基于随机森林算法的 ESG 评价模型及其应用

吴 彪 陈 倩

摘 要 为积极践行我国生态优先、节约集约、绿色低碳的高质量发展道路，探索金融行业的 ESG 管理实践、做好绿色金融大文章，本文借鉴国际主流 ESG 评级框架和信用评级领域建模实践，立足我国国情，创新构建了一套基于随机森林算法的 ESG 评价模型。通过机器学习方法的应用，本文建立起上市企业 ESG 表现与企业财务绩效以及投资组合收益的逻辑框架，为投资组合构建和风险防控提供了一套新的方法。在比较分析多种深度学习算法后，最终采用了随机森林算法来构建模型，预测上市公司 ESG 评分。从实证分析结果来看，本文的 ESG 评价模型具有较好的分层效果和单调性，ESG 优选组合能获得优于业绩基准的超额回报，在一定程度上反映了上市公司追求经济价值与社会价值相统一的发展观，较好地契合经济高质量发展要求。

关键词 绿色金融 ESG 评价模型 随机森林 投资组合管理

一、引言

ESG 与国企可持续发展价值和保险资金长期限价值投资理念高度契合。2022 年 10 月党的二十大报告强调“深化国资国企改革”，作为实现国家战略发展的“排头兵”，央企如何践行可持续发展战略，助力加快构建中国特色 ESG 体系，推动企业提升 ESG 管理实践，促进高质量发展，已成为央企上市公司的一道必答题。ESG 以健全的公司治理为基础，追求更具绿色和社会责任的高质量可持续发展，与新一轮国企改革目标相一致，是国企改革的重要抓手。同时 ESG 与险资投资要求和理念相契合，保险资金负债的长期限特性要求其投资更具长期思维，追求能够穿越周期的价值投资。

基于政策导向和险资长期价值投资需要，为积极探索金融行业的 ESG 管理实践，我们借鉴国际主流 ESG 评级基本框架，立足我国国情，研究构建了一套基于随机森林算法的 ESG 评价模型（以下简称 RFB-ESG 评价模型）。该评价模型基于机器学习方法进行模型构建，有效减少了人为因素干扰，提升了评分客观性和准确性。同时为了解决不同行业 ESG 信息披露差异较大，且对环境因素敏感性不一的问题，该评价模型对上市公司按照环境敏感程度分组建模，按照不同指标权重区分各行业组，以实现行业间和行业内评分可比。RFB-

吴彪（通信作者），中信保诚资产管理有限责任公司风险总监；陈倩，中信保诚资产管理有限责任公司风险控制部。

ESG 评价模型采用三级议题指标体系，其中一级议题为国际通用的基本框架，涵盖环境（E）、社会（S）和治理（G），二级议题延伸为 E、S、G 项下 13 项主题，三级议题为各主题项下主要关键指标，实现了对全部 A 股上市公司的 ESG 表现进行评价，金融机构可基于该评价模型，构建 ESG 投资策略、组合管理策略和风险管理策略，为投资决策的科学性和业务发展的可持续性提供保障支持。

二、ESG 评级体系和建模方法研究述评

（一）ESG 评级体系

国际上 ESG 评级体系发展较为完善，评级机构众多，其中影响力较大的包括 MSCI、汤森路透、富时罗素、道琼斯等。在评级体系的设计上，各机构大多基于联合国责任投资原则提出的 ESG 核心要义构建基本框架，参照国际市场 ESG 信息披露机制，结合机构自身的理解诠释 ESG 价值观，设计各具特色的评级体系。

国内 ESG 评级起步较晚，目前主要包括华证指数、中证指数、Wind 等。各机构在评级体系的设计上主要对标国际主流机构评级框架，并有效结合中国国情对二、三层级议题进行合理调整，设计具有中国特色的指标体系。由于国内 ESG 数据披露率偏低，基于 A 股市场的 ESG 评级体系建立时间整体较短，各评级机构间的评分结果相关性较低。

