



债券违约风险与机器学习方法: 研究进展与未来方向

齐欣林 林奕皓 姜富伟

摘要: 自刚兑打破以来, 我国债券市场违约趋于常态化, 迫切需要加强对信用债违约风险的识别和防控研究。本文系统梳理了债券违约风险的影响因素, 包括宏观经济、金融市场、企业特征、债券特征四大方面, 并总结了适用于债券违约风险预警的模型。进一步, 本文指出有必要将前沿机器学习方法应用于信用风险的精准识别与智能预警, 具体可以使用文本分析深入挖掘债券违约风险影响因素, 运用机器学习算法推动债券违约风险预警模型的迭代创新, 旨在为坚守不发生系统性金融风险的底线, 维护经济金融稳定, 推动金融市场高质量发展提供政策参考。

关键词: 债券市场 违约风险 文献综述 机器学习

一、引言

2022年党的二十大开启了全面建设社会主义现代化国家的新征程, 擘画出以中国式现代化全面推进中华民族伟大复兴的宏伟蓝图, 其中提到要加强和完善现代金融监管, 依法将各类金融活动全部纳入监管, 守住不发生系统性风险底线, 表明中国式现代化要求金融体系具有安全性的特征(何德旭和张雪兰, 2023)。

作为资本市场的重要组成部分, 我国债券市场在近年来蓬勃发展, 成为全球第二大债券市场, 具体表现为市场规模日益增大、参与主体日趋多元、市场活跃度稳步提升、制度框架不断完善。在债券市场

不断发展的同时, 违约常态化伴随的风险也不容小觑。2020年以来, 债券市场违约事件频出, 违约风险持续释放, 各类“信仰”继踵打破。据统计, 2018—2022年末, 共计有超过800只债券发生违约, 违约总金额达到7100余亿元, 平均每年违约金额达1435亿元, 违约规模远超2017年末的312亿元。日趋频发的超预期信用风险事件引发学界与业界对于债券违约风险的广泛关注。

党中央、国务院及各金融监管部门在防范化解系统性金融风险进程中高度重视信用风险防控, 持续夯实政策基础。2014年, 人民银行《中国金融稳定报告(2014)》

齐欣林, 中国工商银行总行金融市场部, Email:qixinlin@icbc.com.cn; 林奕皓, 中央财经大学金融学院博士研究生, Email:cufelyh2021@163.com; 姜富伟, 中央财经大学金融学院教授、博士生导师, Email:jfuwei@gmail.com。本研究得到国家社科基金重大项目(22&ZD063)和国家自然科学基金面上项目(72072193, 71872195)支持。

指出，要“有序打破刚性兑付”；2020年，中央经济工作会议上明确指出，要“完善债券市场法制”“打击各种逃废债行为”；2021年，我国“十四五”规划纲要指出，要“完善市场化债券发行机制，稳步扩大债券市场规模”；2021年8月，人民银行等部门联合发布的《关于推动公司信用类债券市场改革开放高质量发展的指导意见》指出，要“把主动防范化解债券市场风险放在更加重要的位置，筑牢市场准入、早期干预和处置退出三道防线，健全风险监测预警机制”；2022年，人民银行《中国金融稳定报告（2022）》指出，要“进一步完善债券市场法制，及时防范化解债券市场风险”。在债券市场“去刚兑”后违约常态化的现实背景下，加强对债券违约风险的研究具有重要的理论和实践价值。

梳理已有研究，关于债券违约风险的研究主要有两方面，一是挖掘债券违约风险影响因素，二是构建债券违约风险预警模型。为厘清债券违约风险领域的研究现状，本文从影响因素和预警模型两方面对现有文献进行全面、系统的梳理。在此基础上，本文围绕前沿机器学习方法，提出债券违约风险研究领域的未来展望，包括使用文本分析深入挖掘债券违约风险影响因素、使用机器学习算法推动预警模型创新等路径，为信用风险研究领域的有效拓展与延伸提供来自机器学习的全新视角。^①

二、债券违约风险的研究进展

学术界围绕债券违约风险的研究主要分为两类：影响因素和预警模型。本文从这两方面对已有文献进行梳理，致力于厘清债券违约风险的可能影响因素，并梳理适用于债券违约风险预警的模型，涵盖国内外经典及前沿文献的研究成果。

（一）债券违约风险影响因素

本文对债券违约风险影响因素的梳理主要参考债券信用利差相关文献，因为多数研究将信用利差作为债券违约风险的代理变量（史永东等，2021；杨国超和蒋安璇，2022）。此外，债券违约风险是企业债务违约风险的重要表现形式，企业融资成本等特征与违约风险相关性较大，因此在梳理影响因素过程中也借鉴了部分主题为企业债务风险、融资成本等文献的结论。按照“从宏观到微观”“从外部到内部”的梳理思路，本文从宏观经济、金融市场、企业特征、债券特征四方面切入，对代表性研究进行系统的归纳和总结。

1. 宏观经济层面的违约风险影响因素

从已有研究看，很多学者围绕债券违约风险的宏观经济影响因素开展研究，从结论看主要分为以下角度：宏观经济环境（包括宏观经济周期、经济政策不确定性）、宏观经济指标、宏观政策（包括货币政策）、宏观经济其他部门（包括居民

^① 本文后续安排如下：第二节，对债券违约风险的影响因素进行梳理，分为宏观经济、金融市场、企业特征、债券特征四大方面，并梳理适用于债券违约风险预警的模型，包括Altman模型、Merton模型、信用评级模型等；第三节，从机器学习视角对债券违约风险的未来研究进行展望；第四节，对全文进行总结。



部门、政府部门等), 相关研究可总结为表 1。可以发现, 宏观经济指标是较为重要的可观测因素, 货币政策、地方政府救助等调控手段能够反映在全国、地方宏观经济指标中, 从而对债券违约风险带来影响。

2. 金融市场层面的违约风险影响因素

相比于宏观经济环境, 金融市场环境对债券违约风险的作用更为直接, 发债企业所发行的股票和债券的特征都对债券违约风险具有影响。还有学者围绕金融市场其他特征探讨了债券违约风险影响因素。相关研究可总结为表 2。可以发现, 由于

投资者在金融市场的交易行为在一定程度上反映了对企业经营情况的分析和预期, 因此市场交易带来的企业金融资产特征变动能够识别债券违约风险。

3. 企业特征层面的违约风险影响因素

企业自身经营管理质效是影响债券违约风险的重要内部因素, 已有研究从产权性质、财务状况、公司治理角度来探讨企业特征对债券违约风险的影响。相关研究可总结为表 3。可以发现, 企业治理对于识别与防控债券违约风险尤为重要, 表现在内部治理和外部治理水平的提高都有助于降低违约风险。

