

# 应用文本情感分析和主成分分析解决融资融券风险量化问题

单学佳

常州工学院, 江苏 常州

收稿日期: 2022年9月24日; 录用日期: 2022年10月17日; 发布日期: 2022年10月26日

## 摘要

融资融券制度为我国股市交易注入了新的活力, 激发了广大投资者和券商的积极参与。我国监管机构虽然制定了严格的交易制度, 但其系统性风险仍不容忽视。因此, 本文针对如何量化融资融券业务风险问题进行了建模求解。首先通过Python爬虫和数据库下载的方式获取财务信息、股吧评论、新闻公告等与融资融券风险相关的数据, 其中, 针对非结构化数据运用了文本情感分析进行模糊信息的挖掘。其次, 对所有指标与融资融券风险进行相关性分析, 最终选取出13个与融资融券风险相关的指标, 运用主成分分析法计算每个指标影响融资融券风险的权重, 建立了融资融券风险量化模型。研究结果有助于厘清关键财务指标以及以股吧评论为代表的模糊信息影响融资融券风险的路径和机理, 并为有效量化融资融券风险提供了一定的借鉴思路。

## 关键词

Python爬虫, 文本情感分析, 主成分分析, 融资融券风险量化

# Research on Quantitative Risk of Margin Trading Based on Text Analysis and Principal Component Analysis

Xuejia Shan

Changzhou Institute of Technology, Changzhou Jiangsu

Received: Sep. 24<sup>th</sup>, 2022; accepted: Oct. 17<sup>th</sup>, 2022; published: Oct. 26<sup>th</sup>, 2022

## Abstract

The margin trading system has injected new vitality into China's stock market trading and stimu-

lated the active participation of investors and securities traders. Although China's regulatory authorities have formulated a strict trading system, its systematic risks cannot be ignored. Therefore, this paper models and solves the problem of how to quantify the risk of margin trading. First, the data related to the risk of margin trading, such as financial information, stock bar comments, news bulletins, etc., are obtained through Python crawlers and database downloads. Among them, text emotion analysis is used to mine fuzzy information for unstructured data. Secondly, the paper analyzes the correlation between all indicators and the risk of margin trading, finally selects 13 indicators related to the risk of margin trading, calculates the weight of each indicator affecting the risk of margin trading using the principal component analysis method, and establishes a quantitative model of the risk of margin trading. The research results are helpful to clarify the path and mechanism of the impact of key financial indicators and fuzzy information represented by stock bar comments on the risk of margin trading, and provide some reference ideas for effectively quantifying the risk of margin trading.

## Keywords

Python Crawler, Text Emotion Analysis, Principal Component Analysis, Risk Quantification of Margin Trading

Copyright © 2022 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

2010年我国沪深交易首次推出融资融券交易试点,引起了相关领域学者的广泛讨论。在完善的市场体系下,财务杠杆这一天然属性使得两融交易制度可以有效充当股市的“价格稳定器”,然而目前情况来看,许多证券机构没有建立完善的风险对冲机制,融资融券业务中涉及的风险控制数学模型预测滞后,不能及时向客户提供风险警示,可能给证券公司和客户造成损失。因此,在稳步推进融资融券业务发展的前提下,如何建立融资融券业务中的风险控制数学模型,量化研究融资融券业务中的风险,为融资融券业务决策提供服务,成为证券行业面临的急需解决的课题。

