

中国上市公司 ESG 评价体系与企业高质量发展

沈宇捷 致理-物 22 2022012290

摘要:

ESG 体系为企业的绿色转型提供了基本框架,推动企业 ESG 发展是保护环境、实现高质量发展的关键环节。本文基于中国部分上市企业 2017-2021 年数据,采用驱动力-压力-状态-影响-响应(DPSIR)模型构建涵盖环境(E)、社会责任(S)、公司治理(G)三维度的中国上市企业 ESG 评价体系。并采用彭博资讯网的 ESG 评分数据,实证研究 ESG 评分对中国能源类上市公司高质量发展的影响。研究发现,ESG 评分与企业收益并没有显著关联,企业更多地基于环境和社会责任考虑提高 ESG 表现。最后,本研究对 2022 年的能源类上市企业 ESG 评分进行预测。

关键词: ESG 表现; DPSIR 模型; 固定效应模型; 中国能源类上市公司

1 引言

近年来，随着气候变化和环境问题给国际经济、社会发展带来严峻挑战，可持续发展理念日益成为共识。作为经济的行为主体之一，企业是发展绿色、可持续经济的基础。近年兴起的 ESG 体系就为企业的绿色转型提供了基本框架。ESG 是综合考虑企业环境（Environmental）、社会责任(Social)和治理表现（Governance）来实现企业协调发展的新发展理念。

对于 ESG 评价体系的构建和 ESG 表现的影响，已经有丰富的文献讨论。在 ESG 评价体系的构建和对 ESG 评分的综合分析方面，袁家海等基于 PSR 模型和层次分析的指标分类与权重配比方法构建中国大型发电上市公司的 ESG 评级体系（袁家海等，2018）。蔡泽栋基于 MSCI 的 ESG 评级思想，结合熵权-灰色关联分析模型进行 ESG 指标综合评价（蔡泽栋，2020）。在探究 ESG 表现的影响方面，彭百川等以中国 A 股上市公司为样本，构建基准回归模型并进行稳健性检验和异质性分析,发现企业良好的 ESG 表现能显著提高绿色创新效率，中介变量为企业风险承担水平和供应链话语权（彭百川等，2024）。张宏等基于沪深 A 股上市工业企业 2015–2020 年数据，实证检验了 ESG 对企业融资成本的影响，并发现内部控制和媒体关注的调节作用（张宏等，2024）。

本文通过使用 DPSIR 模型，构建包括驱动力、压力、状态、影响和响应五个维度的中国能源类上市企业的 ESG 综合评价指标体系，运用熵值法客观赋权，评价和分析 2017–2021 年我国上市企业的 ESG 发展状况。再聚焦于能源类企业，构建 ESG 评分对企业高质量发展的线性回归模型和时间固定效应模型，探究二者的关联。最后构建线性回归模型，对 2022 年中国上市能源类企业的 ESG 评分进行预测。

2 中国上市公司 ESG 发展评价

2.1 中国上市企业的 ESG 综合评价指标体系

DPSIR 模型是解决环境和社会关系问题的有效工具，包含了驱动力、压力、状态、影响和响应 5 大因素。其中，驱动力（D）代表经济增长和市场需求等作为驱动因素，推动企业的变化和发展；压力（P）主要是指企业经济活动对环境的压力；状态（S）指在经济发展的压力下，企业的环境保护等措施所处的状态水平；影响（I）则是指各种状态水平对企业发展的反馈结果和影响程度；响应（R）是指企业为实现经济与环境的协调发展而采取的各种积极措施和对策。

表 1 基于 DPSIR 模型的中国上市企业 ESG 综合评价指标体系

目标层	因素层	指标层	单位	属性
ESG 综合评价指数	D 驱动力	企业市值	元	+
		市销率 PS1		+
		账面市值比		+
	P 压力	氮氧化物排放量	千克	–
		硫氧化物排放量	千克	–
		颗粒物排放量	千克	–
		废水排放量	吨	–
		温室气体净排放量	千克	–
		COD 排放量	千克	–
	S 状态	能源消耗总量	千克	–

	用水量	吨	-
	环保意识	0 表示无, 1 表示有	+
	节约能源种类数		+
	主要排污种类数		-
I 影响	营业收入	元	+
	总负债	元	-
R 响应	管理费用绿色投资	元	+
	在建工程环保投入	元	+
	废弃物综合利用率		+

2.2 综合分析测算中国能源类上市企业的 ESG 发展状况

本文采用熵值法进行客观指标赋权,进而由指标计算得到 ESG 综合评价指数。熵值法的基本原理为根据指标的信息效用价值确定权重。

(1) 数据标准化

由于不同的指标具有不同量纲和单位,会影响指标之间的可比性。因此首先需要对 ESG 指标的各项原始数据进行标准化处理。构建原始数据矩阵 $X = \{X_{ijt}\}_{m \times n \times p}$, X_{ijt} 表示第 i 个样本的第 j 项指标在第 t 年的值,用极差法进行处理。

$$\text{正向指标: } y_{ijt} = \frac{x_{ijt} - \min\{x_i\}}{\max\{x_i\} - \min\{x_i\}} \quad (1)$$

$$\text{负向指标: } y_{ijt} = \frac{\max\{x_i\} - x_{ijt}}{\max\{x_i\} - \min\{x_i\}} \quad (2)$$

其中 $\max\{x_i\}$ 和 $\min\{x_i\}$ 分别表示第 i 个样本在 n 项指标和 p 年中最大和最小的指标值。

(2) 熵值法获取权重

对于第 j 个指标,其信息熵 E_{ij} 可通过以下公式计算得到:

$$E_j = -\frac{1}{\ln(np)} \sum_{i,t} p_{ijt} \ln(p_{ijt}) \quad (3)$$

其中 p_{ij} 为:

$$p_{ijt} = \frac{y_{ijt}}{\sum_{i,t} y_{ijt}} \quad (4)$$

第 j 个指标的权重 w_j 为:

$$w_j = \frac{1 - E_j}{\sum_{j=1}^m (1 - E_j)} \quad (5)$$

得到的各指标权重结果如下表所示:

表 2 各指标权重		
因素层	指标层	权重
D 驱动力	企业市值	0.0594
	市销率 PS1	0.0674
	账面市值比	0.0743
P 压力	氮氧化物排放量	0.0383
	硫氧化物排放量	0.0255

S 状态	颗粒物排放量	0.0534
	废水排放量	0.0407
	温室气体净排放量	0.0425
	COD 排放量	0.0023
	能源消耗总量	0.0486
	用水量	0.0332
	环保意识	0.0703
	节约能源种类数	0.0716
	主要排污种类数	0.0727
	营业收入	0.0559
I 影响	总负债	0.0574
R 响应	管理费用绿色投资	0.0627
	在建工程环保投入	0.0509
	废弃物综合利用率	0.0730

由此可知，账面市值比、环保意识、节约能源种类数、主要排污种类数、废弃物综合利用率所占权重较大

(3) ESG 综合评价得分

对于每个样本，其 ESG 综合评价得分 S_{it} 为：

$$S_{it} = \sum_{j=1}^m w_j y_{ijt} \quad (6)$$

所有样本分属四种产业：工业，原材料，公用事业和能源。对于每个产业我们选取企业市值最大的公司作为代表进行 ESG 综合评价得分的可视化以观察得分变化趋势并比较不同产业的 ESG 得分。

