夏普比率也应该较高，

表现至少应优于半数策略。如果策略 \bar{n} 的测试集夏普比率排名在后 50%，则很有可能属于回测过拟合。回测过拟合的概率，即为最优策略 \bar{n} 的测试集夏普比率排名位于后 50% 的概率。

本文主要讨论 PBO 在量化交易中的应用。本文安排如下：第 2 节介绍 PBO 算法的基本原理，并在此基础上提出基于 PBO 的动态优化算法；第 3 节测试 PBO 的动态优化算法在比特币和黄金交易策略上的性能；第 4 节对本文工作做出总结。

2. PBO 算法的介绍与应用

PBO 算法能够输出单次回测中，策略的过拟合概率。考虑到量化交易业界进行回测研究主要以实盘应用为目的[2]，得到回测过拟合的结论能降低实盘模型发挥不良的风险，但不能得到在样本外有更好表现的模型。因此，我们建立了基于 PBO 的动态优化算法，能够对量化交易模型的参数进行动态优化。

2.1. PBO 算法的介绍

以下为 PBO 算法的计算步骤：

- 1) 构建矩阵 $M_{T \times N}$ ，每列分别表示第 N 组策略下 T 期的收益率序列。
- 2) 按行切割矩阵 $M_{T \times N}$ ，得到子矩阵 $M_t, t=1, 2, \dots, S$ ；需要注意的是，这里 S 必须为偶数，此时每个子矩阵维度相同，均为 $\frac{T}{S} \times N$ 。

3) 从 S 个子矩阵中任意选出 $\frac{S}{2}$ 为一组，用 C_S 表示所有可能的组合；根据组合原理，这样的组合共有 $C_S^{\frac{S}{2}}$ 种。

4) 对于 C_S 中的任意一组 c ，进行如下操作：

- a) 构建训练集 J ：将 c 中的 $\frac{S}{2}$ 个子矩阵 M_t 按行拼接起来；
- b) 构建测试集 \bar{J} ：即 J 的补集，将不包含在 c 中的子矩阵 M_t 按行拼接起来；
- c) 对于训练集 J ，计算每列的夏普比率，得到夏普比率最高的策略 \bar{n} ；
- d) 对于测试集 \bar{J} ，计算每列的夏普比率，得到策略 \bar{n} 在测试集夏普比率 $SR_{\bar{n}}$ 的绝对排名 $Rank(\bar{n})$ 和

相对排名 ω （均为降序排列），通常取 $\omega = \frac{Rank(\bar{n})}{N+1}$ ， $\omega \in (0,1)$ 。

e) 定义对数概率 $\lambda = \log\left(\frac{\omega}{1-\omega}\right)$ ； λ 随着 ω 的增大而增大，当 $\omega = 0.5$ 时， $\lambda = 0$ ；当 $\omega \rightarrow 1$ 时， $\lambda \rightarrow \infty$ 。

5) 根据第 4 步，对于 C_S 中的任意一组 c ，可计算 λ 。对全部 c 进行遍历，最终得到 $\lambda_m, m=1, 2, \dots, C_S^{\frac{S}{2}}$ 。进一步可得到 λ 的经验分布 $f(\lambda)$ ，PBO 是 $f(\lambda)$ 在区间 $(-\infty, 0]$ 上的定积分：

$$PBO = \int_{-\infty}^0 f(\lambda) d\lambda$$

需要说明的是，上述定积分的前提假设是 S 取无穷大，此时相对排名 ω 和对数几率 λ 均为连续变量。一般而言， S 为有限整数时，第 5 步中计算的对数几率 λ 为离散变量。根据离散变量经验分布的定义：

$$f(\lambda) = \sum_{i=1}^n I_{\{\lambda_i < \lambda\}}$$

有如下推导：

$$PBO = \int_{-\infty}^0 f(\lambda) d\lambda = \frac{\#\{\lambda_m < 0\}}{C_S^{\frac{S}{2}}} = \frac{\#\{\lambda_m < 0.5\}}{C_S^{\frac{S}{2}}}$$

其中 $\#\{\lambda_m > 0\}$ 表示 C_S^2 个 λ_m 中大于 0 的个数, $\#\{\lambda_m > 0.5\}$ 则表示 C_S^2 个 ω_m 中大于 0.5 的个数。当划分完“训练集”和“测试集”后, 只需要计算相对排名 ω , 随后统计相对排名 ω 中大于 0.5 的个数即可。

