

基于 Spearman 相关系数法对金融市场变化影响因素的探究

李臻^{1, a, *}, 胡焱禹^{1, b}

(1. 中国辽宁省葫芦岛市兴城, 125000, 辽宁工程技术大学安全科学与工程学院, 安全工程;

* a. 2513598481@qq.com; b. 2328831850@qq.com)

摘要: 为了更好地研究金融市场的影响因素, 以收集到的数据为研究对象, 首先对数据进行预处理, 即使用最邻近插值法对缺失值进行插补, 在验证数据不符合正态分布以后, 选择使用 Spearman 相关系数法对金融市场变化的影响因素进行分析, 通过热力学图发现环境因素与整体股市表现较强的相关性, 并对最后的相关性结果进行了 Kendall's W 检验。该研究结果可为金融从业者提供较为科学的指导依据。

关键词: 影响因素; 数据; 最邻近插值; Spearman 相关系数法; Kendall's W 检验

引言

在当前全球气候变化不断加剧的背景下, 对环境因素对经济和金融市场的影响日益引起人们的重视。这一趋势并非仅仅停留在环保领域, 而是逐渐渗透到经济学和金融学的研究中。极端天气事件, 如飓风、洪水和干旱, 以及空气质量的变化, 不仅直接影响了人们的生活质量, 更在全球范围内产生了深刻的经济和金融市场影响[1-2]。

首先, 极端天气事件可能导致农作物产量减少, 从而影响食品价格和相关产业的股票表现。投资者不得不考虑气候变化对这些行业的潜在风险, 从而调整他们的投资组合。其次, 较差的空气质量可能导致公共卫生成本增加, 进而影响医疗保健部门和相关保险产品的需求。同时, 需求增加也会影响相关的保险产品, 投资者和保险公司需要考虑这些潜在的风险, 以及它们对医疗和保险行业的影响[3]。

总体而言, 全球气候变化对经济和金融市场的影响已经超越了环保问题, 成为一个需要全球性关注和研究的重要议题。只有通过跨学科的研究和全面的认识, 金融从业者才能更好地理解和应对这一复杂的挑战。

1 数据预处理

首先对收集到的数据进行了预处理, 发现缺失值的存在, 然后对异常值进行整理, 经过查阅大量文献和资料, 发现 AQI 表示空气质量指数, 通常包括对大气中的污染物浓度进行测量和评估, 其值不为 0, 而在附录中发现部分 AQI 的值为 0, 可能是由于测量误差或人员统计失误造成的异常值, 于是对这些异常值首先进行剔除, 然后利用最邻近插值法来弥补缺失, 最邻近插值用于在离散数据中估计未知的点, 这种插值方法的基本思想为: 对于要估计的点, 找到最接近的已知数据点, 最后将该点的值设定为最邻近点的值, 经过插值后的 AQI 值如下表 1 所示:

表 1 插之后的 AQI 值

Date	E	AQI	Stock Index	Tourism Stock	Energy Stock	Oil Price	Gas price	P	C
2019/1/22	0	83.47334286	1128.255534	103.6636105	98.75894774	47.58626381	5.26281827	1	3
2019/8/30	0	33.68536175	1018.898145	102.4821492	96.28609435	48.59295994	5.196316408	0	0
2019/10/18	1	37.15123526	992.1883798	96.41735066	95.96410832	51.84593206	5.013265955	1	2
2021/5/24	0	73.58457124	936.7341111	105.8913912	102.3881125	50.89293264	5.222692904	1	1
2022/7/12	1	47.47347798	1107.995852	103.4693697	100.6480867	53.66056198	4.855419388	0	0
2022/7/31	0	52.70099722	978.8191456	104.1313448	100.4955712	52.60510407	4.910467207	0	1
2022/12/28	0	37.63558837	901.1590663	103.7936537	96.40856689	49.28214002	4.964683503	0	2
2023/3/27	0	38.54716063	973.7923217	95.6293991	109.2464574	44.94785994	5.234366155	0	1
2023/5/19	1	64.1821894	993.5557515	99.07384234	102.9302692	53.03998359	4.855189417	1	4
2023/7/18	0	29.24597486	1053.198677	101.1144792	108.6036715	51.23134705	4.742727177	0	2
2023/1/8	0	51.73804772	1011.462746	103.1586969	105.4942211	47.2724263	4.77317669	0	0