表 1 宏观经济层面的违约风险影响因素

研究角度	作者及年份	主要发现
宏观经济环境	Tang & Yan, 2010	宏观经济周期会影响债券信用利差, 信用利差在经济扩张时期有所下降, 在经济衰退时期有所上升
	周宏等, 2011	宏观经济不确定性上升会显著提高债券信用风险
宏观经济指标	Longstaff & Schwartz, 1995; David, 2008; Giesecke et al., 2011	宏观经济指标会影响债券信用利差, 包括 GDP、股指收益、利率、通货膨胀、税收等
	王安兴等, 2012	在中国债券市场, 信用利差与 GDP 指数和 M1 发行量正相关, 与无风险利率负相关
宏观政策	王雄元等, 2015	货币政策不确定性会正向影响债券信用利差
	于静霞和周林, 2015	货币政策对信用利差的影响在信用等级更低、期限更长的债券中更为显著
	郭晔等, 2016	债券信用利差对未预期货币政策具有显著反应, 这种反应在经济繁荣时期更大
	黄振和郭晔, 2021	2013 年央行担保品管理框架设立后, 被纳入合格担保品框架的债券信用利差显著下降
	徐思等, 2022	“一带一路”倡议的实施能够显著降低支持企业的债券信用利差
宏观经济其他部门	汪莉和陈诗一, 2015; 罗荣华和刘劲劲, 2016	从地方政府角度, 地方政府的救助意愿和救助能力对于债券违约风险较为重要, 原因在于地方政府可能通过地方财力为债券提供隐性担保, 从而影响债券的潜在违约风险
	何德旭和张斌彬, 2021	从居民部门角度, 居民部门加杠杆会加剧企业债务违约风险, 作用机制包括缩短债务期限、减弱偿债能力、加深金融化趋势、增强“僵尸化”程度

表 2 金融市场层面的违约风险影响因素

研究角度	作者及年份	主要发现
股票特征	Merton, 1974	股票的收益和波动性会影响违约风险
	Brogaard et al., 2017	较高的股票流动性可以提高股价信息含量和公司治理水平, 从而降低违约风险
	王安兴等, 2012	公司债信用利差与股票收益波动率正相关, 与换手率变化、零交易天数也有关
债券特征	纪志宏和曹媛媛, 2017; Nagler, 2020	从债券流动性角度, 债券信用利差会随着债券流动性的恶化而增加, 并且这种影响会随着公司债展期风险的增大而增大
	Andreasen et al., 2021	从债券收益角度, 债券信用利差在经济扩张时期与债券收益正相关, 在衰退时与收益负相关, 宽松货币政策是转变的关键驱动因素
金融市场其他特征	杨璐和方静, 2021	投资者按期全额兑付预期的降低会显著提高债券的发行利差
	史永东等, 2021	中债估值跳跃能够显著提高债券信用利差, 下跳相对于上跳对信用利差的作用更大
	Brugler et al., 2022	二级市场透明度越高, 债券违约风险越低

表 3 企业特征层面的违约风险影响因素

研究类别	研究角度	作者及年份	主要发现
基本特征	产权性质	王博森等, 2016	国有产权能够发挥隐性担保的作用, 从而降低国有企业债券违约风险
	财务特征	Molina, 2005	公司盈利能力、偿债能力、现金流水平等因素对违约风险有重要影响, 比如过高的杠杆率会提高违约风险
内部治理	股东治理	史永东等, 2021	控股股东股权质押显著提高了发债企业债券信用利差, 这种现象在控制权转移风险更高的企业中更为明显
	高管治理	林晚发等, 2018	有高管担任过人大代表或政协委员的企业的债券发行成功率更高, 但更有动力盈余管理, 导致更高的发债成本
	董事会治理	Baghdadi et al., 2020	董事会构成的变化可能增加违约风险
外部治理	信息披露	Minnis, 2011	信息披露完整度越高, 债务融资成本越低
		Bonsall & Miller, 2017	信息披露文本可读性越高, 债务融资成本越低
		方红星等, 2013	信息披露文本情绪越积极, 债券信用利差越低
		吴武清等, 2021	债券募集说明书风险披露程度越高, 债券信用利差越高
		林晚发等, 2022	异常积极语调的管理层讨论与分析 (MD&A) 更有可能是操纵的结果, 暗示企业未来风险越高
社会责任承担	周宏等, 2016	企业承担社会责任越多, 债券信用利差越低	
税收征管	林晚发和顾乾坤, 2021	企业纳税评级越高, 债券信用利差越低	

4. 债券特征层面的违约风险影响因素
债券特征方面, 信用评级作为评级机构对债券违约风险的综合评价结果, 具有

一定参考价值。还有学者从其他角度进行了探索, 相关研究可总结为表 4。可以发现, 债券的静态特征 (包括承销商、担保等限



表 4 债券特征层面的违约风险影响因素

研究角度	作者及年份	主要发现
信用评级	何平和金梦, 2010 ; 王雄元和张春强, 2013	债券评级越高, 债券违约风险越低, 债券评级相对于主体评级对债券发行成本的影响更大
	陈关亭 等, 2021	多重信用评级有利于降低债券融资成本, 一致的多重信用评级更有利于降低债券融资成本
限制性契约条款	Bradley & Roberts, 2015	限制性契约条款有助于降低债券风险溢价, 作用机制包括信号传递和缓解道德风险
	钟辉勇 等, 2016	债券的担保条款能够提高债券评级, 但是对降低信用利差没有明显作用
	林晚发 等, 2022	债券的担保条款会增加债券违约风险, “评级包装”模型有助于解释“担保正溢价”之谜
	杨国超和蒋安璇, 2022	附有交叉违约制度的债券具有更高的违约风险, 该债券违约会对企业其他债券产生连锁违约效应
债券期限	王安兴 等, 2012	债券的剩余期限越长, 债券信用利差越低
债券承销商	林晚发 等, 2019	债券承销商评级越高, 债券信用利差越低, 这一现象在信息不对称更高、信用风险更高的企业中更明显

制性契约条款)、动态特征(包括信用评级、剩余期限)都对债券违约风险具有影响。

综上所述, 债券违约风险影响因素的相关研究较为全面, 包括宏观经济、金融市场、企业特征、债券特征等方面。一方面, 现有研究具有积极意义, 表现在丰富的研究结论为监管部门提供充分的理论依据, 拓展了债券违约风险识别需要注意的方面。另一方面, 现有研究也存在一定局限性, 体现在部分国外学者的研究结论是否适用于中国市场有待检验。由于市场环境存在一定差异, 同一因素对国内外债券市场违约风险的作用效果可能是不同的, 不能盲目将国外学者的研究结论迁移到中国, 在未来需要对中国市场开展针对性研究, 为相关领域的探讨提供来自中国资本市场的独特证据。此外, 影响因素的多元

