

基于改进的对数周期幂律模型研究中国股市崩盘预警

吴俊传 唐振鹏 杜晓旭 陈凯杰

Research on Chinese stock market crash early warning based on improved log-periodic power law model

Wu Jun-Chuan Tang Zhen-Peng Du Xiao-Xu Chen Kai-Jie

引用信息 Citation: *Acta Physica Sinica*, 71, 020201 (2022) DOI: 10.7498/aps.71.20201940

在线阅读 View online: <https://doi.org/10.7498/aps.71.20201940>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于格兰杰因果网络的中美贸易战对上证行业冲击的研究

The impact of trade war on Shanghai stock exchange industry based on Granger causality network

物理学报. 2021, 70(7): 078901 <https://doi.org/10.7498/aps.70.20201516>

随机波动率费曼路径积分股指期货定价

Pricing of stochastic volatility stock index option based on Feynman path integral

物理学报. 2019, 68(20): 203101 <https://doi.org/10.7498/aps.68.20190714>

不可压缩幂律流体气-液两相流格子Boltzmann 模型及其在多孔介质内驱替问题中的应用

Lattice Boltzmann model of gas-liquid two-phase flow of incompressible power-law fluid and its application in the displacement problem of porous media

物理学报. 2019, 68(21): 214702 <https://doi.org/10.7498/aps.68.20190873>

一种基于文本互信息的金融复杂网络模型

Financial complex network model based on textual mutual information

物理学报. 2018, 67(14): 148901 <https://doi.org/10.7498/aps.67.20172490>

基于有效介质理论的物理性能计算模型的软件实现

Software realization of physical property calculation model based on effective medium theory

物理学报. 2019, 68(15): 157701 <https://doi.org/10.7498/aps.68.20182273>

基于多径信道模型研究时间反演腔的反演特性

Analysis of time reversal cavity characteristics based on multipath channel model

物理学报. 2022, 71(2): 024101 <https://doi.org/10.7498/aps.71.20210701>

基于改进的对数周期幂律模型 研究中国股市崩盘预警*

吴俊传 唐振鹏[†] 杜晓旭 陈凯杰

(福州大学经济与管理学院, 福州 350108)

(2020 年 11 月 18 日收到; 2021 年 9 月 8 日收到修改稿)

基于金融物理学中著名的对数周期幂律模型 (log-periodic power law model, LPPL) 来预警 2015 年 6 月份中国上证综合指数、创业板指数的崩盘. 鉴于已有采用 LPPL 模型预警市场崩盘的研究均只考虑市场历史交易数据. 本文将投资者情绪因素纳入到 LPPL 模型建模过程, 以改进 LPPL 模型的预警效果. 采用文本挖掘技术结合语义分析方法对抓取的财经媒体的股评报道进行词频统计, 以构建媒体情绪指数. 进一步修改 LPPL 模型中的崩溃概率函数表达式, 将其表示为关于历史交易数据及媒体情绪的函数, 构建 LPPL-MS 组合模型预警股市崩盘. 实证结果表明, 本文所构建的 LPPL-MS 组合模型相比 LPPL 模型具有更高的预警精度, 其预测的大盘见顶的临界时点与上证指数、创业板指数真实的见顶时点更为接近, 并且其拟合结果通过了相关检验.

关键词: 对数周期幂律模型, 媒体情绪, 股市崩盘, 预警

PACS: 02.50.-r, 02.50.Cw, 02.50.Ng, 02.60.Ed

DOI: 10.7498/aps.71.20201940

1 引言

资本市场中投资者之间的拥挤交易行为极易促成资产价格泡沫的形成, 并且其泡沫迅速膨胀过程当中呈现出高度的非线性、自组织等复杂特性. Sornette 等^[1]物理学家提出对数周期幂律模型 (log-periodic power law model, LPPL) 来刻画资本市场处于泡沫阶段的运行规律, 将地震预测模型成功地运用到资本市场泡沫崩溃的预测. 股市崩盘使市场参与者蒙受巨额损失, 将风险输出到相应上市公司. 此外, 股市崩溃往往伴随着系统性风险, 股市系统性风险的频繁爆发极大地危害资本市场的平稳运行, 弱化资本市场的吸引力及竞争力, 降低各类投资主体, 尤其是境外投资者的入市积极性, 严重削弱资本市场的融资及资源优化配置等功

能的进一步发挥, 并最终扰乱市场经济秩序, 对中国经济发展的危害不言而喻. 股票市场作为资本市场最重要的组成部分之一, 对一国经济发展的重要性十分明显, 提前预警并及时管控其系统性风险, 对维护中国经济、金融稳定发展大局具有十分重要的意义.

长期以来, 诸如随机漫步 (random walk)^[2]、有效市场假说 (efficient markets hypothesis, EMH)^[3]等理论均认为股票价格序列之间不存在长记忆效应, 未来的股价无法提前预测. 然而, 大量研究表明股票价格序列并非满足随机游走的假设, 股票价格剧烈波动的背后存在可预测成分^[4]. 股市疯狂暴涨、形成泡沫, 并伴随泡沫破裂的现象时有发生^[5]. 自 Sornette 等^[1]学者以来, 越来越多的研究借鉴统计物理学的方法, 利用对数周期修正的临界点来识别股市泡沫. 该方法后来被称为 JLS (Johansen-

* 国家自然科学基金 (批准号: 71573042, 71973028) 资助的课题.

[†] 通信作者. E-mail: zhenpt@126.com

Ledoit-Sornette) 模型或对数周期幂律奇异性模型 (LPPL). 该模型认为泡沫可以从市场的自然运作中本质上产生, 类似于其他统计物理复杂系统, 金融资产也具备自组织临界特性, 自组织即指复杂系统^[6]某一状态的形成主要由系统内部不同组成部分互相作用所驱使, 而并非由其他外界因素主导; 临界性是指复杂系统当前处于极其敏感的特殊状态, 某一微小的局部变化即可引致整个系统的突变. 自组织临界性可以解释诸如日辉耀斑、地震、泥石流、山体滑坡、雪崩、物种灭绝及金融市场中的资产价格崩溃现象^[7,8]. LPPL 模型认为由于市场上交易者同一时间段互相模仿, 并通过正反馈机制不断强化, 致使资产价格在市场的自然运作过程中逐渐产生泡沫, 并最终崩溃. 潘娜等^[9]和周炜星^[10]认为 LPPL 模型具备两个显著特征: 一是对数周期性振荡, 在线性尺度下, 越接近临界点, 振荡频率越快, 而在对数尺度下, 振荡频率为常数; 二是幂律增长, 即价格的增长率并非常数, 而是单调递增.

