



募集说明书文本与债券发行利差¹

Prospectus text and bond issuance spread

第 7 组 方清源、付煜、毛思文

指导老师：刘晓蕾、刘晓宇

【摘要】绿色债券所具有的绿色属性，在发行时的募集说明书中应当有一定体现。本文围绕债券募集说明书，通过文本分析构建了“环境保护”“政策响应”和“语体特质”等指标，探究其与债券发行利差的关系。研究发现，募集说明书当中具备“环境保护”“政策响应”等相关概念词汇有助于降低发行利差；绿色债券相较于非绿债更重视“环境保护”，相应的风险利差更低，但在公告中强调“环境保护”的边际效果有限；募集说明书语言有一定规范，“与众不同”的措辞反而会加大债券发行的风险利差。本文结论有助于更好地认识债券募集说明书作用。

【关键词】募集说明书；文本分析；发行利差；绿色债券

1 背景

1.1 绿色金融与绿色债券

绿色金融是指为支持环境改善、应对气候变化和资源节约高效利用的经济活动，即对环保、节能、清洁能源、绿色交通、绿色建筑等领域的项目投融资、项目运营、风险管理等所提供的金融服务²。其中，绿色债券是指将所得资金专门用于资助符合规定条件的绿色项目，或为这些项目进行再融资的新型债券工具³。2015 年，中共中央、国务院印发《关于加快推进生态文明建设的意见》，催生了巨大的绿色融资需求；该年 7 月，我国第一只绿色债券成功发行；12 月，中国人民银行出台了《关于在银行间债券市场发行绿色金融债券有关事宜公告》，并配套发布《绿色债券支持项目目录》，对绿色债券的发行进行了引导。在一系列政策文件的支持引导下，我国绿色债券市场蓬勃发展。截至 2020 年末，中国累计发行绿色债券约 1.2 万亿元，位居世界第二。

与一般性债券相比，绿色债券具有以下两个维度的重要特征：

1) 环保相关度高。现有多家机构开展绿色债券的评估认证业务，以第三方认证的形式加强债券“绿色”特征的可信度。

2) 政策关联度大。绿色债券市场发展与我国生态文

明建设的进程紧密挂钩。中央或地方的政策文件均会受到绿色债券发行与交易方的紧密关注

市场是有效整合信息的有效机制。受此启发，我们想要通过文本分析的方式提取以上特征，探究它们对绿色债券定价及风险利差的影响。

1.2 相关文献

1.2.1 绿色债券定价

绿色债券的定价受到很多因素影响。朱家明等[1]通过多元回归分析，发现绿色债券利率的影响因素主要包括市场利率、国内通胀水平、发行主体注册资本、净资产收益率，以及信用评级。许欣怡[2]基于可贷资金模型，得到了类似的实证结果。龚玉霞等[3]通过构建二叉树模型，发现我国绿色债券市场价格低于理论价格，并提出市场流动性不足、消费了解程度有限等解释。倪弋菁[4]发现媒体关注度和分析师关注度与绿色债券发行信用利差有显著的负相关关系，而公众关注度与绿色债券信用利差并没有显著相关性。可以看出，已有文献大多以宏观因子、债券发行主体、评级等角度构建解释变量。而我们的研究关注的是绿色债券自身更本质的特征，填补了研究的空白。

一般认为，绿色债券存在更多的发行成本优势。王玲[5]通过对债券募集说明书、信用评级报告和其他公开资料的分析，将政府补贴和第三方绿色认证确定为绿色

¹ 完整的数据和代码请见 https://github.com/Wins-M/bond_prospectus

² 百度百科 (<https://baike.baidu.com/item/绿色金融/10547233?fr=aladdin>)

³ 百度百科 (<https://baike.baidu.com/item/绿色债券/24155150?fr=aladdin>)

债券的“绿色”因素，得出绿色因素具有显著降低我国绿色公司债券发行成本的效果。与此研究相比，我们提取的“绿色”因素直接来自债券募集说明书的文本，是将非结构化数据纳入债券定价考虑的新尝试。

1.2.2 债券募集说明书与文本分析

债券募集说明书一般包含债券发行情况、主要股东情况、财务会计信息、管理层讨论与分析、募集资金运用等重要信息，是公开发行人公司债券时对投资者最重要的资料。募集说明书属于非结构化数据，随着文本分析方法的发展，近年来出现了一系列以其为语料的债券市场研究。例如，说明书风险披露程度和投资者违约风险感知、债券风险溢价的关系[6]；说明书语调与公司违约风险、债券发行利差的关系[7]；说明书中特质性信息披露与公司未来违约风险的关系[8]等等。