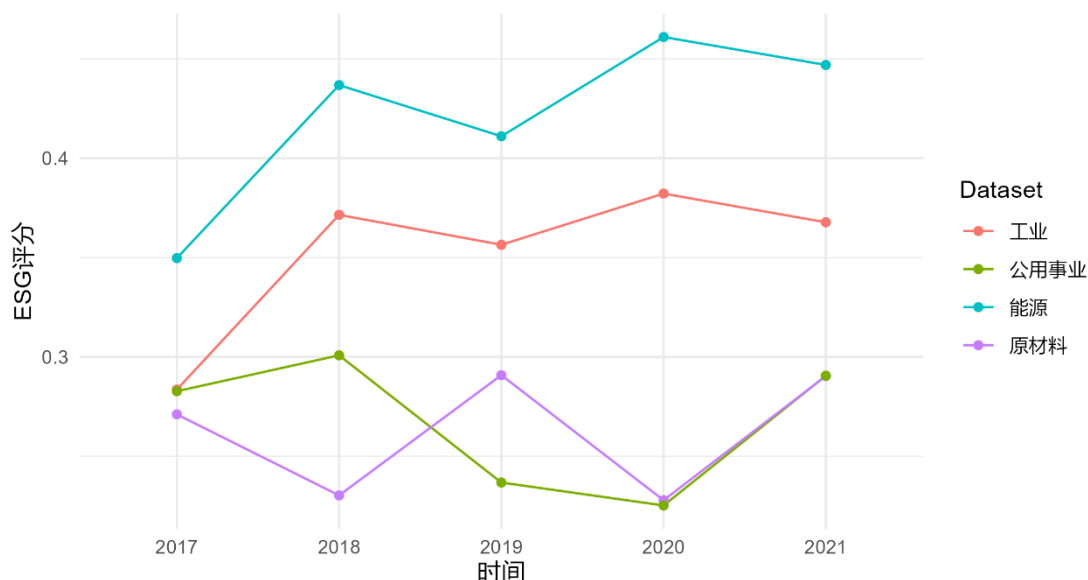


图 1 不同产业的 ESG 评分随时间变化趋势

对于 4 种产业，分别选取部分企业，进行更详细的可视化。由图可知，能源类企业相比其他三种产业 ESG 评分要更高，且更为稳定，整体呈现上升趋势。工业类企业 ESG 评分尚可，虽波动较大，但整体呈上升趋势。原材料类企业和公用事业类企业 ESG 评分波动较大，