性在一定程度上给监管部门带来无形挑战, 如何把众多影响因素纳入整体监管框架、构建更为有效的风险预警模型是值得思考的问题。

(二) 债券违约风险的预警模型

相比于分析债券违约风险影响因素, 学术界和实务界更关注如何准确地对债券违约事件提前预警。从理论上讲, 使用数理模型预测债券违约风险可以归结为如下形式:

$$Bond_Risk_{i,t+1} = E_i[Bond_Risk_{i,t+1}] + \varepsilon_{it} \quad (1)$$

其中: $E_i[Bond_Risk_{i,t+1}] = f(\mathbf{x}_{it})$, $\mathbf{x}_{it} = (x_{1,it}, x_{2,it}, x_{n,it})'$

上述表达式中包含两个重要参数: 指标变量 x_{it} 、函数形式 $f(\cdot)$ 。目前, 国内外尚未有公认有效的债券违约风险预警模型。本文在梳理预警模型时主要参考企业财务预警模型、债券定价模型等和违约相

关性较大的经典模型，这些模型中的财务预警得分、债券风险溢价等指标都能够作为预警债券违约风险的参考变量。信用评级作为评级机构通过内部模型得到的综合评价结果，对于预警债券违约在一定程度上也有效。本部分从经典模型和信用评级模型两个角度梳理适用于债券违约风险的预警模型。

1. 经典债券违约风险模型^①

根据模型函数形式，本文从线性和非线性两个角度梳理经典模型。早期的学者通常使用线性回归构建违约风险与少数几个指标的关系，得到指标系数后代入指标取值即可得到违约风险大小^②。随着研究推进，有学者从违约的经济学原理出发，通过理论推导构建非线性模型来测度违约风险，得到的模型虽然更具有理论基础，但也存在着假设严苛、变量有限的不足。

从非线性模型看，结构模型、强度模型是两大经典模型，侧重于分析债券风险定价问题。其一，结构模型的理论基础是期权定价理论，其代表性模型是Merton模型，通过计算违约距离得到违约风险(Merton, 1974)。在Merton模型的基础

上，Bharath & Shumway (2008) 和KMV公司分别对违约距离的计算方法加以修订，构建出简化Merton模型和KMV模型，由于具有计算简便、数据可得性强的优点得到广泛运用(Kedia et al., 2014；丁志国等, 2021)。但是，结构模型也存在假设严苛的缺陷，比如Merton模型假设只有当债券到期且企业资产价值小于债券面值时，企业才可能发生违约行为，这与实际情况不完全相符。其二，强度模型的理论基础是现金流贴现理论，其假设违约不依赖于企业资产价值，基于风险率模型将复杂的违约机制简化为简单的概率分布，认为违约概率由特定违约强度的泊松过程决定，使用贴现法来计算信用风险(Jarrow & Turnbull, 1995)。相比于结构模型，强度模型虽然原理简单、可拓展性好，但经济意义不足，还可能存在着债券市场不够有效带来的数据不准确问题。进一步，混合模型实现了对结构模型和强度模型的有机结合，综合分析企业财务结构和违约强度来计算违约风险(Giesecke, 2006)，但存在着操作复杂、参数估计困难等问题。

① 综合本部分所述，根据经典模型计算得到的企业财务预警得分、违约风险溢价等指标都可以预测债券违约风险，但也存在着两个问题。一是学术界和业界尚未对最优模型的选择达成共识，不同模型的表达式差异较大，使用不同模型得到的违约风险测算结果可能相差甚远。二是不同模型使用的指标各异、假设各不相同，考虑到债券违约风险影响因素的多样性，所述经典模型很可能存在遗漏重要变量、假设条件难以满足的问题，构建一个变量更为全面、形式更为灵活的预警模型是未来研究的重要方向。

② 从线性模型看，Altman模型最为经典，该模型侧重于分析企业财务预警问题，主要包括Altman Z值模型和ZETA模型(Altman, 1968; Altman et al., 1977)。其中，Z值预警得分当前被广泛地作为企业债务风险的代理变量(李建军和韩珣, 2019; 刘晓光和刘元春, 2019)。Altman Z值的计算方法简单，但选取的指标较少，可能存在指标选取不够全面的问题。还有学者使用Probit或Logit模型来构建违约风险和预测变量之间的关系，这两类模型都可以通过简单变换转化为线性形式，在本质上也是线性模型(潘泽清, 2018)。



2. 债券信用评级模型

信用评级是机构通过内部评级模型分析得到的债券违约风险综合评价指标。低评级债券通常确实具有较大风险，因此债券评级能够作为风险预警的参考指标，具有一定价值。另一方面，由于近年来高等级债券违约事件不断出现，债券信用评级的有效性饱受争议，被认为存在严重的评级膨胀现象，这说明信用评级模型的主要问题是部分对高风险债券的识别能力有限。

已有很多学者围绕信用评级膨胀现象的成因、不同评级模式对评级膨胀现象的影响等问题开展讨论，相关研究可总结为表 5。在此基础上，宋敏等（2019）提出评级机构事后问责、严格准入等有助于缓

解评级膨胀现象的建议。

三、债券违约风险研究的未来方向：机器学习视角

近年来，金融科技迈入高质量发展的新阶段，中国人民银行印发的《金融科技发展规划（2022—2025 年）》指出，要“践行安全发展观，运用数字化手段不断增强风险识别监测、分析预警能力”。在此背景下，以文本分析和前沿算法为代表的机器学习方法为债券违约风险研究提供了新的视角。

（一）利用文本分析技术挖掘影响因素

从现有研究看，多数关于债券违约风险影响因素的研究尚聚焦于结构化数据，

表 5 债券信用评级膨胀相关研究

研究角度	研究内容	主要论述	相关文献
评级膨胀成因分析	评级购买	侧重于研究公司行为，指的是发债公司可向多个评级机构征求评级，然后选择最高的评级作为债券评级	Sangiorgi & Spatt, 2017
	评级迎合	侧重于研究评级机构行为，指的是评级机构为了吸引更多的发债公司与自己合作，可能通过给出超出债券应有评级的评级结果	Bolton et al., 2012 ; Flynn & Ghent, 2018
	评级包装	企业和评级机构可能通过交流和协商来获得更高评级结果	林晚发 等, 2022
不同评级模式对比	评级结果对比	“发行人付费”下评级机构给出的评级结果更为乐观	Skreta & Veldkamp, 2009 ; Jiang et al., 2012 ; Bonsall, 2014 ; Baghai & Becker, 2018
	评级质量对比	发行人付费模式下的评级质量更优 投资人付费模式下的评级质量更优	Bonsall, 2014 Beaver et al., 2006 ; Cornaggia & Cornaggia, 2013 ; 孟庆斌 等, 2018 ; Berwart et al., 2019 ; 吴育辉 等, 2020
评级模式的影响	对评级机构的影响	发行人付费评级机构在“投资人付费”模式的压力下，会调低信用评级，但提高了评级质量	Xia, 2014
	对发债企业的影响	企业在获得投资人付费机构的低评级后，会去寻求能给出更高评级的发行人付费机构，进行评级“对冲”	寇宗来 等, 2020