大多数学者研究基于财务风险的角度探讨企业面临的风险,如李素红(2011)通过对房地产企业特点的分析,找出相关影响因素,运用主成分分析法构造了房地产企业的财务风险综合评价体系,并给出相关建议[1]。陈茜(2017)基于33家林业企业样本,对影响其财务风险的12个因素进行打分和排序,将财务风险归集为良好、一般和高风险三类[2]。也有学者基于外部监督的角度对影响股价波动的因素进行相关性分析。如叶德磊和姚占雷(2017)通过实证分析认为公司新闻数、公司公告数以及股吧的发帖数对股价波动均有显著影响,其中公告数负向影响股价的波动性[3]。关静怡和朱恒(2020)认为股吧评论的分歧越小反而会容易产生股价崩溃风险[4]。王雨(2017)认为投资者恐慌或亢奋的情绪会助推股融资融券交易市场的波动性[5]。由上述分析可知,并未有研究基于全面的角度去探讨融资融券交易市场中的风险量化问题,因此本文将企业的关键财务指标、股吧评论、公司公告、新闻数量以及投资者情绪指数等内外部与融资融券交易风险相关的指标逐个进行相关性分析,最终运用主成分分析法将选取出的13个显著相关的指标,构建出衡量融资融券风险的综合量化指数,以期帮助证券公司及时发现可能存在的问题,并为融资融券制度的制定和优化提供一定的贡献。

## 2. 研究方法及数据处理

### 2.1. 数据来源与样本选择

本文首先按照三大产业划分标准以及行业融资融券余额大小两个标准，最终选取出 30 支标的股票，并以 2019 年 4 月 1 日至 2022 年 3 月 31 日期间该 30 个沪深 A 股上市公司为研究样本。其中，股票交易数据、融资融券交易数据来源于 iFind (同花顺)数据库，所有财务数据来自于 CSMAR (国泰安)数据库。股吧评论、公司公告和新闻报道数据来自于东方财富网、百度新闻网站，主要通过 Python 爬虫的方式收集获取，获得原始数据后，一方面以周为时间节点分别计算评论、公告和新闻报道出现的频次获得周数据，另一方面再运用自然语言处理(SnowNLP)算法程序分析文本内容并计算情绪倾向得分。

本文还对样本做了如下预处理：首先，在筛选 30 家上市公司样本时剔除了各年被 ST 或 ST\*的公司以及证监会行业分类中归属金融和保险业的公司；其次，剔除了有缺失值的样本数据；最后，为降低异常值影响，对连续变量进行上下 1%分位数的 Winsorize 处理。由于财务类指标和股吧评论等反映投资者情绪类指标对标的股票影响的时效性和敏感程度不同[3]，本文分别采用了季度数据和周数据进行实证分析，以确保结果的合理性。以上数据的预处理主要通过 Excel、Stata16.0、SPSS 等软件完成。

### 2.2. 财务指标的选择与构建

首先，我们将标的股票的公司财务数据进行整理，对照 Pearson 相关系数筛选出 15 个与股价显著相关的财务指标，筛选结果如表 1 所示。

**Table 1.** List the selection results of financial indicators

**表 1.** 财务指标选取结果列示

变量名	变量符号	变量定义
融资融券交易余额	$Y_1$	每季度最后一个交易日的融资融券交易余额
每股经营活动产生的现金流量净额	$X_1$	经营活动产生的现金流量净额本期值/实收资本本期期末值
流动比率	$X_2$	流动资产/流动负债
速动比率	$X_3$	(流动资产 - 存货)/流动负债
资产负债率	$X_4$	负债合计/资产总计
营业收入增长率	$X_5$	(营业收入本年本期单季度金额 - 营业收入上一个单季度金额)/(营业收入上一个单季度金额)
每股净资产增长率	$X_6$	(每股净资产本期期末值 - 每股净资产本期期初值)/每股净资产本期期初值
总资产周转率	$X_7$	营业收入/资产总额期末余额
每股收益	$X_8$	净利润本期值/实收资本本期期末值
每股净资产	$X_9$	所有者权益合计期末值/实收资本本期期末值
市净率	$X_{10}$	今收盘价当期值/(所有者权益合计期末值/实收资本本期期末值)
资产报酬率	$X_{11}$	(利润总额 + 财务费用)/资产总额
净资产收益率	$X_{12}$	净利润/股东权益余额
营业利润率	$X_{13}$	营业利润/营业收入
前五大股东持股比	$X_{14}$	前五大股东持股比例和
股权性质	$X_{15}$	国有性质为 1，非国有性质为 0

