

中国数字经济指数与研报情绪间的动态关联

——基于分位数法的实证研究

张兆威

澳门城市大学金融学院

DOI:10.12238/ej.v8i4.2487

[摘要] 本研究运用通义大语言模型对2023年7月1日至2024年6月30日期间数字经济相关研究报告内容进行情绪分析。并采用分位数动态关联法对数字经济相关研报情绪与中证数字经济主题指数收益率之间的连通性进行实证分析。研究发现,在极端市场情况下,研报情绪与数字经济指数收益率之间的连通性显著增强且呈现对称特征,2024年后反向相关分位数连通性多于直接相关分位数连通性。本研究填补了研报情绪与数字经济指数收益率间动态关联研究的空白并为投资者、企业以及政策制定者在关注数字经济发展时提供新视角。

[关键词] 数字经济; 研究报告; 文本分析; 分位数动态关联

中图分类号: F0 **文献标识码:** A

Dynamic connectedness between China's Digital Economy Index and the Sentiment of Research Reports – An Empirical Study Based on the Quantile Connectedness Approach

Zhaowei Zhang

Faculty of Finance, City University of Macau

[Abstract] This study utilizes the Tongyi large language model to conduct sentiment analysis on the content of digital economy-related research reports from July 1, 2023 to June 30, 2024. It employs the Quantile Connectedness Approach to empirically analyze the connectedness between the sentiment of digital economy-related research reports and the yield of the CSI Digital Economy Theme Index. The study finds that in extreme market conditions, the connectedness between the sentiment of research reports and the yield of the digital economy index is significantly enhanced and exhibits symmetrical characteristics. After 2024, reversely-related quantiles connectedness is more than the direct-related quantiles connectedness. This study fills the gap in the research on the dynamic connectedness between the sentiment of research reports and the yield of the digital economy index and provides a new perspective for investors, enterprises, and policymakers when focusing on the development of the digital economy.

[Key words] Digital Economy; Research Report; Textual Analysis; Quantile Connectedness Approach

引言

在当今时代,信息技术的飞速发展深刻地改变了全球经济格局,数字经济应运而生并成为推动经济增长的关键力量。随着5G、人工智能、区块链、云计算、大数据等新技术不断涌现,为数字经济发展提供强有力的支撑,经济活动的各个环节正经历着前所未有的数字化转型。近年来全球数字经济规模持续扩张,其增长速度远超传统经济领域。

数字经济是一种新型经济形态,它以数字化的信息和知识作为关键生产要素,现代信息网络为重要载体,通过数字技术与

实体经济深度融合,驱动决策优化,实现价值创造。在数字经济中,数据成为核心价值源泉。海量的数据被收集、存储、分析和应用,企业和组织能够依据数据洞察市场趋势、消费者偏好,从而精准定位产品和服务,优化资源配置,提升生产效率和经济效益。

近年来,中国数字经济取得了令人瞩目的发展成就。从规模上看,2023年其规模达到53.9万亿元,这彰显了数字经济在我国经济中的地位日益重要。从2012年至2021年,我国数字经济规模从11万亿元增长到超45万亿元,10年间增长3.031倍,年均增长

率达到17.09%; 2022年更是达到50.2万亿元, 占GDP比重提升至41.5%。数字经济的规模增长势头强劲且稳定, 持续为国民经济的发展注入新动力^[1]。企业层面同样成果显著, 中央企业成立数字科技类公司近500家, 展现出大型企业对于数字经济领域的积极布局 and 战略重视。除此之外, 我国工业互联网已覆盖41个工业大类, 培育421家国家级示范工厂等, 这些数据充分说明企业在数字经济发展进程中积极创新、深入实践, 通过数字化转型提升自身竞争力, 同时也推动了数字技术与实体经济的深度融合, 促进了产业升级和创新发展。