结合绿色债券的特征，我们希望通过文本分析方法从债券募集说明书中提取环境保护相关度（衡量绿色债券与环境保护、节能减排等业务的相关性）以及政策响应力度（衡量绿色债券对政策响应的程度与关注度）两个维度的信息。同时，我们需要将文献中已经发现的影响因子（如文本复杂度、区别于其他债券的特异性信息等）作为控制变量，提高回归模型的解释力。

1.2.3 “绿色”词典构建

我们从对绿色专利的研究中获得启发[9-11]，筛选出可以衡量环境保护相关度以及政策响应相关度的词语构建词典。

1.3 研究目的

本文旨在探究公司、企业债券的募集说明书文本特征与其债券风险利差之间的关系。考虑到近年来备受关注的“绿色金融”主题，并结合目前已有的金融文本分析文献，本文选定“环境保护相关度”、“政策响应力度”与“特质性信息含量”三个维度衡量募集说明书的文本特征，探究其在利率决定过程中发挥的作用。

2 研究设计

2.1 数据获取与样本来源

2.1.1 样本选取与文本采集

本研究选取 2016 年 1 月 1 日至 2021 年 12 月 31 日

期间在中国大陆上市发行的全部企业债券与公司债券作为研究样本。同时，我们对其中包括的“绿色债券”抱有特别的关注，本文使用的“绿色债券”定义为：将所得资金专门用于资助符合规定条件的绿色项目，并且获得 Wind 数据库认证收录的债券。在上述五年时间区间内，Wind 数据库共收录 8053 支企业债与公司债券，其中共 542 支为绿色债券。

研究使用的债券募集说明书来自于巨潮资讯网⁴。使用“八抓鱼采集器”脚本，设定日期范围和关键词“募集说明书”进行查询，获得 2015 年至 2021 年标题包含“募集说明书”的公告记录 15830 条。去除“摘要”“更正公告”“变更说明”等记录后，根据债券编号同 Wind 数据进行匹配，获得 2016 年 1 月至 2021 年 12 月发行上市的企业债/公司债募集说明书记录共 4881 条。去除重复和 PDF 文件异常等情况，以及下文变量定义中控制变量缺失的样本，剩余最终文本 4424 份。其中包含绿色债券 174 份。

2.1.2 债券交易数据与公司特征信息

本研究所采用的债券市场交易数据与企业特征数据、企业财务数据均来自于 Wind 数据库。由于研究样本中债券的发行主体企业多为未上市企业，不存在公开披露的准确财务数据，因此我们使用 Wind“债券估值”数据模块中提供的财务信息预测值构建控制变量，具体见本文“变量定义”部分。

2.2 变量定义

2.2.1 解释变量

环境保护 (Eco): 衡量债券募集说明书中体现的与环境保护、节能减排等绿色业务的相关度变量。其构造方法为首先构建“环境保护”概念词库，随后统计债券募集说明书中词库收录词汇出现的频数，并计算其与说明书总词汇数的比值，即定义债券 i 的环境保护相关度为：

$$Eco_i = \frac{\text{环境保护相关词汇出现频数}}{\text{募集说明书总词汇数}}$$

在构建“环境保护”相关概念词库时，我们首先统计绿色债券募集说明书中包含的全部中文词汇，并选取其中能够体现与绿色业务联系的重要词汇，具体构建流程为：

步骤一：分词和停用。对 PDF 中的文字内容进行统一的文本预处理。

步骤二：词频统计。统计绿色债券募集说明书中出现的全部表意中文词汇的出现频数。

步骤三：概念词选取。对全部词汇按照“包含该词

⁴ <http://www.cninfo.com.cn>

汇的募集说明书数量”进行排序，找到出现两次以上的词汇共 1755 个；逐词核对，选取其中“绿色”“环保”“能源”等与环境保护紧密相关的词汇，得到“环境保护概念词库”。其中共包括：['绿色', '持续', '环境', '能源', '环保', '风电', '清洁' ...]等共计 41 个词汇。

政策响应 (Policy)：债券募集说明书中体现的对于国家政策、政府决议进行响应的力度与关注度。构造方法为首先构建“政策响应”概念词库，随后统计债券募集说明书中词库收录词汇出现的频数，计算其与说明书总词汇数的比值。定义债券 i 的政治响应相关度为：

$$Policy_i = \frac{\text{政策响应相关词汇出现频数}}{\text{募集说明书总词汇数}}$$

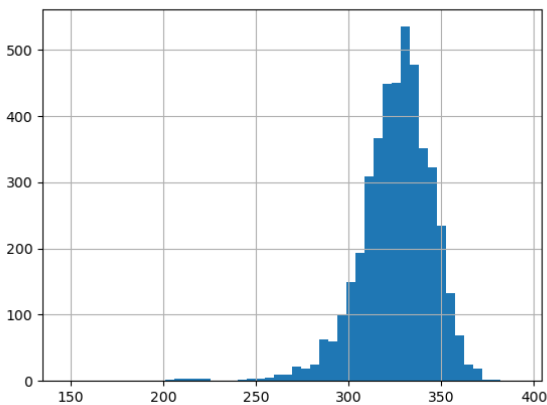
其中“政策响应”概念词库的构建流程与上文“环境保护”词库完全相同，共包括['法律', '国家', '文件', '法规', '政策', '决议', '政府' ...]等共计 39 个词汇。