其次, 利用因子分析法, 通过降维的思想将 15 个财务指标降至 5 个维度, 并按因子特征依次命名为盈利能力、偿债能力、股东获利能力、成长性与营运能力以及股权结构与管理能力。最后, 将 5 个因子加权平均后得到的财务指标综合得分与融资融券交易余额( $Y_1$ )进行回归, 从而探究财务指标与股价之间的相关性。其中, 主成分分析法下的指标构建具体步骤如下所示:

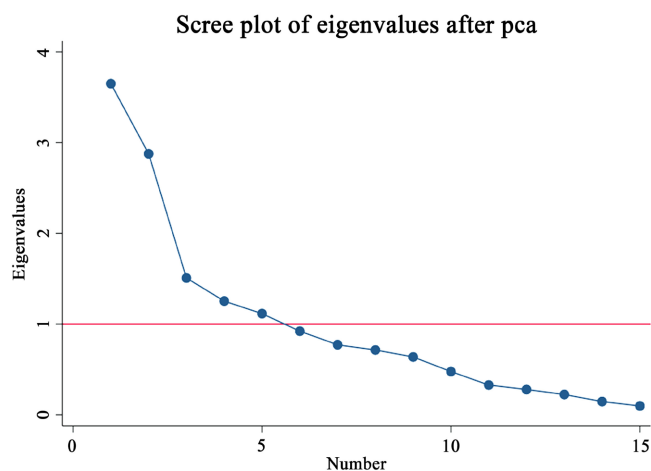
1) 适用性检验。首先需要对原始样本数据进行 KMO & Bartlett 球形检验, 从而保证因子分析法的适用性, 检验结果如表 2 所示。表中检验结果为 0.695, 大于 0.6 则表明适合做因子分析。Bartlett 球形检验用于检验相关阵是否是单位阵, 检验结果为 2420.011, 自由度为 105, Sig 值为 0, 表明检验值较大, 变量相互之间具有相关性, 同样说明适合进行因子分析。

**Table 2.** Results of KMO test and Bartlett sphericity test

**表 2.** KMO 检验与 Bartlett 球形检验结果

检验方法	指标	检验结果
KMO 检验	KMO 值	0.695
	$\chi^2$ 值	2420.011
Bartlett 球形检验	自由度	105
	显著性 Sig	0.000

2) 确定因子变量。通过计算各个变量的特征根, 其中大于 1 的特征根提取主因子的效用较大。



**Figure 1.** Crushed stone chart

**图 1.** 碎石图

如图 1 碎石图显示, 特征值大于 1 的主成分(红线以上)是我们需要的。同时表 3 显示, 其累积解释 67.099%, 对原有变量包含的信息具有抢解释力, 因子分析效果较为理想。因此选取五个主成分。旋转后 5 个因子累计贡献率没有变化、特征值和贡献率发生变化, 即没有影响原始变量的共同度, 但却重新分配了各个因子解释原始变量的方差, 改变了各因子的方差贡献, 使得因子易于解释。

3) 因子命名。由上述分析知本文共选取了 5 个主因子, 为方便区分, 依次将其命名为  $F_1 \sim F_5$ 。基于方差最大的原则, 我们使用主成分法提取以便更好地解释因子变量, 具体做法是利用具有 Kaiser 标准化正交旋转法不断对成分矩阵进行旋转, 结果显示旋转在 7 次迭代后收敛。结果见表 4。

**Table 3.** Extracted principal components  
**表 3.** 提取的主成分

成分	总方差解释								
	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	3.441	22.939	22.939	3.441	22.939	22.939	3.078	20.522	20.522
2	2.791	18.603	41.542	2.791	18.603	41.542	2.578	17.188	37.71
3	1.442	9.615	51.158	1.442	9.615	51.158	1.557	10.38	48.09
4	1.301	8.674	59.831	1.301	8.674	59.831	1.428	9.518	57.608
5	1.09	7.268	67.099	1.09	7.268	67.099	1.424	9.491	67.099
6	0.97	6.468	73.567						
7	0.836	5.574	79.141						
8	0.749	4.991	84.132						
9	0.646	4.305	88.437						
10	0.58	3.864	92.3						
11	0.43	2.87	95.17						
12	0.296	1.97	97.14						
13	0.231	1.542	98.682						
14	0.119	0.792	99.474						
15	0.079	0.526	100						