研究报告是反映市场动态与行业趋势的重要窗口。在数字经济快速发展的进程中, 市场情况瞬息万变, 而透过研究报告传递出的情绪能够及时捕捉行业内的信息。对这些信息的分析有助于企业、投资者以及政策制定者捕捉数字经济领域的发展态势, 进而做出更为科学合理的决策。

为探究研报内容信息对于中国数字经济发展的意义, 本研究将考察数字经济相关研报情绪与数字经济指数收益率之间的动态关联关系。运用通义大语言模型进行文本情绪分析, 并结合分位数动态关联法深入探究二者关系。在理论层面为数字经济领域研究提供新视角与实证依据。

1 数据与方法

1.1 数据

本文研究对象为2023年7月1日至2024年6月30日期间中证数字经济主题指数与数字经济相关研究报告, 来源均为中国经济金融研究数据库CSMAR。选择该时间段的原因是排除新冠疫情对研究的影响, 而世界卫生组织显示新冠疫情结束的时间是2023年5月5日。最终将不完整或者格式错误的数据排除之后获得25444条研究报告数据, 而对于数字经济指数取收盘价的对数收益率作为指标。

1.2 情绪检测方法

自然语言处理任务中, 通常需要对文本进行特征提取和表示, 将其转化为计算机可理解的结构化数据。文本表示方法可分为基于词典的信息提取、基于统计原理的方法、词编码器和句编码器^[2], 而大语言模型有强大的自然语言处理能力^[3]。本文将采用通义大语言模型对数字经济相关研究报告进行文本分析。相比之前的方法, 大语言模型能够精准剖析各类自然语言表达, 无论是复杂的句子结构还是微妙的语义关系都能准确把握。模型还具有较强的适应性和泛化能力, 面对新的、未曾见过的文本数据时, 依然能够进行有效的情感分析。本研究将研究报告数据通过python API注意将研究报告内容提交给通义大语言模型进行分析, 提示词设计如下:

“我将提供中国数字经济相关的研究报告内容, 请你扮演专业的经济学家分析其传达的情感态度, 如果是积极的情绪输出1, 没有明显情绪的输出0, 而消极的情绪输出-1”。

1.3 连通性检验方法

在本研究中, 我们首先采用了Ando等人提出的分位数连通性方法。其主要作用在于能够深入探究变量间冲击传播的分位

数依赖特性, 可捕捉不同分位数(如极端值变化)下的连通性模式, 这有助于我们理解在极端市场条件下变量间的相互影响机制。该方法相较于传统方法能更全面地刻画变量间关系, 不再局限于均值或简单相关性分析而是考虑到了不同分位数下的变化情况, 提供了更丰富的信息。此外, 在处理金融市场数据时, 能够有效应对数据的非正态性和异方差性等问题, 使得分析结果更加稳健可靠。

而后本研究还将采用Gabauera和Stenfors提出的分位数对分位数连通性方法^[4]。它突破了仅关注单一分位数的限制, 能够全面地考虑多个分位数之间的溢出效应。另一方面, 它在处理变量相关性方面具有更高的灵活性, 无论是正相关、负相关还是非线性相关关系, 都能通过该方法得到更有效的评估和分析。

2 实证分析

根据大语言模型分析研究报告内容后返回的情绪值, 本文依据研报发布的时间对研究时段内每日所有研究报告产生的情绪进行平均得到当日的情绪值。情绪值序列和数字经济指数收益率序列如图1所示。