且没有整体的上升趋势。在下一部分，我们选取 ESG 评分较好的能源类企业，探究 ESG 表现对于企业高质量发展的影响。

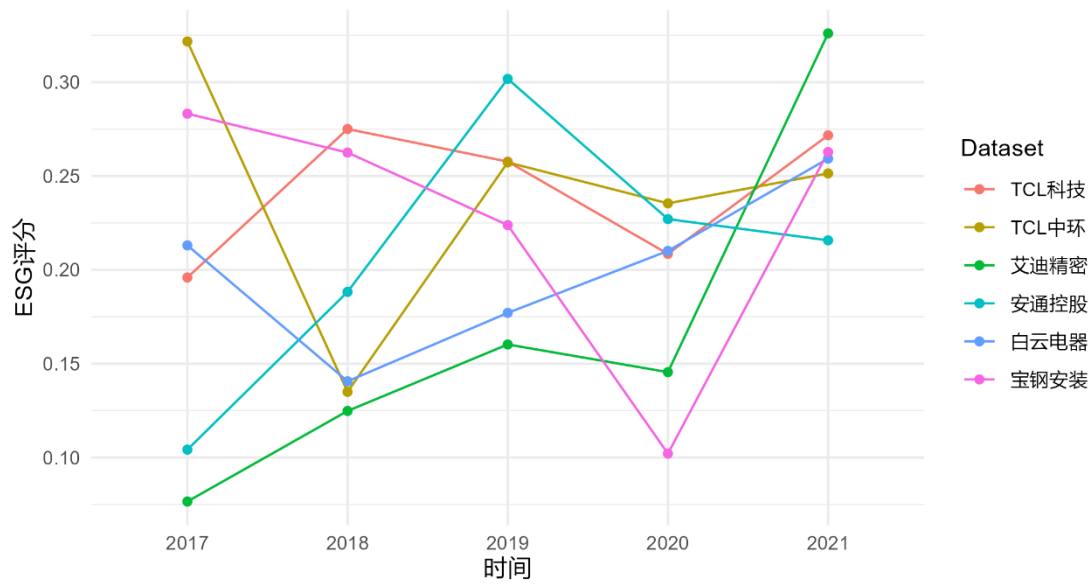


图 2 不同工业类企业的 ESG 评分随时间变化趋势

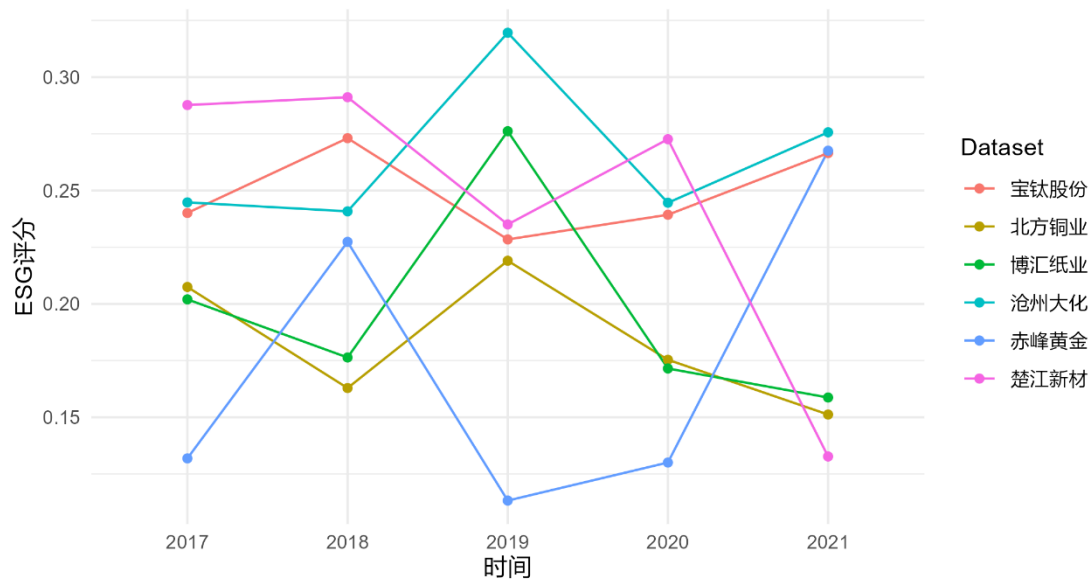


图 3 不同原材料类企业的 ESG 评分随时间变化趋势

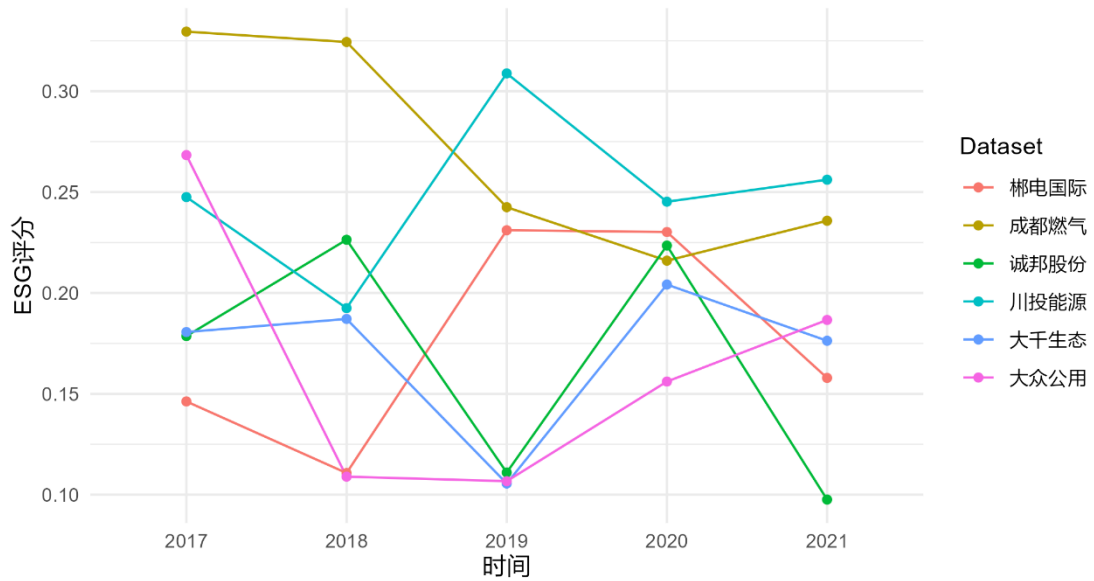


图 4 不同公用事业类企业的 ESG 评分随时间变化趋势

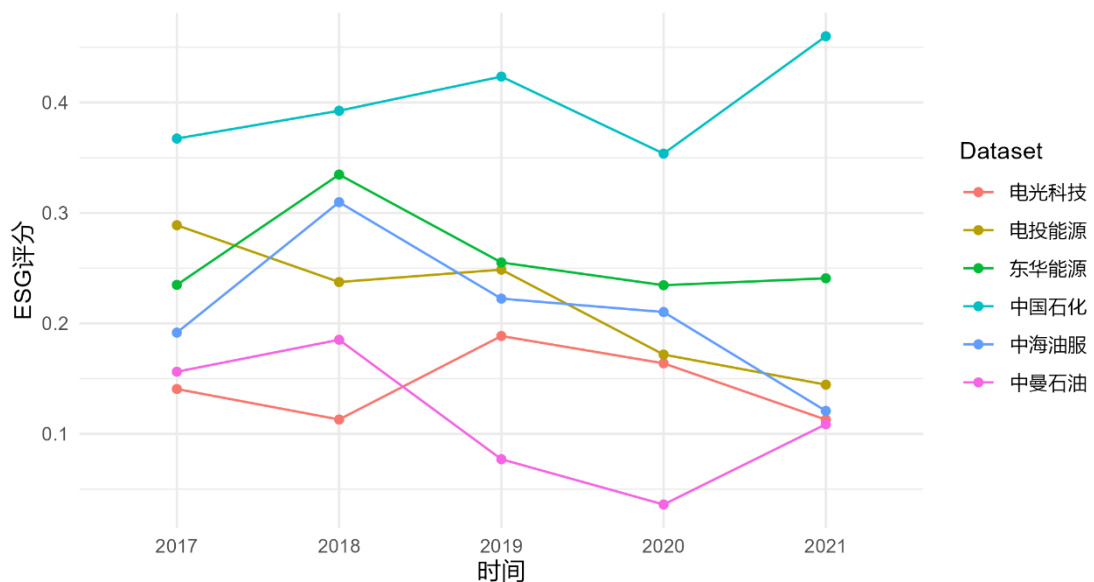


图 5 不同能源类企业的 ESG 评分随时间变化趋势

3 ESG 综合表现对中国能源类企业高质量发展的影响

在第 2 部分，我们获得了中国部分上市企业的 ESG 综合得分。在这一部分，我们建立回归模型，聚焦于能源类企业，探究 ESG 综合表现对企业高质量发展的影响。

3.1 变量选择和定义

结合已有文献（雒京华等，2020），选取企业总资产净利润率（ROA）作为企业高质量发展的指标。选取彭博资讯网的 ESG 评分变量为解释变量。选择 Size, Lev, Cashflow, ATO, FIXED, Growth, Indep 为控制变量，同时需要控制年份变量 Year, 具体变量含义如下表所示。

表 3 各变量名称、符号及定义

变量类型	变量名称	变量符号	变量定义
被解释变量	企业总资产净利润率	ROA	净利润/总资产平均余额
解释变量	ESG 综合评分	Esg	由第 2 部分得到的评分
控制变量	公司规模	Size	年总资产的自然对数
	资产负债率	Lev	年末总负债/年末总资产
	现金流比率	Cashflow	经营活动产生的现金流量净额/ 总资产
	总资产周转率	ATO	营业收入/平均资产总额
	固定资产占比	FIXED	固定资产净额/总资产
	营业收入增长率	Growth	本年营业收入/上一年营业收入 -1
	独立董事比例	Indep	独立董事除以董事人数

首先探究各变量的分布和相关性。被解释变量 ROA 的分布如图所示, 非正态分布, 为尖峰态, 可能导致后续线性回归时不满足基本的正态假设。Growth 变量右偏严重, 因为有几个取值较大的离群值。