注：E是Extreme Weather Event，P是Policy Change，C是Climate Change News

并且利用插值后的数据我们绘制出 2019 年 1 月 1 日——2023 年 12 月 31 日近 5 年的 AQI 时序图，如图 1 所示：

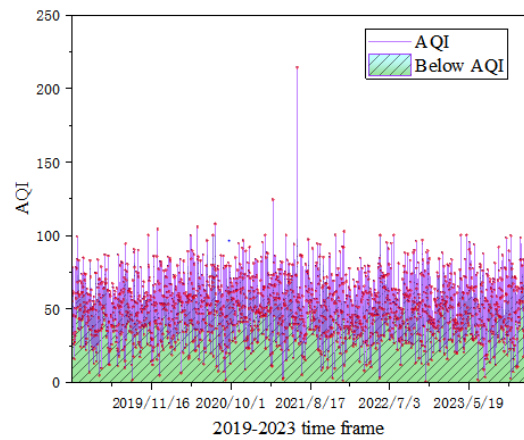


图 1 AQI 时序图

由上图分析发现，大部分 AQI 的值都在[0,150]之间波动，部分值出现了，利用插值之后的数据从而使得 AQI 变化更加完整，增加建模的便捷性与准确性。

2 数据正态化的判断

Q-Q 图是一种用于检验一个数据集是否符合特定分布的统计图表[4]，它通过比较观察值的分位数与理论分位数之间的关系，绘制出环境因素与股市的 Q-Q 图，如图 2 所示，调整置信水平为 95%,计分方法为 Blom。

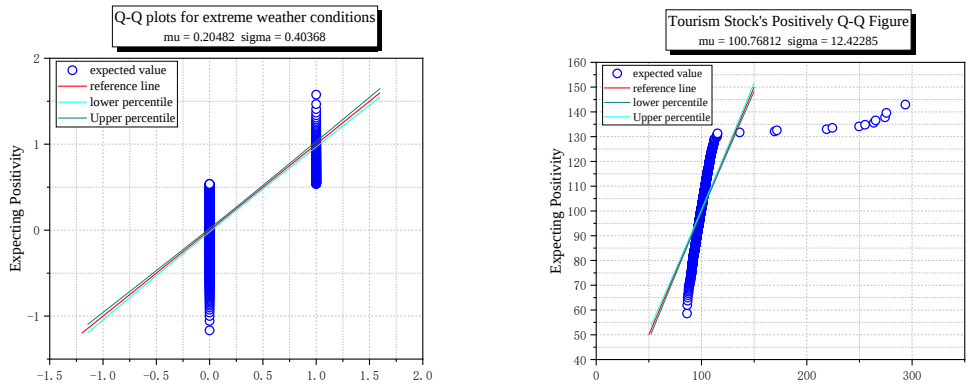


图 2 Q-Q 图检验结果

由上图分析可知，我们团队以极端环境因素与股票旅游为例，绘制 Q-Q 图，发现该列数据在 Q-Q 图显现无规划状态，并不满足 Q-Q 图的检验，并且 σ 与 μ 的值也不适用于 Pearson 相关系数法，于是我们使用范围更广的 Spearman 相关系数法，来对相关性进行分析。

3 模型的建立与求解

Spearman 相关系数法是一种用于衡量两个变量之间单调关系的统计方法[6]，它与 Pearson 相关系数不同，Spearman 相关系数不要求变量满足正态分布，而是基于变量的秩进行计算，其原理如下：

定义： X 和 Y 表示两组数据，其 Spearman 相关系数为

$$r_s = 1 - \sigma \sum_{i=1}^n d_i^2 / n(n^2 - 1)$$

其中： d_i 表示 X_i 和 Y_i 之间的等级差，一个数的等级就是将它所在一系列数按照从大到小排序后，这个数所在的位置（ $-1 \leq r_s \leq 1$ ）

热力图是一种通过颜色编码来表示矩阵中数值的图表，而相关行热力图则是一种特殊类型的热力图，用于可视化变量之间的相关性，下图为上述结果的相关性热力图，如图 3 所示。