（二）建模方法

目前 ESG 评价模型主要采用专家经验的方法进行指标选取和权重设置，通过机器学习方法的实践较少。参考信用风险内部评级模型 / 评分模型（以下简称信用评分模型）建模实践，Logistic 回归是信用评分模型领域的常用方法，其主要优点包括稳定性好、可解释性强、参数少等。于立勇和詹捷辉（2004）使用商业银行贷款数据为样本构建 Logistic 回归模型对企业违约概率进行实证分析，取得了较好的预测效果。伴随金融科技的快速发展，大量弱相关的另类数据被用于审批决策，数据结构多元化，变量成千上万个，而单一使用 Logistic 回归模型无法对成千上万个变量进行特征提取，从而无法对信用主体进行准确评价。根据 Shi et al.（2022）的研究，机器学习方法如 K 近邻、随机森林、支持向量机等，相较统计方法在信用风险评估领域更加有效和更具灵活性。D'Amato（2021）使用随机森林算法，研究了财务数据如何影响构成 STOXX 600 指数成分公司的汤森路透 Refinitiv ESG 评分。其发现资产负债表数据是解释 ESG 评分的关键因素，特别是某些特定的财务指标，如资产规模、负债水平、盈利能力等，对 ESG 评分具有显著的影响。研究结果表明，公司的财务状况和运营表现与其 ESG 表现之间存在密切的关系。Si（2023）同样采用了一种基于随机森林算法来构建 ESG 评分预测模型，通过交叉验证和网格搜索等技术来优化模型的参数，但该方法下因数据集的局限性和模型的可解释性等问题，在 ESG 投资决策及风险管理中的应用可行性有待提升。

（三）建模的挑战与应对

目前，上市公司社会责任报告披露率整体仍较低，数据不足且更新频率低为建模的最大障碍。根据同花顺 iFinD 数据统计，2022 年共有 1 232 家 A 股上市公司发布了 2021 年度企业社会责任报告（剔除重复发行的英文版、北交所），发布公司数

量占全部上市公司的 25.82%。随机森林模型对于缺失值相对不敏感，且对于非线性、高维度数据集具有运算高效、不易过拟合的特点，能够有效提升模型的适用性。

我国上市公司 ESG 管理理念普遍仍处于起步阶段，ESG 管理水平参差不齐，从 ESG 表现到股价表现的传导可能不是一蹴而就的。在国家政策导向、监管要求、资本市场和上市公司自身长期价值追求等多方面因素推动下，ESG 管理已经提上企业日程表，后续随着 ESG 数据的不断完善，我们也将及时进行模型迭代，不断提升 ESG 评价模型的有效性和鲁棒性。

三、RFB-ESG 评价模型构建框架

（一）总体思路

基于国内外优秀实践，ESG 评分模型的构建流程通常包括议题选择、权重设置、评分三步，而其中的难点集中在权重设置上，也是各家评级机构的核心技术所在。我们采用机器学习方法确定第三层级议题指标和指标权重，并逐级分层累加至二级议题、一级议题和 ESG 总分。

（二）模型关键假设

从 A 股上市公司 ESG 表现的三个维度分别来看，企业的环境表现反映企业持续改善其污染防治与资源利用等方面的效率与效果，有助于企业取得竞争优势（杨东宁 等，2004）；企业的社会表现主要强调企业在经营生产过程中加强对利益相关方关系的管理，最大限度地为企业各利益相关方创造多元综合价值（李伟阳 等，2009）；企业的公司治理表现包括企业内部治理和外部治理两个方面，其通过一系列制度安排来协调企业与利益相关方的关系，最终维护企业和各利益相关方的利益（李维安 等，2019）。因此，良好的企业 ESG 表现有助于提升企业绩效，进而反映在股价上，为投资者带来收益。

本文建模的关键假设：ESG 信息作为一种非财务的另类数据，能够为投资带来信息增量，ESG 评分高的公司相较评分低的公司能够为投资者带来更高的中长期股票投资收益。

具体逻辑可通过现金流贴现（Discounted Cash Flow，DCF）模型进行解释，DCF 模型由分子现金流以及分母贴现率组成，从分子端来看，ESG 水平高的公司有较强的同业竞争优势，其盈利能力和股利水平也会更高，进而估值较高；从分母端来看，由于贴现率（资本成本）与股票的风险程度相关，ESG 水平高的公司，其系统性风险和特质性风险相对较低，因而贴现率也会处于较低水平，进而估值较高。

现金流贴现模型：$$P=\sum_{t=1}^{n}\frac{CF_t}{(1+r)^t}$$

（三）模型路径选择

建模方法上，基于关键假设，以量化建模为主，尽可能减少人为因素的干扰，提升评价模型的客观性；以专家经验为辅，确保建模符合业务逻辑。虽然在许多情况下，使用机器学习方法的目的是降低专家设计规则的主观性，但并不意味着应该舍弃该领域的先验知识，例如专家经验能够在特征选择中帮助找出有用的特征。