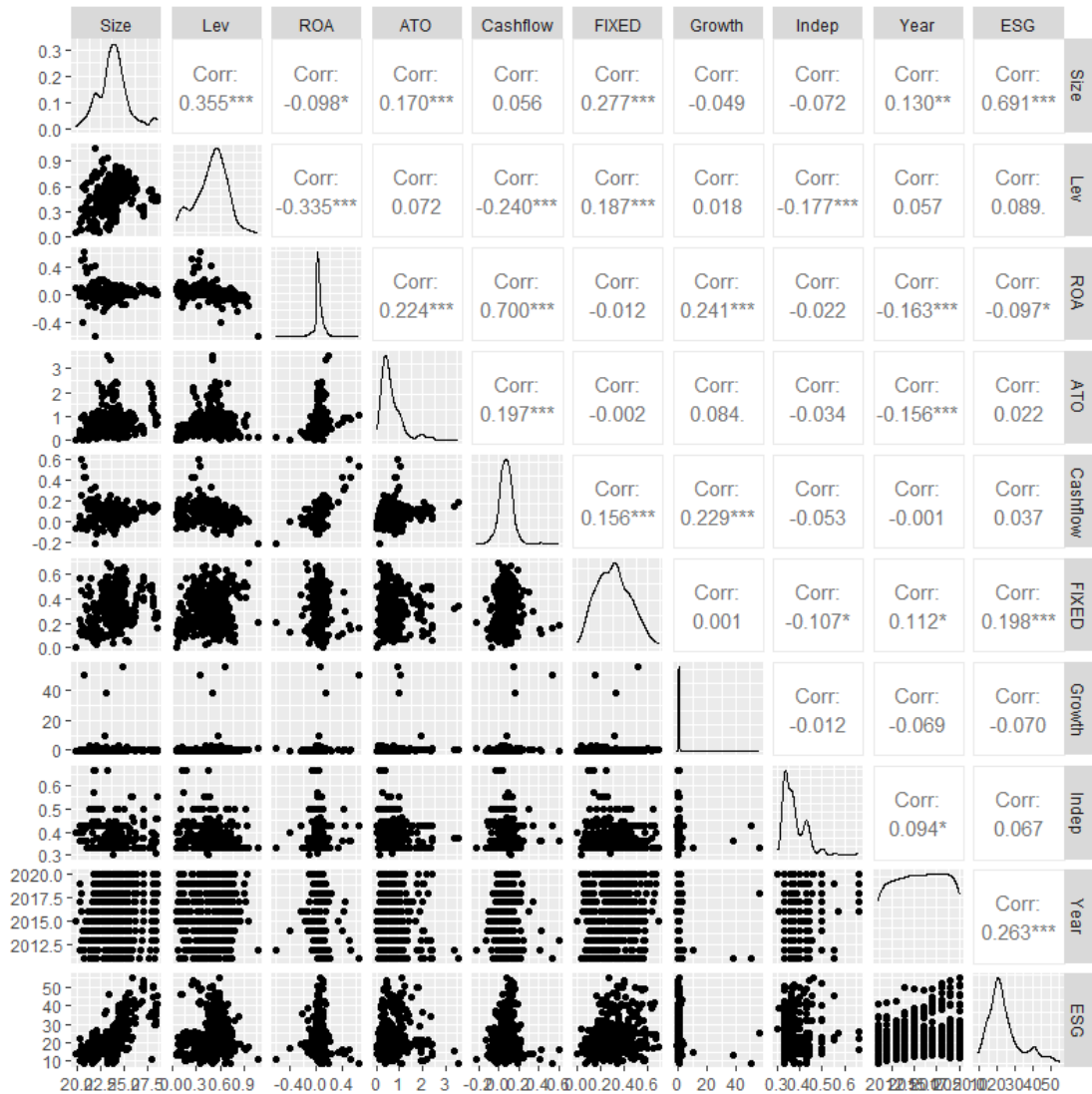


图 6 变量分布和相关性

变量之间的相关性可由相关系数热力图更直观地看出。由图可知，ROA 和 Cashflow、Lev 的相关系数较大。ESG 和 Size 间的相关系数较大，可能会在线性回归中导致共线性问题。

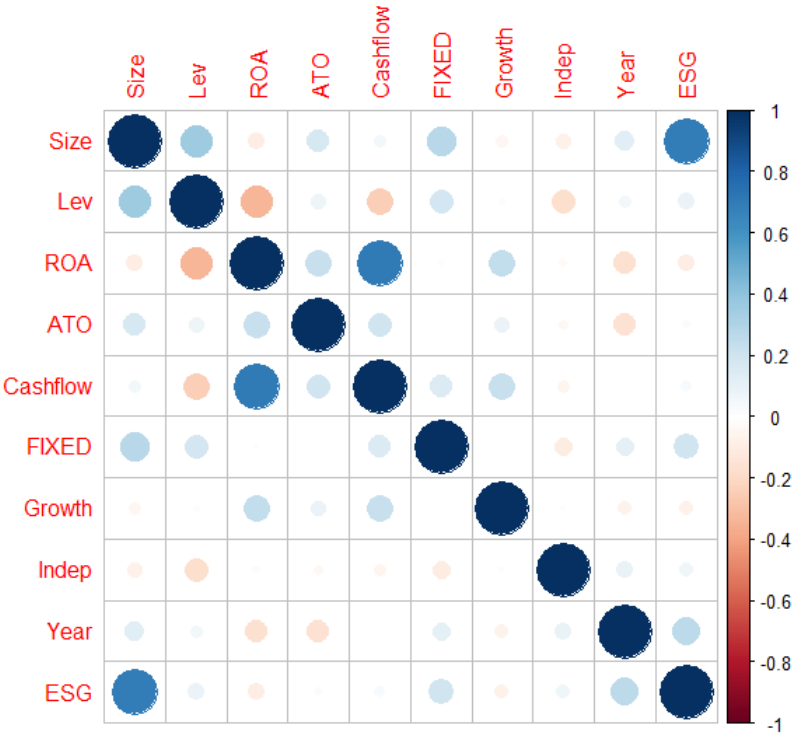


图 7 相关系数热力图

3.2 线性回归模型

为探究 ESG 对 ROA 的影响，首先建立最基础的线性回归模型：
Model1: $ROA \sim Size + Lev + ATO + Cashflow + FIXED + Growth + Indep + Year + ESG$
线性回归结果如下，发现 Lev, Cashflow 和 Year 表现出了较显著的影响，其余因子则作用较弱。ESG 变量的 t 检验的 p 值高达 0.4，不显著。R-squared 为 0.5652，Adjusted R-squared 为 0.5562。Size 和 Year 的系数为负，与直觉不符，因此计算各个变量的方差膨胀因子，观察是否有多重共线性。

表 4 Model1 回归结果

	Estimate	Std.Error	t.value	Pr.value
(Intercept)	6.888469	2.037534	3.380788	0.000788***
Size	-0.00171	0.00259	-0.66005	0.509572
Lev	-0.0725	0.017054	-4.25118	2.61E-05***
ATO	0.015869	0.00581	2.731558	0.006561**
Cashflow	0.709079	0.039699	17.8616	7.28E-54***
FIXED	-0.03167	0.020539	-1.54172	0.123871
Growth	0.001649	0.000695	2.371878	0.018134*
Indep	-0.01127	0.049905	-0.22576	0.821492
Year	-0.00338	0.001011	-3.345	0.000894***
ESG	-0.00037	0.000439	-0.83781	0.402601

表 5 各变量的方差膨胀因子

	Size	Lev	ATO	Cashflow	FIXED	Growth	Indep	Year	ESG
VIF	2.5002	1.4050	1.1281	1.2668	1.1536	1.0796	1.0724	1.1304	2.2049

已知若 VIF 最大超过 10 或平均值远超过 1 则存在严重共线性问题。由表可知，出了 ESG 和 Size 的 VIF 在 2 左右，其他 VIF 均在 1 左右，这也和之前的相关系数热力图中 Size 和 ESG 相关系数较大，其他变量间相关系数较小的结果相符合。

为判断是否需要删除部分控制变量，采用最优子集的方法利用 Cp 和 Adjusted R-squared 准则进行模型选择。发现二者均舍弃了 Size 和 Indep 变量。