包括宏观经济指标、企业财务指标、债券基本指标等信息，少有研究关注文本等非结构化数据。目前，金融文本分析已经被广泛应用于股票市场、宏观经济研究，涉及的文本类型包括公司年报、新闻媒体报道、社交媒体评论、政府报告等。很多学者发现文本情绪等指标对股票市场、宏观经济具有显著影响（表 6）。由于债券市场与股票市场具有联动性，也会受到宏观经济状况的影响，因此文本情绪等指标也很可能会影响债券市场，有必要将上述研究拓展至债券市场，从文本分析这一机器学

表 6 金融文本分析代表性研究

文本类型	作者及年份	主要内容
上市公司 年报	Loughran & MacDonald, 2011	年报中负面词语比例越高，股票未来越有可能出现负回报
	Jiang et al., 2019	利用 LM 词典构建了经理人文本情绪指数，发现该指数能够显著负向预测未来股市收益
	孟庆斌等, 2017	年报 MD&A 的信息含量越高，未来股价崩盘风险越低
	胡楠等, 2022	年报 MD&A 中披露的“短期视域”语言能够反映管理者内在的短视主义特质，管理者短视会导致企业减少资本支出和研发支出
新闻媒体 报道	Tetlock, 2007	华尔街日报相关报道对整体经济走势存在影响，报道中的消极情绪能够显著预测未来道琼斯工业平均指数的下降
	游家兴和吴静, 2012	新闻媒体报道的情绪会使股票价格偏离基本价值，这种影响在信息不透明时更为明显
	汪昌云和武佳薇, 2015	新闻媒体报道的语气会影响 IPO 抑价率
	姜富伟等, 2021	媒体报道中蕴含的文本情绪对于股票市场回报具有显著的预测能力
社交媒体 评论	Antweiler & Frank, 2004	雅虎金融的股票论坛评论的情绪信息对公司股票未来回报具有显著的预测能力
	Huang, 2018	亚马逊网站上对公司产品的评论能够影响公司未来股票回报
	杨晓兰等, 2016	根据东方财富网股吧评论数据，发现本地关注和帖子文本情绪对股票收益具有影响
	部慧等, 2018	股吧评论数据中的投资者情绪对股票收益和交易量存在影响
	钱宇等, 2020	股吧评论中在线用户对未来市场的支持倾向对未来股市收益具有显著的负向影响
	姚加权等, 2021	社交媒体情绪能有效地预测公司股票收益、成交量、波动率和非预期盈余等因素
政府报告	Born et al., 2014	各国央行发布的金融稳定报告中的正面情绪可以产生长期显著为正的股票市场收益
	Correa et al., 2021	金融稳定报告的情绪指数能解释信贷、资产价格、系统性风险和货币政策利率等指标的变动，还能有效预测银行危机
	姜富伟等, 2021	货币政策执行报告文本情绪的改善会引起显著为正的股票市场价格反应，并且文本相似度的增加会显著降低股票市场波动性
	林建浩等, 2021	央行沟通文本中的信息可以提高对中国宏观经济状况的实时预测能力



视视角更深入地挖掘债券违约风险的影响因素。

结合已有研究, 文本分析在债券违约风险识别的应用可从以下三方面开展。

第一, 拓展文本分析素材。少数学者已经围绕上市公司年报、债券募集说明书等文件分析其文本信息对债券违约风险的影响。比如, 吴武清等(2021)使用“风险”一词作为风险识别的关键词, 发现风险披露程度对债券发行溢价具有正向影响, 林晚发等(2021)使用清华大学建立的负面和正面语料词库(一种中文通用情感词典)计算债券募集说明书的文本情绪来识别违约风险。然而, 尚未有学者深入探讨新闻媒体报道、社交媒体评论、政府报告等文本信息对债券违约风险的影响。由于不同类型文本的发布主体不同, 在形式与内容上存在明显差异, 包含的信息也具有一定异质性, 不同文本信息对债券违约风险的影响可能有所差别。因此, 在未来有必要进一步探讨其他类型文本中的特质信息对债券违约风险的影响, 对相关研究结论进行进一步完善和补充。

第二, 构建特色情感词典。已有很多研究通过词典法计算文本情绪来测度文本传达出的积极或消极态度。然而, 由于文字语义具有多样性, 情感词典不具有普适性, 分析不同文本需要使用不同的情感词典。比如, Correa et al. (2021)为分析金融稳定报告中的金融稳定情绪指数, 专门构建了一款英文金融稳定词典。姚加权等(2021)通过词典重组和深度学习算法分别

构建了适用于正式文本与非正式文本的金融领域中文情绪词典。由此可见, 直接使用通用情感词典来测度债券募集说明书的文本情绪具有局限性。比如, “担保”一词在通用情感词典中往往表达正面含义, 但是债券的担保条款暗示着更大的风险(钟辉勇等, 2016; 林晚发等, 2022)。为更有效地识别违约风险, 有必要构建具有特色的情感词典。

第三, 使用更为前沿的情绪提取方法。就词典构建而言, 不同的词典构建思路会得到不同的情感词典。比如, 姜富伟等(2021)在LM词典的基础上通过人工筛选以及word2vec算法扩充的方式构建了一个中文金融情感词典, 范小云等(2022)利用更为灵活的“N-gram”思路开发了一种新的中文金融文本情绪词典。未来的研究可以对比分析不同思路构建得到的词典质量, 进而确定最为合适的情感词典构建方法。此外, RNN、LSTM、Transformers、BERT、Attention机制等各类深度学习方法也正在被应用到文本情绪提取中, 并且取得不错的表现, 未来也可以尝试使用这些新方法来更好地提取文本信息。

(二) 机器学习算法推动预警模型创新

除利用文本分析挖掘影响因素外, 利用前沿机器学习算法构建预警模型也是债券违约风险领域的未来发展方向。学术界目前通常使用Altman Z值模型和Merton模型定量测度违约风险, 这两个模型也可以用于债券违约风险预测。理论上, 使用机器学习模型预测债券违约风险相比于

