

基于推特情感分析预测股指回报率

赵丹阳

中国社会科学院大学国际政治经济学院

DOI:10.12238/ej.v7i6.1611

[摘要] 随着互联网经济的发展,互联网评论渗透在人们生活的方方面面。为了研究Twitter上关于新能源汽车的大量评论情绪是否是TESLA的股价波动产生的原因,本论文假设从Twitter收集到的用户情绪数据与TESLA股票市场价格相关。并采取CS新能源车指数399976和Twitter上的关于新能源汽车的评论情绪数据与CS新能源车指数的股价数据进行格兰杰因果检验。研究结果表明中国投资者情绪是指数价格变化的主要原因,且投资者对新能源汽车指数的正面冲击是短期的,长期来看情绪对股票价格的影响将会消失。

[关键词] 股价预测; 情绪分析; 行为金融学

中图分类号: F421.36 **文献标识码:** A

Predicting stock index returns based on Twitter sentiment analysis

Danyang Zhao

School of International Political Economy, University of Chinese Academy of Social Sciences

[Abstract] With the development of the Internet economy, Internet comments permeate every aspect of people's lives. In order to investigate whether the large amount of comment sentiment on Twitter about new energy vehicles is the cause of the fluctuation of TESLA's stock price, this thesis hypothesises that the user sentiment data collected from Twitter is correlated with the market price of TESLA's stock. We also take the CS New Energy Vehicle Index 399976 and the sentiment data of comments about new energy vehicles on Twitter and the stock price data of CS New Energy Vehicle Index to conduct Granger causality test. The results of the study show that Chinese investor sentiment is the main reason for the index price change, and the positive impact of investors on the new energy vehicle index is short-term, and the impact of sentiment on stock price will disappear in the long run.

[Key words] stock price forecasting; sentiment analysis; behavioural finance

1 文献综述

关于投资者情绪对市场整体影响的研究主要基于心理学和行为学理论,以确定两者之间是否存在互动机制。Lee et al. (2002) 认为^[1], 投资者情绪的变化代表着系统性风险, 应给予相应的补偿。Benhabib et al. (2016) 也发现, 投资者情绪的影响导致持续产出, 从而助长了持续的市场表现^[2]。宋军和吴冲(2003)的研究表明, 投资者在股市中的心理会影响其行为结果。崔晓磊等(2014)、方勇等(2010)也得出了类似的结论。关于投资者情绪对市场预测能力的影响, DeLong et al. (1990) 提出了噪声交易的基本理论模型(DSSW模型), 并指出噪声交易者正是反馈交易者, 利用DSSW模型, 交易者可以对历史收益进行连续预测^[3]。Ergungor (2016) 认为现货市场措施不会影响投资者的风险意识, 但会引导交易行为的变化^[4]。关于投资者情绪对市场收益的影响, 主要有两种观点。一种观点认为投资者情绪对股票收益

有影响。Brown和Cliff (2004年) 对收益相关性进行了系统研究, 发现投资者情绪与近期常用的其他市场收益衡量指标相关。2012年, 他们得出了一个不同的结论, 即投资者情绪对股票回报率没有影响。Salt和Statman (1988年) 发现, 在4周、26周和52周内, 投资者情绪指数与道琼斯工业平均指数之间没有显著相关性。杨洋和王迪攀(2010)、杨墨竹(2013)、迟立旭等(2012) 也对我国投资者情绪对股市收益的影响进行了研究^[5]。在Ben-Refhael和Candel (2012年) 的研究中, 他们发现股票基金与投资者情绪波动呈正相关, 这在小型股票和成长型股票基金中表现得更为明显^[6]。Baker和Wurgler (2006年)、Yu和Yuan (2011年) 的研究也具有一定的参考价值。因此, 国内也开始出现这方面的研究(蒋玉梅和王明照, 2010年; 刘志远和金光辉, 2013年), 并得出了投资者情绪对不同股票收益影响不同的结论^[7]。夏普(Sharpe, 1964年) 和林特纳(Lintner, 1965年) 的资本资产定价

