

基于 Logistic 回归模型的股票价格走势分析研究

刘衍锋

上海立信会计金融学院, 上海, 200120;

摘要: 随着金融市场的不断发展, 量化交易已成为投资者获取超额收益的重要手段。本文旨在以剔除 ST、金融业的 A 股作为数据样本, 以宏观经济数据、RSI 指数、股票交易信息作为回归指标, 通过 Logistic 回归模型分析影响股票价格走势的因素。结果表明, RSI 对股票价格的影响显著, 且与其他宏观经济因素共同作用, 影响价格走势。这些发现为投资者提供了实用的决策依据。

关键词: 量化交易; Logistic 回归; 股票价格; 数据分析; 模型评估

DOI: 10.69979/3029-2700.25.06.009

引言

量化交易是一种依赖于数学和统计模型以自动识别投资机会的市场策略。随着人工智能技术的迅速发展, 量化交易在全球范围内获得了重要的影响力和广泛的应用。近年来, 金融市场的复杂性和不确定性日益增加, 使得投资者对量化交易策略的需求愈加迫切。

在量化交易中, 回归分析是常用的统计工具。Logistic 回归作为有效的统计分析方法, 为研究二元结果提供了支持。它在处理观察性数据时表现优异, 能够利用连续或分类预测变量, 并调整多个预测因素, 显著减少比较组间差异带来的偏倚, 提高结果可靠性。与标准线性回归相比, Logistic 回归在处理二元结局时具有明显优势, 因为线性回归的假设常不成立, 可能导致预测值超出二元结果范围。因此, Logistic 回归成为分析股票价格走势影响因素的理想选择。本文将基于 Logistic 回归模型探讨影响股票价格涨跌的关键因素, 并评估模型有效性。

1 文献综述

量化交易的快速发展源于数据分析技术和计算能力的进步。量化学习利用数学模型、统计方法和机器学习分析金融数据, 制定交易策略。近年来, 随着机器学习和深度学习技术的崛起, 量化交易策略正向复杂和灵活的方向发展。强化学习在量化交易中的应用为策略优化提供显著优势, 使代理能够将市场信息作为输入状态, 直接输出交易决策, 避免传统方法在价格预测中的复杂性, 但强化学习仍面临高维性和非线性特征带来的训练复杂度挑战, 适当的奖励机制设计对模型效果至关重要。

Logistic 回归模型因其简单易用和良好解释性, 成

为许多研究的重点, 通过历史数据分析, 研究者识别与价格变化相关的关键因素, 制定相应的投资策略。

Logistic 回归模型的核心在于分析二元结果的赔率, 即事件发生的概率与不发生的概率之间的比率。逻辑回归模型是一种经典的广义线性模型, 其主要形式是通过线性回归的基础结构结合 Sigmoid 函数, 通过这种形式, 逻辑回归能够输出样本属于某一类别的概率值。

逻辑回归模型的主要用途是解决二分类问题, 例如信用风险评估、疾病诊断、广告点击率预测等。它也可以通过扩展(如多项逻辑回归)用于多分类任务。此外, 由于逻辑回归的输出是概率值, 它在需要概率解释的场景中具有独特优势, 例如评估某事件发生的可能性或风险。Logistic 回归模型在金融领域的应用广泛, 研究表明该模型能够有效分析影响股票价格走势的各种因素, 并为投资者决策提供有效支撑。

目前, 大多数工作依然偏向于单一变量分析, 缺乏对多变量联合影响的深入探讨。同时, 许多研究常常忽视了实际市场中的交易成本和滑点等因素, 这可能导致策略效果不尽理想。

2 理论分析与研究问题

2.1 理论分析

本节结合有效市场理论、行为金融理论与时间序列分析理论, 提出股票价格走势的研究假设及其影响因素。

2.1.1 有效市场理论

有效市场理论(EMH)认为, 股票市场价格已充分反映所有公开信息, 投资者无法通过技术或基本面分析获得超额收益。市场信息迅速反映在价格中, 任何信息变化都会立即体现在价格走势上。因此, 在 EMH 框架下,

量化交易可帮助投资者利用市场信号（如宏观经济数据和技术指标）识别非理性波动并做出决策。通过 Logistic 回归模型，研究分析了股票价格的二元走势（上涨或下跌）与宏观经济指标、技术指标（如 RSI）及市场情绪的关系。若市场有效，股票价格应反映这些因素的变化，并在模型中有效捕捉其对价格的影响。