经典模型有三点优势。

第一，充分拓展模型使用的信息集。债券违约风险的影响因素是丰富的，宏观经济、金融市场、企业特征、债券特征等都可能影响债券违约风险，但是经典模型中使用的预测变量缺乏多样性，没有体现出影响因素的丰富性。比如，Altman Z 值模型中的预测变量均属财务特征变量，Merton 模型中仅用到少数财务特征和股票特征。在债券违约风险影响因素不断涌现的背景下，如果在预警模型中仅使用少数特征，就很可能存在变量遗漏问题。机器学习算法允许模型输入大量预测变量，能够尽可能地利用丰富的预测变量信息，进而规避变量遗漏问题。Altman 的研究历程也验证了上述分析的正确性，ZETA 模型选取的预测变量相比于 Z 值模型进行了补充和完善，新模型稳定性和预测准确度更高的实证结果进一步说明只使用少数变量进行预测具有局限性。

第二，有效规避现有模型的局限性。一方面，线性模型在预测指标库大量拓展时会陷入过度拟合和维度陷阱，不再具有良好的样本外预测能力，这是因为高维数据难免存在多重共线性等问题，导致线性回归 (OLS) 的基本假设不再满足。前沿机器学习算法能够通过参数正则化等方法对数据的非线性性、高维特性实现充分分析，兼顾样本内拟合优度和样本外预测能力，进而提升模型预测效果。另一方面，结构模型、强度模型等经典非线性模型虽然在函数形式上有所创新，但也存在模型

假设严苛的问题。机器学习是一种自组织、自适应的非参数方法，对数据分布、模型形式不作特别限制，能够有效缓解非线性模型的不足。最后，国内债券市场的信用评级因受到“发行人付费、发行人选择”选聘机制的影响，一定程度上持续存在数据粉饰、信息失称程度高的问题，在评级购买、评级迎合的作用下易导致评级膨胀现象。机器学习模型中能够输入充分全面的客观数据，尽可能避免数据选择性披露和人为因素带来的负面影响。

第三，自然引入变量系数动态时变性。经典模型使用的参数通常是固定不变的，缺乏动态时变性。以 Z 值模型为例，从空间维度上看，由于中美两国市场环境具有差异，基于美国市场得出的变量系数在中国市场不一定适用。从时间维度上看，同一变量的系数在不同时期也会有所变化，Z 值是在 20 世纪 60 年代提出的，是否适用于我国的“去刚兑”背景也有待商榷。因此，有必要在模型中引入变量系数动态时变特性。机器学习算法在股票收益预测中常用的滚动窗口建模方法能够通过参数动态更新调整的方式得到模型在各时期的最优参数，保证模型在每一期都具有最优的预测能力。

综上可知，使用机器学习模型预测债券违约风险具有理论意义，有助于构建一个变量更为全面且能够灵活调整的模型。已有大量研究验证了机器学习算法能够显著提高对金融市场资产收益的预测能力 (表 7)。Bali et al. (2022) 是机器学习在债券市场应用的代表性文献，首次参考异象因



子已有研究构建了美国债券市场高维预测指标库, 具体包括债券基本特征、风险特征、流动性特征、收益特征共计 43 个指标。总体上看, 机器学习在股票、债券市场的应用在方法上具有相似之处, 但是债券市场的相关研究要滞后于股票市场, 导致这一区别的可能原因是股票市场异象因子的研究较为成熟, 大量因子构成的高维指标天然地可以作为机器学习算法的输入变量。然而, 债券市场因其 OTC 交易特征, 在基础研究数据采集、数据库构建, 尤其是针对市场连续交易数据的获取方面, 较股票市场有显著的难度。机器学习在美国债券市场研究的蓬勃发展主要得益于以 TRACE 为代表的高质量数据库建设, 相比之下, 我国尚未有机构开发出类似于 TRACE 的债券市场权威数据库。因此, 将机器学习应用于中国债券市场的难点在于构建具有中国特色的债券市场预测指标库, 研究者需要花费大量精力考虑研究数据的可得性、可追溯性以及数据质量等诸多问题。

为开展基于大数据和机器学习的债券违约风险预警模型研究, 本文对债券违约风险影响因素进行了全面梳理, 这些因素可以自然地考虑在机器学习算法的输入指标里。具体地, 可充分借鉴债券违约风险影响因素相关文献, 构建包含宏观经济、金融市场、企业特征、债券特征等在内的原始预测指标库, 这一过程中还可以在指标库中加入文本分析提取的影响因素。在此基础上, 可参考 Gu et al. (2020) 构建宏微观交互指标, 得到宏观经济-微观企业混合大数据集, 并系统性地使用主流机器学习算法构建债券违约风险与预测指标之间的非线性关系, 分析比较不同算法的预测能力差异, 确定更为合适的机器学习算法。在运用机器学习模型时, 还可以尝试人为将经济学先验信息变量和机器学习方法相结合, 使模型更具经济学意义。进一步, 可以对机器学习预测背后的经济机制开展讨论, 比如分析机器学习算法在哪些债券子样本上具有更强预测能力, 分

表 7 金融机器学习代表性研究

研究角度	作者及年份	主要内容
机器学习与股票收益预测	Light et al., 2017	利用偏最小二乘法 (PLS) 实证发现公司特征对股票收益具有预测能力
	Gu et al., 2020	系统地使用多种机器学习算法对股票收益进行预测, 发现决策树和神经网络的表现更好, 非线性模型具有重要作用
	李斌等, 2019	实证检验机器学习在我国股票市场收益预测的有效性
	姜富伟等, 2021	构建基于机器学习的动态 CAPM 模型, 发现其能够解释我国股票市场低风险定价异象, 并且非线性机器学习表现更好
	马甜等, 2022	利用生成对抗网络 (GAN) 预测我国股票市场收益, 发现深度学习具有优势
机器学习与债券收益预测	Bianchi et al., 2021	实证检验了机器学习算法对于美国国债收益预测的有效性
	Bali et al., 2022	实证检验了机器学习算法对于美国公司债收益预测的有效性