语体特质 (Info, Stdd, Heter)：衡量债券募集说明书文本信息包含的无法被其他可比债券的募集说明书所解释的比重。具体测算方法参考 Hanley 与 Hoberg (2010) [12]，采用文本向量化方法计算各债券募集说明书文本的词频向量：

$$D = (w_1, w_2, \dots, w_n, 0, 0, \dots, 0)$$

一个词频向量 D 对应于一篇募集说明书文档。 w_s 为表意词汇 s 出现频数占文档 D 总表意词汇数的权重。为限制向量规模，规定最小权重需要大于 $1e-3$ ，调整各词汇权重之和为 1。募集书文档所对应的词频向量非零元素数量分布如下图 1 所示：

图 1 文档词频向量非零维数



划定每一支债券发行前 90 日内的其他债券为“可比债券”。随后，对于每一支发行的债券 i ，将其词频向

量与其全部“可比债券”的文本词频向量的算术平均值 \tilde{D}_i 回归。记回归估计系数为“规范语体”指标 $Stdd$ ，衡量文本措辞与“可比债券”募集说明书平均措辞的相关度；记回归模型残差的绝对值之和为“特殊语体”指标 $Info$ ，衡量文本措辞中不同于“可比债券”的特殊部分。以上两个变量来源于 Hanley 与 Hoberg 的文献 (2010)。此外，我们基于回归残差的平方和 SSE ，构造“文档异质性”指标 $Heter$ 。计算公式如下：

$$D_i = (w_{i1}, w_{i2}, \dots, w_{in}),$$

$$\tilde{D}_i = (\tilde{w}_{i1}, \tilde{w}_{i2}, \dots, \tilde{w}_{in})$$

$$w_{is} = \beta_i \tilde{w}_{is} + \varepsilon_{is}$$

$$Stdd_i = \beta_i,$$

$$Info_i = \sum_s |\varepsilon_{is}|,$$

$$Heter_i = \ln \left(\frac{SSE}{SSR} \right) = \ln \left(\frac{1 - R_i^2}{R_i^2} \right)$$

2.2.2 被解释变量

本研究选择债券发行时的风险利差作为被解释变量，其定义为债券 i 的票面利率与发行时无风险利率之差，本研究选用上海银行间同业拆放利率 (Shibor) 作为无风险利率标准，即：

$$Spread_i = \text{Coupon Rate} - \text{Shibor}$$

其中 Shibor 的具体数值为债券发行前五日上海银行间同业拆放利率的平均值，取年化收益率。

2.2.3 控制变量

参考吴武清、林晚发 (2021) 等，由于本研究采取的债券样本多由未上市企业发行，因此本文在企业财务指标方面进行了一定的简化，最终从债券特征、环境参数与财务信息指标三个角度选择了共计十五个控制变量加入到回归模型之中，变量的具体定义与计算参见表格 2。

在以上控制变量中，企业财务信息变量与固定效应环境参数与本文研究目的之间的相关度较弱，仅为排除无关因素影响。因此，在后续的回归模型中，默认该 9 个变量 (Company_Control) 加入在模型中，但其回归系数不予展示和分析。

以上变量详细定义见下表 1。

变量类型	变量名称	变量简称	变量定义
解释变量	环境保护相关度	Eco	募集说明书对环保相关业务及绿色项目的阐述程度

(续表)

	政策响应力度	Policy	募集说明书对国家政策进行阐述与响应的程度
	特质性信息含量	Heter	债券募集说明书中含有的特质性信息含量
	“规范语体”含量	Stdd	募集说明书中含有的与可比债券相似的文本信息量
	“特殊语体”含量	Info	募集说明书中含有的与可比债券不同的文本信息量
被解释变量	风险利差	Spreads	债券票面利率与发行时上海银行同业拆放利率之差
发行主体 控制变量	销售收入	Income	Wind 预估的企业在债券发行年度的销售收入
	净利润	Profit	Wind 预估的企业在债券发行年度的净利润
	现金流量	Cashflow	Wind 预估的企业在债券发行年度的现金流净流量
	是否上市	Listed	若发行主体是上市公司则为 1，否则为 0
	存续年限	Log_Age	债券发行时债务主体已经存续的年限的自然对数
	发行人属性	Com_Type	发行主体的最终控制人属性，民营企业=0，地方国有企业=1，中央国有企业=2
债券特征 控制变量	绿色属性	Green	是否为 Wind 收录的绿色债券，若是记为 1，否为 0
	利率类型	Int_Type	若债券利率为累进利率则为 0，为固定利率则为 1
	债券类型	Bond_Type	债券类型为企业债则记为 0，公司债则记为 1
	债券期限	Maturity	债券期限
	期权属性	Convert	债券属于含期权债券则取值为 1，否则取值为 0
	发行规模	Volume	债券发行总规模的自然对数
	信用评级	Rating	第三方债券信用评级，当债券评级低于 AA-时取值 0，等于 AA-时取值为 1，评级为 AA 时取值为 2，评级为 AA+时取值为 3，评级为 AAA 时取值为 4。
固定效应 环境参数	年份	Year	债券发行年份，2016 年至 2021 年分别设置为 0-5
	地区	Province	发债企业所属省份
	行业	Industry	债务主体企业所属于的 Wind 行业分类

