

# 基于深度神经网络的股票预测系统

邢艳雅<sup>1</sup>, 李东喜<sup>2</sup>

<sup>1</sup>太原理工大学数学学院, 山西 晋中

<sup>2</sup>太原理工大学大数据学院, 山西 晋中

Email: 1342491536@qq.com, dxli0426@126.com

收稿日期: 2021年4月25日; 录用日期: 2021年5月8日; 发布日期: 2021年5月28日

## 摘要

股票市场受经济市场、政策等因素的影响, 其内部变化规律极其复杂。随着中国股票市场的快速发展和投资者规模的扩大, 股票市场产生了大量的交易数据, 使获取有价值的信息变得更加困难。由于深度神经网络善于处理数据量大, 非线性映射关系复杂的预测问题, 本文基于深度神经网络设计了一个股票智能预测系统。主要工作内容如下: 股票选股模型研究: 从股票财务指标和股票变化趋势层面, 研究了多因子量化选股的问题, 提出一种基于股票趋势识别算法构建选股模型的方法。股票价格预测模型研究: 针对股票交易数据量大, 非线性关系复杂, 难以准确预测股票价格等问题, 基于LSTM (长短时记忆神经网络) 深度神经网络对股票价格的预测进行研究, 从股票选股模型的输出结果中选择可投资的股票。

## 关键词

股票选股, LSTM, 股票智能预测系统

# Stock Forecasting System Based on Deep Neural Network

Yanya Xing<sup>1</sup>, Dongxi Li<sup>2</sup>

<sup>1</sup>College of Mathematics, Taiyuan University of Technology, Jinzhong Shanxi

<sup>2</sup>College of Big Data, Taiyuan University of Technology, Jinzhong Shanxi

Email: 1342491536@qq.com, dxli0426@126.com

Received: Apr. 25<sup>th</sup>, 2021; accepted: May 8<sup>th</sup>, 2021; published: May 28<sup>th</sup>, 2021

## Abstract

The stock market is affected by various factors such as economic markets, political, which lead to

文章引用: 邢艳雅, 李东喜. 基于深度神经网络的股票预测系统[J]. 应用数学进展, 2021, 10(5): 1721-1727.

DOI: 10.12677/aam.2021.105183

the complex of its internal changes. With the rapid development of China's stock market and the expansion of investor scale, a large amount of transaction data of the stock market has been generated, from which it's difficult to obtain valuable information. The deep neural network has certain advantages in dealing with large amount of data and complex nonlinear mapping. Therefore, based on deep network technology, an intelligent stock forecasting system has been designed. The main work is summarized as follows: Stock selection model: The problem of multi-impact factor quantitative stock selection is studied based on the stock financial indicators and stock change trend, and a stock trend identification algorithm is proposed to construct stock selection model. Stock price forecasting model: Because of the large amount of stock trading data and the complex nonlinear relationship, it forecasts stock price based on LSTM (Long Short Term Memory) deep neural network. We select investable stocks from the output of stock selection model.

## Keywords

Stock Selection, LSTM, Stock Intelligent Prediction System

Copyright © 2021 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

自 1985 年 11 月 18 日中国发行第一支名为“飞跃音响”的股票[1], 中国股票市场就开始不断扩大, 目前已成为全世界最重要的股票市场。然而中国股票市场有许多零散投资者, 非理性投资状况严重, 又易受到金融政策、公司经营状况、投资者心理等众多因素的影响, 经常引起股票价格大幅上涨或大幅下跌, 使得股票市场内部变化规律复杂, 变化周期不稳定, 从而很难直接获取有价值的信息。随着深度神经网络技术在各个领域的快速发展和广泛应用, 许多学者采用深度神经网络对股票市场进行股票预测研究。

