

基于时间序列回归与主成分分析的绿色发展预测及评估

摘要

“稻花香里说丰年，听取蛙声一片。”寥寥几句诗，一片自然和谐共生的景象浮现脑海。在人类历史发展进程中，人们越来越清晰认识到，经济社会快速发展决不能以环境的破坏、资源的浪费为代价。在这一背景下，本文讨论以下三个问题：（1）预测中国在 2025-2035 年内的森林面积。（2）构建数学模型分析退耕还林政策对国家工业水平和国民经济的影响（3）对世界各国经济水平、工业水平和森林植被面积进行评价。

对于问题一，我们通过箱图对附件中数据进行异常值检验发现未出现异常值。在此基础上基于时间序列模型首先通过时序图、自相关图、Daniel 检验进行数据平稳性检验，结果显示附件数据非平稳。随后根据这一结果进行一次、二次差分构建平稳序列并检验其平稳性，其中二次差分显示平稳。再根据数据的自相关图与偏相关图特征选择参数并通过 SPSS24.0 建立 ARIMA (2, 2, 1) 模型，并基于模型残差图对模型稳定性进行检验，结果显示模型平稳。在此基础上我们用 SPSS24.0 对结果进行预测估计，并将部分数据预测结果与实际数据进行比较，结果显示数据模拟结果与附件已知实际数据结果拟合度极高。最后，我们在所得估计值基础上对 2025-2035 中国森林面积（平方公里）预测值进行 95%置信区间估计，以考虑进其它因素导致的不可见误差来使得估计结果更为准确可信。

对于问题二，先用拉以达准则算法剔除异常数据，再对剔除后的数据作统计分析。随后，将就业效益和潜在效益作为衡量退耕还林对国家工业水平影响的指标，将经济效益和生物多样性作为衡量退耕还林对国民经济影响的指标，并在此基础上构建相关数学表达式。接着，对所构建数学表达式中的参数进行讨论与确定。将所搜集历年数据求平均值后代入表达式，在此基础上得到退耕还林对国家工业水平和国民经济影响的标准值。再将历年数据代入表达式，将得到的数值与标准值进行比较，进而得到不同年份退耕还林对国家工业水平和国民经济影响的折合成绩。对结果进行分析可知：（1）退耕还林政策的实施对带动第二、三产业就业水平影响总体呈现上升态势。（2）退耕还林政策对国家工业水平的潜在价值影响在稳步增长中呈现一定周期性波动规律。（3）退耕还林政策的实施给国家工业水平发展注入了更多动力与活力。（4）退耕还林政策实施有力提升了经济效益。（5）森林面积增长与生物多样性增长呈现边际效应递减。（6）自 2003 年退耕还林政策实施后，国民经济发展态势愈发向好。

对于问题三，经过权衡，选取四个指标——二氧化碳排放量、GDP、工业增加值、森林面积，对其 1990-2020 年数据的平均值进行主成分分析，得到综合评价结果。随后，绘制综合评价结果图，以直观显示综合评价结果。最终，对结果分析可知：（1）综合评价结果中排名第一、第二位国家分别为中国、美国。（2）中低收入国家综合评价结果普遍低于中高等国家。这可能是因为低收入国家目前亟需解决温饱、医疗等基本社会问题，其 GDP 还未达到理想发展水平。（3）欧洲大部分国家都拥有较高的综合评价结果。

对问题二中森林面积和生物多样性的关系式进行灵敏度分析，当 β 的值每改变 0.3（相当于 5.7%左右），其结果变化小于 3%，体现出模型较稳定。

关键词：平稳性 Daniel 检验，时间序列回归，主成分分析，灵敏度分析

一、问题重述

近年来，全球范围内植被大量破坏，对生态造成了极大的负面影响。世界各国大力发展工业科技，碳排放量也与日俱增，而减少的植被面积又导致森林固碳能力下降，解决生态问题刻不容缓。习近平总书记说，“绿水青山就是金山银山”。保护环境对国家发展具有重要意义。相对于世界水平，我国的植被面积却在逐步升高。