**Table 4.** Rotation component matrix  
**表 4.** 旋转成分矩阵

财务指标	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	单值性
每股经营活动产生的 现金流量净额	0.181	-0.091	0.827	0.138	-0.002	0.256
流动比率	-0.030	0.936	-0.057	-0.015	0.029	0.118
速动比率	0.009	0.946	-0.030	0.026	0.075	0.098
资产负债率	-0.199	-0.850	0.152	0.089	0.007	0.207
营业收入增长率	0.106	0.088	-0.068	0.761	-0.018	0.398
每股净资产增长率	0.583	0.076	0.292	-0.043	-0.060	0.563
总资产周转率	0.051	-0.194	0.289	0.694	0.092	0.386
每股收益	0.835	0.003	0.344	0.055	-0.086	0.175
每股净资产	0.173	-0.146	0.848	-0.058	-0.106	0.215
市净率	0.058	-0.112	-0.202	0.411	-0.568	0.451
资产报酬率	0.912	0.048	0.103	0.128	-0.049	0.137
净资产收益率	0.808	0.073	-0.074	-0.110	0.069	0.319
营业利润率	0.686	0.051	0.062	0.121	-0.055	0.505
前五大股东持股比	0.061	-0.062	0.432	0.118	0.518	0.524
股权性质	-0.092	0.099	-0.232	0.077	0.822	0.246

如表 4 所示, 各个变量的单值性均小于 0.6, 表示选取指标正常。同时, 我们可以得到各个因子的对原始变量的解释力度: 1) 因子  $F_1$  在每股净资产增长率、每股收益资产报酬率、净资产收益率、营业利润率中有较大的负载, 它们用于反映公司的盈利能力, 因此  $F_1$  命名为盈利能力; 2) 因子  $F_2$  在流动比率、速动比率、资产负债率中有较大的负载, 它们代表了公司的偿债能力, 因此  $F_2$  命名为偿债能力; 3) 因子  $F_3$  在每股经营活动产生的现金流量净额、每股净资产中有较大的负载, 它们反映了股东的获利能力, 因此  $F_3$  命名为股东的获利能力; 4) 因子  $F_4$  在营业收入增长率、总资产周转率中有较大的负载, 它们分别反映了公司的成长性与营运能力, 因此  $F_4$  命名为成长性与营运能力; 5) 因子  $F_5$  在市净率、前五大股东持股比、股权性质中有较大的负载, 前五大股东持股比、股权性质反映了公司的股权构架与产权性质, 属于结构性指标, 市净率越低代表公司风险越低, 一定程度体现了公司的管理能力, 因此  $F_5$  命名为股权结构与管理能力。具体见表 5。

**Table 5.** Results of factor naming and interpretation

**表 5.** 因子命名与解释结果

因子名称	因子符号	变量
盈利能力	$F_1$	每股净资产增长率
		每股收益
		资产报酬率
		净资产收益率
		营业利润率
偿债能力	$F_2$	流动比率
		速动比率
		资产负债率
股东获利能力	$F_3$	每股经营活动产生的 现金流量净额 每股净资产
成长性与营运能力	$F_4$	营业收入增长率
		总资产周转率
股权结构与管理能力	$F_5$	市净率
		前五大股东持股比 股权性质

4) 因子得分与建立因子模型。首先, 根据因子得分系数矩阵计算各公共因子的因子得分函数。

**Table 6.** Component score coefficient matrix

**表 6.** 成分得分系数矩阵

变量符号	财务指标	成分				
		$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
$X_1$	每股经营活动产生的 现金流量净额	-0.075	0.060	0.449	0.054	-0.027
$X_2$	流动比率	-0.075	0.389	0.090	0.040	-0.048