据中国证监会统计, 截至 2018 年 4 月 7 日沪深 A 股已达 3478 支。如此多支股票, 如何从中选择出可投资的股票, 成为股票市场迫切需要解决的问题, 使得选股投资研究变得越来越重要。到目前为止, 虽然国内已经出现了许多投资选股模型, 但是从股票市场的收益效果来看, 这些选股模型的收益效果并不稳定。现有的选股模型考虑的影响因素较为单一, 没有充分与其他影响因素结合起来构建选股模型, 从而使得选股模型收益率较低[2]。相比于欧美的股票市场, 国内票市场成长时间短, 发展速度快, 存在很多非理性的投资现象, 因此通过结合多影响因素构建的选股模型可以在国内股票市场获得稳定收益[3][4][5]。本文通过获取股票的众多财务指标数据, 结合股票趋势识别算法标识财务数据样本, 采用主成分分析法和 BP 神经网络相组合的方法进行投资选股研究, 从而构建出适合股票市场的选股模型, 指导投资者进行合理的投资选股, 进而获得高额且稳定的收益。本文构建的选股模型可以改变目前选股模型影响因素不足的缺点, 并充分考虑了股票市场趋势, 为多影响因素、多组合模型的选股投资研究提供了新的想法和视角。

在股票市场中常采用统计学方法和机器学习方法对股票价格进行建模预测。统计学方法通过建立统计模型观察股票价格的波动趋势, 从而预测股票价格[6]。这类预测模型选取的股票数据特征简单, 未能充分挖掘股票市场的规律, 还带有严重的滞后性, 且影响股票价格变动的因素众多, 使得基于统计学方法构建的股票价格预测模型的预测结果并不准确。而基于机器学习方法构建的预测模型, 在股票市场中

取得了较好的预测效果。因为深度神经网络在处理样本数据量大, 非线性关系复杂的问题上具有很强的优势, 使得基于深度神经网络构建的股票预测模型效果最优。LSTM (Long Short Term Memory) 神经网络是一种改进的递归神经网络, 它通过循环反馈的结构处理时间序列数据, 进而拟合出时间序列数据的时序性, 还通过对网络中门开关的设置进行选择性的信息存储, 改善了递归神经网络中常见的梯度消失的缺点, 特别适合处理股票市场中股票价格预测的问题[7]。因此本文选择 LSTM 神经网络进行股票价格的预测研究, 通过对影响预测模型效果的时间序列长度和网络结构等因素进行实验分析, 使股票价格预测效果得到很大的提高, 为投资者提供了更准确的投资意见。

## 2. LSTM 简介

LSTM 是为了解决 RNN 的长期依赖问题而专门设计出来的, 所有的 RNN 都具有一种重复神经网络模块的链式形式。在标准 RNN 中, 这个重复的结构模块只有一个非常简单的结构, 例如一个  $\tanh$  层。LSTM 同样是这样的结构, 但是重复的模块拥有一个不同的结构。不同于单一神经网络层, 而是有四个, 以一种非常特殊的方式进行交互。

图 1(a) 是 RNN 的输入输出图, 它只有一个传递状态  $h'$ 。图 1(b) 是 LSTM 的输入输出图, 它有两个传递状态, 一个是  $c'$  (cell state), 另一个是  $h'$  (hidden state)。其中传递下去的  $c'$  改变得很慢, 通常输出的  $c'$  是上一个状态传过来的  $c^{t-1}$  加上一些数值。这里需要注意, RNN 的一个传递状态  $h'$  对应于 LSTM 的  $c'$ 。