“节能减排”和发展绿色经济作为国家重大发展战略，习近平主席也在第七十五届联合国大会一般性辩论上宣布，“中国将提高国家自主贡献力度，采取更加有力的政策和措施，二氧化碳排放力争于 2030 年前达到峰值，努力争取 2060 年前实现碳中和”。与此同时，根据过去中国森林面积数据预测 2025-2035 森林面积将有助于“碳达峰、碳中和”政策更好落地实施、为政策制定者提供未来数据预测支持，有助于其根据未来预测数据，调整目前相关政策力度、方向、手段，最终助力推动经济社会绿色转型与系统性深刻变革。

我们团队要解决以下问题：

- （一）根据给定数据，预测我国 2025-2035 的森林面积。
- （二）分析退耕还林政策对我国工业水平和国民经济的影响。
- （三）研究分析世界各国的工业水平、经济水平、森林面积三项指标的联系。
- （四）撰写一份报告书递交联合国，内容包括我们团队的发现和环境问题解决措施。

二、问题分析

（一）问题一的分析

问题 1 属于预测类数学问题，就此类问题特征而言，即要求在观察历史数据变化规律基础上预测未来的值及预测值可能存在的区间，其重要前提假设是事物的过去能够在正常发展情况下延续到未来，符合时间序列分析特点。

附件中所提供数据是各时间点上形成的数值序列，且题目要求预测 2025-2035 年内的森林面积，即 2025—2035 年间各年的森林面积（平方公里）。

基于以上原因，我们将首先创建时间序列，建立时间序列模型。随后，对比已有预测值与实际值并在此基础上对 2025—2035 年间各年的森林面积（平方公里）展开预测。

（二）问题二的分析

该问题要求构建数学模型分析退耕还林政策对国家工业水平和国民经济的影响。本题的要点在于评价指标的选定与参数值的选定。考虑到自行查找的数据可能存在异常值，因此首先需要对原始数据进行特异值的检验及剔除，然后在新得到的数据上进行分析。用就业效益和潜在效益来衡量退耕还林对国家工业水平的影响。用经济效益和生物多样性来衡量退耕还林对国民经济的影响。通过构造数学表达式，将搜集到的数据代入表达式，进而计算折合成绩，分析退耕还林对国家工业水平和国民经济的影响。

（三）问题三的分析

该问题属于评价类问题，可以对附件的部分数据进行主成分分析，得到综合评价结果。然后对所得结果进行分析。可以选取二氧化碳排放量、GDP、工业增加值、森林面积四个指标，取各指标 1990-2020 年数据的平均值进行主成分分析。通过绘图的方法使综合评价结果更加直观。

三、模型假设

- 1. 假设附件所给数据经二阶差分处理后是完全平稳的
- 2. 假设题目所给数据真实可靠
- 3. 假设预测的未来时间段无重大事故
- 4. 假设搜集到的数据真实可靠

四、定义与符号说明

符号	定义	单位
d_1	流向一产业的劳动力的占比	
d_2	流向二产业的劳动力的占比	
d_3	流向三产业的劳动力的占比	
x_i	第一产业的贡献率	
y_i	第二产业的贡献率	
z_i	第三产业的贡献率	
X_1	农业总产值	亿元
X_2	林业总产值	亿元
D	工业增加值	
d	工业增加值的平均值	