Continued

$X_3$	速动比率	-0.064	0.392	0.096	0.070	-0.010
$X_4$	资产负债率	-0.038	-0.321	0.012	0.032	0.048
$X_5$	营业收入增长率	0.001	0.059	-0.082	0.594	0.012
$X_6$	每股净资产增长率	0.165	0.025	0.090	-0.078	-0.027
$X_7$	总资产周转率	-0.047	-0.022	0.102	0.518	0.093
$X_8$	每股收益	0.251	-0.013	0.064	-0.023	-0.022
$X_9$	每股净资产	-0.075	0.038	0.472	-0.105	-0.113
$X_{10}$	市净率	-0.002	-0.014	-0.131	0.306	-0.411
$X_{11}$	资产报酬率	0.318	-0.031	-0.094	0.044	0.026
$X_{12}$	净资产收益率	0.330	-0.055	-0.175	-0.123	0.114
$X_{13}$	营业利润率	0.238	-0.016	-0.078	0.054	0.006
$X_{14}$	前五大股东持股比	-0.013	-0.010	0.205	0.079	0.394
$X_{15}$	股权性质	0.057	-0.040	-0.166	0.104	0.650

由上表 6，依次建立各个因子( $F_1 \sim F_5$ )的得分模型：

$$F_1 = -0.075X_1 - 0.075X_2 - 0.064X_3 - 0.038X_4 + 0.001X_5 + 0.165X_6 - 0.047X_7 + 0.251X_8 - 0.075X_9 - 0.002X_{10} + 0.318X_{11} + 0.330X_{12} + 0.238X_{13} - 0.013X_{14} + 0.057X_{15} \quad (2-1)$$

$$F_2 = 0.060X_1 + 0.389X_2 + 0.392X_3 - 0.321X_4 + 0.059X_5 + 0.025X_6 - 0.022X_7 - 0.013X_8 + 0.038X_9 - 0.014X_{10} - 0.031X_{11} - 0.055X_{12} - 0.016X_{13} - 0.010X_{14} - 0.040X_{15} \quad (2-2)$$

$$F_3 = 0.449X_1 + 0.090X_2 + 0.096X_3 + 0.012X_4 - 0.082X_5 + 0.090X_6 + 0.102X_7 + 0.064X_8 + 0.472X_9 - 0.131X_{10} - 0.094X_{11} - 0.175X_{12} - 0.078X_{13} + 0.205X_{14} - 0.166X_{15} \quad (2-3)$$

$$F_4 = 0.054X_1 + 0.040X_2 + 0.070X_3 + 0.032X_4 + 0.594X_5 - 0.078X_6 + 0.518X_7 - 0.023X_8 - 0.054X_9 + 0.306X_{10} + 0.044X_{11} - 0.123X_{12} + 0.054X_{13} + 0.079X_{14} + 0.104X_{15} \quad (2-4)$$

$$F_5 = -0.027X_1 - 0.048X_2 - 0.010X_3 + 0.048X_4 + 0.012X_5 - 0.027X_6 + 0.093X_7 - 0.022X_8 - 0.013X_9 - 0.411X_{10} + 0.026X_{11} + 0.114X_{12} + 0.006X_{13} + 0.394X_{14} + 0.650X_{15} \quad (2-5)$$

其次，以旋转后各公因子的方差贡献率为权重，又可以计算各因子综合评价得分  $F$ ：

$$F = 0.22939 * F_1 + 0.18603 * F_2 + 0.09615 * F_3 + 0.08674 * F_4 + 0.07268 * F_5 \quad (2-6)$$