表 1 主要变量

2.3 假设提出与回归模型

2.3.1 研究假设

近年来，由于我国政府大力推进可持续发展与绿色

经济建设，开创“绿色债券”等新型金融工具，试图降低环保项目在融资时面临的成本；因此我们认为，债券的募集说明书与环保主题相关度越高、对政策号召的响应力度越强，则其发行主体企业融资成本应当越低，体

现为该债券较低的风险利差，即：

H1: 债券募集说明书中对环境保护的重视程度(Eco)越高，则债券的风险利差 (Spreads) 越低。

H2: 债券募集说明书中对政策响应的关注度(Policy)越高，则债券的风险利差 (Spreads) 越低。

同时，根据现有文献资料，债券募集说明书中的特质性含量能够在一定程度上反应融资项目的质量与募集资金使用规划的真实性。特质性信息含量越高，往往意味着该债务主体希望用特有的、异质性的语言和内容引起投资者的特别关注；相反，大量使用模板化语言、特质性信息匮乏的说明书则尽量避免引起投资者的关注与分析。因此，我们提出假设三：

H3: 债券募集说明书中含有的特质性信息含量 (Heter) 越高，则债券的风险利差 (Spreads) 越低。特别地，当募集说明书中“规范语体”的比例 (Stdd) 越低，“特殊语体”的比例 (Info) 越高时，其债券的风险利差越低。

此外，“绿色债券”作为近年来政府支持的创新型金融工具，其发行目的为解决生态环境问题，加速企业转型进程，是政策支持与国家规划的具体实现，因此我们推测，被分类为绿色债券能够显著地降低利差，减少企业融资负担：

H4: 债券是否被第三方机构归类为“绿色债券”(Green)能够显著地影响风险利差(Spread)。在控制全部文本信息特征与控制变量的前提下，被归类为“绿色债券”的债券利差较低，不被认为是“绿色债券”的利差较高。

2.3.2 回归模型

为检验上述假设，本文采用以下四个回归模型，对比其回归结果。为了公式的简洁，此处记图表 2 中“财务信息控制变量”与“固定效应环境参数”包括的共 9 个变量为 Com_Control，记利率类型、债券期限、期权属性、发行规模四个变量为 Bond_Control。根据以上规则，本文建立的回归模型分别为：

模型 1: 不控制变量间的交叉项，使用 Heter 变量表征文本特质性信息含量：

$$\begin{aligned} Spreads_i = & \alpha + \beta_1 Eco_i + \beta_2 Policy_i + \beta_3 Heter_i \\ & + \beta_4 Green_i + \sum \beta_k \times Bond_Control \\ & + \sum \beta_l \times Com_Control + \varepsilon_{i,t} \end{aligned}$$

模型 2: 不控制变量间的交叉项，使用 Stdd 与 Info 变量表征文本语言特征：

$$\begin{aligned} Spreads_i = & \alpha + \beta_1 Eco_i + \beta_2 Policy_i + \beta_3 Stdd_i \\ & + \beta_4 Info_i + \beta_5 Green_i \\ & + \sum \beta_k \times Bond_Control \\ & + \sum \beta_l \times Com_Control + \varepsilon_{i,t} \end{aligned}$$

模型 3: 引入解释变量 (Eco, Policy, Heter) 与是否为绿色债券(Green)之间的交叉项，使用 Heter 变量表征文本特质性信息含量：

$$\begin{aligned} Spreads_i = & \alpha + \beta_1 Eco_i + \beta_2 Eco_i \times Green_i \\ & + \beta_3 Policy_i + \beta_4 Policy_i \times Green_i \\ & + \beta_5 Heter_i + \beta_6 Heter_i \times Green_i \\ & + \beta_7 Green_i + \sum \beta_k \times Bond_Control \\ & + \sum \beta_l \times Com_Control + \varepsilon_{i,t} \end{aligned}$$