五、模型的建立与求解

第一部分：问题 1 的模型

问题 1 的思路流程图如下
思路流程图

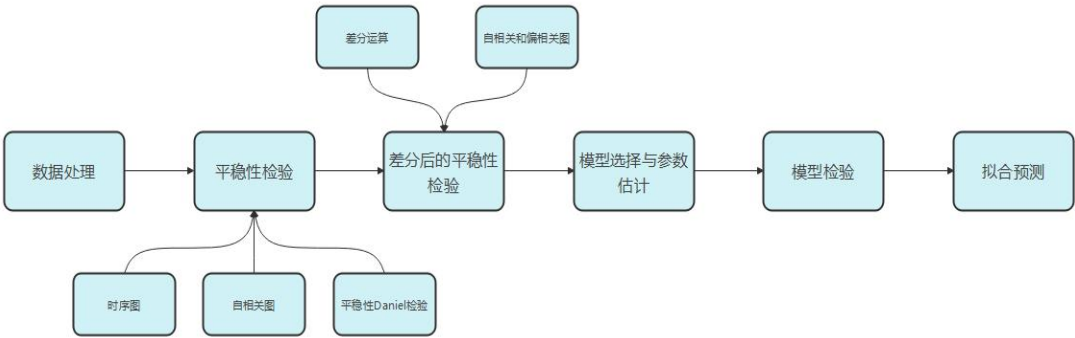


图 1 思路流程图

- (一) 数据处理
- (1) 检测是否存在特异值

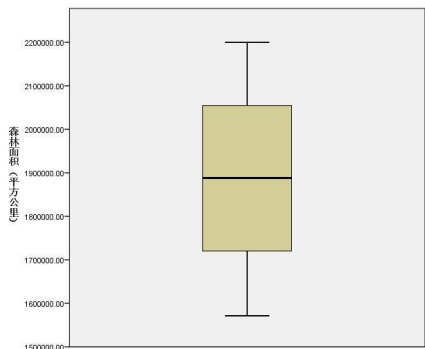


图 2 原始数据箱图

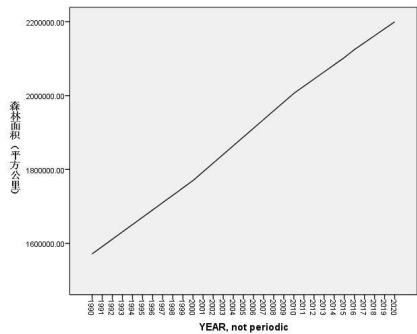


图 3 1990-2020 中国森林面积（平方公里）时间序列图

根据 SPSS24.0 对所给数据的箱图处理结果来看并未显示出特异值,因此判定不存在特异值。

(二) 平稳性检验

1. 时序图

由图 3 可知, 1990-2020 年中国森林面积（平方公里）呈上升趋势, 由 1990 年的 157.1 万平方公里上升到 2020 年的 220.0 万平方公里（见上数据展示）, 年平均增长率为 1.13%。所以它有很大可能不是平稳序列。

2. 自相关图

从图 6 可以看出, 在 7 阶后才落入区间内, 并且自相关系数长期大于 0, 显示出很强的自相关性。

3. 平稳性 Daniel 检验

相关性				
		森林面积（平方公里）		年份
斯皮尔曼 Rho	森林面积（平方公里）	相关系数	1.000	1.000**
		显著性（双尾）	.	.
		个案数	31	31
	年份	相关系数	1.000**	1.000
		显著性（双尾）	.	.
		个案数	31	31

**、在 0.01 级别（双尾），相关性显著。

由此可知对 $\alpha = 0.01$. 由

$$T = \frac{q_s \sqrt{n-2}}{\sqrt{1-q_s^2}} \tag{1}$$

由上表知 $q_s = 1$, 计算得统计量 $T = +\infty$, 由于上 $\alpha/2$ 分位数的值 $t_{\alpha/2}(29) = 2.0452$, 所以 $|T| > t_{\alpha/2}(n-2)$, 故认为序列是非平稳的; 且鉴于 $q_s > 0$, 因此序列有上升趋势。

(三) 差分后的平稳性检验

1. 差分运算

鉴于上述平稳性检验结果为数据非平稳, 因此我们拟对该序列作差分运算以构造平稳序列。差分运算即后一时间点减去当前时间如 $y_{(t)} - y_{(t-1)}$, 用 D 表示, 定义为 $Dy_{(t)} = y_{(t)} - y_{(t-1)}$ 。

那么 k 阶差分可表示为： $y_{(t)} - y_{(t-k)} = D_{(k)} * y_{(t)} = (1 - L^k) * y_{(t)} = y_{(t)} - (L^k) * y_{(t)}$ ，L 为滞后算子，定义为 $L * y_{(t)} = y_{(t-1)}$ ，则 k 阶之后算子定义为 $(L^k) * y_{(t)} = y_{(t-k)}$ 。