最终，我们得到用于构建融资融券风险的财务指标包括五个因子( $F_1 \sim F_5$ )和一个财务综合评价得分  $F$ 。

### 2.3. 非结构化数据的指标与构建

本文还需要对非结构性数据进行处理。因为目前证券市场不乏信息披露不规范的情况，所以投资者开始借助东方财富、雪球、和讯等越来越多的财经网站获得即时的股票相关信息，并进行解答与互动，从而降低信息的检索、核实成本。而我们从东方财富网和百度资讯上获取的公司公告、新闻报道以及股吧评论等数据是以文本形式存在的。这些非结构化数据本质上是一种模糊信息，因此需要挖掘其中的内在信息。

根据现有学者的研究，本文借助了 Python 中的 SnowNLP 对数据进行文本情感分析。SnowNLP 是一个 python 写的类库，主要用于处理中文文本，可实现分词、词性标注、情感分析、汉字转拼音、繁体转





$$Y_{i,t+1} = \beta_0 + \beta_1 F_{i,t} + \alpha_i + \lambda_t + \varepsilon_{i,t} \tag{3-1}$$

其中,  $i$  表示企业个体,  $t$  表示年份,  $\alpha_i$  表示个体固定效应,  $\lambda_t$  表示时间固定效应,  $\varepsilon_{i,t}$  为随个体和时间而改变的扰动项。

### 3.1.3. 相关性分析

表 8 列示了财务指标综合得分( $F$ )与融资融券交易余额( $Y_1$ )的回归结果。如表所示, 删除缺失值后得到 340 个观测值。财务指标综合得分与融资融券交易余额的相关系数为 0.623, 且在 1%水平上显著正相关, 这表明企业的财务指标越高, 越有利于融资融券的交易额显著提高。 $R^2$  为 0.105, 表明财务综合得分可以解释标的股票融资融券交易余额 10.5%的变动。

**Table 8.** Regression results of comprehensive score of financial indicators and balance of margin trading

**表 8.** 财务指标综合得分与融资融券交易余额回归结果

$Y_2$	Coef.	St.Err.	t-value	p-value	[95% ConfInterval]		Sig
$F$	0.623	0.104	6.02	0	0.42	0.827	***
Constant	21.233	0.027	790.30	0	21.18	21.286	***
Mean dependent var	21.241		SD dependent var		0.885		
R-squared	0.105		Number of obs		340		
F-test	36.239		Prob > $F$		0.000		
Akaike crit. (AIC)	458.025		Bayesian crit. (BIC)		465.683		

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ ,  $p < 0.1$ 。

## 3.2. 针对非结构化数据的相关性分析

### 3.2.1. 变量解释

借鉴叶德磊(2017), 股价运行的情况可以从股价波动程度衡量, 股价波动程度反映的是股票价格在某一时间段内上下震荡的幅度, 我们将各样本股一周内的最高价对最高价与最低价所构成的中位值的偏离程度作为衡量股价波动程度的指标[3]。具体计算公式为:

$$VOLA = \frac{P_H}{(P_H + P_L) \div 2}$$

其中,  $VOLA$  表示股价波动程度,  $P_H$  表示股票的最高价,  $P_L$  表示股票的最低价。因变量与自变量的定义如表 9 所示。

**Table 9.** Variable definitions

**表 9.** 变量定义

	变量名称	符号	变量定义
因变量	股价波动程度	$VOLA$	见上述定义
自变量	公告关注度	Notice	每周发布公告的累计频次
	公告情绪指数	Notice_index	运用 snowNLP 文本情绪分析得出
	新闻关注度	News	每周发布新闻的累计频次
	新闻情绪指数	News_index	运用 snowNLP 文本情绪分析得出
	评论关注度	Comment	每周发表评论的累计频次
	评论情绪指数	Comment_index	运用 snowNLP 文本情绪分析得出

### 3.2.2. 模型建立

我们从股价波动幅度角度反映股价运行情况，依次探讨各解释变量与该指标之间的关系。按照情绪指数与关注度两类分别构建如下回归模型：

$$VOLA = \alpha_0 + \alpha_1 \text{comment\_index} + \alpha_2 \text{notice\_index} + \alpha_3 \text{news\_index} + \varepsilon \quad (3-2)$$

$$VOLA = \alpha_0 + \alpha_1 \text{comment} + \alpha_2 \text{notice} + \alpha_3 \text{news} + \varepsilon \quad (3-3)$$