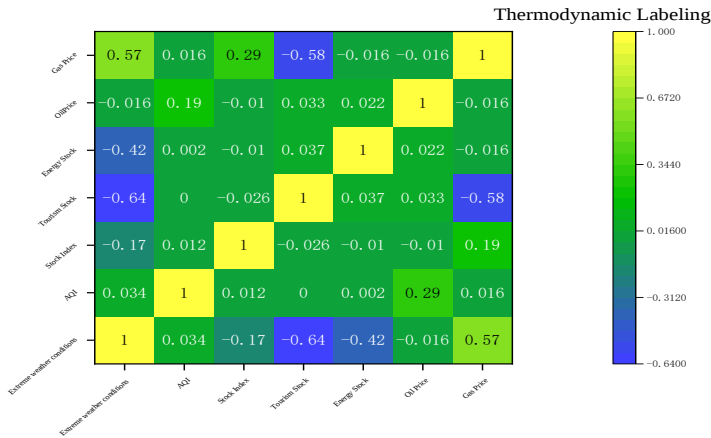


图 3 相关性热力图

由该图分析可知，其中极端天气条件与股票指数、旅游指数、能源股票、天然气价格全呈现负相关，与油价呈现正相关，并与旅游股票呈负相关性最大，这与实际情况相符，极端天气因素会影响人们出行，其中对旅游业影响最大，并且油气储存由于极端天气影响，运输困难，导致价格上涨，AQI 与股票整体都呈现正相关，而 AQI 代表空气质量的优劣程度。因此，我们团队得出结论，环境因素与整体股市表现较强的相关性。

4 显著性检验

肯德尔一次性检验是表示多列等级变量相关程度的一种方法，适用于多面评价的综合分析，肯德尔 W 系数又称为肯德尔和谐系数。我们的团队利用该方法对上述 Spearman 相关系数结果进行检验，利用下述基本公式

$$W = S/[1/12K^2(N^2 - N)]$$

式中：

$$S = \sum(R_i - \frac{\sum R_i}{N})^2 = \sum R_i^2 - (\sum R_i)^2/N$$

表 2 Kendall's W 分析结果

Kendall's W 分析结果					
名称	秩平均值	中位数	Kendall's W 系数	X ²	P
极端天气环境	1.001	0			
AQI	3.515	50.579			
Stock_Index	7	1001.886			
Tourism_Stock	5.482	99.897	0.96	10513.553	0.000***
Energy_Stock	5.491	99.848			
Oil_Price	3.504	50.044			
Gas_Price	2.007	5.013			

注：***、**、*分别代表 1%、5%、10%的显著性水平

由表 2 分析可知，其中 kendall’sW 级系数为 0.96→1，表示上述评价结果完全一致，且 p<0.05 表示上述 Spearman 相关系数显著的异于可以接受的上述结果。

5 结论

- 1) 对收集到的金融股票数据进行预处理，使用最邻近插值法来插补缺失值，插值后的数据表现状态较好，方便后续模型的建立。
- 2) 在发现数据不符合正态分布后，选择使用 Spearman 相关系数法来对各个影响之间的相关性展开深入分析，研究发现环境因素与整体股市表现较强的相关性，最后并通过了肯德尔一次性检验。

参考文献

[1] 吴晓求,方明浩,许荣.中国资本市场的功能转型与估值校正[J/OL].财贸经济,1-16[2024-06-17].<https://doi.org/10.19795/j.cnki.cn11-1166/f.20240613.005>.

[2] 李文瑶,程瑞琼,龙松伟,等.自然灾害冲击与居民家庭金融资产配置[J].经营与管理,2024,(06):210-216.DOI:10.16517/j.cnki.cn12-1034/f.2024.06.009.

[3] 江娟.淮安市绿色金融助力“双碳”目标研究[J].合作经济与科技,2024,(15):58-61.DOI:10.13665/j.cnki.hzjjykj.2024.15.058.

[4] 张栋,李春,彭京启,等.分段插值算法在注塑机背压控制中的应用[J].液压与气动,2022,46(12):145-151.

[5] 丁浩川,蓝永庭,解远航,等.激光熔凝改性 Q235 钢表面的微观组织和力学性能演变分析[J].钢铁研究学报,2024,36(05):640-649.

[6] 于忠涛,别业军,张威,等.高职院校学生体质健康、心理健康、社会适应能力现状及斯皮尔曼等级相关性研究[J].体育视野,2023,(12):7-11.