模型 4: 引入解释变量 (Eco, Policy, Stdd, Info) 与是否为绿色债券(Green)之间的交叉项，使用 Stdd 与 Info 变量表征文本语言特征：

$$\begin{aligned} Spreads_i = & \alpha + \beta_1 Eco_i + \beta_2 Eco_i \times Green_i \\ & + \beta_3 Policy_i + \beta_4 Policy_i \times Green_i \\ & + \beta_5 Info_i + \beta_6 Info_i \times Green_i \\ & + \beta_7 Stdd_i + \beta_8 Stdd_i \times Green_i \\ & + \beta_9 Green_i + \sum \beta_k \times Bond_Control \\ & + \sum \beta_l \times Com_Control + \varepsilon_{i,t} \end{aligned}$$

3 实证结果分析

3.1 描述性统计

下表 2 展示了主要变量的描述性统计结果。一共 4224 支债券中，绿色债券有 174 支，约占 4%。相比于非绿色债券，绿色债券的风险利差更低，对环境保护的重视程度、对政策响应的关注度更高，“规范语体”及“特殊语体”比例均更低；绿色债券更有可能是企业债而非公司债，平均债券期限更；发行绿色债券的公司平均销

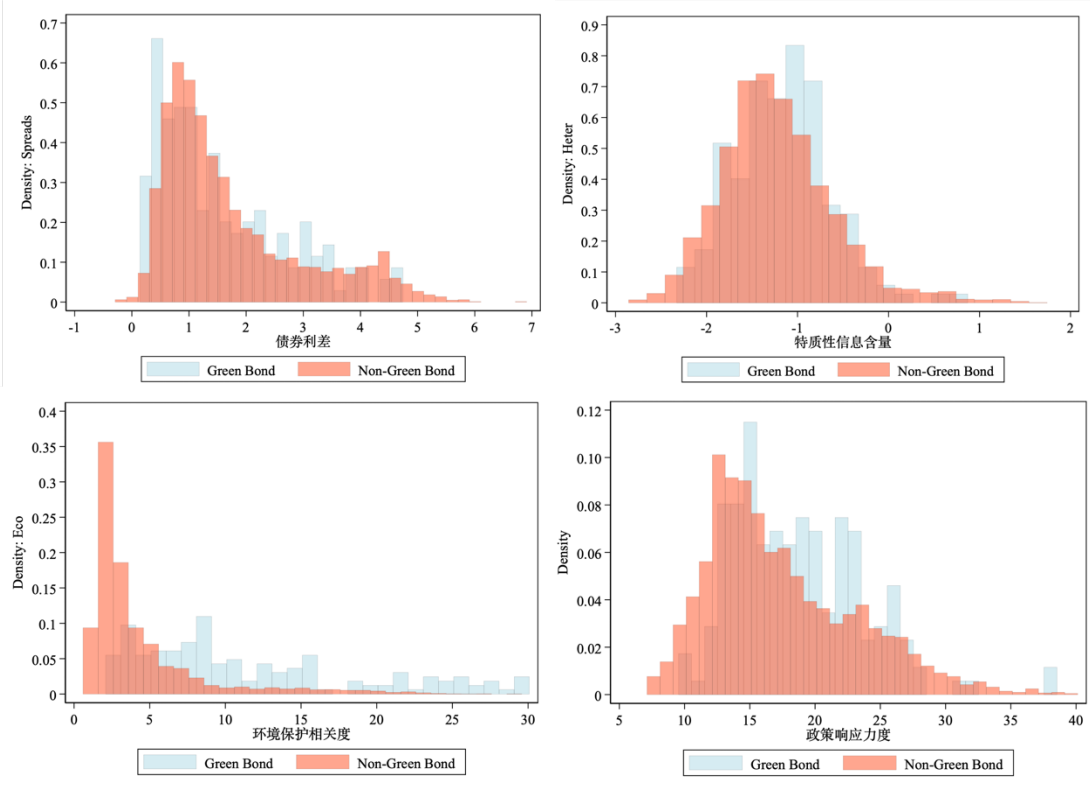
售收入更低。二者特质性信息含量、其他的发行主体或债券特征的控制变量的均值则无显著差异。