金融资产价格泡沫在趋近崩溃边缘时, 呈现出对数周期幂律振荡特性. 因此, 采用 LPPL 模型可有效识别金融市场是否处于泡沫状态, 并对大盘见顶, 泡沫崩溃时点进行提前预警. Sornette 和 Johansen^[11]使用了 20 世纪两次最大的历史崩溃, 即 1929 年 10 月和 1987 年 10 月的崩溃, 并表明该方法通过分析泡沫发生前 8 年间隔内的股指走势可有效识别这两次崩溃. Sornette 和 Johansen^[12]提出了一个简单的股票市场交易员等级模型, 其研究表明, 不管是模型中处于最高等级的国家层面, 还是处于次高级的国家内部的银行等金融机构, 其交易行为均表现出明显的羊群效应, 这种不同层级的相互模仿足以使股票价格产生对数周期幂律振荡, 并最终引发市场的崩溃. Sornette 和 Zhou^[13]提出了一个系统的算法预先检测到道琼斯工业平均指数 (DJIA) 和香港恒生综合指数 (HSI) 泡沫的崩溃. Sornette 等^[14]利用 LPPL 方法分析了 2006—2008 年石油价格的上涨情况, 并证明了石油价格的演变过程也呈现出类似泡沫的状态. Kyriazis 等^[15]采用 LPPL 方法成功的预测到比特币、以太坊等加密货币 2013 年、2017 年的泡沫崩溃. Omer^[16]详细梳理了 LPPL 模型用于预警金融市场崩溃的成功案例. Sirca 和 Omladic^[17]成功地运用 LPPL 模型用于预测全球 33 个主流股票市场

泡沫的崩溃. Filimonov 和 Sornette^[18]提出了一种改进的 LPPL 公式, 将其从 3 个线性参数和 4 个非线性参数的函数转化为 4 个线性参数和 3 个非线性参数, 实证结果表明, 这种转换显著降低了拟合过程的复杂性, 并提高了其稳定性. 国内学者在这方面亦展开了广泛的研究. Jiang 等^[19]利用 LPPL 模型检测了 2005 年 5 月至 2009 年 7 月期间中国股市 2 个重要指数——上证综合指数和深证成分指数的 2 个泡沫和随后的市场崩溃, 并成功地预测了 2 个指数的崩溃. Yan 等^[20]利用 LPPL 模型, 开发了一种基于先进模式识别方法的预警指数用于探测市场泡沫, 并对市场崩溃和反弹进行预测. 在 10 个主要的全球股票市场上取得了较好的效果. 李东^[21]针对 LPPL 模型提出了基于网格的 Nelder-Mead Simplex 预搜索方法优化 LPPL 模型的估计参数, 并对恒生指数的泡沫崩溃进行提前预警, 取得了较好的效果. 周伟和何建敏^[22]采用 LPPL 模型成功识别到后金融危机时代不同类型的金属期货价格的集体泡沫, 做出了归因分析, 并验证了其结论. 李斯嘉等^[23]借鉴金融物理学方法, 分析上证综指和深证成指的崩盘前后动力学特征, 采用 LPPL 模型分别预测 2007 年、2015 年上证综指、深证成指的崩盘, 发现股市崩盘的特征符合 G-R 规律与 Omori 定律. 李伦一和张翔^[24]使用 LPPL 模型对中国房地产市场交易价格泡沫进行测度. 文章基于 2010 年 6 月至 2017 年 11 月中国 100 个主要城市的房地产交易数据对各城市房地产价格泡沫进行识别. 研究表明, LPPL 模型能够有效甄别各城市房地产价格正向泡沫、反向泡沫. 陈卫华和蔡文靖^[25]运用改进的 LPPL 参数估计方法预测 2015 年上证和创业板指数的崩盘, 并对其进行 Lomb 谱分析和 O-U 过程检验, 取得了稳健的效果. 赵磊和刘庆^[26]采用 LPPL 模型拟合比特币价格, 结果发现, 比特币价格呈现出 LPPL 模型中所隐含的超指数膨胀、对数周期幂律振荡特性, 并预测比特币价格可能在 2020 年下半年出现泡沫.

纵观已有文献, LPPL 模型在预警金融资产泡沫方面已经取得了丰硕的研究成果, 然而, 该模型并未将金融资产交易价格以外的因素, 诸如金融资产价格基本面、政策层面及市场情绪等因素纳入建模过程. 为此, 本文尝试将市场情绪层面的因素纳入 LPPL 模型建模过程, 对改进 LPPL 模型的泡沫预警效果.

随着 20 世纪八九十年代金融市场上一系列金融异象, 如股票溢价之谜、股价过度波动之谜、股价长期反转效应等, 被逐渐揭示. 行为金融理论 (behavioral finance) 借助行为科学、心理学的研究成果, 能有效解释这类金融异象. 行为金融理论认为市场上交易者并非完全理性, 其交易决策会受到各类心里因素, 如投资者情绪等非市场因素影响^[27]. 国内外大量实证研究表明, 投资者情绪在股票价格的决定中起着重要的作用. Baker 和 Wurgler^[28,29] 将投资者情绪定义为对未来现金流和投资风险的信念, 而这些信念并不是由基本面决定的. 他们的研究表明, 投资者情绪会影响股票市场的回报, 即积极情绪对股价具有正面影响, 消极情绪对股价具有负面影响, 并且随着投资者意识到情绪与基本面的偏离时, 这种影响在接下来的交易日会发生逆转. García^[30] 的研究表明, 投资者情绪对道琼斯工业平均指数的短期走势的影响符合上述发现. 新闻媒体作为投资者获取重要信息的窗口, 其对股票市场信息的报道对投资者指定交易决策具有重要影响. 媒体报道通过对股票市场的表现状况做出评判及预测, 会直接诱导投资者未来预期, 影响投资者情绪, 进而影响市场未来走势^[31,32]. Solomon^[33] 的研究发现, 比起负面报道, 媒体更热衷于正面报道, 即新闻媒体的报道内容往往掺杂着媒体商主观意见, 并非完全基于市场上已有的客观事实. Nardo 等^[34] 通过综合采用文本挖掘、情感分析及机器学习技术基于大量互联网新闻报道预测未来股票市场走势, 发现互联网新闻报道有助于预测市场未来走势, 然而其实际回报很少超过 5%. Fisher 和 Statman^[35] 的研究显示个人投资者的情绪受新闻媒体撰稿人的情绪影响较为显著, 二者相关性较强, 并且投资者的情绪与未来股市收益率负相关. 国内学者方面, 武佳薇等^[36] 基于前景理论框架, 考虑投资者非理性预期因素, 构建投资组合决策模型, 其采用某券商 2007 年至 2009 年近 177 万个人股票账户交易数据进行实证分析, 实证结果表明, 投资者情绪与处置效用具有负相关关系. 李合龙和冯春娥^[37] 采用 EEMD 分解技术提取投资者情绪及股指价格在不同时间尺度下的波动特征, 研究结果发现, 投资者情绪与股指价格序列在不同时间尺度下呈现出不同的波动关系. 由于中国股票市场中, 散户投资者占据了绝大部分比例, 其活跃程度远超机构投资者, 截止 2020 年 8 月底, 中国股市