模型(资本资产定价模型)假定投资者是同质的、无差别的,每个投资者的投资策略都是相同的^[8]。但实际上,投资者在信念、偏好、收益和其他特征方面是异质的。因此,传统的资产定价模型无法真实反映现实市场,也导致其无法对“股票溢价之谜”、“利率之谜”等金融异象做出合理解释。在投资者异质性的基础上,行为金融学研究者对投资者异质性与资产定价之间的关系进行了研究。Campbell(1999)将投资者异质性分为四类:异质性偏好、异质性约束、异质性收益和异质性信念^[9]。其中,异质信念是研究投资者异质性的基础。本文对异质信念调整对投资者情绪影响机制的探讨还很初步,没有涉及其本质。可以看出,虽然所研究的论文都初步探讨了投资者情绪和异质信念在资产定价中的作用,但仍存在一些不足。首先,在这些研究论文中,投资者情绪和异质信念仍然只是影响资产定价,而没有考虑它们之间的相互作用。其次,这些论文主要讨论了投资者情绪对异质信念或异质信念对投资者情绪的单向影响。事实上,两者之间的关系应该是更为复杂的互动影响关系。最后,投资者情绪和异质信念的交互作用对资产定价的复杂影响,应该表现出不同于这两个因素分别对资产定价影响的机制和方法。理解综合影响的机制和方法显然不应该是单个影响机制和方法的简单叠加,需要考虑一些新的因素和结构。基于以上分析,我们认为市场信息会影响异质投资者信念的调整,而异质投资者信念的调整又会导致投资者情绪的变化。反过来,投资者情绪的变化也会影响投资者异质信念的调整^[10]。投资者异质信念与投资者情绪之间的这种动态互动机制共同形成了对资产定价的影响。

2 理论分析

2.1 情绪指数 (IS)

投资者情绪指数是在收集网络能反映投资者情绪的上亿条金融文本大数据的基础上,使用深度学习方法,度量文本信息反映的中国投资者情绪。学术研究中对于投资者情绪 (Investors' Sentiment) 的定义主要有两类: (1) 噪声交易者关于股票未来股价预期偏离理性套利者信念的程度 (DeLong等, 1990)。 (2) 投资者基于对资产未来现金流和投资风险的预期而形成的一种信念 (Baker和Wurgler, 2006)。国外对于投资者情绪指标的选取主要包括以下三类: (1) 客观指标。包括: 封闭式基金折价, IPO发行量及首日收益, 交易量, 共同基金净赎回, 股票发行与债券发行比例等。 (2) 主观指标。包括: 个体投资者协会指数, 投资者智能指数, 证券分析师情绪指数, 消费者信心指数。 (3) 复合指标。采用主成分分析法, 将客观指标和主观指标等结合在一起。本文采用第三种方法。通过主成分分析构建的投资者情绪指数可以用来预测股市的高峰和底部。本文以特斯拉股评信息为研究对象, 利用SnowNlp文本分析库对股评标题进行情感分析。在扫描数据的基础上, 结合股票的市场表现, 对股评舆情与股票市场表现之间的关系进行回归分析。结果显示, 股评标题得分与股市表现的相关性最高。股票成交量、换手率和成交笔数越高, 阅读量与成交笔数呈正相关, 而评论数量与股市表现无显著相关性。本文的研究对投资者的量化投资决策具有一定的指导意义。

2.2 BAPM理论

BAPM即行为资产定价模型 (Behavioral Asset Pricing Model), 是Shefrin和Statman在1994年挑战资本资产定价模型, 提出了行为资产定价模型。1999年, 两人又挑战资产组合理论, 提出了行为组合理论。BAPM将投资者分为信息交易者和噪声交易者两种类型。信息交易者即CAPM下的投资者, 他们从不犯认知错误, 而且不同个体之间表现有不错的统计均方差性; 噪声交易者则是那些处于CAPM框架之外的投资者。他们时常犯认知错误, 不同个体之间具有显著的异方差性。在BAPM模型中, 证券的预期收益由其“行为贝塔”决定。非理性交易者, 即噪音交易者, 并不具备投资者理应具备的知识和行为储备。他们并不偏好均值方差, 而且经常偏离CAPM。因此, 在BAPM中, 与CAPM不同, 决定证券预期收益的系数 β 主要与行为有关。这种行为 β 与均值和方差的切线有效组合有关, 而不是与市场组合有关。BAPM模型可以用如下公式表示:

$$r_{it} = r_{ft} + (\beta_i B + NTR) (r_{mt} - r_{ft}) \quad (1)$$

其中, r_{mt} 为市场组合收益率, 通常以市场指数的收益率代替。NTR表示噪声交易者风险, 上式说明由于NTR的存在, 传统贝塔 β_{iC} 要高于行为贝塔 β_{iB} , 并存在关系式 $NTR = \beta_{iC} - \beta_{iB}$ 。其中NTR为额外的系统性风险——噪声交易者风险 (Noise Trader Risk, 简称NTR)。