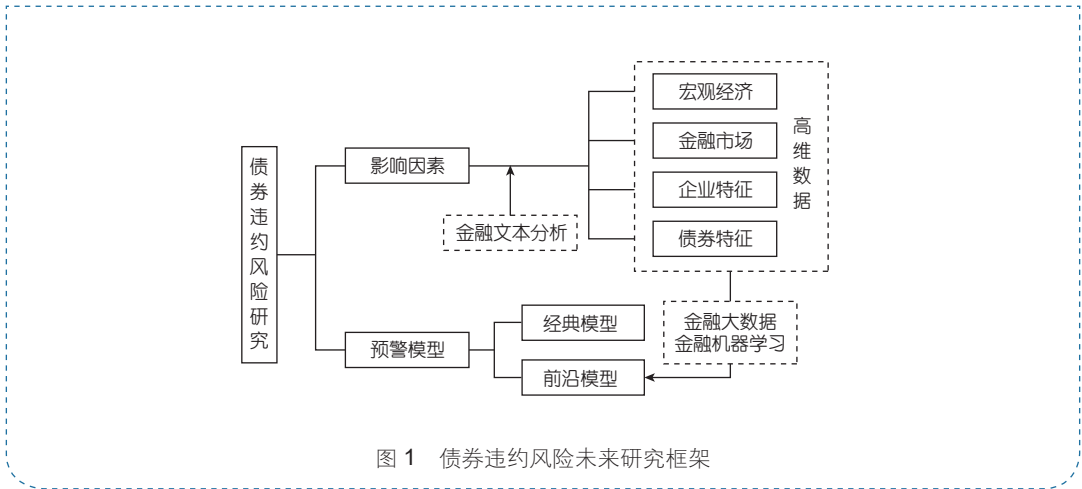


图 1 债券违约风险未来研究框架

析哪些预测指标在机器学习预测中具有更高重要性，还可以分析机器学习预测得到的不同风险债券在哪些特征上具有更为明显的差异，上述问题的结论可以为债券违约风险的精准识别与智能防控提供来自机器学习的新视角。在经典机器学习算法的基础上，还可以在模型中使用 Huang et al. (2022) 提出的缩放主成分分析 (Scaled PCA) 等更为前沿的算法，利用技术创新提高对债券违约风险的预警能力。

四、总结

综上所述，债券违约风险的影响因素较为丰富，包括宏观经济、金融市场、企业特征、债券特征等方面，但从文本数据视角还有探索空间。另一方面，债券违约风险预警模型的研究相对滞后，Altman 模型、Merton 模型等经典模型使用的指标数量有限，也存在着模型假设严格的不足。鉴于此，本文从机器学习方法视角对未来

研究方向提出展望，指出可以通过文本分析进一步挖掘债券违约风险的影响因素。在此基础上，可构建包含大量影响因素的债券违约风险高维预测指标库，并将这些指标作为机器学习模型的输入变量，运用机器学习算法推动债券违约风险预警模型的创新。研究框架参见图 1。

基于机器学习视角的研究落地预计能够为企业经营者、机构投资者及市场监管者前瞻性研判债券违约风险、相机应对路径与政策决策等工作提供更为丰富的科学方法论依据，同时为各类型市场主体尤其是监管机构精准防控违约风险、合理应对信用风险事件提供更前沿的技术支持。这对于深入推进我国债券市场高质量发展，牢牢守住不发生系统性金融风险的底线，充分发挥金融体系在实体经济健康发展过程中的稳定器作用，助力我国经济实现高质量发展均具有较强的理论价值和实践意义。^[N]

学术编辑：卢超群



参考文献:

- [1] 部慧,解峥,李佳鸿,吴俊杰.基于股评的投资者情绪对股票市场的影响[J].管理科学学报,2018,21(04):86-101.
- [2] 陈关亭,连立帅,朱松.多重信用评级与债券融资成本——来自中国债券市场的经验证据[J].金融研究,2021,(02):94-113.
- [3] 丁志国,丁垣竹,金龙.违约边界与效率缺口:企业债务违约风险识别[J].中国工业经济,2021,(04):175-192.
- [4] 范小云,王业东,王道平,郭文璇,胡焯翊.不同来源金融文本信息含量的异质性分析——基于混合式文本情绪测度方法[J].管理世界,2022,38(10):78-101.
- [5] 方红星,施继坤,张广宝.产权性质、信息质量与公司债定价——来自中国资本市场的经验证据[J].金融研究,2013,(04):170-182.
- [6] 郭晔,黄振,王蕴.未预期货币政策与企业债券信用利差——基于固浮利差分解的研究[J].金融研究,2016,(06):67-80.
- [7] 何德旭,张斌彬.居民杠杆与企业债务风险[J].中国工业经济,2021,(02):155-173.
- [8] 何德旭,张雪兰.中国式现代化需要怎样的金融体系[J].财贸经济,2023,44(01):18-29.
- [9] 何平,金梦.信用评级在中国债券市场的影响力[J].金融研究,2010,(04):15-28.
- [10] 胡楠,薛付婧,王昊楠.管理者短视主义影响企业长期投资吗?——基于文本分析和机器学习[J].管理世界,2021,37(05):139-156+11+19-21.
- [11] 黄振,郭晔.央行担保品框架、债券信用利差与企业融资成本[J].经济研究,2021,56(01):105-121.
- [12] 纪志宏,曹媛媛.信用风险溢价还是市场流动性溢价:基于中国信用债定价的实证研究[J].金融研究,2017,(02):1-10.
- [13] 姜富伟,胡逸驰,黄楠.央行货币政策报告文本信息、宏观经济与股票市场[J].金融研究,2021,(06):95-113.
- [14] 姜富伟,马甜,张宏伟.高风险低收益?基于机器学习的动态CAPM模型解释[J].管理科学学报,2021,24(01):109-126.
- [15] 姜富伟,孟令超,唐国豪.媒体文本情绪与股票回报预测[J].经济学(季刊),2021,21(04):1323-1344.
- [16] 寇宗来,千茜倩,陈关亭.跟随还是对冲:发行人付费评级机构如何应对中债资信的低评级?[J].管理世界,2020,36(09):26-39.
- [17] 李斌,邵新月,李玥阳.机器学习驱动的基本面量化投资研究[J].中国工业经济,2019,(08):61-79.
- [18] 李建军,韩珣.非金融企业影子银行化与经营风险[J].经济研究,2019,54(08):21-35.
- [19] 林建浩,陈良源,罗子豪,张一帆.央行沟通有助于改善宏观经济预测吗?——基于文本数据的高维稀疏建模[J].经济研究,2021,56(03):48-64.
- [20] 林晚发,方梅,沈宇航.债券募集说明书文本信息与债券发行定价[J].管理科学,2021,34(04):19-34.
- [21] 林晚发,顾乾坤.纳税信用评价的债券市场反应[J].经济管理,2021,43(10):119-137.
- [22] 林晚发,刘岩,赵仲匡.债券评级包装与“担保正溢价”之谜[J].经济研究,2022,57(02):192-208.
- [23] 林晚发,刘颖斐,赵仲匡.承销商评级与债券信用利差——来自《证券公司分类监管规定》的经验证据[J].中国工业经济,2019,(01):174-192.
- [24] 林晚发,赵仲匡,宋敏.管理层讨论与分析的语调操纵及其债券市场反应[J].管理世界,2022,38(01):164-180.
- [25] 林晚发,钟辉勇,李青原.高管任职经历的得与失?——来自债券市场的经验证据[J].金融研究,2018,(06):171-188.
- [26] 刘晓光,刘元春.杠杆率、短债长用与企业表现[J].经济研究,2019,54(07):127-141.
- [27] 罗荣华,刘劲劲.地方政府的隐性担保真的有效吗?——基于城投债发行定价的检验[J].金融研究,2016,(04):83-98.
- [28] 马甜,姜富伟,唐国豪.深度学习与中国股票市场因子投资——基于生成式对抗网络方法[J].经济学(季刊),2022,22(03):819-842.
- [29] 孟庆斌,杨俊华,鲁冰.管理层讨论与分析披露的信息含量与股价崩盘风险——基于文本向量化方法的研究[J].中国工业经济,2017,(12):132-150.
- [30] 孟庆斌,张强,吴卫星,王宇西.中立评级机构对发行人付费评级体系的影响[J].财贸经济,2018,39(05):53-70.