中, 自然人投资者数目达到 1.72 亿, 而非自然人投资者数目仅为 40.28 万 (数据来源于中国证券登记结算有限责任公司). 相比机构等专业投资者, 广大散户, 尤其是刚入市的新股民, 其专业知识及实操经验较为缺乏, 其对市场运行状况的评价较依赖于各类新闻媒体的股评观点. 随着近些年, 互联网技术的迅猛发展, 互联网步入 Web 3.0 时代, 诸如微博、论坛等各种新兴媒体的崛起, 投资者可轻而易举地从互联网上获取各类海量并充斥着强烈情感色彩的股评文本信息. 游家兴和吴静^[38] 通过选取 2004 年至 2010 年间国内 8 家财经报纸新闻, 运用文本分析方法综合报道基调、曝光程度及关注水平三个维度信息构建媒体情绪指数, 发现新闻媒体情绪能够显著引致股价波动. Wu 等^[39] 从新浪财经网站抓取股评短文, 采用 SVM 训练媒体情绪, 实证结果表明, 媒体股评情绪与股价未来走势具有显著相关关系, 并且其在成长股中表现尤为明显. 戴德宝等^[40] 通过抓取东方财富网络论坛文本结合情感词典分析方法生成投资者情绪序列数据, 并采用 SVM 学习机将生成的情绪指数纳入训练过程, 其实证结果表明, 相比基准模型, 考虑情绪指数的 SVM 学习器具有更高的预测精度. 由于每日均有海量的新闻报道产生, 投资者受限个人有限的注意力, 无法逐一对每日新闻报道内容进行细致阅读^[41], 作为新闻报道内容的凝练与总结, 新闻标题可迅速吸引读者的注意力, 甚至有媒体为了博取投资者的关注, 故意为媒体报道冠上夸张的标题^[42]. Huang 等^[43] 研究发现企业管理层可以充分利用投资者的有限注意力, 在新闻标题中嵌入更多积极信息, 以煽动投资者的乐观情绪. 由上可得, 采用文本挖掘技术剖析新闻媒体评论稿标题可提取媒体情绪, 进而为下文分析媒体情绪对股市未来走势的影响作铺垫.

综上, 关于 LPPL 模型用于预警股市崩盘方面已经取得了丰硕成果, 然而已有研究未将交易历史数据以外的因素纳入建模过程. 此外, 现有研究均表明投资者情绪可在一定程度上解释未来股价走势的变动. 本文尝试将投资者情绪纳入 LPPL 建模过程, 以改进 LPPL 模型股市崩溃预警效果. 本文将从以下几个方面开展研究工作: 首先, 采用文本挖掘技术结合情感词典构建媒体情绪指数; 其次, 对现有 LPPL 模型进行改进, 将媒体情绪指数纳入建模过程; 最后, 通过实证数据检验改进模型与基准模型的预测效果.

2 模型介绍

2.1 媒体情绪指数的构建

2.1.1 中文金融情感词典介绍

本文主要基于 Sentiment 等^[44]构建的中文金融词典对获取的媒体评论文本进行情感标注: 首先, 采用知网 HOWNET、大连理工大学 DLUTSD 及台湾大学 NTUSD 作为基础词典; 然后, 使用在线路演纪要, 业绩说明电话会议纪要, IPO 招股说明书及公司年报构建基础语料库; 最后, 基于算法和人工删选, 使用多阶段剔除法来构建中文金融情感词典 CFSD. 具体步骤如下.

步骤 1 通过合并 HOWNET, DLUTSD 和 NTUSD 中的肯定和否定术语并删除重复项, 创建基础情感词典.

步骤 2 收集 1411 个在线路演纪要, 7138 个电话会议纪要, 2043 个 IPO 招股说明书和 29737 份年度财务报告. 运用 jieba 分词包 (Python 分词工具) 用于分割这些文档以创建基本语料库.

步骤 3 根据步骤 1 中创建的基础情感词典, 计算基本语料库中每个积极词和消极词的词频. 排除频率为零的词. 在创建基本的中国金融情感词典 (CFSD0.0) 之前, 将与财务文档无关的术语手动排除.

步骤 4 CFSD 0.0 中的所有术语均来自 HOWNET, DLUTSD 和 NTUSD. 由于这 3 个词典在财务中不包括一些常用的否定词和肯定词, 因此手动添加了 100 个常用的财务否定词和 100 个肯定词以生成字典 CFSD 0.1.

步骤 5 Gensim(Python 软件包) 用于将基本语料库转换为单词向量模型. 单词向量模型用于找出 CFSD 0.1 中每个术语的前 50 个同义词. 在手动删除与财务、金融无关的所有同义词之后, 创建 CFSD 0.2.

步骤 6 将 CFSD0.0, CFSD0.1 和 CFSD0.2 合并以删除重复项以获得 CFSD. CFSD 中有 1488 个否定词和 1107 个肯定词.

2.1.2 媒体情绪指数构建

利用前述构建的 CFSD 词典对媒体评论进行情感标注, 在情感标注过程中, 对词汇所表达的情感强度不作进一步区分, 仅通过统计词汇出现的次

数的高低来描述情感得分. 每一自然日出现积极、消极情绪得分的表达式如 (1) 式所示:

$$\text{Senti}_{t,k} = \sum_{j=1}^{m_k} n_{t,j} \quad (k = \text{Pos}, \text{Neg}), \quad (1)$$

式中, $k = \text{Pos}, \text{Neg}$ 分别表示积极、负面情绪, $\text{Senti}_{t,k}$ 为第 t 个自然日的媒体股评文章标题中情绪类型 k 的得分, m_k 为隶属于情绪类型 k 的词汇出现的次数, $n_{t,j}$ 为第 t 个自然日股评文章标题中第 j 个词汇出现的次数.

由于每一天的股评标题里可能同时包含积极词汇、消极词汇, 为了综合评价某一天的情绪值, 本文参考王晓丹等^[45]的处理方法, 得到综合的媒体情绪指数, 其表达式为

$$\text{sentiment}_t = \ln[(1 + \text{pos}_t)/(1 + \text{neg}_t)], \quad (2)$$

式中, pos_t 表示第 t 天的积极情感得分, neg_t 表示第 t 天的消极情感得分. 当积极情感得分大于消极情感得分时, 判定相互抵消后的综合情绪值大于 0, 表明情绪偏乐观; 反之, 当积极情感得分小于消极情感得分时, 综合情绪值小于 0, 表明情绪偏悲观.

2.2 LPPL-MS 组合模型的构建

2.2.1 LPPL 模型介绍

Sornette 等^[1]认为金融市场具有典型的复杂系统的特性. 发生在复杂系统中的极端事件往往是由于系统内部相互作用、自组织等自然演化的结果, 并非由外部冲击驱使. Johansen 和 Sornette^[2]认为金融市场上由于投资者之间的羊群效应及正反馈机制造成的资产价格泡沫的膨胀与破裂可由 LPPL 模型进行刻画. LPPL 模型具有如下特性:

- 1) 投资者之间相互模仿, 呈现集体追涨的交易局面, 造成对数资产价格呈超指数暴涨;
- 2) 随着资产价格泡沫破裂的风险逐渐扩大, 市场需要更高的回报作为补偿;
- 3) 资产价格泡沫在临界点 t_c 具有最大的崩盘概率;
- 4) 越接近临界点 t_c , 资产价格对数周期性振荡频率越高.