图 2 展示了风险利差、环境保护、政策响应、特质性信息含量在绿色债券、非绿色债券样本中的分布图。

	All				Green	Non-Green	Mean-Diff
	mean	sd	min	max	mean	mean	
Eco	4.70	0.07	0.59	50.23	13.02	4.34	8.68***
Policy	17.52	0.09	7.12	58.07	18.76	17.47	1.29***
Heter	-1.22	0.01	-2.86	4.57	-1.16	-1.22	0.06
Stdd	0.79	0.00	0.43	2.29	0.76	0.80	-0.04***
Info	0.24	0.00	0.15	0.47	0.24	0.24	-0.01***
Spreads	1.73	0.02	-0.29	6.74	1.58	1.74	-0.16*
Income	57.95	2.01	-4.05	1833.52	36.33	58.87	-22.54***
Profit	3.66	0.12	-10.98	66.65	3.24	3.67	-0.43
Cashflow	1.01	0.17	-176.58	274.95	1.14	1.00	0.14
Listed	0.15	0.01	0.00	1.00	0.17	0.15	0.02
Log_Age	2.76	0.01	0.00	4.26	2.80	2.76	0.04
Com_Type	1.21	0.01	0.00	3.00	1.21	1.21	0.00
Int_Type	0.42	0.01	0.00	2.00	0.43	0.42	0.01
Bond_Type	0.70	0.01	0.00	1.00	0.42	0.71	-0.29***
Maturity	5.31	0.03	1.00	20.00	6.07	5.28	0.80***
Convert	0.76	0.01	0.00	1.00	0.76	0.76	-0.00
Volumn	6.83	0.01	3.40	9.47	6.85	6.82	0.03
Rating	3.56	0.01	0.00	4.00	3.61	3.56	0.06
N	4224				174	4050	

表 2 主要变量描述性统计

图 2 核心变量分布



3.2 变量相关性分析

表 3 展示了核心变量之间的相关性。可以看到，债券绿色属性与环境保护、政策响应正相关，与“规范语体”及“特殊语体”比例均相关，与债券利差负相关，但相关系数不显著。

结合表 1、表 3，我们尝试检验环境保护（Eco）、政策响应（Policy）与债券绿色属性（Green）以及其他文本特征（Heter, Stdd, Info）之间的关系。回归模型见下式。其中，因变量 Y_i 为 Eco 或 Policy。回归结果见表 4。

$$Y_i = \alpha + \beta_4 \text{Green}_i + \sum \beta_k \times \text{Bond_Control} + \sum \beta_l \times \text{Com_Control} + \varepsilon_{i,t}$$

$$Y_i = \alpha + \beta_5 \text{Green}_i + \beta_3 \text{Stdd}_i + \beta_4 \text{Info}_i + \sum \beta_k \times \text{Bond_Control} + \sum \beta_l \times \text{Com_Control} + \varepsilon_{i,t}$$

由表 4 可以看出，在控制了一系列债券、发行主体变量及固定效应后，债券文本说明书体现的环境保护程度与债券绿色属性正相关，与异质性信息含量负相关；债券的政策响应程度与异质性信息含量负相关，与债券是否为绿债无显著关系。其中列（1）～列（4）均控制了销售收入、净利润、现金流量、是否上市、存续年限、发行人属性等 6 个发行主体变量，同时控制了发行年份、地区、行业固定效应。回归系数标准差按地区聚类。

	Green	Eco	Policy	Heter	Stdd	Info	Spreads
Green	1.00						
Eco	0.36***	1.00					
Policy	0.04**	0.01	1.00				
Heter	0.02	-0.06***	-0.01	1.00			
Stdd	-0.13***	-0.13***	-0.19***	-0.40***	1.00		
Info	-0.03*	-0.09***	-0.03	0.82***	-0.05**	1.00	
Spreads	-0.03	-0.14***	0.23***	0.09***	-0.09***	0.11***	1.00

表 3 变量相关性

	Y : Eco		Y : Policy	
	(1)	(2)	(3)	(4)
Green	8.91*** (0.85)	8.57*** (0.73)	-0.17 (0.33)	-0.18 (0.32)
Heter	-0.78*** (0.21)		-0.90*** (0.24)	
Stdd		-11.01* (5.77)		1.79 (5.03)
Info		-17.51*** (4.14)		-10.52*** (2.31)
Int_Type	-0.96** (0.41)	-0.85** (0.37)	-1.67*** (0.25)	-1.71*** (0.24)
Bond_Type	0.07 (0.25)	0.79** (0.38)	-5.33*** (0.44)	-5.19*** (0.45)
Maturity	-0.02 (0.05)	-0.01 (0.04)	0.16*** (0.06)	0.17*** (0.06)
Convert	0.06 (0.39)	0.06 (0.36)	0.16 (0.21)	0.21 (0.21)
Volumn	0.23* (0.12)	0.21 (0.13)	0.27*** (0.09)	0.27*** (0.10)
Rating	0.46*** (0.16)	0.44*** (0.14)	-0.73*** (0.18)	-0.74*** (0.18)
R^2	0.307	0.330	0.638	0.635
N	4224	4224	4224	4224

表 4 环境保护、政策响应与债券属性、文本信息变量的关

3.3 回归结果分析

3.3.1 结果说明与初步分析

表 5 中展示了四组模型的回归结果，表中（1）至（4）列分别对应第二章研究设计部分给出的模型（1）至（4），其中模型的因变量均为风险利差（Spreads）。在模型（1）与（3）中，使用 Heter 变量表征文本特质性信息含量；在模型（2）与模型（4）中则使用 Std 与 Info 变量表征文本语言特征。