其中， $\alpha_0$  为个体截距项， $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$  为相关系数， $\varepsilon$  为随机误差项。

### 3.2.3. 股价涨跌与关注度的相关性分析

由表 10 可知，共获得 4533 个合并周数据的观测值。其中，股价波动程度(VOLA)与评论关注度(Comment)的相关系数在 1% 水平上显著正相关，即评论关注度越高，股价的波动程度越大。我国的股票市场始终弥漫着浓厚的投机气氛，因此较高的讨论度可能造成投资者在买卖股票时存在盲目跟风的心理，进而影响股价的平稳性。公告关注度(Notice)与股价波动程度(VOLA)的相关系数在 1% 水平上显著负相关，即公告关注度越高，股价波动程度越小，这可理解为公告的发布在一定程度上起到了平抑股价波动的作用。新闻关注度(News)与股价波动程度(VOLA)的相关系数在 10% 水平上显著正相关，即发帖数或新闻量越多(意味着对股票的关注程度越高)，股价波动幅度越大。

**Table 10.** Correlation analysis table of stock price fluctuation and attention

**表 10.** 股价涨跌与关注度的相关性分析表

VOLA	Coef.	St.Err.	t-value	p-value	[95% ConfInterval]		Sig
Comment	0.0012646	0.0001793	7.05	0.000	0.0009131	0.0016161	***
Notice	-0.0008889	0.0001897	4.69	0.000	0.0005171	0.0012606	***
News	0.0008107	0.000464	1.75	0.081	-0.0000987	0.0017201	*
Constant	1.013201	0.001152	879.52	0.000	1.010943	1.015459	***
Mean dependent var	1.022		SD dependent var		0.012		
Overall r-squared	0.016		Number of obs		4533		
Chi-square	88.313		Prob > chi2		0.000		
R-squared within	0.022		R-squared between		0.000		

\*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1。

### 3.2.4. 股价涨跌与情绪指数的相关性分析

由表 11 知，共得到 4521 个合并周数据的观测值。其中，股价波动程度(VOLA)与评论情绪指数(Comment\_index)的相关系数在 1% 水平上显著正相关；与公司公告情绪指数(Notice\_index)的相关系数在 5% 水平上显著相关；与新闻情绪指数(News\_index)的相关系数并未呈现显著性，但系数符号为正。这进一步验证了上述结论，其中新闻指数未得到验证，可能是因为新闻发布者多以客观的语气进行描述，且存在公司买热量的现象。

## 3.3. 融资融券风险量化体系构建

### 3.3.1. 评价指标设计

已有文献知，融资融券业务风险深受公司财务绩效、市场行情、股民关注度、新闻监督等等因素的

影响。财务指标的提高具有长期性，是任何企业最终追求的，而非财务指标的提高具有短期性，是股价波动的驱动因素，因此需要对各方面都要兼顾。我们结合前述问题的结果，设计指标如表 12 所示。

综上，令风险指数为  $Z$ ，指标值为  $P_i$ ，对应的指标权重为  $D_i$ ，将上述 13 个指标值与权重相乘并相加即可得出证券公司融资融券业务风险指数，具体评价体系数学模型(3-4)如下公式：

$$Z = \sum_{i=1}^{13} P_i D_i \quad (i = 1, 2, 3, \dots, 13) \tag{3-4}$$

**Table 11.** Correlation analysis table of stock price fluctuation and sentiment index  
**表 11.** 股价涨跌与情绪指数的相关性分析表

VOLA	Coef.	St.Err.	t-value	p-value	[95% ConfInterval]		Sig
Comment_index	3.06e-06	3.47e-07	8.82	0.000	2.38e-06	3.74e-06	***
Notice_index	-0.0000855	0.0000423	2.02	0.043	2.63e-06	0.0001684	**
News_index	0.0001852	0.0001983	0.93	0.350	-0.0002036	0.0005739	
Constant	1.020288	0.0004381	2329.14	0.000	1.01943	1.021147	***
Meandependent var	1.022		SD dependent var		0.012		
Overall r-squared	0.019		Number of obs		4521		
Chi-square	86.061		Prob > chi2		0.000		
R-squared within	0.019		R-squared between		0.020		