模型选择上，基于 ESG 数据特征和应用需求，拟优先选择可解释性强的模型，如逻辑回归、决策树、随机森林、XGBoost、支持向量机等模型，尽量回避深度学习模型，即在模型效果和模型可解释之间寻求平衡。Fernández-Delgado et

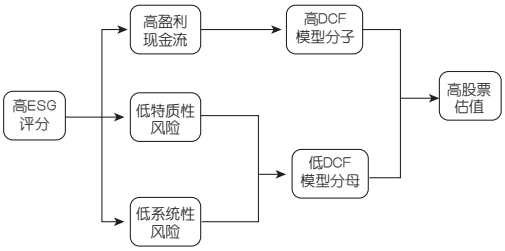

图 1　ESG 评分至股票估值的传导机制

al.（2014）在 UCI 121 个数据集[①]上进行 179 种分类学习算法的评价比价，结果显示随机森林算法和支持向量机分别名列第一、第二，且两者差异不大。

参考信用评分模型建模领域，较为传统和广泛使用的是 Logistic 回归模型，主要源于其在模型稳定性和可解释性方面优势显著，但随着数字经济的发展，越来越多的信贷业务实现线上审批，在此过程中，大量另类数据被应用到审批决策中，而 Logistic 回归模型主要适用于低维度和强信用风险相关数据，此时随机森林模型由于其更好的灵活性、更高的预测准确性和对大样本的适应性，开始被广泛应用。

在 ESG 评价模型的构建过程中，基础数据覆盖了大量另类数据，包括污染物排放、员工福利、市场监管处罚等，这些信息与股价走势不存在明确的线性关系。使用随机森林模型能够纳入更多的变量，且在一定程度上能够对不太稳定的变量具有较好的鲁棒性，尤其对于弱相关变量，随机森林能够使其在局部的预测能力得到发挥。

ESG 议题权重设置是评价模型构建的核心，在选择模型中，模型预测的准确率是需要重点考虑的。同时模型的可解释性也是 ESG 评分的重要组成，相较于深度神经网络等的“黑箱”，随机森林所用决策树内在的可解释性提升了整体模型的可解释性，基于特征重要性排序进行议题权重设置是较为合理且有效的。

根据本文建模实践，对高环境敏感等级行业（分组 1）和中低环境敏感等级行业（分组 2）分别采用随机森林和 Logistic 回归、支持向量机等方法建模，在两组分组下，随机森林模型效果均显著优于其他模型。因此从平衡精准性和可解释性考虑，本文选取随机森林模型。

四、评价模型构建

（一）行业分组

我国 A 股上市公司的 ESG 信息披露分为强制性和自愿式报告，两种报告方式并存，前者主要是根据证券交易所发布的披露指引，对针对涉及环境污染的污染物等信息进行强制披露，后者为上市公司在年报和社会责任报告中的自愿披露。国内社会责任报告的披露情况虽然逐年改善，但整体披露率不足 30%，仍然较低，同时由于缺乏强制性审核，以及各企业使用的参考标准和指标统计口径不同，导致报告在可用性和可比性方面仍有较大提升空间。

考虑到各行业、各公司 ESG 数据披露情况差异较大，尤其是环境排放类指标各行业的披露情况差异较大，其中高碳行业由于多涉及环境污染的污染物强制披露而

① UCI数据库是加州大学欧文分校提出的用于机器学习的数据库。

整体披露质量好于其他行业，图 2[①] 展示了环境排放类指标数据缺失率情况。并且不同行业在环境（E）因素的敏感程度上存在较大差别，因此为构建稳健、准确、有效的评价体系，有必要针对不同行业进一步细分模型。

基于以上两点，我们按照环境敏感等级将申万一级行业划分为高、中、低三个子集（表 1）。在实际模型构建过程中，由于低环境敏感等级行业样本较少，为提升模型效果，将低环境敏感等级与中环境敏感等级行业进行合并建模，高环境敏感等级行业单独建模。在入模特征的选择中，不区分行业子集，而通过不同权重加以区分，以保证 ESG 评分行业间可比。