2.2. 基于 PBO 的动态优化算法

以 BOLL 模型为例, 这是一个由移动平均线 MA, 压力线 $BOLL_{upper}$ 和支撑线 $BOLL_{lower}$ 构成的预测模型:

$$\begin{cases} MA = \frac{1}{n} \sum_i p_{t-i} \\ BOLL_{upper}(t) = MA(t) + a \times \sigma_n(t) \\ BOLL_{lower}(t) = MA(t) - a \times \sigma_n(t) \end{cases}$$

其中 P_{t-i} 是 $t-i$ 时刻的价格, a 为一个常数, σ_n 为是价格在 n 个单位周期内的标准差。根据投资中的技术分析理论[6], 当价格上涨突破压力线时, 模型预测下跌; 当价格下降跌破支撑线时, 模型预测上涨。

显然, 该模型有参数 a 和 n , 需要随着市场的变化而动态优化。在动态优化算法中的求解过程如下:

1) 确定参数的取值范围: 以 BOLL 为例, 参数 n 的取值范围为从 3 到 60 的整数, 参数 a 的取值范围为 1 到 3 步长为 0.1 的实数。

2) 设置回测的时间范围为时刻 τ 到时刻 t 。记录预测模型在每一组参数组合中的年化夏普比率, 得到年化夏普比率序列 $\{K_i\}$:

若 $\max\{K_i\} < 0$, 则认为当前回测区间中该预测模型无解, 求解结束;

若 $\min\{K_i\} \geq 0$, 则将夏普比率平均分为 6、8、10、12 组, 输入 PBO 算法, 得到的 4 个 PBO 值 $P_i, i=1,2,3,4$ 和相对排名。

3) 在第 2 步的基础上, 进行如下计算:

若 $\text{mean}\{P_i\} > 0.5$, 说明过拟概率较高, 删除最优策略在 $\{K_i\}$ 中对应的样本, 并返回第 2 步, 直到满足 $\text{mean}\{P_i\} \leq 0.5$;

若 $\text{mean}\{P_i\} \leq 0.5$, 说明过拟概率较低, 接受最优策略的参数为 t 时刻的最优参数。

4) 将第 3 步得到的参数带入预测模型, 对 $t+1$ 时刻的市场做出预测, 并返回第 1 步。

考虑到 BOLL 的参数 n 有从 3 到 60 共 58 个可取值, 参数 a 有从 1 到 3 共 21 个可取值, 做笛卡尔积后得到 1218 个参数组合, 如果直接对所有参数组合进行遍历, 将面临过拟合风险。但通过 PBO 算法, 能够排除过拟合概率较高的参数, 使遍历求解的方式更加有效。

3. 算法性能的测试

为了测试算法的性能, 除 BOLL 以外, 我们还基于传统技术分析理论[6]建立了 4 个量化交易预测模型, 分别为 MA (Moving Average, 移动平均) [7]、EMA (Exponential Moving Average, 指数移动平均) [6]、TRIX (Triple Exponentially Smoothed Average, 三重指数平滑平均线) [6]、MACD (Moving Average Convergence/Divergence, 异同移动平均线) [8]。

下图 1 中, 黑色实线显示了 MA 模型在黄金上进行交易的回测收益, 截止 2019 年 3 月 13 日(黑色垂直虚线), 为样本内的策略收益, 之后为样本外收益。灰色虚线为基准收益曲线, 即买入并始终持有黄金的收益曲线。

图 1 显示, 在样本内, 该策略的性能是稳定的, 而在样本外, 随着新数据的引入, 策略的收益逐渐和基准持平, 而从 2021 年开始, 策略远远落后于基准。显然模型对样本内数据过度拟合, 导致了模型在样本外表现不佳。