3 实证分析

根据Twitter对新能源汽车行业的评价, 正面情绪占 89%。新能源汽车行业的暴涨很可能受到推特上投资者积极情绪的影响。为了证实这一假设, 我们还将对Twitter情绪是否影响中证新能源汽车3999796指数进行实证分析。首先, 利用Twitter API收集2021年10月15日至2022年1月30日中国人在Twitter 上对“新能源汽车”的评论信息, 利用SnowNlp进行情绪分析, 得到投资者情绪指数, 并收集相应时间段的中证新能源汽车指数收盘价和中国消费者信心指数信息。然后, 利用VAR模型, 将投资者情绪指数和中国消费者信心指数作为自变量, 将中证新能源汽车股票指数作为因变量。进行格兰杰因果检验, 建立预测模型。

3.1 稳定性检验

如图1所示, VAR模型的所有特征根都小于1, 这意味着单位根落在单位圆内。因此, 模型通过了稳定性检验, 可以进行脉冲分析和方差分解。

Variables	ADF	P	Стационарность
Index	-2.983	0.0365	✓
SI	-10.634	0.0000	✓
Confidence	-10.317	0.0000	✓

图1 建立VAR模型

对所有变量进行单位根检验后, 拒绝了初始假设。所有p值均小于0.1, 证明了这一点。因此, 因变量和自变量, 即控制变量数据是静态的。

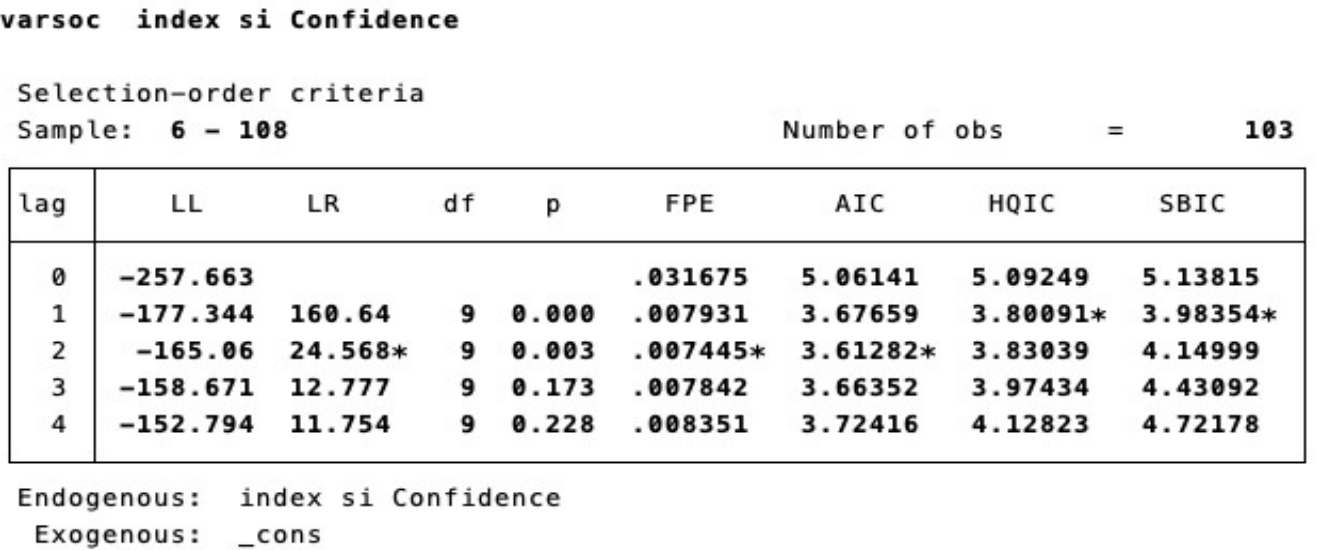


图 2 滞后的最佳顺序

3.2最佳滞后阶数的确定

根据AIC准则、SBIC准则、LR准则等的联合验证,确定lny、lnx所建VAR模型的最优滞后阶数为2。结果见图2。

3.3方差分解

首先,我们提前一步预测指数,预测方差来自指数本身。然而,对第10期的预测显示,虽然指数的大部分预测方差来自指数本身,但随着时间的推移,预测方差会逐渐减少。从图3中可以看出,方差si对指数的贡献系数逐渐增大,因此si会对指数产生影响。