在 LPPL 模型中, 股票价格在泡沫阶段的演化过程如 (3) 式所示:

$$\frac{dp}{p} = \mu(t) dt + \sigma(t) dW - \kappa dj, \quad (3)$$

式中 p 是股票价格; $\mu(t)$ 为趋势项; W 为满足均值为 0, 方差为 1 的维纳过程; dj 为非连续的跳跃, 泡沫崩溃以前, $dj = 0$, 崩溃之后 $dj = 1$. 与发生崩溃对应的价格损失幅度由参数 κ 确定. 假定 $h(t)$ 为崩溃发生的概率, 则有: $E_t[dj] = 1 \times h(t) dt + 0 \times (1 - h(t)) dt = h(t) dt$.

在 LPPL 模型中, 假设股价崩盘的概率 $h(t)$ 如 (4) 式所示^[2]:

$$h(t) = B'(t_c - t)^{m-1} + C'(t_c - t)^{m-1} \times \cos(\omega \ln(t_c - t) - \phi), \quad (4)$$

式中 B', C' 均大于 0; t_c 为金融资产价格见顶的时刻, 也称为崩溃临界点; m 为幂增长指数, 代表股价增长的加速度, m 越小, 股价增长的加速度越快, 其临近泡沫的可能性也越大. ω 为振荡的角频率, ϕ 为振荡的初始相位.

由 $E_t[dj] = h(t)dt$, $E_t[dW] = 0$ 及满足市场不存在无风险套利条件, 即 $E_t[dp] = 0$, 对 (3) 式两端取期望, 可得 $\mu(t) = \kappa h(t)$, 进一步, 假定股价崩盘已经产生, (3) 式简化为

$$\frac{dp}{p} = \mu(t) dt + \sigma(t) dW = \kappa h(t) dt + \sigma(t) dW, \quad (5)$$

进一步, 对 (5) 式两端取期望, 得

$$E_t \left[\frac{dp}{p} \right] = \kappa h(t) dt. \quad (6)$$

(6) 式取积分结合 (4) 式即得 LPPL 方程^[2]:

$$\ln(p(t)) = A + B(t_c - t)^m + C(t_c - t)^m \times \cos(\omega \ln(t_c - t) - \phi), \quad (7)$$

式中 $A > 0$, 为股票在临界点的价格, 也即股票见顶时的价格; $B = -\kappa B'/m$; $C = -\kappa C'/\sqrt{m^2 + \omega^2}$. 并且, $B(t_c - t)^m$ 反映了股价由正反馈机制驱使而出现的幂律增长特性, $C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi)$ 刻画了股价在临近见顶并崩盘时呈对数周期性振荡, 是对股价幂律增长行为的修正. 图 1 以中国上证指数 2015 年泡沫期间走势为例, 展示了 LPPL 模型中呈现的对数周期幂律振荡现象.

如图 1 所示, 在线性尺度下, 越接近崩盘的临界点, 其振荡的频率越快, 其表现为图中指数振荡的间隔周期越来越短, 如图 1 中振荡的周期依次为 116 (由余弦函数的波峰或波谷数值相减可得, 如第一个周期为 200 减去 84 为 116 个交易日, 以此类推), 67, 38, 22, 13 个交易日. 而在对数尺度

下, 其振荡频率为常数, 如图 1 中所示, 分别对线性尺度下的振荡周期取对数, 则其周期为 $\ln(116) - \ln(67) = \ln(67) - \ln(38) = \ln(38) - \ln(22) = \ln(22) - \ln(13) = 0.54$. 并且股价越是接近临界点, 股价增长速度越快, 即股价增长率呈现单调递增的特点.

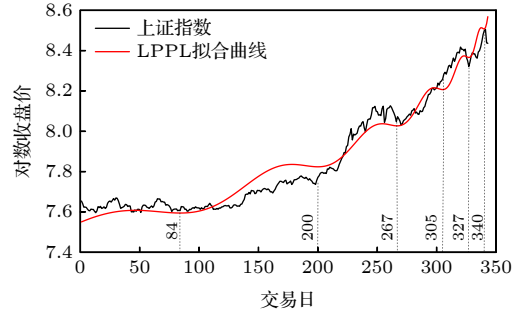


图 1 上证指数 2015 年泡沫期间 LPPL 拟合图

Fig. 1. LPPL fitting diagram of the Shanghai stock exchange (SSE) index during the 2015 bubble.

2.2.2 LPPL-MS 模型的构建

基础的 LPPL 模型未将投资者情绪因素纳入到建模过程当中, 本文创新性地对 LPPL 模型中的崩溃概率公式 $h(t)$ 进行修改, 以刻画媒体情绪因素对崩溃风险的影响. 如前文所述, 已有研究均已表明, 投资者情绪对股价具有显著的影响, 因此, 本文假定在股市泡沫阶段, 媒体积极评论将进一步助推市场情绪的高涨, 加速羊群效应, 放大市场正反馈机制, 驱使股价加速膨胀, 导致其崩盘概率增大. 反之, 媒体负面评论则可能平抑过热的市场情绪, 延缓羊群效应, 抑制市场正反馈机制, 延缓崩盘时刻的到来. 考虑媒体情绪 (media sentiment, MS) 影响后的股价崩溃函数 $h(t)$ 如 (8) 式表示:

$$h(t) = B'(t_c - t)^{m-1} (1 + \text{sentiment}_{t-1})^\lambda + C'(t_c - t)^{m-1} \cos(\omega \ln(t_c - t) - \phi), \quad (8)$$

式中 sentiment_{t-1} 为 $t-1$ 时刻媒体情绪值; $\lambda > 0$, 则表明积极情绪会放大正反馈效应, 消极情绪会抑制正反馈效应, 否则不成立. 同上, 与 2.2.1 节中情形类似, 可得到考虑媒体情绪后的 LPPL 方程的表达式:

$$\begin{aligned} \ln(p(t)) &= A + B(t_c - t)^m (1 + \text{sentiment}_{t-1})^\lambda \\ &+ C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi). \end{aligned} \quad (9)$$

3 实证分析

3.1 数据来源

Johansen 等^[46]提到, LPPL 模型刻画的股市崩盘具有典型的 3 个特征: 1) 股价达到明显的峰值; 2) 股价至少在 6 个月内维持上涨的态势; 3) 股价达到峰值后在短期内便迅速下降. 由于自 2014 年 1 月 1 日起, 新闻来源众多, 而 2013 年及之前的网站来源少、新闻数据量少, 结合 LPPL 模型所满足的 3 个特征, 本文选择上证指数 2014 年 1 月 2 日至 2015 年 5 月 29 日、创业板指数 2014 年 1 月 2 日至 2015 年 5 月 20 日收盘价数据及对应新闻媒体评论数据作为样本进行拟合. 所有样本数据均来源于东方财富旗下 Choice 金融终端, 通过在 Choice 金融终端抓取关键词“*A 股*”、“*大盘*”相关的各类媒体评论, 汇总每一自然日新闻财经媒体对当下股票市场整体运行状况的评价及未来走势预判. 其数据覆盖面较为全面, 其数据均来源于诸如, 沪深交易所、中国基金报、券商中国、澎湃新闻、21 世纪经济报道、中证网、中金网等国内主流媒体. 部分媒体评论稿标题信息如图 2 所示.

采用 2.1 节中构建的情感词典扫描每一条评论的标题, 统计每条评论对应的积极词汇、消极词汇出现的次数, 再结合 (2) 式, 便得到每一天的情绪值. 本文所参照的 CFSD 金融情感词典部分内容如表 1 所示.