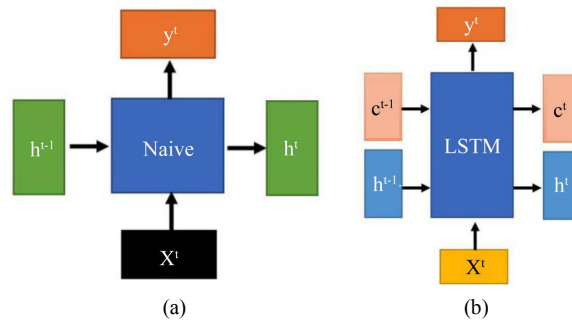


Figure 1. The difference between RNN and LSTM  
图 1. RNN 和 LSTM 的区别

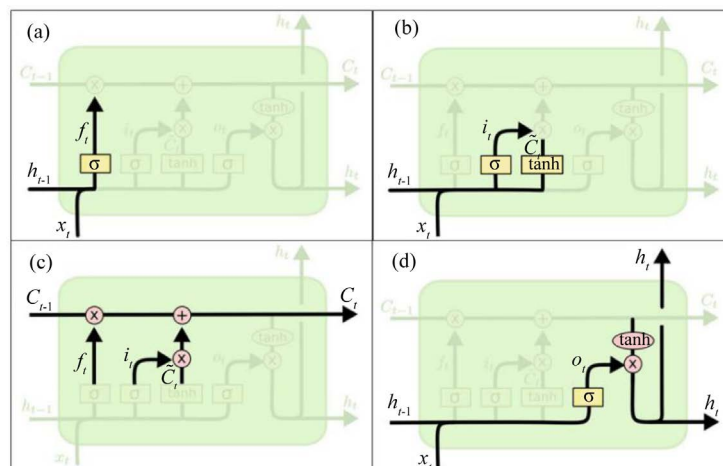


Figure 2. LSTM information flow split diagram  
图 2. LSTM 信息流向拆分图

LSTM 核心思想在于细胞的整个状态(绿色的图表示的是一个 cell), 和穿过细胞的那条水平线。细胞状态类似于传送带, 直接在整个链上运行, 只有一些少量的线性交互, 信息在上面流传保持不变会很容易。若只有上面的那条水平线是没办法实现添加或者删除信息的, 而是通过一种叫做门的结构来实现的。门可以实现选择性地让信息通过, 主要是通过一个 sigmoid 的神经层和一个逐点相乘的操作来实现的。sigmoid 层输出(是一个向量)的每个元素都是一个在 0 和 1 之间的实数, 表示让对应信息通过的权重(或者占比)。比如, 0 表示“不让任何信息通过”, 1 表示“让所有信息通过”。LSTM 通过三个这样的本结构来实现信息的保护和控制, 这三个门分别输入门、遗忘门和输出门。

图 2 所示就是 LSTM 的内部结构。通过门控状态来控制传输状态, 记住需要长时间记忆的, 忘记不重要的信息, 而不像普通的 RNN 那样只能够“呆萌”地仅有一种记忆叠加方式, 对很多需要“长期记忆”的任务来说, 尤其好用。但也因为引入了很多内容, 导致参数变多, 也使得训练难度加大了很多。

### 3. 传统模型的不足以及新模型的简介

本文选择以股票财务数据为主, 结合一些技术指标组合生成选股指标集, 从而构建股票选股模型。由于选股指标之间存在相关性, 因此选用主成分分析法对选股指标进行特征提取。

#### 3.1. 传统选股指标的问题

传统的选股指标常选用一些交易技术类指标或十几个财务技术指标构成选股指标集, 选股指标比较少不能全面反映股票市场的变动[8] [9]。2016 年王淑燕等人提出基于市盈率、市销率、市净率、ROA、EPS 等 16 个财务数据指标的选股指标集[10], 虽然相比于之前的选股指标集较全面, 但通过研究其他财务指标和股票技术指标发现, 它们与股票市场的变动有很强的相关性。因此本文在王淑燕的选股指标集基础上, 增加了一些技术性指标和其他财务数据指标, 并结合股票趋势来研究股票选股模型。传统的股票趋势识别算法一般采用滑动窗口提取时间序列, 通过设置几个指定位置的数值, 计算序列的涨跌幅, 再与预定的阈值比较从而判断股票趋势。这种方法没有完全考虑序列中的所有数值, 判断的趋势不是特别准确。

#### 3.2. 基于股票趋势识别算法[11] [12]的股票选股模型设计

BP 神经网络结构设计通常需要确认网络层数, 隐层神经元数, 输入输出节点数以及激活函数这些结构参数来完成神经网络结构的设计。① 确定网络的层数、② 确定隐层的神经元数、③ 确定输入输出节点数、④ 激活函数的选取。

基于股票趋势识别算法对选股指标数据进行趋势识别, 采用主成分分析方法对选股指标进行相关性分析, 进而选用 BP 神经网络构建股票选股模型。本节从改进的趋势识别算法对数据趋势处理和基于主成分分析对数据进行相关性分析两方面来验证所对应的选股模型的优劣。经过主成分分析的选股指标数据, 由于消除了选股指标间的相关性, 简化了 BP 神经网络结构, 也使得选股效果得到提升。对影响因素众多的股票市场提供了一个有效的选股方法, 并为股票价格预测模型的研究提供了研究对象。