首先对附件中相应经处理的时间序列进行 1 阶差分运算，得出如下图所示的趋势图。

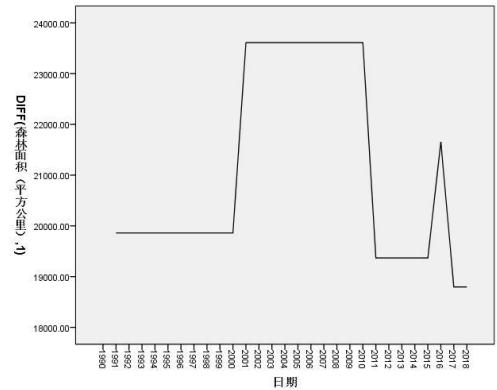


图 4 中国森林面积一阶差分时间序列图

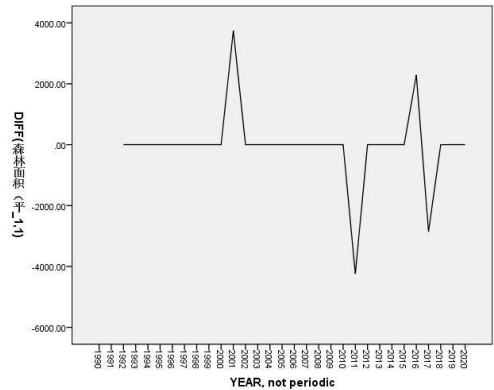


图 5 中国森林面积二阶差分时间序列图

上图显示，1 阶差分处理后的数据增减趋势较原始数据更为平稳，但是未达到进行时间序列分析所需平稳性要求，因此对 1 阶差分后的时间序列进行 2 阶差分运算。

由图可知，较一阶相比，二阶差分运算后的时间序列符合进行时间序列分析所需平稳性要求，故时间序列分析相关模型中参数 d 取 2。

2. 自相关图和偏相关图

在确定差分阶数 d 之后，接下来对时间序列模型中参数 p 与 q 进行选择。由于时间序列的自相关系数与偏自相关系数可以帮助判断参数 p 与 q，因此绘制数据的自相关系数图与偏相关系数图如下。

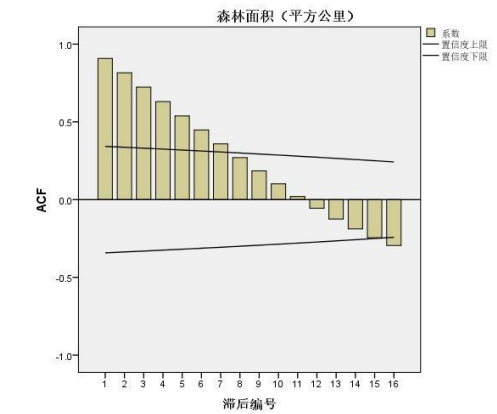


图 6 中国森林面积（平方公里）自相关图

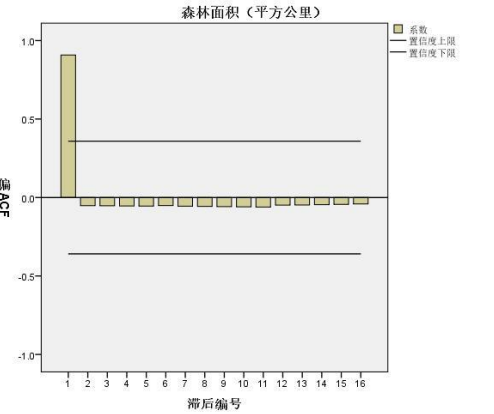


图 7 中国森林面积（平方公里）偏相关图

由图像可知自相关函数为拖尾，偏自相关函数为截尾，因此选择 ARIMA(1) 模型。

（四）模型选择与参数估计

选择 ARIMA (2, 2, 1) 模型进行预测估计。

模型统计							
模型	预测变量数	模型拟合度统计		杨-博克斯 Q(18)			
		平稳 R 方	R 方	统计	DF	显著性	离群值数
森林面积（平方公里）	1	.149	1.000	15.990	15	.383	0