（二）建模

本文按照数据准备、探索性数据分析、数据清洗与处理、特征工程、模型构建、模型评估的流程进行建模。

1. 数据准备

（1）数据提取

ESG 指标数据基于秩鼎 ESG 数据库，来源于上市公司披露、政府监管、第三方平台等，行业、收益率等数据基于聚源数据库和 Wind 终端。

（2）好坏样本定义

采用监督建模方法，需定义目标变量，同时为了提升好坏样本的区分度，减少边界附近样本量，还需定义好坏样本的标准。本文以超额收益率作为目标变量，目标变量及好坏样本标准定义如下：

a. 好样本：超额收益率 ≥3%

b. 坏样本：超额收益率 ≤–3%

其中，超额收益率 = 个股收益率 –Wind 全 A 同期收益率

2. 探索性数据分析

本文在建模前进行探索性数据分析，进一步了解原始数据的缺失值、异常值、数据类别和分布等。图 3~5 分别展示了数据缺失分布样例、异常值分布样例、各指标类别分布样例。

3. 数据清洗与预处理

为实现入模数据的可用性，在建模前需要进行数据清洗与预处理，通常包括数据清洗、缺失值处理、异常值处理、删除唯一值变量、删除长尾变量以及根据专家经验删除不相关变量等。以下结合本文主要实践进行简要介绍。

（1）缺失值处理

本文首先计算了各变量的缺失率情况，对缺失值具体处理如下。

a. 对于缺失率超过 80% 的变量（全口径）直接进行删除。

b. 在缺失值填补方面，结合变量类型

表 1　行业划分

环境敏感等级	申万一级
高	钢铁、公用事业、基础化工、建筑材料、交通运输、煤炭、石油石化、有色金属
中	电力设备、电子、房地产、纺织服饰、国防军工、环保、机械设备、家用电器、建筑装饰、农林牧渔、汽车、轻工制造、食品饮料、通信、医药生物、综合、美容护理
低	商贸零售、传媒、非银金融、计算机、社会服务、银行

① 图2~8，见增强出版，中国知网—《金融市场研究》。

和具体变量，分别按照众数、行业均值等方法进行填补。大多数连续变量优先按照申万三级行业均值进行填补，如果申万三级行业无相关变量数据，则按照申万一级行业均值进行填补[如颗粒物排放密度（营收）、人均基本养老保险等变量]。

（2）异常值处理

异常值处理通常是为了防止过拟合[①]并进而导致的模型泛化能力降低，本文考虑采用 3σ 方法对连续变量进行异常值处理，在实际建模中，本文采用的随机森林方法对异常值不敏感。

（3）变量替换

根据专家经验，主要选取密度、比值等变量进行建模，替换金额类变量（如近 12 个月环境处罚金额）、总量类变量（如颗粒物年排放量）等。

4. 特征工程

好的数据和特征是模型和算法发挥更大作用的前提，通过特征工程能够筛选出更好的数据特征并进一步提升模型训练效果。

本文在特征工程步骤完成评价模型构建的第一个关键步骤，即议题选择。需要注意的是，在实际应用中，仅仅从数据挖掘的角度进行特征选择是不够的，需要选择的特征具有业务解释性，脱离了具体业务而得到的特征选择结果，即使可以达到较好的预测效果，也需要进一步推敲。本文的特征工程包含特征编码和特征选择两个步骤。需要指出的是，此处的随机森林特征选择与下文模型构建过程中采用的随机森林模型是基于同一思想和方法，但不是同一个模型。

（1）特征编码

本文对离散变量采用 One-hot（独热）编码，如在治理维度下，以内控审计-意见类别变量为例，One-hot 编码如表 2[②] 所示。

（2）特征选择

本文主要采用相关性分析、聚类分析、随机森林相结合的方法进行特征选择。

相关性分析：按照 E、S、G 维度分别计算指标相关性，删除相关系数 0.6 以上的特征，在删除时优先选择不同簇中的特征以及随机森林特征重要性排序中重要性靠后的特征。图 6 展示了特征相关性分析样例。聚类分析：本文采用特征聚集方法（Feature Agglomeration），应用凝聚聚类将行为相似的特征分组在一起。随机森林：采用随机森林做特征选择时并没有给出选择多少特征，而是侧重给出每个特征的重要性排序。

本文基于上述三种方法，选定不同维度的入模特征，也即评价模型的三级指标体系。

5. 模型构建及评估

（1）模型构建

本文在模型构建步骤完成评价模型构建的第二个关键步骤，即权重设置，具体采用随机森林进行模型构建。

① 过拟合是指机器学习模型在训练样本中表现得过于优越，导致在验证数据集以及测试数据集中表现不佳，其通常发生在模型过于复杂、参数过多的情况下，导致模型过度记忆了训练数据的噪声和细节，而忽略了数据的整体趋势和泛化能力。