#### 3.3. 基于 LSTM 深度神经网络的股票价格预测模型研究

由于中国股票市场复杂多变、交易频繁, 使得对股票价格预测成为股票投资中的重大挑战, 而 LSTM 深度神经网络善于处理数据量大、非线性关系复杂的时间序列数据, 并在自然语言处理, 图像识别等领域取得了较好的结果。本节基于 LSTM 深度神经网络, 对股票交易基本数据和股票技术指标数据的进行学习, 拟合出影响股票价格变化的因素与股票价格的非线性映射关系。并对影响股票预测模型性能的输入特征、时间序列长度和网络结构进行验证分析, 从而设计出较高预测精度的股票价格预测模型。

基于股票交易指标设计 LSTM 神经网络股票预测模型时, 需要对股票预测模型的输入和输出节点个数, 股票预测模型的隐藏层个数和隐藏层节点个数和优化方法的选取进行详细的分析设计, 从而提高股票预测模型性能。分为如下三步: ① 股票预测模型的输入和输出节点设计、② 股票预测模型的隐藏层和隐藏层节点设计、③ 股票预测模型的优化方法选取。

## 4. 实验研究和结果分析

### 4.1. 实验数据

从选股模型的选股结果中选择可投资的个股, 进行股票收盘价预测研究。股票的交易基本数据通过爬虫程序从东方财富数据源[13]中获得, 由于中国股市制度在 2005 年进行了改革, 本文选取 2006 年初至 2018 年底的股票交易基本数据, 以 600000-浦发银行为样例进行详细预测分析。表 1 为股票交易基本数据样本:

**Table 1.** Basic data sample of stock trading

**表 1.** 股票交易基本数据样本

code	date	open	high	close	low	volume	price_change	p-change
600000	2018/10/31	10.99	11.03	10.89	10.87	217371.1	0.03	0.27

在股票市场中影响股票价格的因素很多, 本文选取与股票交易相关的交易基本数据和交易技术指标数据来构建股票预测模型。通过这两类股票数据设计以下 3 个股票价格预测模型, 如下表 2, 从而选择最佳的输入特征。

**Table 2.** Forecast results

**表 2.** 预测结果

股票预测模型	输入特征维数	股票交易指标选取	RMSE	R <sup>2</sup>
M1	6	股票交易基本数据	0.874	88.9%
M2	32	股票交易技术指标	0.988	85.8%
M3	38	股票交易基本数据、股票交易技术指标	0.900	88.2%