同时，相较于列（1）（列（3）），列（2）（列（4））加入了文本变量与债券绿色属性（Green）的交叉项。列（1）～列（4）均控制了 6 个发行主体控制变量（销售收入、净利润、现金流量、是否上市、存续年限、发行人属性）与 3 个固定效应环境参数（发行年份、地区、

行业固定效应）。回归系数的标准差按地区聚类。

从图表 5 可以得出，在全部四个模型中，债券特征变量 Maturity, Convert, Int_type, Bond_type, Volumn, Rating 均表现出对风险利差的显著作用。债券期限越短、信用评级越好、发行规模越大的债券具有相对更低的风险利差，这与现有文献资料以及理论模型相吻合。此外，虚拟变量的显著性说明：不含期权选项的债券利率显著低于可赎回债券，这说明公司愿意为赎回选项付出额外的利率成本；同时，公司债券的利率显著低于企业债券，造成这一现象的原因可能是我国体制对公司债与企业债的发行主体具有不同要求；而在票面利率相同时，执行累进利率实际上对应更高的利息水平，因此在其他条件相同时，累进利率债券具有更低的名义利差。

	(1)	(2)	(3)	(4)
Green	-0.04 (0.06)	-0.03 (0.05)	-0.55** (0.26)	-1.57** (0.74)
Eco	-0.02** (0.01)	-0.01* (0.01)	-0.02** (0.01)	-0.02** (0.01)
Policy	-0.03*** (0.01)	-0.03*** (0.01)	-0.03*** (0.01)	-0.03*** (0.01)
Heter	0.15*** (0.05)		0.15*** (0.05)	
Std		0.24 (0.38)		0.17 (0.40)
Info		3.74*** (0.89)		3.70*** (0.90)
Eco_Green			0.01*** (0.00)	0.01*** (0.00)
Policy_Green			0.01 (0.01)	0.01 (0.01)
Heter_Green			-0.07 (0.09)	
Std_Green				1.18 (0.79)
Info_Green				1.14 (1.37)
Int_Type	-0.48*** (0.07)	-0.50*** (0.08)	-0.48*** (0.07)	-0.50*** (0.08)
Bond_Type	-0.42*** (0.08)	-0.42*** (0.09)	-0.42*** (0.08)	-0.42*** (0.09)
Maturity	0.05*** (0.01)	0.05*** (0.01)	0.05*** (0.01)	0.05*** (0.01)
Convert	0.63*** (0.07)	0.63*** (0.07)	0.63*** (0.07)	0.63*** (0.07)
Volumn	-0.29*** (0.06)	-0.29*** (0.06)	-0.29*** (0.06)	-0.29*** (0.06)
Rating	-0.69*** (0.05)	-0.69*** (0.05)	-0.69*** (0.05)	-0.69*** (0.05)
R ²	0.539	0.545	0.539	0.545
N	4224	4224	4224	4224

表 5 回归结果

3.3.2 研究假设验证

在四个模型中，Eco 与 Policy 变量的回归系数均显著为负，这证明了本研究提出的假设 H1 与 H2，即债券募集说明书中对环境保护的重视程度越高，对政策法规的分析与相应越多，则该债券的风险利差将会越低。这一结果体现出我国政府对债券市场定价的引导作用，与当前绿色金融取得的实际效果，越能够符合政府预期并贴合绿色项目的项目，能够获得越低的融资成本。

分别对比模型 1) 与 2); 3) 与 4)，两种衡量文本特质性信息的变量 Heter 与 Info 均对债券利差具有显著的正向作用，这与本研究提出的假设 H3 相矛盾。造成这一现象的原因可能有两点：其一，特质性信息越多的债券募集说明书，越可能引起买方机构的警觉与审视，这不仅会加深投资者对于债券的了解程度，更可能反而导致对债券募集目的的怀疑与不信任。其二，目前市场上多数债券募集说明书已经形成了相对成熟、完善的写作模式与模板，轻易做出更改并使用特质性的语言结构产生的负面作用高于其正面作用。

分别对比模型 1) 与 3); 2) 与 4)，可见在控制自变量与 Green 变量的交叉项之后，绿债类型变量 Green 的显著性有极大的改善，并且其回归系数的绝对值显著增大，体现出对于风险利差的重要负向作用。这证明了本研究提出的假设 H4，说明债券被分类为“绿债”这一行为本身就对其风险利差的降低有极大作用。相同质量的企业发行的相同类型债券，仅因为资金募集用途为环保相关项目，便可以产生约 0.5%-1.5% 的融资成本差距；一方面，这说明政府推进创立的“绿色债券”能够有效帮助企业完成环保转型，加速绿色经济建设；但另一方面，这也很可能产生巨大的监管套利空间，应注意完善对“绿债”的分类评级监管。