② 表2见增强出版，中国知网—《金融市场研究》。

随机森林模型[①]是Bagging框架下最著名的一种机器学习模型，其本质上是许多决策树的集合，其核心思想是希望“每棵树好而不同”，集成后使其具有较高的预测性能和鲁棒性。图 7 展示了随机森林模型示意图。

为了实现这一策略，通常需要构造许多棵决策树，每棵树都应该对目标值做出可以接受的预测，还应该与其他树不同，即将随机性添加到树的构造中。随机森林中树的随机性体现在两方面，一是用于构造树的样本，二是每次划分的特征。构造随机森林模型，需要确定用于构造的树的个数，例如想要构造 100 棵树，这些树在构造时彼此独立，算法对每棵树进行不同的随机选择，以确保树和树之间的差异性。构造一棵树，首先对样本进行自助采样（bootstrap sampling），即在包含 M 个样本的数据集中，有放回地进行 m 次随机采样操作，得到包含 m 个样本的采样集。接下来基于该采样集构造决策树，在决策树的每个结点处，先从该结点的特征集合中随机选择一个包含 k 个特征的子集，然后再从这个子集中选择一个最优特征用于划分。在随机森林模型的建模过程中，不仅通过自助采样法引入了数据层扰动，增加了样本的多样性，还通过在每棵树的训练中随机选择特征，从而增加了特征层的扰动，这使得模型最终集成的泛化性能可通过基学习器（决策树）之间差异度的增加而进一步提升。在随机森林中，每棵决策树都是一个分类器，随机森林的最终输出结果由决策树通过投票的方式产生。基于 CART 算法的随机森林算法，具体如下。

①应用 Bootstrap 方法从原始训练集 X 有放回地随机抽取子样本集 Xi (i=1, 2, ..., N)。

②对每个子样本集 Xi , 重复如下步骤，直到生成 T 棵决策树。

第一步：假设共输入 M 个特征属性，随机抽取其中的 m 个特征属性作为当前决策树分裂的属性集，其中 m 远小于 M。

第二步：在决策树每个内部节点，从 m 个特征属性中，选择基尼系数最小的特征及其对应的切分点作为最优特征与最优切分点，并依最优特征与最优切分点对该节点进行分裂，直至子样本集 Xi 的基尼系数小于预定阈值：

$$\mathrm{Gini(p)}=\sum_{k=1}^{k}p_k(1-p_k)=1-\sum_{k=1}^{k}p_k^2=1-\sum_{k=1}^{k}\left(\frac{|C_k|}{D}\right)^2$$

其中，p_k 表示选中的样本属于 k 类别的概率，C_k 表示数据集中属于类别 k 的样本数量，D 表示数据集中样本总数。

③将通过上述两个步骤生成的 T 棵决策树组合即为随机森林，每棵决策树采用投票的方式参与表决，所得票数最多的类别将作为随机森林的输出结果：

$$H(x)=\arg\max_{y\in y}\sum_{t=1}^{T}\mathrm{I}\left(h_t(x)=y\right)$$

其中，$H(x)$ 表示组合分类模型，$h_t(x)$ 表示决策树模型，y 表示目标变量，I 为指示性函数，$\arg\max_{y\in y}$ 表示使 $\sum_{t=1}^{T}\mathrm{I}\left(h_t(x)=y\right)$ 取最大值时 y 的取值。

① 随机森林算法是Leo Breiman (2001) 提出的一种分类和预测模型，具有参数调整较小、对高维变量的多重共线性不敏感，能高效处理大样本数据、能较好估计特征重要性，以及较强的抗噪音能力等特点，在很多任务中展现出强大的性能。

随机森林模型的主要参数包括决策树数量、决策树最大深度、拆分内部节点所需最少样本数、叶子节点所需最小样本数、不纯度衡量指标（本文采用 Gini 系数），本文采用了带交叉验证的网格搜索进行调参，以得到对模型泛化性能的更好估计。在调参过程中，为进一步提升模型泛化能力，本文同时尝试了是否对数据进行标准化、归一化、降维等方法，以及是否进行上采样、下采样以处理样本不均衡问题。

本文基于 python 语言，使用机器学习领域主流的开源工具包 scikit-learn，并基于前述行业子集分别构建子模型，输出随机森林特征重要性排序。在随机森林中，每棵决策树都可以输出自己的特征重要性，特征重要性是为每个特征对树的决策的重要性进行排序，对于每个特征来说，它都是一个介于 0~1 之间的数字，其中 0 代表没有用到该特征，1 代表完美预测目标值，全部特征的特征重要性求和始终为 1。随机森林特征重要性的计算方法是将森林中所有树的特征重要性求和并取平均。图 8 展示了随机森林中决策树样例。