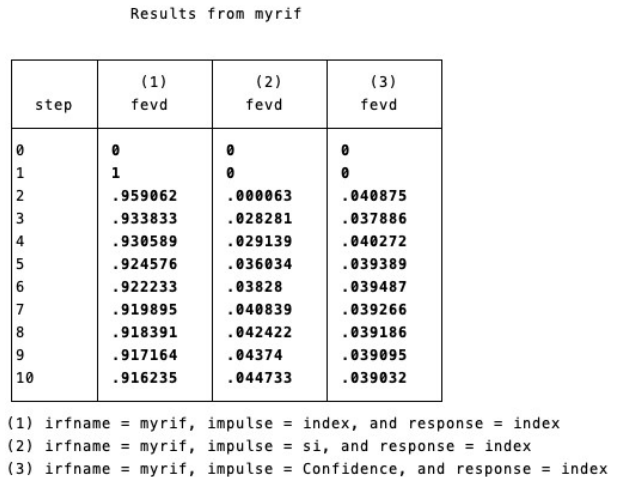


图3 方差分解

3.4格兰杰因果检验

中证新能源汽车指数与推特上的中国新能源汽车评论具有更强的因果关系。其中,推特情绪是新能源汽车指数的主因,而新能源汽车指数并非推特情绪的主因。如图4所示,在5%的显著性水平下,证实了中国投资者情绪是指数价格变化的主要原因,但新能源汽车(EV)指数不是影响投资者情绪的主要原因。

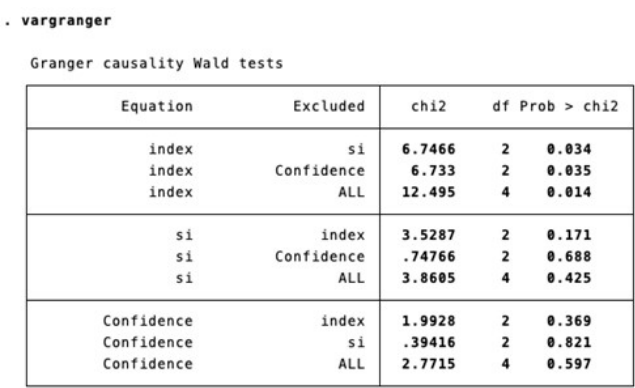


图4 情绪、信心与EA指数价格之间的关系

3.5预测

如图5所示,所有自变量和因变量之间都存在相关性。

在延迟2期的情况下,可以构建以下基于情绪的股指价格预测模型(2):

$$Index = 0.5084691 * Index(t - 1) + 0.3629554 * Index(t - 2)) + 26.82837 * SI(t - 1) + 359.5886 * SI(t - 2) - 68362.39 * Confidence(t - 1) + 3259.347 * Confidence(t - 2) + 662.6559. (2)$$

从一步样本外预测(见图6)可以看出,该样本无论显示的是上一区间的预测还是下一区间的预测,均为正值,因此该模型具有较强的预测能力。

根据预测可以看出,由于投资者情绪上升,投资者信心增强,新能源汽车产业指数将随之上升。由于市场景气指数和投资者信心指数均为正数,且二阶滞后,所以预期是向上的,新能源汽车产业指数会随着市场景气度的上升而继续上升,会随着投资者信心的上升而继续上升。因此,预期向好,新能源汽车产业投资应持乐观态度。投资新能源汽车指数是非理性的。舆论鼓励许多投资者买入股票。新能源汽车行业正在形成一个巨大的泡

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
index	index						
	L1.	.5084691	.0998802	5.09	0.000	.3127076	.7042306
	L2.	.3629554	.0986408	3.68	0.000	.169623	.5562877
	si						
	L1.	26.82837	139.5691	0.19	0.848	-246.722	300.3788
	L2.	359.5886	138.4532	2.60	0.009	88.22537	630.9519
Confidence							
	L1.	-68362.39	26463.9	-2.58	0.010	-120230.7	-16494.1
	L2.	3259.347	23615.43	0.14	0.890	-43026.05	49544.74
_cons		662.6559	283.3908	2.34	0.019	107.2202	1218.092

图5 预测模型

	t	Confidence	index_f	si_f	Confidence_f	index_f_L	index_f_U	index_f_se	si_f_L	si_f_U	si_f_se	Confidence~L	Confidence~U	Confidence~e
102	102	0	*	*	*	*	*	*	*	*	*	*	*	*
103	103	0	*	*	*	*	*	*	*	*	*	*	*	*
104	104	0	*	*	*	*	*	*	*	*	*	*	*	*
105	105	0	*	*	*	*	*	*	*	*	*	*	*	*
106	106	0	*	*	*	*	*	*	*	*	*	*	*	*
107	107	0	*	*	*	*	*	*	*	*	*	*	*	*
108	108	0	5139.02	.13672383	0	*	*	*	*	*	*	*	*	*
109	109	*	5261.026	.12074521	-.00036116	4644.3497	5877.7024	314.63657	-.30553704	.54702745	.21749494	-.00291749	.00219517	.00130427