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 。

**Table 12.** Specific index design  
**表 12.** 具体指标设计

目标层	准则层	方案层		
		变量名称	指标值	指标权重
融资融券业务风险指标体系	资本规模	收盘价	P1	D1
		VOLA	P2	D2
	财务水平	$F_1$	P3	D3
		$F_2$	P4	D4
		$F_3$	P5	D5
		$F_4$	P6	D6
		$F_5$	P7	D7
	业务规模	融资融券余额	P8	D8
		融资买入额	P9	D9
		融券卖出额	P10	D10
	外部监督	Comment	P11	D11
		Notice	P12	D12
		News	P13	D13

### 3.3.2. 评价指标的权重设置

通过查阅文献可知,计算权重的方法主要有专家评分法、主成分分析法与聚类分析法。为了更准确、真实、科学反映评价指标体系,借鉴叶争辉(2019),我们使用了主成分分析法,设置最终权重[6]。首先计算子阵各主成分对应标记值,其次将所有标记值叠加计算,各指标权重  $D_i$  则是前述数值之比,最后建立指标模型。主成分分析计算原理与步骤在建立财务指标时已作阐述,在此不过多赘述。通过 Stata16 得出融资融券业务风险评价指标权重计算结果如表 13 所示。将表中权重带入模型(3-4),并将相应季度的指标值输入,即可获得该季度的风险量化指标。如 2019 年第一季度融资融券市场的风险指数为 66.10。

**Table 13.** Risk evaluation index weight table of margin trading business

**表 13.** 融资融券业务风险评价指标权重表

变量名称	指标值	指标权重
收盘价	P1	0.441
VOLA	P2	0.156
$F_1$	P3	0.147
$F_2$	P4	0.134
$F_3$	P5	0.123
$F_4$	P6	0.117
$F_5$	P7	0.111
融资融券余额	P8	0.103
融资买入额	P9	0.065
融券卖出额	P10	0.056
Comment	P11	0.042
Notice	P12	0.031
News	P13	0.020

## 4. 结论

本文创新性地凝聚了基于企业内部的财务绩效和基于外部监督的股吧评论、公司公告以及新闻报道等因素构建融资融券风险量化指标,有利于对融资融券业务风险进行更好、更全面的分解,实现融资融券业务较为准确的预警;采用主成分分析法确定权重,更加客观,最终确定风险指标,易于理解,数据可获得性和操作性较强。

## 参考文献

- [1] 李素红, 陈立文. 基于因子分析法的房地产上市公司财务风险评价[J]. 河北工业大学学报, 2011, 40(6): 101-106.
- [2] 陈茜, 田治威. 林业上市企业财务风险评价研究——基于因子分析法和聚类分析法[J]. 财经理论与实践, 2017, 38(1): 103-108.
- [3] 叶德磊, 姚占雷, 刘小舟. 公司新闻、投资者关注与股价运行——来自股吧的证据[J]. 华东师范大学学报(哲学社会科学版), 2017, 49(6): 136-143+172.
- [4] 关静怡, 朱恒, 刘娥平. 股吧评论、分析师跟踪与股价崩溃风险——关于模糊信息的信息含量分析[J]. 证券市场导报, 2020(3): 58-68.
- [5] 王雨. 融资融券对我国股票市场收益率波动性的影响研究[D]: [博士学位论文]. 北京: 对外经济贸易大学, 2017.
- [6] 叶争辉. 中航证券公司融资融券业务风险评价指标体系研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2019.