Figure 1. Backtesting example of overfitting
图 1. 回测过拟合示例

3.1. 黄金交易策略测试

使用基于 PBO 的动态优化算法对 BOLL、EMA、MACD、MA、TRIX 五个预测模型进求解，并统计它们在黄金交易策略中的表现，如下图 2 所示。

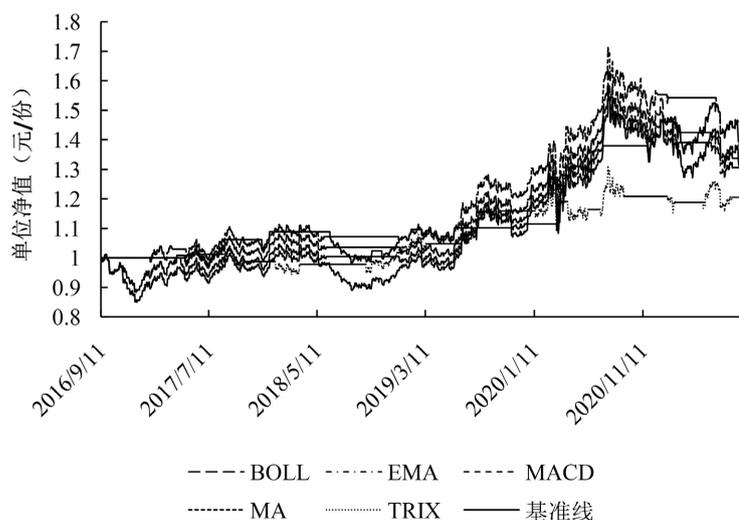


Figure 2. Profit curves of prediction model under dynamic optimization (gold)
图 2. 动态优化下预测模型回测收益曲线(黄金)

根据图 2，我们可以得出以下结论：在 2018 年之前，5 种策略都优于基准(黑色线)；2018 年至 2019 年，除 TRIX 策略的收益率曲线小幅低于基准外，其他指标的收益率曲线也优于市场。从 2019 年到 2021 年，EMA、MA 和 MACD 策略的收益率曲线均优于市场。为了测试基于 PBO 的动态优化算法的性能，需要将其与全样本全局优化的结果进行对比。如下图 3 为全局优化下预测模型的回测表现。

为了方便对比，将动态优化得到的模型收益减去全局优化得到的模型收益，如下图 4 所示。

观察图 4 不难发现，动态优化的模型表现全部优于全局优化。全局优化选取全样本收益最高的参数作为优化结果，模型在整个回测区间中的参数恒为定值。考虑到市场是动态变化的，在全局表现最优的参数不一定在局部表现最优，因此动态优化优于全局优化的结果是合理的，且说明基于 PBO 的动态优化算法具有良好的性能。

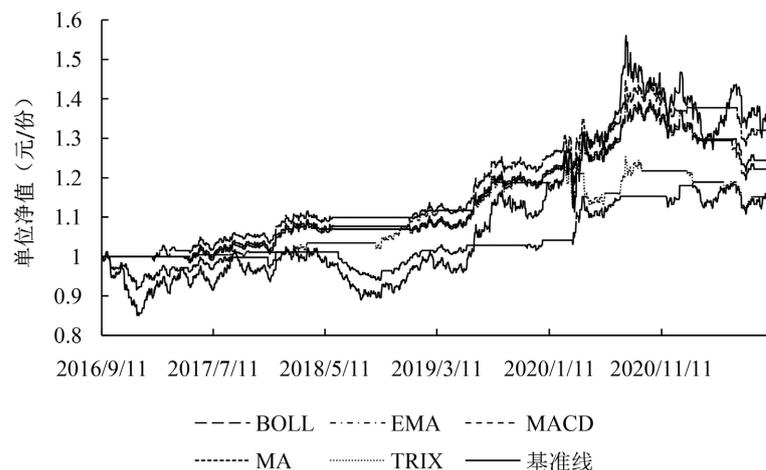


Figure 3. Profit curves of prediction model under global optimization (gold)

图 3. 全局优化下预测模型回测收益曲线(黄金)

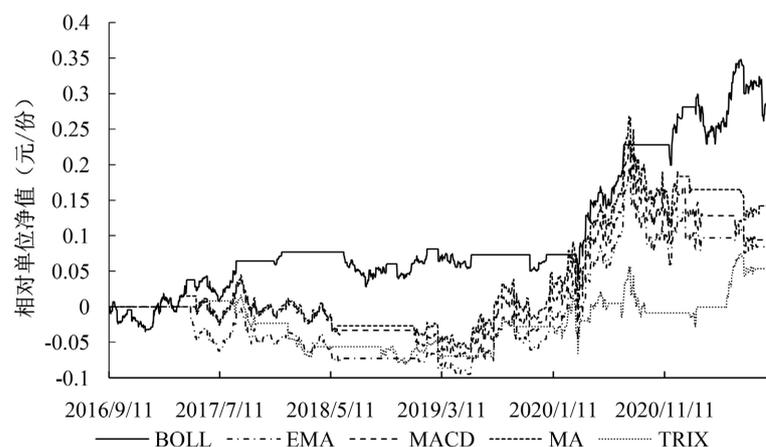


Figure 4. Relative profit curves of prediction model under dynamic and global optimization (gold)

图 4. 动态优化与全局优化相对收益曲线(黄金)