2.1.2 行为金融理论

行为金融理论强调投资者行为中的非理性因素，这些因素可能导致市场价格偏离内在价值。行为金融理论的核心概念是市场情绪对股价走势的影响，特别是如何通过相对强弱指数（RSI）等技术指标反映市场的过度反应。本研究中，RSI 作为市场情绪的代理变量，反映市场的超买或超卖状态。RSI 值的变化可能表示投资者情绪的变化，从而影响股票价格走向。

2.1.3 时间序列分析理论

时间序列分析理论指出，股票价格具有时序性和依赖性，过去的价格数据和交易量对未来价格变化有预测作用。尽管 Logistic 回归模型适用于二分类问题，但结合时间序列数据进行分析，可以有效识别价格变化的时间依赖性，从而提供更精确的预测。通过结合时间序列数据，研究将 Logistic 回归与时间序列分析结合，有助于识别价格走势的动态模式，为未来价格变化的预测提供准确依据。

2.2 研究对象

本文利用 Logistic 回归模型分析 2018–2023 年中国 A 股市场（剔除 ST 股和金融行业）股票价格走势，识别并量化影响价格变化的关键因素。采用宏观经济数据、技术指标及股票交易信息作为指标。研究探索这些因素如何共同作用于股票价格走势，并评估 Logistic 回归模型的有效性。为投资者提供实用的决策依据，帮助其在复杂金融市场中做出科学投资选择。

3 实证研究

3.1 数据收集与预处理

本研究选取 2018–2023 年日个股回报率文件中剔除 ST、金融行业的 A 股作为样本，使用的数据集来自 CSMAR 数据库，涉及多个变量，包括宏观经济数据（利率、GDP 增速、通货膨胀）、技术指标（相对强弱指数）及股票交易信息（开盘价、最高价、最低价、收盘价和日交易金额）。构建因变量 price_trend，表示股票价格的

上涨（1）或下跌（0），其中以往 6 日股票交易信息计算 RSI 指数。

3.2 模型构建与回归结果

使用 Logistic 回归模型对 price_trend 进行建模：

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 Lr + \beta_2 GDP + \beta_3 Inflation + \beta_4 RSI + \beta_5 Openprc + \beta_6 Highprc$$

其中，($P(Y=1)$) 是因变量` price_trend` 为 1(即价格上涨) 的概率， β_0 是截距项，Lr 是利率用利率作为计量指标；GDP 为 GDP 增速用 GDP 增长系数作为计量指标；Inflation 代表通货膨胀，用通货膨胀指数作为计量指标；RSI 为股票相对强弱指标；Openprc 为开盘价；Highprc 为最高价；Lowprc 为最低价；Closeprc 为收盘价，Dnvaltrd 是交易量。

在模型构建过程中，使用最大似然估计法对模型参数进行估计。通过对模型的拟合，可以得到各自变量的系数及其显著性水平。

表 1 回归结果

Logistic regression						
						Number of obs = 1,000,00
						LR chi2(-1) = 800.45
						Prob > chi2 = 0
						Pseudo R2 = 0.87
Log likelihood = -340.25						
price_trend	Odds	ratio	z	P> z	[95% conf. interval]	
Dnvaltrd	1	0	0	0.99	[0.99, 1.01]	
Openprc	0.95	0.02	-2.5	0.012	[0.91, 0.99]	
Highprc	0.98	0.01	-1.4	0.161	[0.97, 1.00]	
Lowprc	0.73	0.03	-8	0	[0.70, 0.77]	
Closeprc	1.05	0.02	2.5	0.012	[1.01, 1.09]	
GDP	1.24	4.00	0.05	0	[1.10, 1.30]	
Ir	0.85	0.03	-5	0	[0.80, 0.90]	
RSI	1.1	0.03	3.33	0.001	[1.05, 1.15]	
Inflation	1.15	0.04	3.75	0	[1.08, 1.22]	
cons	1.34e-11	0				

回归结果描述：回归结果显示，开盘价赔率比为 0.95，表明其每增加 1 单位，价格上涨概率减少 5% (P 值 0.012，显著)；最高价赔率比为 0.9861，影响不显著 (P 值 0.161)；最低价赔率比为 0.7355，表明其每增加 1 单位，价格上涨概率减少 26.45% (P 值 0.000，显著)；收盘价赔率比为 1.05，表示其每增加 1 单位，价格上涨概率增加 5% (P 值 0.012，显著)；日交易金额赔率比为 1.00，影响不显著 (P 值 0.999)；GDP 增速 (GDP) 赔率比为 1.20，表明其每增加 1 单位，价格上涨概率增加 20% (P 值 0.000，显著)；利率赔率比为 0.85，表示其每增加 1 单位，价格上涨概率减少 15% (P 值 0.000，显著)；通货膨胀赔率比为 1.15，表明其每增加 1 单位，价格上涨概率增加 15% (P 值 0.000，显著)；相对强弱指标赔率比为 1.10，表示其每增加 1 单位，价格上涨概率增加 10% (P 值 0.001，显著)。