### 4.2. 实验预测

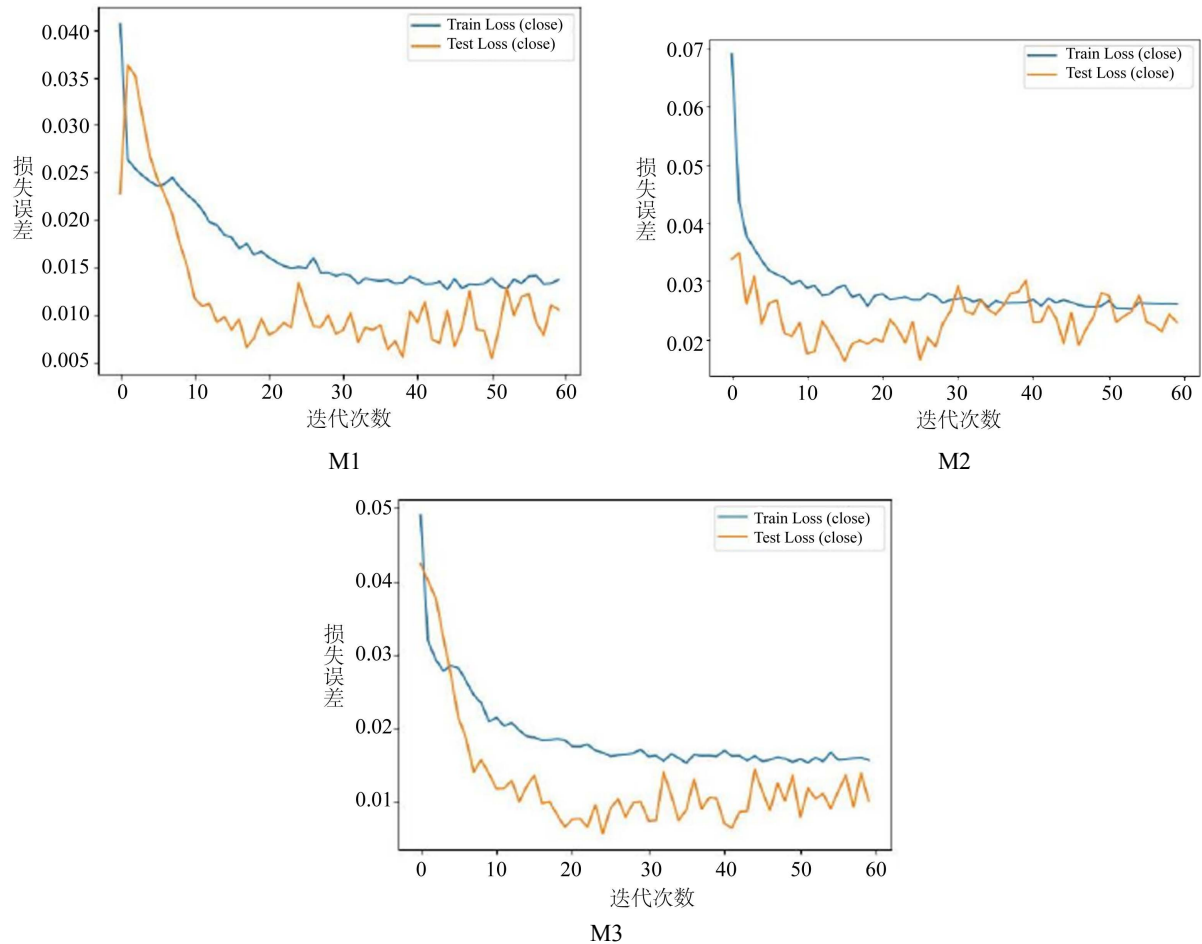
从表 2 可以看出, 以股票交易基本数据为输入特征的预测模型 M1 的决定系数为 88.9%, 均方根误差为 0.874, 所以使用股票交易基本数据进行股票价格预测是有效的; 以股票交易技术指标为输入特征的预测模型 M2 的模型决定系数为 85.8%, 均方根误差为 0.988, 与 M1 相比, 预测模型 M2 的预测效果明显下降。

模型 M1、M2、M3 的训练效果如图 3 所示。

### 4.3. 预测结果比较与分析

从图 3 可以看出, M2 相比于 M1 模型的损失误差大, 且对于训练数据预测模型很快达到稳定而验证数据比较不稳定。由股票市场理论可知, 股票交易技术指标含有的信息量有限, 且我国股票市场对股票价格有 10% 的涨跌幅度限制, 而股票交易技术指标没有这样的限制, 使得其波动幅度大。因此股票交易技术指标包含的信息有限和自身变幅度较大, 导致 M2 模型效果不好。以股票交易基本数据和股票交易

技术指标为输入特征的预测模型 M3 的决定系数为 88.2%，相比于预测模型 M1 稍微有所欠缺。对比后可以明显看出，输入特征维度较大的预测模型 M3 损失误差稍微增大。由此可推断出股票输入特征数据信息冗余时，会使预测模型训练过程学习能力降低。



**Figure 3.** Effect picture of model training  
**图 3.** 模型训练效果图

因此对于 LSTM 神经网络训练股票价格预测模型来说，适当增加股票的有效信息可以提高预测模型预测效果，但当增加的股票数据维度过大，股票信息冗余时，严重影响股票的预测效果。

本节将股票交易基本数据和处理后的股票交易技术指标数据[14]进行组合，作为预测模型的输入数据。选用 LSTM 神经网络构建股票价格预测模型，从而预测股票的价格。并对影响股票预测模型效果的输入特征、时间序列长度和网络结构等因素进行了详细的实验论证分析，进而能够训练出适应于股票市场且预测效果好的股票价格预测模型。从而为股票智能预测系统的智能预测模块，提供了准确的股票价格预测模型。

## 5. 结论

本文结合股票投资领域的理论知识[15]、机器学习方法以及软件开发技术[16]，设计了一个基于深度神经网络的股票智能预测系统。具体工作分为如下三点：

针对多因子量化选股的量化指标不足和忽略股票涨跌趋势的问题, 提出了一种基于股票趋势识别算法构建选股模型的方法。该方法选取股票的财务数据指标和技术性指标构建了选股指标集, 并根据改进的趋势识别算法对选股指标集进行趋势标识。