图6 一步式样本外预测

沫。何时破灭无法预测。虽然跟风投资可以获利,但从长远来看,新能源汽车行业的股价还是会回归其价值,股价会下跌。

4 结论

本研究论文探讨了如何利用市场参与者的情绪状态信息来预测股指回报率的问题。论文主题的现实意义在于,当前新能源汽车股价的快速变化并非理性投资的结果。新能源汽车的股价在很大程度上受舆论操纵。微博上关于新能源汽车的评论释放了许多正面信息,导致新能源汽车股价上涨。由此可见,社交媒体上的评论情绪会影响股票价格。论文研究的理论意义如下:首先,实证表明市场是无效率的,情绪是预测股指的重要因素。其次,利用情感分析技术对微博评论进行情感分析可以获得股价预测的替代数据

[参考文献]

[1]Author Index/Текст:электронный // Applied Hydro - Aeromechanics in Oil and Gas Drilling / Book Authors: _:n138. — Hoboken, NJ, USA: John Wiley & Sons,Inc., 2010.— С.431 - 432.—URL:https://onlinelibrary.wiley.com/doi/10.1002/9780470542392.indauth(дата обращения:02.05.2022).

[2]Bagheri H. Twitter Sentiment Analysis / H. Bagheri, M. J.Islam.—Open Science Framework,2017.—URL:https://osf.io/6x

c4y(дата обращения:02.05.2022).— Текст:электронный.

[3]Bing L. Public Sentiment Analysis in Twitter Data for Prediction of a Company' s Stock Price Movements / L. Bing, K.C.C.Chan,C.Ou.— Текст:электронный// 2014 IEEE 11th International Conference on e-Business Engineering 2014 IEEE 11th International Conference on e - Business Engineering(ICEBE).—Guangzhou,China:IEEE,2014.— С.232-239.—URL:http://ieeexplore.ieee.org/document/6982085/(дата обращения:02.05.2022).

[4]Chen L.—P.Using Machine Learning Algorithms on Prediction of Stock Price / L. - P. Chen // Journal of Modeling and Optimization.—2020.— Т.12.— №2.— С.84 - 99.

[5]Comparing Methods for Twitter Sentiment Analysis: / E. Psomakelis, K.Tserpes, D. Anagnostopoulos, T. Varvarigou. — Текст:электронный//Proceedings of the International Conference on Knowledge Discovery and Information Retrieval International Conference on Knowledge Discovery and Information Retrieval.—Rome,Italy:SCITEPRESS - Science and and Technology Publications, 2014. — Comparing Methods for Twitter Sentiment Analysis.— С.225-232.URL:http://www.

scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/000507530225 0232 (д а т а о б р а щ е н и я:02.05.2022).

[6]Dinsoreanu M.Unsupervised Twitter Sentiment Classification:/M.Dinsoreanu,A.Bacu.—Т е к с т : э л е к т р о н н ы й //Proceedings of the International Conference on Knowledge Management and Information Sharing International Conference on Knowledge Management and Information Sharing. — Rome,Italy:SCITEPRESS – Science and and Technology Publications,2014.—Unsupervised Twitter Sentiment Classification.— С . 220 – 227.—URL: <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005079002200227>(д а т а о б р а щ е н и я :02.05.2022).

[7]El Kalush Y. What is the impact of the crude oil price index on the performance of oil and gas firms? : Master of Business Administration/Y.El Kalush. — University of Northern British Columbia,2009.—URL: <http://unbc.arcabc.ca/islandora/object/unbc%3A16425>(д а т а о б р а щ е н и я : 02.05.20

22).—Т е к с т : э л е к т р о н н ы й .

[8]Gupta R. Forecasting oil and stock returns with a Qual VAR using over 150years off data / R. Gupta, M. Wohar // Energy Economics.—2017.—Т .62.— С .181–186.

[9]Hatch O. G. The Politic of Supply – Side Economics / O. G.Hatch.—Т е к с т : э л е к т р о н н ы й //Review.—1981.— Т .63.—URL:<https://research.stlouisfed.org/publications/review/1981/05/01/the-politic-of-supply-side-economics>(д а т а о б р а щ е н и я :02.05.2022).

[10]Xiao W. Sentiment analysis in twitter: M.Phil. / W. Xiao.—Clear Water Bay, Kowloon, Hong Kong : The Hong Kong University of Science and Technology,2015.—URL:<http://lbezone.ust.hk/bib/b1514648> (date accessed: 02.05.2022).—Text: electronic.

作者简介:

赵丹阳(1994--),女,汉族,河北石家庄市人,中国社会科学院大学国际政治经济学院,博士研究生在读,研究方向:世界经济。