指标权重设置基于随机森林特征重要性排序结果，并对权重过高和过低的指标进行调整，以使每个指标的权重均不高于 5%，且不为 0%。

（2）模型评估

本文采用准确率、召回率、f1- score、AUC 等指标，对模型的泛化性能进行评估。子模型 1（高环境敏感等级行业）测试集准确率 0.73，召回率 0.74，f1-score 0.74，AUC 0.64；子模型 2（中低环境敏感等级行业）测试集准确率 0.72，召回率 0.76，f1-score 0.74，AUC 0.61。图 9[①] 展示了子模型 1 和子模型 2 的 ROC 曲线。

本文中初始好坏样本比例为 1:1.75，面临样本不均衡问题，虽然我们通过数据层上采样的方法来改善该问题，但 AUC 指标整体仍略有偏低，一定程度上也印证了我们前文所说的“我国上市公司 ESG 管理理念普遍仍处于起步阶段，ESG 管理水平参差不齐，从 ESG 表现到股价表现的传导可能不是一蹴而就的”。

6. 模型比较

本文重点选取逻辑回归、支持向量机、决策树、k 近邻等可解释性强的算法与随机森林进行对比，从评估结果来看，随机森林算法整体优于其他算法，表 3 和图 10 展示了子模型 1 各算法比较结果，表 4[②] 和图 11 展示了子模型 2 各算法比较结果。

（三）生成评价模型

评价模型构建的第三个关键步骤即为评分。本文结合国际主流评价标准和中国特色，拟定三级议题指标体系，其中一级

表 3　子模型 1 各算法比较

算法	准确率	召回率	f1–score	AUC
随机森林	0.73	0.74	0.74	0.64
逻辑回归	0.62	0.67	0.69	0.61
支持向量机	0.63	0.77	0.69	0.51
决策树	0.66	0.65	0.65	0.54
k 近邻	0.62	0.57	0.60	0.50

① 图9~11，见增强出版，中国知网—《金融市场研究》。
② 表4见增强出版，中国知网—《金融市场研究》。

表 5　议题节选

一级议题	二级议题	三级议题（节选）
环境	环境管理	重点管理排污单位数量、环境处罚金额占营收比重
	环境排放	颗粒物排放密度（营收）、CO_2/SO_2 排放密度（营收）
	气候变化	供应链气候风险等级
社会	产品责任	质量管理体系认证
	健康与安全	安全生产费成本占比
	人力资本	人均职工福利费
	社会资本	扶贫资金支持额占营收比重
	业务创新	研发投入占营收比重
治理	股东	近 12 个月股东累计增持数量占总股本比重、股利支付率
	合规	行业监管处罚金额占营收比重
	审计	内控审计-是否存在非财务报告内部控制重大缺陷
	信息披露	信息披露等级
	治理结构	独立董事占比、董监高薪酬占营收比重

议题 3 个，二级议题 13 个，三级议题 48 个。

环境议题重点关注企业在生产经营过程中对外部环境的影响，即环境排放因素，同时考察企业在环境管理方面所做的努力和成效，最后考察企业供应链所面临的气候灾害潜在风险。

社会议题考虑中国国情政策，引入中国特色指标，全面反映企业对员工、客户、社区及其他利益相关方所承担社会责任的履行情况，重点关注业务创新和人力资本，同时覆盖产品责任、健康与安全和社会资本。

治理议题着眼于公司的可持续经营，重点关注企业股东和治理结构，同时覆盖合规、审计情况以及企业的信息披露水平。

1. 打分方法

①对三级议题进行指标方向和评价范围划分，依据指标反映公司状况的正负向，在指标方向上分为正向、负向、中性和类别指标，在评价范围上分为全市场和行业内指标。

②针对不同方向的指标，在评价范围内进行排序，以排序值作为得分基础。

③将指标得分与对应权重相乘，作为该项指标的最终得分。

④逐级向上加总各级议题得分，并得出 ESG 总分。

2. ESG 评分实证分析

本文基于 2021 年 ESG 指标数据，通过 RFB-ESG 评价模型对 4 637 家 A 股上市公司 ESG 表现进行评分，评分采用 0~100 分制，得分集中在 60~70 分，评分结果整体呈钟形分布（图 12），具有较好的区分度。同时我们也注意到，上市公司 ESG 评分的整体分值偏低，一方面，或与引入了较多负向指标且依赖排序值打分相关；另一方面，由于底层数据披露率偏低，对于大多数连续变量的缺失值我们采用申万三级行业中位数进行填补，可能一定程度上也拉