3.2. 比特币交易策略测试

考虑到单一品种的回测具有偶然性, 因此测试算法在比特币交易策略上的性能来评估算法性能的稳定性。另外, 由于 BOLL 在传统的分类中属于反转类的交易策略, 并不适合比特币这种高波动率的投资品种[6], 所以只测试 EMA、MACD、MA、TRIX 四个模型在比特币交易上的表现。结果如下图 5 所示:

观察上图发现, EMA、MACD 和 MA 策略的收益率曲线远高于基准。而 TRIX 策略的收益率曲线始终低于基准, 表现不佳。对比算法在黄金交易策略上的结果(图 2)发现, 4 个重叠策略在黄金上的相对排名依次为 MA、MACD、EMA、TRIX, 而比特币结果类似, 为 MA、EMA、MACD、TRIX。另外, 比特币于 2020 年至 2021 年的牛市行情为各策略提供了极高的市场波动, 同一策略在比特币上的绝对收益都显著高于在黄金上的绝对收益。为了测试动态优化算法的性能提升, 我们将其与全局优化的结果进行对比。

根据图 6, 在全局优化下, 比特币回测中仅有 EMA 和 MA 模型最终跑赢了基准, 且 EMA 模型仅小幅跑赢。同样地, 将动态优化得到的模型收益减去全局优化得到的模型收益, 如下图 7 所示。

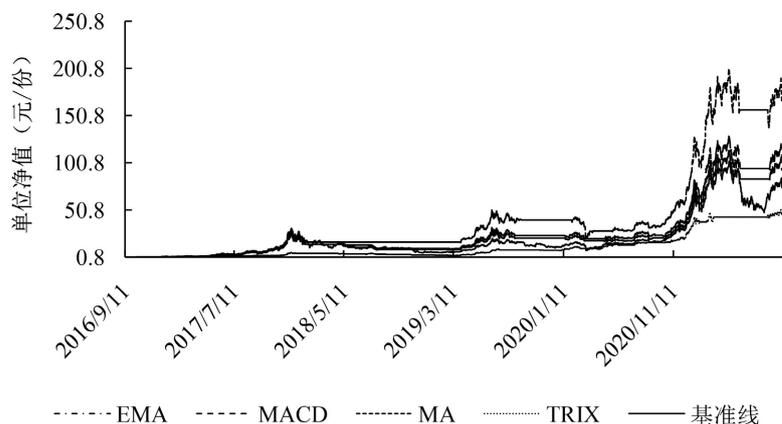


Figure 5. Profit curves of prediction model under dynamic optimization (Bitcoin)
图 5. 动态优化下预测模型回测收益曲线(比特币)

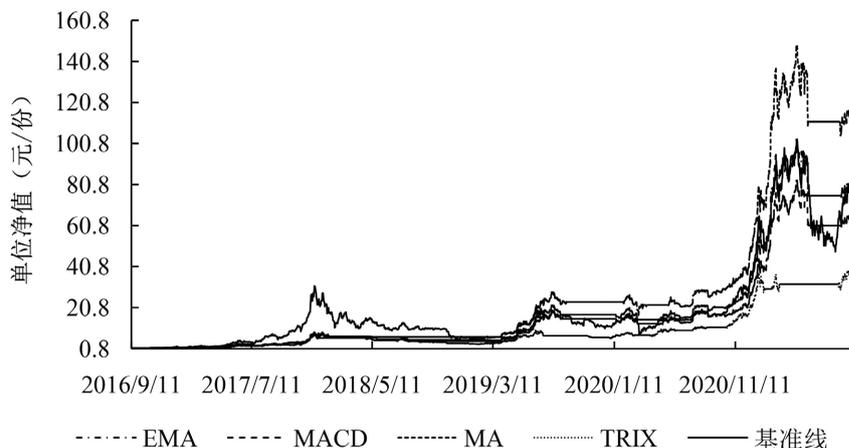


Figure 6. Profit curves of prediction model under global optimization (Bitcoin)
图 6. 全局优化下预测模型回测收益曲线(比特币)