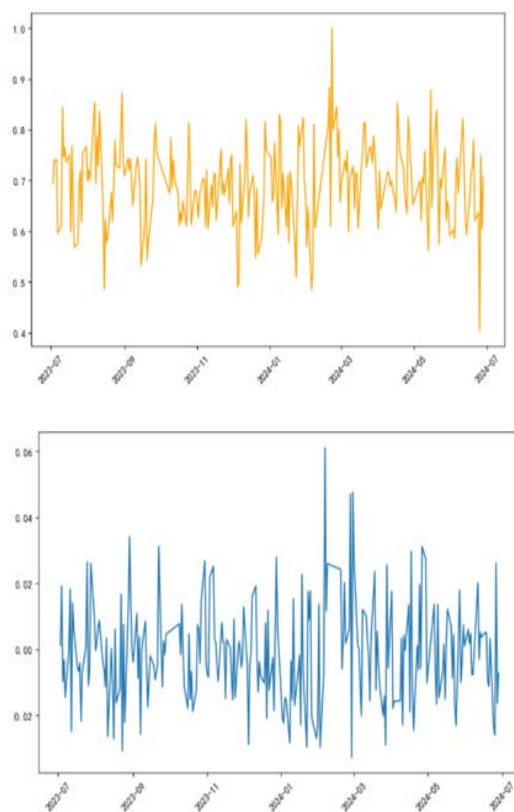


图1 研报情绪值(左)和数字经济指数收益率(右)变化图

在使用分位数动态关联法建立模型之前需要情绪与指数回报率数据均平稳且选择最优阶数。本研究根据ADF法进行检验的结果如图2所示, 确定两列研究数据平稳。并根据贝叶斯信息准则(图3)选取最优阶数为滞后一阶。

使用分位数动态关联法对两列序列进行研究, 图4展示了研报情绪与数字经济指数之间的动态总连通性结果。图中右侧的

位数水平。当两个变量同时处于较高(低)分位数时称为直接相关分位数连通性, 而当一个变量处于较高(低)分位数而另一个处于较低(高)分位数时称为反向相关分位数连通性。在图5中, 我们可以看到不同分位数组合下二者之间的连通性表现, 极端值情况下连通性较高而平均水平下较低。例如, 当研报情绪处于较低分位数(0.15分位数以下)而数字经济指数处于较高分位数(0.85分位数以上)时反向相关分位数连通性指数为78.8, 这意味着这种情况下, 研报情绪的负向极端变化与数字经济指数的正向极端变化之间有着较为紧密的关联。反向相关分位数连通性让分位数补充了仅关注直接相关分位数连通性忽略的因负相关而在反向极端情况下产生的强连通性现象。

Augmented Dickey-Fuller test for unit root

Variable: **log_return**

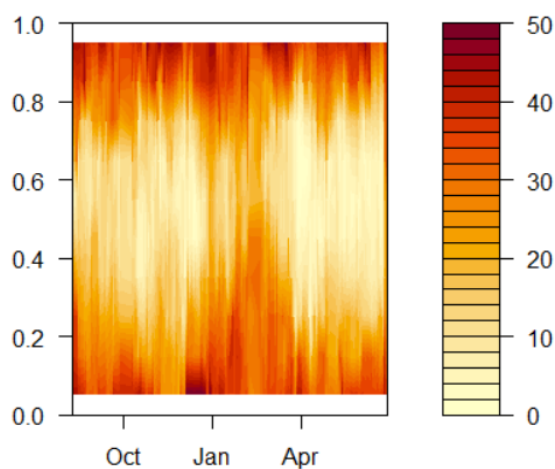
Number of obs = **239**

Number of lags = **1**

H0: Random walk without drift, d = 0

Test statistic	Dickey-Fuller critical value			
	1%	5%	10%	
Z(t)	-10.549	-3.464	-2.881	-2.571

MacKinnon approximate p-value for Z(t) = **0.0000**.



```
Augmented Dickey-Fuller test for unit root
```

Variable: **sentiment** Number of obs = 239
Number of lags = 1

H0: Random walk without drift, d = 0

	Test statistic	1%	5%	10%
Z(t)	-9.001	-3.464	-2.881	-2.571

MacKinnon approximate p-value for Z(t) = 0.0000.