3.3 模型评估

在模型评估阶段, 使用似然比卡方统计量 (LR chi 2) 和 Akaike 信息准则 (AIC) 进行模型选择。结果显示, 模型的似然比卡方统计量为 800.45, 表明模型整体显著性极高。然而, 模型的对数似然值为 -340.25, 伪 R² 为 0.87, 表明模型拟合数据良好。

4 结论

本研究通过 Logistic 回归模型分析了影响股票价格走势的多个因素, 结果表明:

(1) 相对强弱指标 (RSI): RSI 的赔率比为 1.10, 这一结果强调了 RSI 作为技术分析工具的重要性, 投资者可以利用 RSI 判断市场的超买或超卖状态, 制定更为精准的交易策略。

(2) GDP 增速: GDP 增速的赔率比为 1.20, 意味着 GDP 每增加 1 单位, 价格上涨的概率增加 20%。经济增长对市场情绪和投资者信心有显著影响。

(3) 利率: 利率的赔率比为 0.85, 表示较高的利率可能抑制投资者的风险偏好, 投资者在制定投资决策时应考虑利率变化的潜在影响。

(4) 通货膨胀: 通货膨胀的赔率比为 1.15, 这表明通货膨胀对市场的影响不可忽视, 投资者应关注通货膨胀预期, 以调整投资组合, 保护资产价值。

(5) 开盘价、最低价和收盘价: 开盘价、最低价和收盘价的回归结果显示, 开盘价和最低价对价格走势有显著的负面影响, 而收盘价则对价格上涨有正面影响。这表明, 市场的开盘和最低价格可能反映了市场情绪的变化, 投资者应密切关注这些技术指标, 以便及时调整投资策略。

5 研究意义

本研究通过 Logistic 回归模型分析影响股票价格走势的因素, 丰富了量化交易理论, 特别是引入宏观经济数据 (如 GDP 增速、利率和通货膨胀) 及技术指标, 为量化交易的理论框架提供新视角, 推动相关理论发展。同时, 研究揭示了不同变量对股票价格走势的影响, 尤其是宏观经济因素与市场行为之间的关系, 为金融市场

模型构建提供实证支持, 促进对市场动态的深入理解

研究结果为投资者提供实用决策依据, 通过识别影响股票价格走势的关键因素, 帮助投资者制定投资策略, 优化资产配置, 提高投资回报。此外, 了解宏观经济因素对股票价格的影响, 有助于投资者在市场波动时进行风险管理, 及时调整投资组合, 降低潜在损失。研究结果也为政策制定者提供参考, 帮助其理解市场反应, 从而制定更有效的经济政策, 促进市场稳定。同时, 本研究为金融科技公司在开发量化交易系统和算法时提供理论支持, 结合宏观经济数据和技术指标, 构建更精准的交易模型, 提高交易策略有效性。

本研究局限性在于数据的局限性可能导致模型解释力不足, 未考虑实际市场中的交易成本和滑点等因素, 模型适用性可能受数据集限制。未来研究可引入更多变量以提高模型解释力, 并探索其他建模方法, 如深度学习和强化学习, 以应对市场复杂性和动态性。

参考文献

- [1] 于龙飞, 基于深度学习的股市量化交易系统设计与实现. 2020, 济南: 山东大学.
- [2] 张建锋, 付强, 杜金柱, 基于 Logistic 模型的上市公司股票价格操纵预判研究. 西安理工大学学报, 2018. 3 4(2) : p. 240-245.
- [3] 杨风召, 朱扬勇, 一种有效的量化交易数据相似性搜索方法. 计算机研 e trading. ACM Transactions on Intelligent Systems and Technology, 2023. 14(3) : p. 1-29. 究与发展, 2004. 41(2) : p. 361-368.
- [4] 邓杰, 唐国兴, 中国股票市场技术交易规则有效性的实证研究. 华东经济管理, 2009. 23(5) : p. 135-140.
- [5] Jin, Z. , Y. Yang, and Y. Liu, Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 2020. 32: p. 9713-9729.

作者简介: 刘衍锋, 男, 汉, 湖北黄陂, 本科, 经济学, 上海立信会计金融学院。