ARIMA 模型参数								
				估算	标准误差	t	显著性	
森林面积 （平方公里）-模型 _1	森林面积（平方公里）	不转换	常量		47945.887	35151.495	1.364	.185
			AR	延迟 1	.686	.316	2.174	.000
				延迟 2	.033	.239	.138	.001
			差异		2			
			MA	延迟 1	.998	10.312	.097	.924
	年份	不转换	分子	延迟 0	-23.915	17.521	-1.365	.185

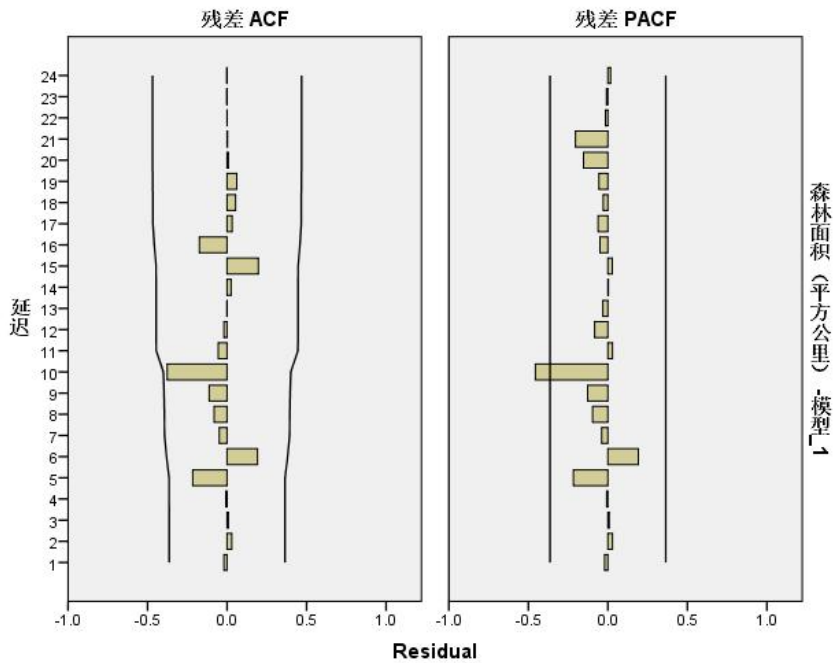


图8 自相关与偏相关残差分析图

由以上残差的ACF和PACF图，可以判断模型基本平稳，因此ARIMA (2, 2, 1)是合理、稳定的。
由上述ARIMA模型参数表可知，ARIMA模型结果为：

$$\Delta X_t = 47945.887 + 0.686X_{t-1} + 0.033X_{t-2} + \varepsilon_t \tag{2}$$

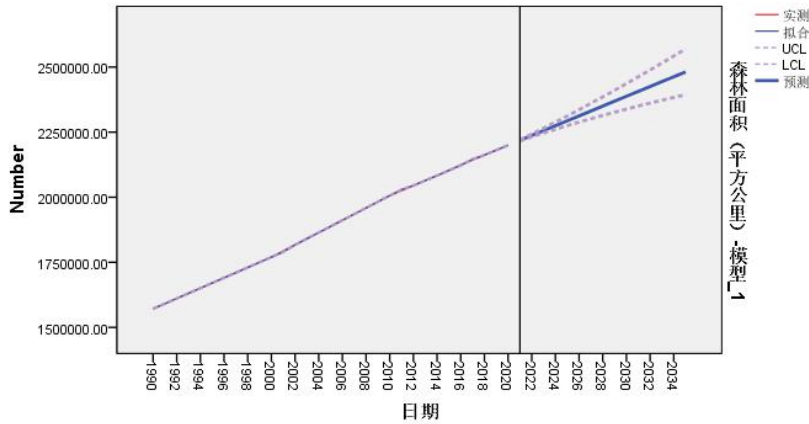


图9 2025-2035中国森林面积（平方公里）预测及结果95%置信区间图

最后进行拟合预测，可以看到拟合效果很好。
因此

表1 2025-2035中国森林面积（平方公里）预测结果表

年份	预测值	95%置信区间
2025	2293760.30	(2261272.30, 2288656.90)
2026	2312556.00	(2275229.67, 2312290.93)
2027	2331351.70	(2288728.08, 2336383.92)
2028	2350147.40	(2301803.95, 2360899.45)
2029	2368943.10	(2314486.27, 2385808.53)
2030	2387738.80	(2326798.79, 2411087.41)
2031	2406534.50	(2338761.43, 2436716.17)
2032	2425330.20	(2350391.20, 2462677.80)
2033	2444125.90	(2361702.87, 2488957.53)
2034	2462921.60	(2372709.40, 2515542.40)