Model2: $ROA \sim Lev + ATO + Cashflow + FIXED + Growth + Year + ESG$

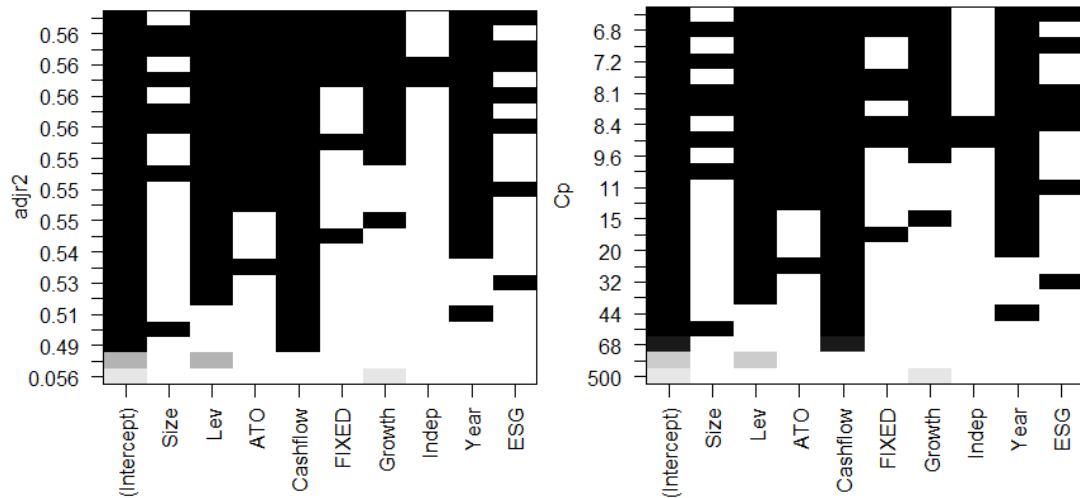
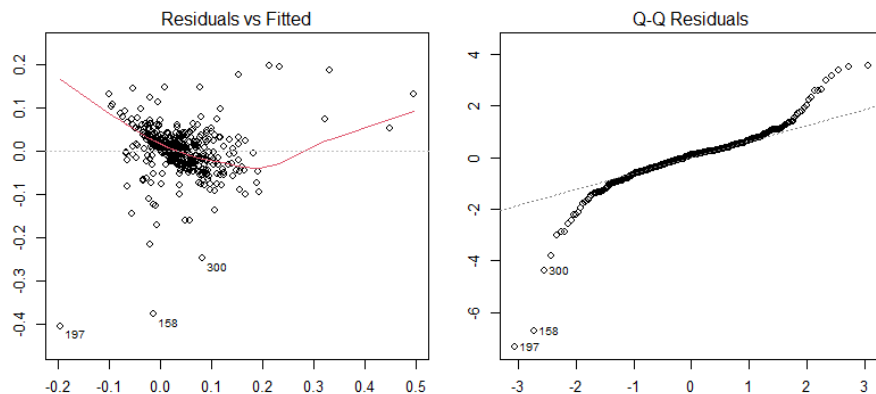


图 8 基于 Adjusted R-squared 和 Cp 准则的模型选择结果

对 Model2 进行线性回归，发现 R-squared 为 0.5648，Adjusted R-squared 为 0.5578，Adjusted R-squared 和 Model1 相差无几。Lev, ATO, Cashflow 和 Year 表现出了较显著的影响，其余因子则作用较弱。ESG 变量的 t 检验的 p 值为 0.066，虽相比 Model1 有大幅下降，但仍不显著。

再结合在已有文献中 Size 是理论上比较重要的变量，不能轻易删去，因此还是保留 Model1。

对 Model1 进行模型诊断。由残差图，存在非线性关系，可能有平方项，存在异方差问题；由 Q-Q 图可知，分布不满足正态，存在重尾现象。由 Cook's Distance 图，离群值不是很明显。



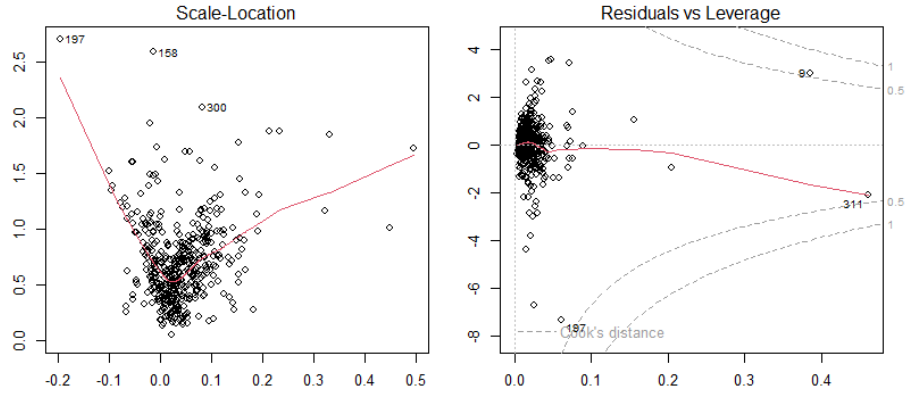


图 9 Model11 模型诊断图

采用 box-cox 变换，尝试解决这些问题。使用 R 中的 boxcox 函数要求被解释变量大于 0，因此以下讨论仅使用于 ROA 大于 0 即净利润大于 0 没有亏损的企业，样本中有 397 家（能源类企业总样本为 443 家）。

首先判断变换的 λ 值。由图结合解释性，可取 λ 为 0.5。因此构建模型：

$$\text{Model3: } \sqrt{\text{ROA}} \sim \text{Size} + \text{Lev} + \text{ATO} + \text{Cashflow} + \text{FIXED} + \text{Growth} + \text{Indep} + \text{Year} + \text{ESG}$$

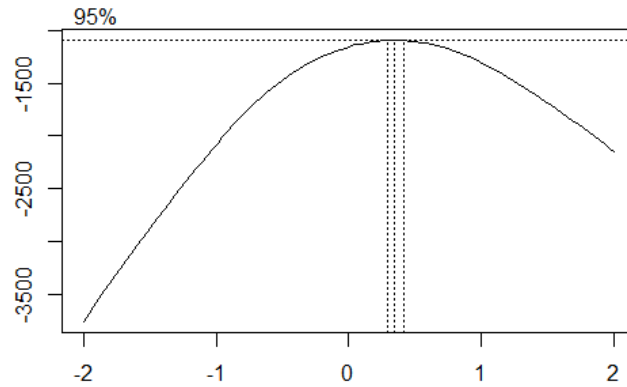
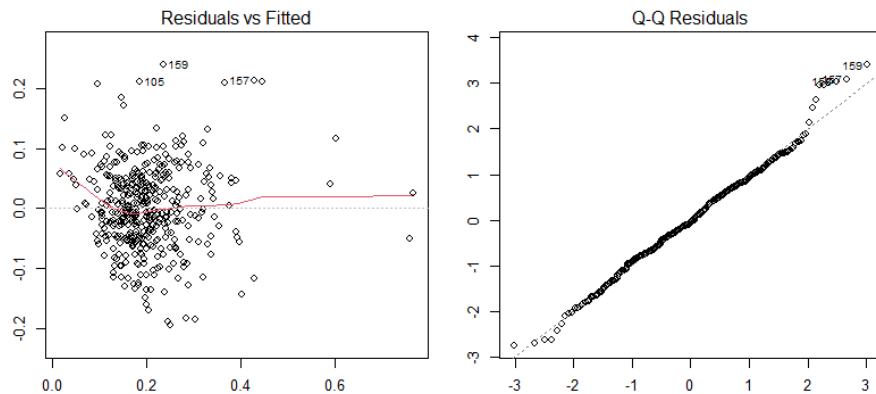