- [31] 潘泽清.企业债务违约风险Logistic回归预警模型[J].上海经济研究,2018,(08):73-83.
- [32] 钱宇,李子饶,李强,袁华.在线社区支持倾向对股市收益和波动的影响[J].管理科学学报,2020,23(02):141-155.
- [33] 史永东,宋明勇,李凤羽,甄红线.控股股东股权质押与企业债权人利益保护——来自中国债券市场的证据[J].经济研究,2021,56(08):109-126.
- [34] 史永东,郑世杰,袁绍锋.中债估值识别了债券信用风险吗?——基于跳跃视角的实证分析[J].金融研究,2021,(07):115-133.
- [35] 宋敏,甘煦,林晚发.债券信用评级膨胀:原因、影响及对策[J].经济学动态,2019,(03):134-147.
- [36] 汪昌云,武佳薇.媒体语气、投资者情绪与IPO定价[J].金融研究,2015,(09):174-189.
- [37] 汪莉,陈诗一.政府隐性担保、债务违约与利率决定[J].金融研究,2015,(09):66-81.
- [38] 王安兴,解文增,余文龙.中国公司债利差的构成及影响因素实证分析[J].管理科学学报,2012,15(05):32-41.
- [39] 王博森,吕元稹,叶永新.政府隐性担保风险定价:基于我国债券交易市场的探讨[J].经济研究,2016,51(10):155-167.
- [40] 王雄元,张春强.声誉机制、信用评级与中期票据融资成本[J].金融研究,2013,(08):150-164.
- [41] 王雄元,张春强,何捷.宏观经济波动性与短期融资券风险溢价[J].金融研究,2015,(01):68-83.
- [42] 吴武清,甄伟浩,杨洁,蔡宗武.企业风险信息披露与债券风险溢价——基于债券募集说明书的文本分析[J].系统工程理论与实践,2021,41(07):1650-1671.
- [43] 吴育辉,翟玲玲,张润楠,魏志华.“投资人付费”vs.“发行人付费”:谁的信用评级质量更高?[J].金融研究,2020,(01):130-149.
- [44] 徐思,潘昕彤,林晚发.“一带一路”倡议与公司债信用利差[J].金融研究,2022,(02):135-152.
- [45] 杨国超,蒋安琪.债券投资者的“保护盾”还是债务违约的“多米诺”——对债券交叉违约制度的分析[J].中国工业经济,2022,(05):140-158.
- [46] 杨璐,方静.适应性预期与债券发行信用溢价[J].国际金融研究,2021,(08):76-86.
- [47] 杨晓兰,沈翰彬,祝宇.本地偏好、投资者情绪与股票收益率:来自网络论坛的经验证据[J].金融研究,2016,(12):143-158.
- [48] 姚加权,冯绪,王赞钧,纪荣嵘,张维.语调、情绪及市场影响:基于金融情绪词典[J].管理科学学报,2021,24(05):26-46.
- [49] 游家兴,吴静.沉默的螺旋:媒体情绪与资产误定价[J].经济研究,2012,47(07):141-152.
- [50] 于静霞,周林.货币政策、宏观经济对企业债券信用利差的影响研究[J].财政研究,2015,(05):49-57.
- [51] 钟辉勇,钟宁桦,朱小能.城投债的担保可信吗?——来自债券评级和发行定价的证据[J].金融研究,2016,(04):66-82.
- [52] 周宏,建蕾,李国平.企业社会责任与债券信用利差关系及其影响机制——基于沪深上市公司的实证研究[J].会计研究,2016,(05):18-25+95.
- [53] 周宏,徐兆铭,彭丽华,杨萌萌.宏观经济不确定性对中国企业债券信用风险的影响——基于2007—2009年月度面板数据[J].会计研究,2011,(12):41-45+97.
- [54] Altman Edward I, Haldeman Robert G, Narayanan P. Zeta analysis: A new model to identify bankruptcy risk of corporations[J]. Journal of Banking & Finance, 1977, 10: 29-54.
- [55] Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy[J]. The Journal of Finance, 1968, 23(4): 589-609.
- [56] Andreasen M M, Engsted T, Møller S V, Sander M. The yield spread and bond return predictability in expansions and recessions[J]. The Review of Financial Studies, 2021, 34(6): 2773-2812.
- [57] Antweiler W, Frank M Z. Is all that talk just noise? The information content of internet stock message boards[J]. The Journal of Finance, 2004, 59(3): 1259-1294.
- [58] Baghai R P, Becker B. Non-rating revenue and conflicts of interest[J]. Journal of Financial Economics, 2018, 127(1): 94-112.
- [59] Baghdadi G A, Nguyen L H G, Podolski E J. Board co-option and default risk[J]. Journal of Corporate