媒体情绪指数描述性统计分析如表 2 所示. 从表 2 中可看出, 本文所选取的媒体情绪样本中积极情绪指数明显高于消极情绪指数, 表明各类财经媒

体的评论具有明显的倾向性, 其倾向于对市场做出积极正面的评论. 受央行多次降准降息, 市场无风险利率不断下探, 叠加国家实施积极财政政策及国企改革等一系列“利好”因素的推动下, 2015 年上半年 A 股市场节节攀升, 上证指数在 2015 年 6 月 12 日最高上涨到 5178.19 点, 创业板指数在 2015 年 6 月 5 日最高上涨到 4037.96 点, 随后随着市场泡沫的破裂, 仅两个月时间, 上证指数从高位滑落至 2850.71 点, 创业板则滑落至 1779.18 点. 上证指数、创业板指数泡沫前后的走势如图 3 所示.

表 1 CFSD 词典部分积极、消极词汇
Table 1. Some positive and negative words in CFSD dictionary.

序号	积极词汇	消极词汇
1	突破	拖累
2	增长	不涨反跌
3	高涨	分化
4	火爆	沉重打击
5	赚钱	恐慌
6	上升	下降
7	一枝独秀	狂跌
8	蒸蒸日上	跌停板
9	活跃	退市
10	提速	暴涨暴跌

表 2 (MS) 序列描述性统计分析
Table 2. Descriptive statistical analysis of media sentiment (MS) sequence.

最大值	最小值	均值	标准差	偏度	峰度
1.794	-0.998	0.252	0.446	0.037	3.158

序号	时间	标题	来源	作者	网址
34001	2015-03-22	新电改方案已内部下发 A股谁最受益?	每日经济新闻	—	http://finance.eastmoney.com/news/1350,20150322488693836.html
34002	2015-03-22	政策支持力度前所未有 A股正式进入大牛市中期	证券市场红周刊	—	http://finance.eastmoney.com/news/1345,20150322488691035.html
34003	2015-03-21	连续四日破万亿 A股春季躁动的四大关注	大众证券报	—	http://stock.eastmoney.com/news/1406,20150321488637494.html
34004	2015-03-21	让A股更好地发挥直接融资功能	金融投资报	—	http://stock.eastmoney.com/news/1406,20150321488633312.html
34005	2015-03-21	高溢价背后的诱惑: 境外公司主动退市奔A股	21世纪经济报道	—	http://finance.eastmoney.com/news/1344,20150321488576880.html
34006	2015-03-20	沪市活跃股公开信息(A股)(03/20)	上海证券交易所	—	http://stock.eastmoney.com/news/1403,20150320488473234.html
34007	2015-03-20	深市活跃股公开信息(A股)(03/20)	深圳证券交易所	—	http://stock.eastmoney.com/news/1404,20150320488470957.html
34008	2015-03-20	A股再现放量上行 沪指“八连阳”站稳3600点	新华网	—	http://stock.eastmoney.com/news/1406,20150320488449017.html
34009	2015-03-20	深南电A股东抛631万股 占股本1.05%	全景网	王汉卓	http://finance.eastmoney.com/news/1354,20150320488351772.html
34010	2015-03-20	连续三日现万亿成交 A股短期或将高位震荡	新快报	曹勇	http://stock.eastmoney.com/news/1406,20150320488327273.html
34011	2015-03-20	亚太股市整固先前涨势 A股盘中再触3600点	FX168	—	http://stock.eastmoney.com/news/1406,20150320488324036.html
34012	2015-03-20	资金跑步入市为A股调整减压	金融投资报	—	http://stock.eastmoney.com/news/1406,20150320488300520.html
34013	2015-03-20	如意集团遭公开叫板 A股做空时代来临?	证券时报	—	http://stock.eastmoney.com/news/1406,20150320488292251.html
34014	2015-03-20	证监会“开门红” 互联网金融企业A股上市“躁动”	金陵晚报	储伟伟	http://stock.eastmoney.com/news/1406,20150320488281848.html
34015	2015-03-20	回归A股? 李宇说: 一切皆有可能	中国经营网	—	http://finance.eastmoney.com/news/1354,20150320488280679.html
34016	2015-03-20	六间房为互联网公司回归A股铺路	证券时报	—	http://finance.eastmoney.com/news/1349,20150320488106851.html
34017	2015-03-20	A股保险中介第一股IPO在即 华泰经纪启动全球化布局	21世纪经济报道	—	http://finance.eastmoney.com/news/1354,20150320488105510.html
34018	2015-03-20	巴菲特指教你掂量A股上行空间	每日经济新闻	—	http://finance.eastmoney.com/news/1345,20150320488102863.html
34019	2015-03-20	做多中国 A股ETF在美受追捧	上海证券报	—	http://stock.eastmoney.com/news/1406,20150320488101426.html
34020	2015-03-19	A股强震暗示动能前弱 权重股或将扭转乾坤	中国经济网	—	http://stock.eastmoney.com/news/1406,20150319488041567.html
34021	2015-03-19	沪市活跃股公开信息(A股)(03/19)	上海证券交易所	—	http://stock.eastmoney.com/news/1403,20150319488021746.html
34022	2015-03-19	深市活跃股公开信息(A股)(03/19)	深圳证券交易所	—	http://stock.eastmoney.com/news/1404,20150319488015611.html
34023	2015-03-19	A股将迎来保险中介第一股	证券时报网	—	http://stock.eastmoney.com/news/1406,20150319487982759.html
34024	2015-03-19	A股迎震荡 牛市亦有“插曲”	中国证券报	王威	http://stock.eastmoney.com/news/1406,20150319487954616.html
34025	2015-03-19	资金进入股市供央行肯定 A股慢牛行情“理直气壮”	通信信息报	陈湘	http://stock.eastmoney.com/news/1406,20150319487952729.html
34026	2015-03-19	沪指攻克历史次高点 A股已在改革的风口上	通信信息报	郭培湘	http://stock.eastmoney.com/news/1406,20150319487950241.html
34027	2015-03-19	A股突破近几年大底形态	羊城晚报	黄智华	http://stock.eastmoney.com/news/1406,20150319487919616.html
34028	2015-03-19	多重利好共振A股任性上涨 2元股已绝迹	新快报	曹勇	http://stock.eastmoney.com/news/1406,20150319487857817.html

图 2 部分财经媒体评论稿标题信息样本

Fig. 2. Some financial media commentary title information samples.

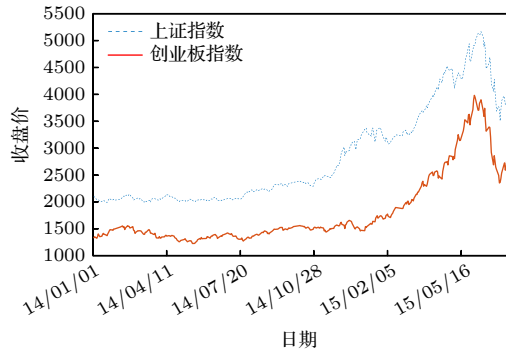


图3 上证指数、创业板指数泡沫前后走势

Fig. 3. Trend of SSE index and growth enterprise market (GEM) index before and after bubble.