图 10 boxcox 变换的 λ 取值

对 Model13 进行模型诊断。由残差图，非线性问题和异方差问题得到了很大的改善；由 Q-Q 图可知，分布基本满足正态。由 Cook's Distance 图，离群值不是很明显。因此最终选定 Model13。



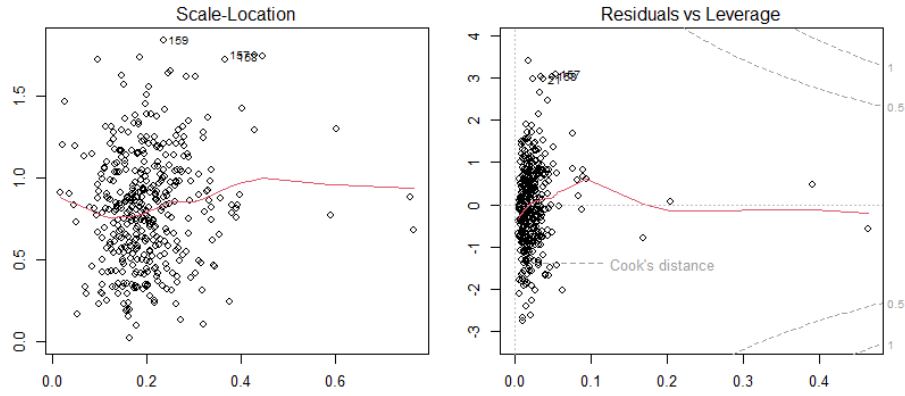


图 11 Model13 模型诊断图

Model13 各变量的系数值和 t 检验结果如下表所示。R-squared 为 0.6047, Adjusted R-squared 为 0.5955, 相比 Model11 均有所提高。Lev, Cashflow 和 Year 表现出了较显著的影响, 其余因子则作用较弱。ESG 变量的 t 检验的 p 值为 0.5, 仍不显著。

表 6 Model13 回归结果

	Estimate	Std.Error	t.value	Pr.value
(Intercept)	15.97041	2.636008	6.058559	3.26E-09***
Size	-0.01127	0.003437	-3.27777	0.001141**
Lev	-0.11694	0.02371	-4.93217	1.21E-06***
ATO	0.01731	0.007443	2.325544	0.020559*
Cashflow	0.916791	0.052156	17.578	2.89E-51***
FIXED	-0.01414	0.027518	-0.51375	0.607722
Growth	0.001285	0.000872	1.473278	0.141489
Indep	-0.11734	0.069928	-1.67803	0.094148
Year	-0.00768	0.001307	-5.87958	8.89E-09***
ESG	0.000374	0.000572	0.654339	0.513282

在以上尝试的模型中, ESG 对于 ROA 的影响均不显著, 因此本研究认为 ESG 综合指数对于企业的利润提高没有太大相关性。

为提高结果的稳健性, 我们采用 Model13 并将 ESG 换为本研究第 2 部分自行得到的 ESG_Score 进行回归分析。发现 R-squared 为 0.4806, Adjusted R-squared 为 0.4615, 相比 Model11 均有所提高。Lev, Cashflow 表现出了较显著的影响, 其余因子则作用较弱。ESG 变量的 t 检验的 p 值为 0.53, 仍不显著。

表 7 回归结果

	Estimate	Std.Error	t.value	Pr.value
(Intercept)	2.841054	6.722423	0.422624	0.672941
Size	-0.01009	0.004572	-2.20607	0.028308*
Lev	-0.11831	0.029433	-4.01983	7.76E-05***
ATO	-0.00898	0.010843	-0.82852	0.408185
Cashflow	0.965912	0.073425	13.15506	2.94E-30***
FIXED	0.033241	0.034939	0.951407	0.342335
Growth	0.000869	0.001322	0.657101	0.511733
Indep	-0.0217	0.096067	-0.22593	0.821441
Year	-0.00121	0.003333	-0.36226	0.717472
ESG_Score	0.056644	0.090819	0.623701	0.533405

3.3 固定时间效应模型

由于数据包含时间维度，为面板数据，因此考虑采用固定时间效应模型。

Model4: $ROA \sim Size + Lev + ATO + Cashflow + FIXED + Growth + Indep + ESG$ (Fixed Year)

对 Model4 进行回归分析,发现 R-squared 为 0.5344, Adjusted R-squared 为 0.4437,。Lev, Cashflow 和时间效应表现出了较显著的影响, 其余因子则作用较弱。ESG 变量的 t 检验的 p 值为 0.29, 不显著。

表 8 Model4 回归结果

	Estimate	Std.Error	t.value	Pr.value
Size	-0.00378	0.009449	-0.39971	0.689604
Lev	-0.17704	0.029726	-5.95567	6.03E-09***
ATO	0.028383	0.009282	3.057714	0.002393**
Cashflow	0.488682	0.046684	10.46792	1.24E-22***
FIXED	-0.02903	0.035937	-0.80787	0.419685
Growth	0.001652	0.000642	2.573629	0.010452*
Indep	0.042266	0.06071	0.696192	0.486746
Year2012	-0.02589	0.010867	-2.38271	0.01769*
Year2013	-0.0286	0.011123	-2.57081	0.010536*
Year2014	-0.04134	0.011586	-3.56826	0.000407***
Year2015	-0.05184	0.011979	-4.32788	1.94E-05***
Year2016	-0.06042	0.012042	-5.01727	8.15E-07***
Year2017	-0.04526	0.011988	-3.77527	0.000186***
Year2018	-0.05546	0.012392	-4.47595	1.01E-05***
Year2019	-0.0545	0.012987	-4.19655	3.4E-05***
Year2020	-0.05457	0.013351	-4.08728	5.36E-05***
ESG	0.00061	0.000571	1.066811	0.286753

3.4 分析与结论

我们采用了线性回归模型和固定时间效应模型,发现 ESG 对于 ROA 的影响均不显著。因此得出结论:企业的 ESG 表现对于高质量发展没有太大关联。但需要注意的是,虽然根据已有文献将 ROA 作为企业高质量发展的评价指标,但“高质量”的含义不应局限于经济收益上的“利润”,而需要包括社会责任和环境保护等多个方面,更符合新时代可持续发展的需要。

因此,“ESG 和高质量发展没有太大关联”的结论也是可解释的。能源类企业通常会对环境产生更多污染,因此 ESG 评分高多是因为对于能源类企业有更多环境方面的政策约束(例如控制排放量),或是能源类企业会更加重视环境保护。所以能源类企业并不是出于提高利润的收益角度去增强环境保护和社会责任方面的建设。

4 预测中国能源类企业在未来一年的 ESG 的变化情况

在这一部分,我们试图预测中国能源类企业在未来一年的 ESG 的变化情况。由于难以预测 2022 年控制变量或其他变量的情况,因此在综合考虑后选择采用最简单的线性回归模型,将第 2 部分的 ESG 评分关于时间进行回归,得到 2022 年的 ESG 评分预测。选取部分企业进行可视化,如下图所示。可以看到,采用最简单的线性回归模型可能会导致预测的 2022 年 ESG 评分趋于集中,在后续研究中可考虑使用时间序列模型等。