Lag-order selection criteria

Sample: 11 thru 241						Number of obs = 231		
Log	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	880.481				1.7e-06	-7.60589	-7.59387	-7.57609
1	897.449	33.937	4	0.000	1.5e-06	-7.71817	-7.68211*	-7.62876*
2	901.567	8.2354	4	0.083	1.5e-06	-7.71919	-7.65909	-7.57017
3	903.796	4.4588	4	0.347	1.5e-06	-7.70386	-7.61971	-7.49523
4	910.061	12.53	4	0.014	1.5e-06*	-7.72347*	-7.61528	-7.45523
5	911.037	1.951	4	0.745	1.6e-06	-7.69729	-7.56505	-7.36944
6	914.573	7.0735	4	0.132	1.6e-06	-7.69328	-7.537	-7.30582
7	917.202	5.2571	4	0.262	1.6e-06	-7.6814	-7.50108	-7.23434
8	922.158	9.9126*	4	0.042	1.6e-06	-7.68968	-7.48532	-7.18301
9	926.129	7.9404	4	0.094	1.6e-06	-7.68942	-7.46102	-7.12314
10	929.671	7.0854	4	0.131	1.6e-06	-7.68547	-7.43302	-7.05957

```
* optimal lag
Endogenous: Log_return sentiment
Exogenous:  cons
```

A heatmap illustrating the relationship between Sentiment (X-axis) and Log_return (Y-axis). The color intensity represents the TCI (Tactical Clustering Index) value, with a legend on the right showing a scale from 20 (lightest blue) to 60 (darkest blue). Numerical TCI values are displayed within each cell of the heatmap.

Log_return \ Sentiment	0.0	0.3	0.6	0.9
0.9	78.8	62.6	45.3	47
0.6	40.4	30.3	23.5	24.8
0.3	19.8	19.1	19.8	16.2
0.0	48.7	34.9	25.9	25.8
0.0	70.2	50.6	35.3	39.2

图6展示了研究样本期间反向相关分位数连通性(红色)和直接相关分位数连通性(绿色)以及两者之差(蓝色)随时间变化情况。从图中可以看到在2023年7月至2024年1月期间研报情绪与数字经济指数之间的直接相关分位数连通性较多而进入2024年后反向相关分位数连通性占主导地位。从宏观环境来看,2024年以来尽管全球经济形势复杂,但数字经济仍保持快速发展。中

国在数字经济领域的发展十分显著，相关政策不断出台，各地积极探索数据治理、确权及使用原则的政策创新。说明在复杂的经济形势下，研报情绪与中国数字经济指数间的反向分位数连通性值得引起广泛关注。



图6 2023. 7. 1至2024. 6. 30期间直接相关和反向相关分位数连通性变化图

3 结论

本研究运用分位数法对研报情绪与中国数字经济指数进行

实证分析。研究显示，研报情绪与数字经济指数在极端分位数时连通性极强且呈对称性。同时，通过分位数对分位数连通性方法发现，反向相关分位数在2024 年后占据主导位置值得引起关注。投资者在市场情况面对反向关联占主导时需更加谨慎，关注数字经济指数与研报情绪联动带来的风险与机会；企业应重视不同市场状态下研报情绪对自身数字经济业务的影响，依据市场连通性特点优化战略决策；政策制定者制定政策时需考虑市场极端情况及动态变化，灵活调整政策导向，推动数字经济健康发展。

[参考文献]

[1]谢易和.我国数字经济发展特征问题与对策建议[J].西部财会,2024(10):76-79.

[2]李合龙,任昌松,柳欣茹,等.金融市场文本情绪研究综述[J].数据分析与知识发现,2023,7(12):22-39.

[3]林建浩,孙乐轩.大语言模型与经济金融文本分析:基本原理、应用场景与研究展望[J].计量经济学报,2025,5(1):1-34.

[4]Gabauer,D.,&Stenfors,A.(2024).Quantile-on-quantileconnectednessmeasures:Evidence from the US treasury yield curve.Finance Research Letters,60,104852.

作者简介：

张兆威(1996--),男,汉族,天津人,博士生在读,金融学专业,研究方向为数字经济,社交媒体与经济。