3.2 LPPL-MS 拟合结果及分析

为了得到模型的参数估计结果, 本文采用遗传算法 (genetic algorithm, GA) 来获取 LPPL、LPPL-MS 参数估计值. 其中, 上证指数的拟合样本区间为 2014 年 1 月 2 日至 2015 年 5 月 29 日; 创业板指数的拟合样本区间为 2014 年 1 月 2 日至 2015 年 5 月 20 日. GA 算法的适应度函数为拟合值与实际值的均方误差, 其种群规模设为 20, 交叉概率设为 0.5, 变异率设为 0.05, 迭代次数设为 100 次, 为了使预测结果更加稳健, 选取适应度最佳的 3 组拟合结果进行分析. LPPL 及 LPPL-MS 模型的 GA 算法估计结果如表 3、表 4 所示.

表 3、表 4 中 MSE 指标为 LPPL, LPPL-MS

模型拟合值与实际值的均方误差, 其余各参数的意义同 2.2 节中的介绍. 所有模型拟合得到的幂增长指数 m 介于 0 到 1 之间, 表明市场崩溃的概率 $h(t)$ 具有不断加速的特征, 满足产生崩盘的条件. 此外, LPPL-MS 模型中, λ 均大于 0, 表明媒体情绪对正反馈效应具有显著影响, 即积极情绪会放大正反馈效应, 消极情绪会抑制正反馈效应. 若 $N \leq t_c < N+1$, 则认为模型预测的股价见顶时点为第 $N+1$ 个交易日. 如上证指数 LPPL 情形 1 估计的临界时点为 368.952, 则认为其估计的大盘见顶的临界时点为 369, 其崩盘时点为 370, 即大盘见顶后的下一个交易日, 其他情形类似. 上证指数实际见顶的时刻为 353, 崩盘时点为 354, 对应交易日为 2015 年 6 月 15 日; 创业板指数见顶时刻为 346, 崩盘时点为 347, 对应交易日为 2015 年 6 月 4 日. 表 5 进一步对比不同模型预测的大盘见顶的临界点与大盘实际见顶时间的差值的绝对值, 若模型预测的见顶时点与真实的见顶时点越接近, 则表明模型预警效果越好^[47].

从表 5 中可看出, LPPL-MS 无论是在上证指数还是创业板指数情形下, 其预测的见顶时点均更接近真实值, 其预测的崩溃临界时点与真实见顶时点的最低绝对误差、最高绝对误差、平均绝对误差均小于 LPPL 模型. 图 4 直观展示了不同模型的预测表现.

表 3 上证指数 LPPL, LPPL-MS 参数估计结果

Table 3. Parameter estimation results of SSE index LPPL and LPPL-MS.

	MSE	A	B	C	m	ω	ϕ	λ	t_c
LPPL	0.997	10.131	-0.864	-0.017	0.188	12.098	1.748	—	369.022
	1.078	10.161	-0.828	-0.015	0.197	11.991	1.840	—	379.036
	1.112	9.949	-0.805	0.014	0.188	10.368	2.029	—	361.668
LPPL-MS	0.948	9.639	-0.562	-0.014	0.226	12.398	6.686	0.008	362.729
	0.991	9.597	-0.509	-0.013	0.240	12.911	3.717	0.007	365.669
	1.038	9.552	-0.508	-0.012	0.237	12.953	3.671	0.010	360.949

表 4 创业板指数 LPPL、LPPL-MS 参数估计结果

Table 4. Estimation results of LPPL and LPPL-MS parameters of gem index.

	MSE	A	B	C	m	ω	ϕ	λ	t_c
LPPL	2.429	8.444	-0.334	0.006	0.239	13.988	2.241	—	335.923
	2.443	8.654	-0.436	-0.015	0.217	13.009	3.989	—	341.826
	2.477	8.517	-0.355	-0.011	0.236	13.039	3.998	—	340.435
LPPL-MS	2.703	8.617	-0.395	0.011	0.229	15.242	0.288	0.015	344.122
	2.819	9.218	-0.766	0.016	0.174	16.445	5.320	0.010	352.216
	2.904	8.922	-0.559	0.010	0.200	21.717	0.256	0.010	353.759

表 5 上证指数、创业板指数预警误差比较分析
Table 5. Comparative analysis of early warning errors between SSE index and GEM index.

		最低绝对误差	最高绝对误差	平均绝对误差
LPPL	上证指数	9	27	18
	创业板指	4	10	7
LPPL-MS	上证指数	8*	13*	11*
	创业板指	1*	8*	6*

注: *表明预测的见顶时点最接近真实值.

图 4 中, 不同颜色的线条分别对应不同的拟合情形, 所有图中上证指数、创业板指数实际的见顶时点所在的时刻均用“^”符号标出. “*, +, <”分别对应情形 1, 2, 3 下预测得到的见顶时点. 其中, 上证指数对应的大盘见顶时点为 353, 对应的交易日期为 2015 年 6 月 12 日, 创业板指数对应的大盘见顶时点为 346, 对应的交易日期为 2015 年 6 月 3 日. 上证指数 LPPL 三种拟合情形预测的大盘见顶时点分别为 370, 380, 362, 即分别比实际的见顶时点晚 17, 27, 9 个交易日. 上证指数 LPPL-MS 模型三种拟合情形预测的大盘见顶时点分别为 363, 366, 361, 即分别比实际的见顶时点晚 10, 13, 8 个

交易日. 从图中也可看出, 上证指数 LPPL-MS 模型预测的大盘见顶时点相比较 LPPL 模型更为接近真实值. 同理, 创业板指数 LPPL 三种拟合情形预测的大盘见顶时点分别为 336, 342, 341, 其分别比实际的见顶时点提前 10, 4, 5 个交易日. 创业板指数 LPPL-MS 三种拟合情形预测的大盘见顶时点分别为 345, 353, 354, 其分别比实际见顶时点提前 1 个交易日、迟延 7 个交易日、延迟 8 个交易日, 其相比 LPPL 模型预测的见顶时点更为接近真实值. 图 4 的直观展示结果亦表明 LPPL-MS 组合模型预警效果优于 LPPL 模型.

3.3 LPPL, LPPL-MS 拟合结果 O-U 过程检验

为了进一步验证 LPPL, LPPL-MS 拟合结果的稳健性, 对其拟合残差进行 Ornstein-Uhlenbeck (O-U) 过程检验. Lin 等^[48]指出, 如果股票市场在泡沫期间的对数收盘价可以用 LPPL 模型进行有效解释, 则其拟合的残差应该满足 O-U 过程. 其拟合的残差若满足平稳性, 则表明其符合 O-U 过程. 表 6 列出了上述模型不同拟合情形下残差的单位根检验结果.