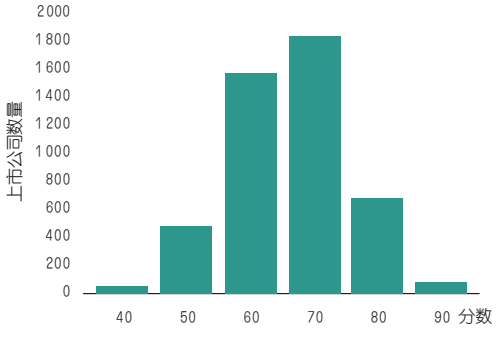

图 12　ESG 评分分布

图 13　基于 ESG 优选策略的组合收益率走势图

低了分值。表 6[①] 展示了 ESG 评分主要统计量。

五、ESG 评分返回测试

为验证 RFB-ESG 评价模型的有效性，我们构建优选策略进行返回测试。优选策略按照沪深 300 成分股行业权重（本文选取年末权重）构建投资组合，其中各行业均选取 T 年 RFB-ESG 得分排名前 20% 上市公司，在行业内进行等权重配置，构建 ESG 优选组合并计算 T+1 年收益率，回测时间区间（即收益率区间）为 2018 年 1 月 1 日至 2022 年 12 月 31 日。图 13 展示了基于 ESG 优选策略的组合及沪深 300 指数收益率走势情况。

绩效指标方面，ESG 优选组合年化收益率和夏普比率均明显优于同期沪深 300 指数。风险指标方面，ESG 优选组合波动率略高于沪深 300 指数，最大回撤显著低于沪深 300 指数（表 7）。历年涨跌幅方面，除 2018 年未超越沪深 300 指数外，2019—2022 年 ESG 优选组合收益率均超越沪深 300 指数，优选策略超额收益明显（表 8）。

六、研究结论与展望

（一）结论

在 ESG 评价模型构建中，指标权重设置是关键，本文创新性地基于随机森林模型来设定指标权重，并在随机森林建模过程的特征选择步骤完成评价模型三级议题指标选择，保证了评价模型构建关键环节的一致性和评价模型的有效性。

由于评价模型构建主要基于量化模型，减少了人为干扰因素，考虑到目前国内外主流 ESG 评级机构主要基于专家经验

① 表6见增强出版，中国知网—《金融市场研究》。

来选择议题和拟定权重，但由于目前缺少监管统一的标准和规则，各家机构的评级差别较大，因此从客观性、有效性和实用性角度出发，我们认为采用基于实证研究的量化分析进行建模是当下金融机构较好的选择。

从评分结果来看，A 股上市公司 ESG 得分整体呈钟形分布，在统计上具备相对较好的区分度和可比性。从回测结果来看，ESG 评分具有较好的分层效果和单调性，优选组合持续带来正回报。

综上所述，本文基于随机森林模型构建 ESG 评价模型，有效丰富了 ESG 评分评级体系建立方法的实践，也是保险行业参与绿色金融工作的一次有益尝试和探索。

（二）研究展望

2024 年 2 月 8 日，上海证券交易所、深圳证券交易所和北京证券交易所发布《上市公司持续监管指引第 X 号——可持续发展报告（征求意见稿）》（以下简称《信息披露指引》）。《信息披露指引》规定了以治理，战略，影响、风险和机遇管理，指标与目标为核心要素构建披露框架，明确了环境、社会、公司治理议题，其中环境议题覆盖应对气候变化，污染防治与生态系统保护，资源利用与循环经济等，社会议题覆盖乡村振兴与社会贡献，创新驱动、供应商与客户，员工等，公司治理议题覆盖可持续发展相关治理机制，防范商业贿赂与不正当竞争等。结合《信息披露指引》未来将正式出台及不断完善的背景，我们提出进一步研究展望。