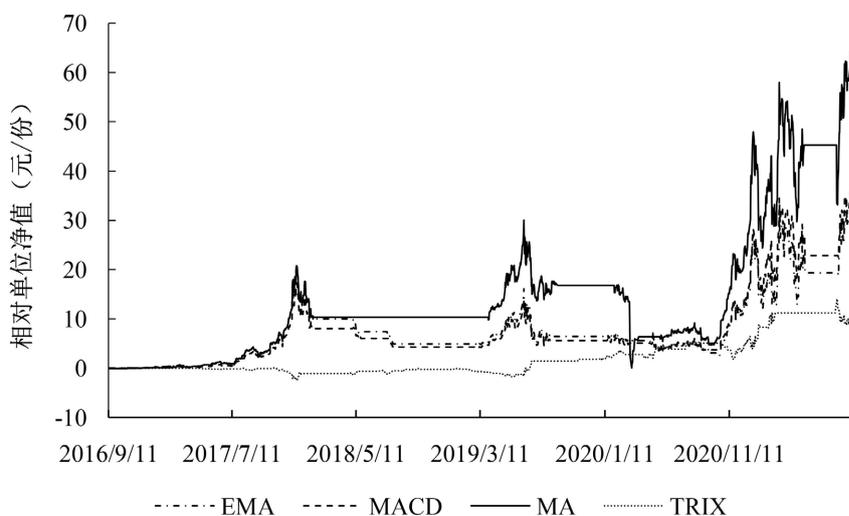


Figure 7. Relative profit curves of prediction model under dynamic and global optimization (Bitcoin)
图 7. 动态优化与全局优化相对收益曲线(比特币)

如上图, 动态优化的模型表现全部优于全局优化, 再一次说明基于 PBO 的动态优化算法具有良好的性能与稳定性。

4. 总结与推广

本文首先强调了控制过拟合在量化交易领域中至关重要, 并基于 PBO 算法为量化交易研究者们提供了一种新颖且有效的算法来改进交易模型。通过测试, 我们发现基于 PBO 的动态优化算法, 在黄金和比特币的交易策略上, 5 个交易模型的结果都优于局部优化, 说明我们提出的算法能够使量化交易模型的性能提升。

实际上, 本文只针对目前量化交易在业界面临的过拟合问题设计了应对算法。该算法只能在具有较为完善的量化交易模型时才能发挥作用, 若没有成熟的量化交易策略作为优化对象, 则未必能表现出较好的性能。然而, 一个量化交易模型一经公开发表, 其样本外的效益将会远远落后于样本内[9]。因此, 成熟且能够稳定盈利的量化交易模型往往具有极高机密性, 不便在公开渠道获取。本研究仅选择公开渠道可获取的量化交易模型进行测试, 存在一定的局限性。如果能够以更为复杂且成熟的量化交易模型作为测试对象, 则文章提出的算法将更有说服力。

基金项目

江苏省大学生创新创业训练计划项目“基于量化投资方法中多因子模型的改进研究”(202211463115Y); 江苏省自然科学基金项目“模型结构及其应用”(BK20211358)。

参考文献

- [1] Narang, R.K. (2011) *Inside the Black Box: The Simple Truth about Quantitative Trading*. John Wiley & Sons, Hoboken.
- [2] De Fusco, R.A., Mc Leavey, D.W., Pinto, J.E. and Runkle, D.E. (2015) *Quantitative Investment Analysis*. John Wiley & Sons, Hoboken.
- [3] Lo, A.W. and Mackinlay, A.C. (2011) *A Non-Random Walk down Wall Street*. Princeton University Press, Princeton. <https://doi.org/10.1515/9781400829095>
- [4] Suhonen, A., Lennkh, M. and Perez, F. (2017) Quantifying Backtest Overfitting in Alternative Beta Strategies. *The Journal of Portfolio Management*, **43**, 90-104. <https://doi.org/10.3905/jpm.2017.43.2.090>
- [5] Bailey, D.H., Borwein, J.M., de Prado, M.L. and Zhu, Q.J. (2017) The Probability of Backtest Overfitting. *Journal of Computational Finance*, **20**, 39-69. <https://doi.org/10.21314/JCF.2016.322>
- [6] Colby, R.W. (2003) *The Encyclopedia of Technical Market Indicators*. 2nd Edition, McGraw Hill, New York.
- [7] 张茂军, 饶华城, 南江霞, 王国栋. 基于决策树的量化交易择时策略[J]. 系统工程, 2022, 40(2): 118-130.
- [8] 孔傲, 朱洪亮, 郭文旌. 一个基于技术指标规则的启发式量化择时系统[J]. 系统工程, 2019, 37(1): 111-122.
- [9] Drobetz, D.W., Neuhierl, A. and Wendt, V.S. (2021) Data Snooping in Equity Premium Prediction. *International Journal of Forecasting*, **37**, 72-94. <https://doi.org/10.1016/j.ijforecast.2020.03.002>