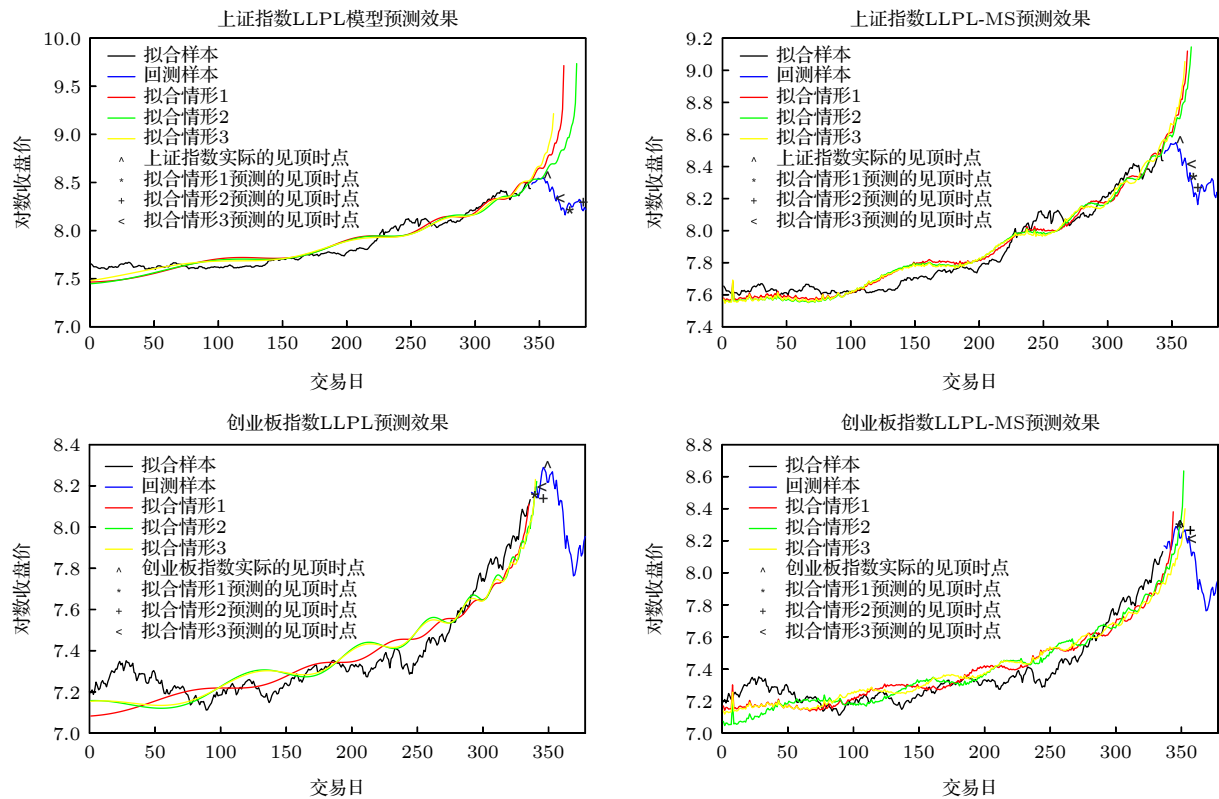


图 4 上证指数、创业板指数不同模型预测表现比较分析

Fig. 4. Comparative analysis of forecasting performance of different models of Shanghai Stock Exchange Index and Growth Enterprise Market Index.

表 6 上证指数、创业板指数不同拟合情形下残差平稳性检验

Table 6. Residual stationarity test of SSE index and GEM index under different fitting conditions.

		拟合情形1	拟合情形2	拟合情形3
LPPL	上证指数	0.014**	0.021**	0.013**
	创业板指	0.0428**	0.073*	0.0959*
LPPL-MS	上证指数	0.001***	0.001***	0.012**
	创业板指	0.019**	0.041**	0.074*

注: *表明在10%水平下显著; **表明在5%水平下显著; ***表明在1%水平下显著.

表 6 中的数值为 LPPL, LPPL-MS 不同拟合情形下残差 Dickey-Fuller 检验的 P 值. 从表 6 可看出, 上述 LPPL, LPPL-MS 模型不同拟合情形下残差均通过了平稳性检验, 表明其拟合残差满足 O-U 过程, 且 LPPL-MS 模型相比 LPPL 模型具有更小的 P 值, 其拟合结果更为稳健. 本文所采用的 LPPL、LPPL-MS 模型可有效解释 2015 年上证指数、创业板指数泡沫期间的运行情况.

4 结 论

本文采用文本挖掘技术生成财经媒体情绪指数 MS, 通过将 MS 指数纳入传统的 LPPL 模型, 构建 LPPL-MS 组合模型对中国上证指数、创业板指数 2015 年 6 月份的崩盘进行预警. 实证分析得出以下结论.

1) 通过文本挖掘技术结合大数据分析, 抓取财经媒体对 A 股的评论信息, 通过分析其评论措辞的情感倾向可有效衡量媒体情绪. 并且, 财经媒体倾向于对市场做出乐观的评价, 以煽动其他投资者的情绪. 由于中国股市中个人投资者基数庞大, 尤其是其中的新进投资者, 其缺乏交易经验, 往往容易受财经媒体具有煽动性报道内容的影响. 以致于媒体情绪一定程度上能影响股票市场的表现.

2) 中国 A 股市场在 2015 年上半年开启单边暴涨模式, 并出现严重的泡沫, 其泡沫阶段的运行情况符合 LPPL 模型中描述的超指数膨胀及对数周期幂律加速振荡特征. 采用 LPPL 模型可有效预警 2015 年 6 月份上证指数、创业板指数的崩溃.

3) 在股票市场泡沫期间, 媒体频繁的带有煽动性正向情感倾向的股评报道可引起市场的极度亢奋情绪, 使市场的崩盘概率进一步加速. 通过在传统的 LPPL 模型中纳入媒体情绪指标 MS, 构建

LPPL-MS 组合模型预警上证指数、创业板指数 2015 年 6 月份的崩盘, 具有更高的预测精度, 其预测大盘见顶的临界点相比 LPPL 模型更接近真实值. 并且其拟合残差单位根检验的 P 值小于 LPPL 模型, 其模型拟合结果更为稳健.

本文的研究结果有助于政府相关监管当局提前预警股市系统性风险, 为其提前制定风险管控政策做铺垫. 同时, 投资者也借助该模型提前优化交易策略, 及时规避系统性风险, 减少投资损失. 本文将投资者情绪因素纳入到 LPPL 模型, 以改进 LPPL 模型预警效果, 然而, 股票市场本身作为一个复杂系统, 其市场运行规律受内外部多维复杂因素影响, 其不仅仅受市场情绪影响, 还受到诸如市场基本面因素、市场技术面特征、外围市场冲击等因素影响, 未来的研究可将上述因素纳入到模型当中, 以进一步优化模型预警效果.