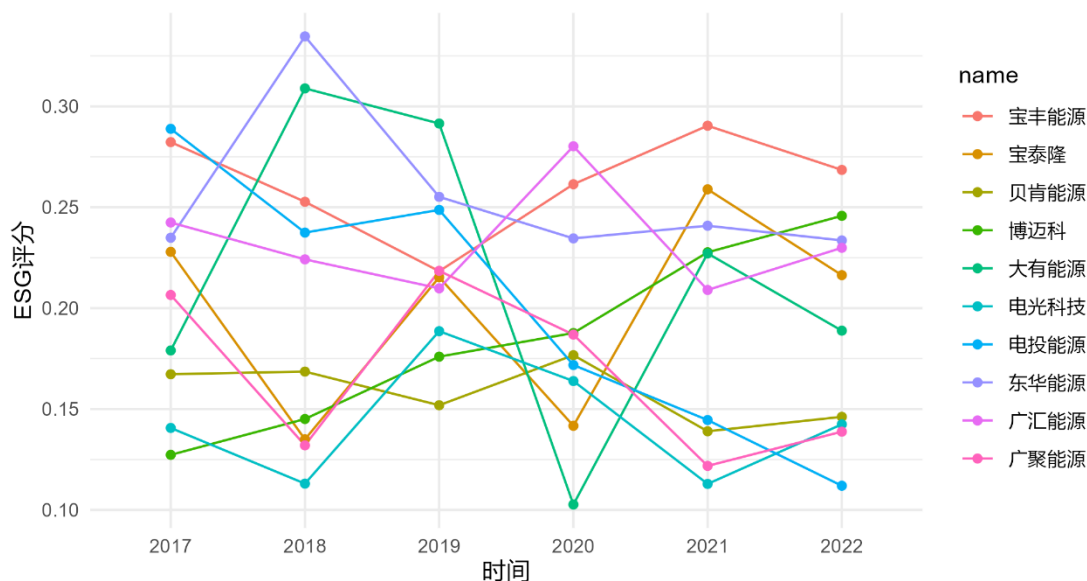


图 12 不同能源类企业 2022 年 ESG 评分预测

5 总结、政策建议与不足

本文通过使用 DPSIR 模型，构建包括驱动力、压力、状态、影响和响应五个维度的中国能源类上市企业的 ESG 综合评价指标体系，运用熵值法客观赋权，评价和分析 2017-2021 年我国上市企业的 ESG 发展状况。再聚焦于能源类企业，构建 ESG 评分对企业高质量发展的线性回归模型和时间固定效应模型，探究二者的关联。最后构建线性回归模型，对 2022 年中国上市能源类企业的 ESG 评分进行预测。

通过分析，本文得到以下主要结论和政策建议：

2017-2021 年，在 ESG 评分上，中国上市能源类企业表现好于原材料类、工业类、公用事业类企业。能源类企业 ESG 评分整体比较稳定且呈上升趋势，而其他产业的企业波动较大。这可能是因为能源类企业排放污染较多，对能源类企业在环境保护方面有较严格的政策，因此能源类企业对环境方面较为重视。建议政策在关注能源类企业的同时，规范本身对环境危害不大的企业。

2012-2020 年，ESG 评分数据和企业高质量发展间没有显著相关性，说明在短时间内 ESG 表现无法为企业经济收益带来显著提高，同时企业重视 ESG 评级的表现也更多是出于未来发展、社会责任和环境保护等方面的考虑。

本文上述分析也存在一定的问题。首先，在构建 ESG 评分模型选取指标时多关注环境部分的变量，而缺少了社会和治理部分；采用熵权法获得指标权重的方法过于客观，无法保证权重的合理性，需要结合专家意见进行主观调整。此外，在实证分析部分，只采用 ROA 作为企业高质量发展的评判指标，比较单一，且只关注了收益部分，后续可考虑采用其他被解释变量提高稳健性；考虑更多的控制变量再进行模型选择。最后，在预测 2022 年 ESG 评级数据时只采用了简单的线性回归，没有考虑其他变量的影响，较为粗糙，后续可考虑引入时间序列模型或神经网络模型。

6 参考文献

[1] 袁家海, 郭宇. 中国大型发电上市公司 ESG 评价体系开发与分值研究[J]. 中国环境管

- 理, 2018, 10(05):50-58. DOI:10.16868/j.cnki.1674-6252.2018.05.050.
- [2] 蔡泽栋. 宏观 ESG 综合评价体系构建与研究——以华东、华中、华南 10 地为例[J]. 经济研究导刊, 2020(25):58-62.
- [3] 彭百川, 张颖, 王治. 企业 ESG 表现对绿色创新效率的影响研究[J]. 统计与决策, 2024, 40(05):172-176. DOI:10.13546/j.cnki.tjyjc.2024.05.031.
- [4] 张宏, 王宇婷, 林慧. 内外双“管”下 ESG 表现对企业融资成本的影响研究[J]. 产业经济评论, 2024(01):41-56. DOI:10.19313/j.cnki.cn10-1223/f.2024.01.002.
- [5] 简冠群, 董小红, 杨忠海. 黄河流域资源依赖型企业 ESG 评价体系的构建及应用[J]. 财会月刊, 2024, 45(07):43-50. DOI:10.19641/j.cnki.42-1290/f.2024.07.006.
- [6] 杜帆, 李立国. 中国博士生教育规模增长预测分析——基于 1996—2018 年省际面板数据的实证研究[J]. 学位与研究生教育, 2020(06):55-63. DOI:10.16750/j.adge.2020.06.009.