①模型迭代视角下，后续可根据《可持续发展信息披露指引》要求，调整、完善三级议题指标和权重，在指标构建上充分体现可持续发展理念的指导思路。②投资视角下，基于“国际可持续发展准则理事会”（ISSB）和“全球报告倡议组织”（GRI）等国际机构以及《可持续发展信息披露指引》提出的“双重重要性”原则，将 ESG 因子作为现有财务因子库的有效补充，基于上市公司 ESG 得分从非财务信息中挖掘优质标的，获取超额 Alpha 收益。③组合管理视角下，结合账户特点和投资策略跟踪组合 ESG 评分、组合碳强度等计量方法，采用 ESG 正面筛选和负面排除策略对投资组合进行正面引导和反向约束，积极探索降低投资组合碳强度管理。④风险管理视角下，基于上市公司 ESG 得分和变化趋势探索构建 ESG 风险规避策略和风险预警策略，对具有严重 ESG 尾部风险的标的持续进行跟踪和预警，前瞻防范化解风险。N

表 7　基于 ESG 优选策略的组合风险绩效指标（2018—2022）

名称	年化收益率	年化波动率	年化夏普比率	最大回撤
ESG 优选组合	7.25%	20.73%	0.39	32.52%
沪深 300	–0.83%	20.60%	–0.01	39.58%

表 8　基于 ESG 优选策略的组合年度收益率

年度	优选组合	沪深 300	超额
2018	–27.05%	–25.31%	–1.75%
2019	38.38%	36.07%	2.31%
2020	29.33%	27.21%	2.12%
2021	18.98%	–5.20%	24.18%
2022	–13.96%	–21.63%	7.67%

学术编辑：韦燕春

参考文献

[1] 李井林,阳镇,陈劲,等.ESG促进企业绩效的机制研究——基于企业创新的视角[J].科学学与科学技术管理,2021,42(9):71-89.

[2] 李维安,郝臣,崔光耀,等.公司治理研究40年:脉络与展望[J].外国经济管理,2019(12):161-185.

[3] 李伟阳,肖红军.基于社会资源优化配置视角的企业社会责任研究—兼对新古典经济学企业社会责任观的批判[J].中国工业经济,2009(4):116-126.

[4] 王青天,孔越.Python金融大数据风控建模实战:基于机器学习[M].北京:机械工业出版社,2020.

[5] 杨东宁,周长辉.企业环境绩效与经济绩效的动态关系模型[J].中国工业经济,2004(4):43-50.

[6] 于立勇,詹捷辉.基于Logistic回归分析的违约概率预测研究[J].财经研究,2004,30(9):15-23.

[7] 周毓萍,陈官羽.基于机器学习方法的个人信用评价研究[J].金融理论与实践,2019(12):1-8.

[8] 周志华.机器学习[M].北京:清华大学出版社,2016.

[9] Dingwen Si.An Efficient Predicative Approach of ESG Invest Scoring Using Random Forest Algorithm[J]. BCP Business & Management,Volume 45(2023) .

[10] Leo Breiman.RANDOM FORESTS[J].Machine Learning,2001,Vol.45:5-32.

[11] Manuel Fernández-Delgado,Eva Cernadas,Senén Barro,Dinani Amorim.Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?[J].Journal of Machine Learning Research,2014,15:3133-3181.

[12] Si Shi,Rita Tse,Wuman Luo,Stefano D' Addona,Giovanni Pau,Machine Learning Driven Credit Risk:A systemic Review[J].Neural Computing and Applications,2022,34:14327-14339.

[13] Valeria D' Amato,Rita D' Ecclesia,Susanna Levantesi.ESG Score Prediction through Random Forest Algorithm[J].Computational Management Science,2022,19:347-373.

An Efficient Environmental, Social and Governance (ESG) Rating Modeling and Application Using Random Forest Algorithm

WU Biao　CHEN Qian

(CITIC-Prudential Asset Management Co., Ltd.)

Abstract　Sustainable and green finance has become a major objective for financial institutions that are regularly dealing with the measurement and management of ESG risks. This paper seeks to build quantitative ESG metrics, examining equity investment and risk management in practice. After comparing five deep learning algorithms, we adopt a Random Forest algorithm to build a model and predict the ESG ratings of 4,637 A-share listed companies. Furthermore, this paper establishes a framework for ESG performance and the financial track record of public companies along with ESG-driven portfolios based on China's national conditions. From the results of empirical analysis, the ESG prediction model displays a relatively good discriminating ability and a high degree of accuracy. It can also be noted that the ESG-driven portfolio can obtain excess returns that are better than the benchmark. In addition to contributing to the ESG modeling and portfolio management, the research in this study also supports the compatibility of pursuing economic value and social responsibility at Chinese public firms.

Keywords　ESG Evaluation Model, ESG Investment, Random Forest, Portfolio Management

JEL Classification　C51　G11　G32