此外，模型 3) 与 4) 显示，Eco×Green 的交叉变量对于债券利差有显著正向作用，且其回归系数绝对值接近 Eco 自身系数。这说明募集说明书的环境保护相关度在全部公司、企业债券样本中能够起到降低利差的作用，但是在绿色债券这一样本区间内效果会大幅降低。造成这一现象的原因可能是稀释作用与边际收益递减，当一支债券已经被分类为绿色债券后，买方机构将会默认其资金募集用途为环保相关与绿色转型；此时企业再次在募集说明书中做出强调则不能起到正面作用，反而会被投资者认为属于冗余信息，损害投资者对于发行主体的信任与认识，造成在绿债样本中 Eco 变量的效果大

大减弱。

综上，本研究提出的假设 H1、H2、H4 得到验证，假设 H3 被否定。

4 总结与展望

本文主要围绕债券募集说明书文本中的词频统计结果，提取出文档中环境保护、政策响应关联度指标，以及文档措辞的特质指标。我们特别关注到公司债、企业债当中具有“绿色”标签的债券，并且联想到债券募集说明书中对环境政策等的强调是否有助于传递有效信息，降低发行利差。在词频统计的过程中，我们也发现募集说明书文本高频词具有较高的重复率，因此参考和改进了文献[12]中的方法衡量文档措辞的规范性与特异性。根据我们的初步结论，控制其他变量后，以上文本特征在样本内对债券发行利差影响显著，一定程度上说明募集说明书在债券募集发行过程中起到了信息传导的作用。

研究的值得改进之处有：1) 样本不全面，受限于募集说明书可得性，最终的样本仅覆盖 2016-2021 年发行的企业、公司债中 55% 左右，后续可以考虑加入更多来源；2) 文本分析过程可以优化，如改用中科院 NLPIR 分词工具，加入百度百科关键词词库等，以增加词频统计可靠性；3) 尝试其他文本特征，包括我们在研究计划的“模型拓展”部分中提到的 Word2vec 加权、主题词相似度等。囿于时间限制，相关内容有待进一步探索。

文本分析类研究普遍存在流程不规范，结论难以复现等问题。为此我们将研究完整的数据和代码在 Github 开源，地址已在报告开头脚注中列出。

5 参考文献

- [1] 朱家明, 祝静远. 基于线性回归的绿色债券票面利率影响因素的实证分析[J/OL]. 辽宁石油化工大学学报, 2019, 39(04): 87 - 92[2022 - 05 - 07].
- [2] 许欣怡. 我国绿色债券发行利率影响因素实证分析——基于可贷资金模型[J/OL]. 经营与管理, 2019(06): 21 - 23[2022 - 05 - 07].
- [3] 龚玉霞, 滕秀仪, 赛尔沃, 等. 绿色债券发展及其定价研究——基于二叉树模型分析[J/OL]. 价格理论与实践, 2018(07): 79 - 82[2022 - 05 - 07].

- [4] 倪弋菁. 社会关注度对绿色债券定价影响研究[D/OL]. 上海交通大学, 2019[2022 - 05 - 07].
- [5] 王玲. 绿色因素对我国绿色公司债券发行定价影响效果研究[D/OL]. 四川大学, 2021[2022 - 05 - 07].
- [6] 吴武清, 甄伟浩, 杨洁, 等. 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/OL]. 系统工程理论与实践, 2021, 41(7): 1650 - 1671.
- [7] 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/OL]. 管理科学, 2021, 34(4): 19 - 34.
- [8] 吴育辉, 唐浩博. 风险信息披露能预测债券违约风险吗? ——来自债券募集说明书的经验证据[J/OL]. 财务研究, 2021(5): 15 - 28.
- [9] Xiang X, Liu C, Yang M, et al. Confession or justification: The effects of environmental disclosure on corporate green innovation in China[J/OL]. Corporate Social Responsibility and Environmental Management, 2020, 27(6): 2735-2750[2022-05-07].
- [10] Messeni Petruzzelli A, Maria Dangelico R, Rotolo D, 等. Organizational factors and technological features in the development of green innovations: Evidence from patent analysis[J/OL]. Innovation, 2011, 13(3): 291 - 310[2022 - 05 - 07].
- [11] Li D, Huang M, Ren S, et al. Environmental Legitimacy, Green Innovation, and Corporate Carbon Disclosure: Evidence from CDP China 100[J/OL]. Journal of Business Ethics, 2018, 150(4): 1089-1104[2022-05-07].
- [12] Hanley, K. W., & Hoberg, G. (2010). The Information Content of IPO Prospectuses. Review of Financial Studies, 23(7), 2821-2864.