7 附录

本研究所使用的 R 代码如下:

```
library(readxl)
# 读取所有子表格
all_sheets <- lapply(excel_sheets("DPSIR.xlsx"), function(sheet) {
  read_excel("DPSIR.xlsx", sheet = sheet)
})

# 给每个子表格赋予名称
names(all_sheets) <- excel_sheets("DPSIR.xlsx")

D = all_sheets$`D-驱动力`
a = D[D$产业=="工业",]
b = D[D$产业=="原材料",]
c = D[D$产业=="公用事业",]
d = D[D$产业=="能源",]
which.max(d$`企业市值(元)`)#d 333 ;a 5500; b 403; c 442

#####
DPSIR = cbind(all_sheets$`D-驱动力`,all_sheets$`P-压力`[,5:10],
              all_sheets$`S-状态`[,5:9],all_sheets$`I-影响`[,5:6],
              all_sheets$`R-响应`[,5:7])
m = 1825 #样本
n = 19 #指标
p = 5 #年数
df = DPSIR[,5:23]
#极差法标准化
df_normalized = as.data.frame(lapply(df, function(x) (x - min(x)) /
(max(x) - min(x))))
#信息熵
entropy_values = lapply(df_normalized, function(x) {
```

```

    p_ij <- x / sum(x)
    p_ij <- p_ij[p_ij > 0] # 移除概率为 0 的元素
    -sum(p_ij * log2(p_ij))
  })
#权重
weights <- lapply(entropy_values, function(x) {
  (1 - x) / sum(1 - unlist(entropy_values))
})
# 将列表转换为数值向量
weights <- as.numeric(unlist(weights))
# 计算综合评分
esg_scores <- as.matrix(df_normalized) %*% weights
# 将综合评分添加到数据框中
DPSIR_final <- cbind(DPSIR, ESG_Score = esg_scores)

#可视化
time_points <- c("2017", "2018", "2019", "2020", "2021") # 时间点
esg_scores1 <- esg_scores[8576:8580,] # 中国石油
esg_scores2 <- esg_scores[8531:8535,] # 中国建筑
esg_scores3 <- esg_scores[1361:1365,] # "福莱新材"
esg_scores4 <- esg_scores[6471:6475,] # "通源环境"

# 将评分和时间点组合成数据框
df_scores1 <- data.frame(Time = time_points, ESG_Score = esg_scores1,
  Dataset = "能源")
df_scores2 <- data.frame(Time = time_points, ESG_Score = esg_scores2,
  Dataset = "工业")
df_scores3 <- data.frame(Time = time_points, ESG_Score = esg_scores3,
  Dataset = "原材料")
df_scores4 <- data.frame(Time = time_points, ESG_Score = esg_scores4,
  Dataset = "公用事业")

# 将数据框合并为一个大的数据框
df_scores <- rbind(df_scores1, df_scores2, df_scores3, df_scores4)

library(ggplot2)
# 绘制折线图
ggplot(df_scores, aes(x = Time, y = ESG_Score, color = Dataset)) +
  geom_line(aes(group = Dataset)) + # 按数据集分组绘制线条
  geom_point() +
  labs(x = "时间", y = "ESG 评分", title = "不同产业的 ESG 评分随时间的变化趋势") + # 设置坐标轴标签和标题
  theme_minimal() # 使用简洁的主题

```

```

ggsave("ESG_score.png", width = 7, height = 4, units = "in",bg="white")

#####
library(readxl)
control = read_xlsx("Control.xlsx")
pbESG = read_xls("彭博 ESG.xls")
# 能源
power = control[control$Industry=="B",c(3,5,6,7,9,10,13,14,17,37)]#多证
券代码 27,28,37
power = na.omit(power)
names(power)[names(power) == "证券代码"] <- "code"
power_esg = pbESG[grepl("B",pbESG$"行业代码"),]
names(power_esg)[names(power_esg) == "year"] <- "Year"
power_esg = power_esg[,1:3]
power_esg$Year = as.numeric(power_esg$Year)
#合并 control 和 pbesg
library(dplyr)
# 使用 inner_join 函数找出两个 dataframe 中完全相同的行
common_rows <- power %>%
  inner_join(power_esg, by = c("code", "Year"))
#合并 control 和 esg
names(DPSIR_final)[names(DPSIR_final) == "证券代码"] <- "code"
names(DPSIR_final)[names(DPSIR_final) == "年份"] <- "Year"
common_rows2 = power %>% inner_join(DPSIR_final[,c(1,2,24)],by =
c("code","Year"))

#变量
library(corrplot)
corrplot(cor(common_rows[, -1]))

#回归
fit1 = lm(ROA~.,data = common_rows[,c(-1)])
par(mfrow=c(2,2), mar=c(2,2,2,2))
plot(fit1)

library(car)
vif_values = vif(fit1)

#固定效应
library(plm)
fit2 = plm(ROA~.,data = common_rows,index=c("code","Year"))
fit3 = plm(sqrt(ROA)~.,data = power_boxcox,index=c("code","Year"))

rows2020 = common_rows2[common_rows2$Year==2020,]

```



```
fit3 = lm(ROA~.,rows2020[,c(-1,-12,-10)])
```

#boxcox

```
library(MASS)
power_boxcox = common_rows[common_rows$ROA>0,]
power_boxcox2 = common_rows2[common_rows2$ROA>0,]
fit4 = boxcox(ROA~.,data=power_boxcox[,c(-1)])#找 lambda
fit5 = lm(sqrt(ROA)~.,data = power_boxcox[,c(-1)])
fit6 = lm(sqrt(ROA)~.,data = power_boxcox[c(-107,-60,-61,-9),c(-1)])
fit7 = lm(ROA^(1/4)~.,data = power_boxcox2[,c(-1)])
```

```
fit = lm(ROA~.,data = common_rows[,c(-1,-2,-9)])
```

#存入xlsx

```
library(xlsx)
write.xlsx(vif_values,"output.xlsx")
write.xlsx(summary(fit2)$coefficients,"output.xlsx")
```

选择的最佳模型

```
library(leaps)
leapSet = leaps(x=common_rows[,c(-1,-4)],y=common_rows$ROA,nbest = 3,method = "adjr2")
leapSet$which[which.max(leapSet$adjr2),]
leapSet2 = leaps(x=common_rows[,c(-1,-4)],y=common_rows$ROA,nbest = 3,method = "Cp")
leapSet2$which[which.min(leapSet2$Cp),]

leapSet3 = regsubsets(x=common_rows[,c(-1,-4)],y=common_rows$ROA,nbest = 3)
```

公用

```
public = control[control$Industry=="N",c(3,5,6,7,9,10,13,14,17,37)]#多证  
券代码 27,28,37
public = na.omit(public)
names(public)[names(public) == "证券代码"] <- "code"
common_rows3 = public %>% inner_join(DPSIR_final[,c(1,2,24)],by =  
c("code","Year"))
fitp = lm(ROA~.,data = common_rows3[,c(-1)])
```

#原材料

```
public = control[control$Industry=="C",c(3,5,6,7,9,10,13,14,17,37)]#多证  
券代码 27,28,37
public = na.omit(public)
names(public)[names(public) == "证券代码"] <- "code"
```

```

common_rows3 = public %>% inner_join(DPSIR_final[,c(1,2,24)],by =
c("code","Year"))
fitp = lm(ROA~.,data = common_rows3[,c(-1)])

#####
predict_power = DPSIR_final[DPSIR_final$ 产业==" 能源 ",c("Year"," 名称
","ESG_Score")]
# 加载 tidyr 包
library(tidyr)

# 使用 pivot_wider 函数转换数据框
df_wide <- predict_power %>%
  pivot_wider(
    names_from = Year, # 使用 time 列作为新列的名字
    values_from = ESG_Score # 使用 score 列作为新列的值
  )
time = 2017:2021
df_wide$"2022" = NA
for(i in 1:72){
  model = lm(as.numeric(df_wide[i,c(-1,-7)])~time)
  new_data <- data.frame(time = c(2022))
  predicted_scores <- predict(model, newdata = new_data)
  df_wide$"2022"[i] = predicted_scores
}

# 使用 pivot_longer 函数逆转转换
df_original <- df_wide %>%
  pivot_longer(
    cols = -"名称",
    names_to = "Year",
    values_to = "ESG_Score"
  )
colnames(df_original) = c("name","Year","ESG_Score")
ggplot(df_original[1:60,], aes(x = Year, y = ESG_Score, color = name)) +
  geom_line(aes(group = name)) +
  geom_point() +
  labs(x = "时间", y = "ESG 评分", title = "不同能源类企业的 ESG 评分 2022 年
预测") +
  theme_minimal()
ggsave("predict.png", width = 7, height = 4, units = "in",bg="white")

```