针对股票交易数据量大, 非线性关系复杂, 从而难以准确预测股票价格的问题, 本文采用 LSTM 深度神经网络对股票价格的预测进行研究。通过研究影响股票价格变动的因素, 基于股票交易基本数据利用技术指标公式获得股票技术指标数据。

以上述两点所提到的模型为基础, 设计并实现了一个基于深度神经网络的股票智能预测系统。根据需求分析和功能设计, 将该系统分为五个模块, 运用 Django 框架、scikit-learn 机器学习库和 Keras 深度学习库完成股票预测系统各模块功能。

未来将围绕以下方面对股票智能预测系统进行进一步的完善和提升: 在股票价格预测模型研究中, 本文选择交易技术指标数据和股票基本数据, 利用 LSTM 深度神经网络构建股票预测模型。虽然 LSTM 深度神经网络对于时间序列的预测是有效的, 能有效解决传统 RNN 的梯度消失的问题, 但是很容易造成模型过拟合。若加入其他有效的信息可能会进一步地改善股票预测模型的预测效果, 比如将政策或消息面的数据进行量化处理, 与股票数据结合作为股票价格预测模型的输入变量。

## 致 谢

衷心地感谢我的研究生导师李东喜老师, 李老师从生活和学习各个方面都给予了我关怀和引导。从论文阅读到论文撰写, 都有老师的把关与帮助。其次, 感谢给予转载和引用权的文献及研究思想和设想的所有者们, 正是借助你们的肩膀, 我才能够更好地完成论文的撰写。

## 参考文献

- [1] 陈晓云. 中国股票市场[M]. 上海: 商务印书馆, 1997: 1-30.
- [2] 兰强太. 基于主成分分析和 BP 神经网络算法的综合选股实证研究[D]: [硕士学位论文]. 广州: 暨南大学, 2017.
- [3] 马树才, 赵丰义. 中国沪深股市可预测性研究[J]. 当代经济管理, 2007, 29(3): 103-106.
- [4] 李国平. 中国股票市场的可预测性研究[J]. 高职论丛, 2006(3): 5-11.
- [5] 利亚涛. 上市公司股票估值与 A 股市场实证研究[D]: [博士学位论文]. 北京: 中国社会科学院, 2010.
- [6] 刘子婷. 基于 EGARCH 模型的股票价格预测[J]. 高师理科学刊, 2018, 38(6): 21-26.
- [7] 毛景慧. 基于 LSTM 深度神经网络的股市时间序列预测精度的影响因素研究[D]: [硕士学位论文]. 广州: 暨南大学, 2017.
- [8] Piotroski, J.D. (2000) Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research*, **38**, 1-41. <https://doi.org/10.2307/2672906>
- [9] Tudor, C. (2012) Active Portfolio Management on the Romanian Stock Market. *Procedia—Social and Behavioral Sciences*, **58**, 543-551. <https://doi.org/10.1016/j.sbspro.2012.09.1031>
- [10] Romo, M.J. (2014) Investment Decisions with Financial Constraints. Evidence from Spanish Firms. *Quantitative Finance*, **14**, 1079-1095. <https://doi.org/10.1080/14697688.2013.785635>
- [11] 骆桦, 秦艳艳. 中国股市动量与反转效应模型的研究[J]. 浙江理工大学学报(自然科学版), 2011, 28(4): 643-646.
- [12] 林德发, 杨潇宇. 跑赢沪深 300 指数的成分股组合构建——基于多因素模型的实证分析[J]. 中国商论, 2014(2): 83-84.
- [13] 柯原, 郑双阳. 价值投资与行业轮动相结合的量化选股策略研究[J]. 福建金融管理干部学院学报, 2014(1): 3-10.
- [14] 田凯, 刘永睿. 创业板基于 logistic 模型量化选股[J]. 现代商贸工业, 2017(1): 96-98.
- [15] 刘洋, 夏思雨, 胡思瑞, 等. GARP 数量化选股及马尔科夫链择时策略研究[J]. 金融与经济, 2016(5): 66-71.
- [16] Das, S.R. and Goyal, M. (2015) Computing Optimal Rebalance Frequency for Log-Optimal Portfolios in Linear Time. *Quantitative Finance*, **15**, 1191-1204. <https://doi.org/10.1080/14697688.2014.926020>