- Finance, 2020, 64: 101703.
- [60] Bali T G, Goyal A, Huang D, Jiang F, Wen Q. Predicting corporate bond returns: merton meets machine learning[J]. Georgetown McDonough School of Business Research Paper, 2022.
- [61] Beaver W H, Shakespeare C, Soliman M T. Differential properties in the ratings of certified versus non-certified bond-rating agencies[J]. Journal of Accounting and Economics, 2006, 42(3): 303-334.
- [62] Berwart E, Guidolin M, Milidonis A. An empirical analysis of changes in the relative timeliness of issuer-paid vs. investor-paid ratings[J]. Journal of Corporate Finance, 2019, 59: 88-118.
- [63] Bharath S T, Shumway T. Forecasting default with the Merton distance to default model[J]. The Review of Financial Studies, 2008, 21(3): 1339-1369.
- [64] Bianchi D, Büchner M, Tamoni A. Bond risk premiums with machine learning[J]. The Review of Financial Studies, 2021, 34(2): 1046-1089.
- [65] Bolton P, Freixas X, Shapiro J. The credit ratings game[J]. The Journal of Finance, 2012, 67(1): 85-111.
- [66] Bonsall Iv S B. The impact of issuer-pay on corporate bond rating properties: Evidence from Moody's and S&P's initial adoptions[J]. Journal of Accounting and Economics, 2014, 57(2-3): 89-109.
- [67] Bonsall S B, Miller B P. The impact of narrative disclosure readability on bond ratings and the cost of debt[J]. Review of Accounting Studies, 2017, 22(2): 608-643.
- [68] Born B, Ehrmann M, Fratzscher M. Central bank communication on financial stability[J]. The Economic Journal, 2014, 124(577): 701-734.
- [69] Bradley M, Roberts M R. The structure and pricing of corporate debt covenants[J]. The Quarterly Journal of Finance, 2015, 5(02).
- [70] Brogaard J, Li D, Xia Y. Stock liquidity and default risk[J]. Journal of Financial Economics, 2017, 124(3): 486-502.
- [71] Brugler J, Comerton-Forde C, Martin J S. Secondary market transparency and corporate bond issuing costs[J]. Review of Finance, 2022, 26(1): 43-77.
- [72] Cornaggia J, Cornaggia K J. Estimating the costs of issuer-paid credit ratings[J]. The Review of Financial Studies, 2013, 26(9): 2229-2269.
- [73] Correa R, Garud K, Londono J M, Mislang N. Sentiment in central banks' financial stability reports[J]. Review of Finance, 2021, 25(1): 85-120.
- [74] David A. Inflation uncertainty, asset valuations, and the credit spreads puzzle[J]. The Review of Financial Studies, 2008, 21(6): 2487-2534.
- [75] Flynn S, Ghent A. Competition and credit ratings after the fall[J]. Management Science, 2018, 64(4): 1672-1692.
- [76] Giesecke K. Default and information[J]. Journal of Economic Dynamics and Control, 2006, 30(11): 2281-2303.
- [77] Giesecke K, Longstaff F A, Schaefer S, Strebulaev I. Corporate bond default risk: A 150-year perspective[J]. Journal of Financial Economics, 2011, 102(2): 233-250.
- [78] Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning[J]. The Review of Financial Studies, 2020, 33(5): 2223-2273.
- [79] Huang D, Jiang F, Li K, Tong G, Zhou G. Scaled PCA: A new approach to dimension reduction[J]. Management Science, 2022, 68(3): 1678-1695.
- [80] Huang J. The customer knows best: The investment value of consumer opinions[J]. Journal of Financial Economics, 2018, 128(1): 164-182.
- [81] Jarrow R A, Turnbull S M. Pricing derivatives on financial securities subject to credit risk[J]. The Journal of Finance, 1995, 50(1): 53-85.
- [82] Jiang F, Lee J, Martin X, Zhou G. Manager sentiment and stock returns[J]. Journal of Financial Economics, 2019, 132(1): 126-149.

- [83] Jiang J X, Stanford M H, Xie Y. Does it matter who pays for bond ratings? Historical evidence[J]. *Journal of Financial Economics*, 2012, 105(3): 607-621.
- [84] Kedia S, Rajgopal S, Zhou X. Did going public impair Moody' s credit ratings?[J]. *Journal of Financial Economics*, 2014, 114(2): 293-315.
- [85] Light N, Maslov D, Rytchkov O. Aggregation of information about the cross section of stock returns: A latent variable approach[J]. *The Review of Financial Studies*, 2017, 30(4): 1339-1381.
- [86] Longstaff F A, Schwartz E S. A simple approach to valuing risky fixed and floating rate debt[J]. *The Journal of Finance*, 1995, 50(3): 789-819.
- [87] Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks[J]. *The Journal of Finance*, 2011, 66(1): 35-65.
- [88] Merton R C. On the pricing of corporate debt: The risk structure of interest rates[J]. *The Journal of Finance*, 1974, 29(2): 449-470.
- [89] Minnis M. The value of financial statement verification in debt financing: Evidence from private US firms[J]. *Journal of Accounting Research*, 2011, 49(2): 457-506.
- [90] Molina C A. Are firms underleveraged? An examination of the effect of leverage on default probabilities[J]. *The Journal of Finance*, 2005, 60(3): 1427-1459.
- [91] Nagler F. Yield spreads and the corporate bond rollover channel[J]. *Review of Finance*, 2020, 24(2): 345-379.
- [92] Sangiorgi F, Spatt C. Opacity, credit rating shopping, and bias[J]. *Management Science*, 2017, 63(12): 4016-4036.
- [93] Skreta V, Veldkamp L. Ratings shopping and asset complexity: A theory of ratings inflation[J]. *Journal of Monetary Economics*, 2009, 56(5): 678-695.
- [94] Tang D Y, Yan H. Market conditions, default risk and credit spreads[J]. *Journal of Banking & Finance*, 2010, 34(4): 743-753.
- [95] Tetlock P C. Giving content to investor sentiment: The role of media in the stock market[J]. *The Journal of Finance*, 2007, 62(3): 1139-1168.
- [96] Xia H. Can investor-paid credit rating agencies improve the information quality of issuer-paid rating agencies?[J]. *Journal of Financial Economics*, 2014, 111(2): 450-468.

Bond Default Risk and Machine-Learning Methods: Research Progress and Future Orientation

QI Xinlin¹ LIN Yihao² JIANG Fuwei²

(1.Industrial and Commercial Bank of China;2.School of Finance, Central University of Finance and Economics)

Abstract Default events have occurred frequently in the bond market since the system of rigid payments was abandoned. This has created new requirements for default risk early warning capabilities. This paper examines the factors that affect bond default risk, including the macro-economic situation, financial market conditions and enterprise and bond characteristics. Additionally, the paper summarizes the warning models suited to identifying bond default risk. In addition, it is noted that the application of machine-learning methods is useful in creating an effective early warning model for determining bond default risk. For example, text analysis can be used to further explore the factors influencing bond default risk, and machine-learning algorithms can be used to promote innovation in bond default early warning models. The above measures are conducive to maintaining financial stability and promoting high-quality development of the financial market.

Keywords Bond Market, Default Risk, Literature Review, Machine Learning

JEL Classification G12 G20 G30