参考文献

- [1] Sornette D, Johansen A, Bouchaud J P 1995 *J. Phys. I* **6** 167
- [2] Johansen A, Sornette D 1999 *Risk* **12** 91
- [3] Jensen M C 1967 *Financ. Anal. J.* **23** 77
- [4] Fama U 1965 *J. Bus.* **38** 34
- [5] Jiang Z Q, Tian J W, Zhou W X 2019 *J. Manag. Sci. China* **22** 92 (in Chinese) [蒋志强, 田婧雯, 周炜星 2019 *管理科学学报* **22** 92]
- [6] Xu X, Zhu C, Zhu X Q 2021 *Acta Phys.* **70** 088901 (in Chinese) [徐翔, 朱承, 朱先强 2021 *物理学报* **70** 088901]
- [7] Gurkaynak R S 2008 *J. Econ. Surv.* **22** 166
- [8] Bak P, Tang C, Wiesenfeld K 1987 *Phys. Rev. Lett.* **59** 381
- [9] Pan N, Wang Z J, Zhou Y 2018 *Chin. J. Manag. Sci.* **26** 25 (in Chinese) [潘娜, 王子剑, 周勇 2018 *中国管理科学* **26** 25]
- [10] Zhou W X 2007 *Introduction of Econophysics* (Shang-hai: Shanghai University of Finance and Economics Press) pp17-33 (in Chinese) [周炜星 2007 *金融物理学导论* (上海: 上海财经大学出版社) 第17—33页]
- [11] Sornette D, Johansen A 1997 *Physica A* **245** 411
- [12] Sornette D, Johansen A 1998 *Physica A* **261** 581
- [13] Sornette D, Zhou W X 2006 *Int. J. Forecast* **22** 153
- [14] Sornette D, Woodard R, Zhou W X 2009 *Physica A* **388** 1571
- [15] Kyriazis N, Papadamou S, Corbet S 2020 *Res. Int. Bus. Financ.* **54** 101254
- [16] Omer O 2020 *Eur. Phys. J. Spec. Top.* **229** 1715
- [17] Sirca S J, Omladic M 2020 *ARS Math. Contemp.* **13** 63
- [18] Filimonov V, Sornette D 2013 *Physica A* **392** 3698
- [19] Jiang Z, Zhou W X, Sornette D, et al. 2010 *J. Econ. Behav. Organ.* **74** 149
- [20] Yan W, Rebib R, Woodard R, Sornette D 2012 *IJPAM* **1** 59
- [21] Li D 2012 *J. Huaihai Insti of Tech. (Natural Science Edition)* **21** 4 (in Chinese) [李东 2012 *淮海工学院学报(自然科学版)* **21** 4]
- [22] Zhou W, He J M 2011 *J. Financ. Res.* **375** 65 (in Chinese) [周伟, 何建敏 2011 *金融研究* **375** 65]
- [23] Li S J, Li D X, Wang L Y 2017 *Shanghai J. Econ.* **7** 42 (in

- Chinese) [李斯嘉, 李冬昕, 王粟昶 2017 上海经济研究 **7** 42]
- [24] Li L Y, Zhang X 2019 *J. Financ. Res.* **474** 169 (in Chinese) [李伦一, 张翔 2019 金融研究 **474** 169]
- [25] Chen W H, Cai W J 2018 *Statistics and Decision* **497** 143 (in Chinese) [陈卫华, 蔡文靖 2018 统计与决策 **497** 143]
- [26] Zhao L, Liu Q 2020 *Statistics and Decision* **558** 128 (in Chinese) [赵磊, 刘庆 2020 统计与决策 **558** 128]
- [27] Kahneman D 2003 *Am. Econ. Rev.* **93** 1449
- [28] Baker M, Wurgler J 2006 *J. Financ.* **61** 1645
- [29] Baker M, Wurgler J 2007 *J. Econ. Perspect* **21** 129
- [30] García D 2013 *J. Financ.* **68** 1267
- [31] Fang L, Peress J 2009 *J. Financ.* **64** 2023
- [32] Barber B M, Odean T 2008 *Rev. Financ. Stud.* **21** 785
- [33] Solomon D H 2012 *J. Financ.* **67** 599
- [34] Nardo M, Petracco-Giudici M, Naltsidis M 2016 *J. Econ. Surv.* **30** 356
- [35] Fisher K L, Statman M 2000 *Financ. Anal. J.* **56** 16
- [36] Wu J W, Wang C Y, Chen Z L, Jie M G 2020 *J. Financ. Res.* **476** 147 (in Chinese) [武佳薇, 汪昌云, 陈紫琳, Jie Michael Guo 2020 金融研究 **476** 147]
- [37] Li H L, Feng C E 2014 *System Eng. Theor. Prac.* **34** 2495 (in Chinese) [李合龙, 冯春娥 2014 系统工程理论与实践 **34** 2495]
- [38] You J X, Wu J 2012 *Econ. Res. J* **534** 141 (in Chinese) [游家兴, 吴静 2012 经济研究 **534** 141]
- [39] Wu D D, Zheng L, Olson D L 2014 *IEEE T. SYST Man CY-S.* **44** 1077
- [40] Dai D B, Lan Y S, Fan T J, Zhao M 2019 *China Soft Sci.* **340** 166 (in Chinese) [戴德宝, 兰玉森, 范体军, 赵敏 2019 中国软科学 **340** 166]
- [41] Martin G 1971 *In Computers, Communication, and the Public Interest* (Baltimore, MD: The Johns Hopkins Press) p37
- [42] Da Z, Engelberg J, Gao P 2011 *J. Financ.* **66** 1461
- [43] Huang X, Nekrasov A, Teoh S H 2018 *Account. Rev.* **93** 231
- [44] Sentiment Analysis, Bian S B, Jia D K, Li F <http://dx.doi.org/10.2139/ssrn.3446388> [2020-11-10]
- [45] Wang X D, Shang W, Wang S Y 2019 *System Eng. Theor. Prac.* **39** 3038 (in Chinese) [王晓丹, 尚维, 汪寿阳 2019 系统工程理论与实践 **39** 3038]
- [46] Johansen A, Ledoit O, Sornette D. 2000 *Int J Theor Appl Finan.* **3** 219
- [47] Yan W F, Woodard R, Sornette D 2014 *Quant. Financ.* **14** 1273
- [48] Lin L, Sornette D 2013 *Eur. J. Financ.* **19** 344

Research on Chinese stock market crash early warning based on improved log-periodic power law model*

Wu Jun-Chuan Tang Zhen-Peng[†] Du Xiao-Xu Chen Kai-Jie

(School of Economics and Management, Fuzhou University, Fuzhou 350108, China)

(Received 18 November 2020; revised manuscript received 8 September 2021)

Abstract

This paper is based on the famous log-periodic power law model (LPPL) in financial physics to warn of the collapse of China's Shanghai Composite Index and GEM Index in June 2015. In view of the existing research using the LPPL model to warn of market crash, only the historical trading data of the market are considered. For the first time, investor sentiment factors are incorporated into the modeling process of LPPL model to improve the early warning effect of LPPL model. Using the text mining technology combined with semantic analysis methods to grasp the financial media's stock evaluation report for word frequency statistics, in order to build the medium sentiment index. The further modified expression of the crash probability function in the LPPL model is represented as a function of historical trading data and medium sentiment, and thus constructing an LPPL-MS combination model to warn of stock market crash. The empirical results show that the LPPL-MS combination model constructed in this paper has higher warning accuracy than the LPL model, and its prediction crash time is closer to the actual crash time of the Shanghai Index and GEM Index, and its fitting results have passed the relevant test.

Keywords: LPPL-MS model, medium sentiment, stock market crash, warning

PACS: 02.50.-r, 02.50.Cw, 02.50.Ng, 02.60.Ed

DOI: 10.7498/aps.71.20201940

* Project supported by the National Natural Science Foundation of China (Grant Nos. 71573042, 71973028).

[†] Corresponding author. E-mail: zhenpt@126.com