2035	2481717.30	(2383422.28, 2542420.92)

第二部分：问题 2 的模型

建立退耕还林对国家工业水平和国民经济的影响的表达式，把不同年份的数据代入表达式折合成某个标准级别的成绩，对折合成绩进行比较，得出退耕还林政策对国家工业水平和国民经济的影响。

思路流程图：

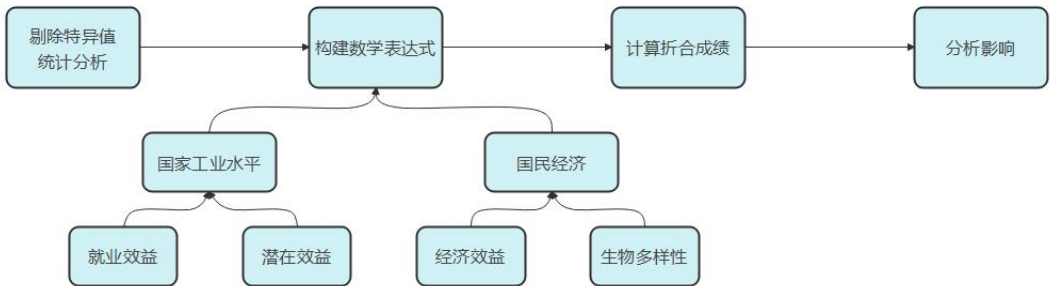


图 10 思路流程图

（一）数据处理

为了提高数据的质量，采用拉以达准则算法进行异常数据的剔除。以定置信概率为 99.7%

为标准，以三倍测量列标准偏差极限为依据，凡超过此界限的误差，就认为它不属于随机误差的范畴，而是粗大误差，即需要被剔除的异常数据。

用 SPSS 对剔除前后的样本分别作统计分析，得到的描述统计量如表 10 所示（见附录）。

（二）对国家工业水平的影响的表达式的构建

1. 就业效益

退耕还林（草）是一场土地利用方式的变革。由种粮食到植树种草，土地的利用性质发生了变化，必然带来生产要素的流动和劳动力投向的变化。实施退耕还林工程解放了农村劳动力，推动了第二、三产业的发展。

$$M_1 = k * (d_1 * x_i + d_2 * y_i + d_3 * z_i) \quad (3)$$

$$d_1 + d_2 + d_3 = 1 \quad (4)$$

式中， k 为系数， d_1 、 d_2 、 d_3 分别为流向一、二、三产业的劳动力的占比， x_i 、 y_i 、 z_i 分别表示一、二、三产业的贡献率（贡献率指各产业增加值增量与 GDP 增量之比）。

2. 潜在效益

我们认为退耕还林对国家工业水平的潜在价值主要与工业增加值 D （占 GDP 的百分比）有关。 D 越大，潜在效益越强。并认为，当 $D=1$ 时，潜在效益 M_2 最大，为 2。为此，我们构建了如下潜在价值的数学表达式：

$$M_2 = e^{\frac{D}{d/r}} \quad (5)$$

式中， D 为工业增加值， d 为 1980–2020 年工业增加值的平均值， r 为修正因子。

用偏好性系数加权法将多目标问题转化为单目标问题，其表达式为：

$$M = k_1 M_1 + k_2 M_2 \quad (6)$$

这里表达式体现了综合体现了退耕还林政策在就业效益和潜在效益方面对国家工业水平的影响，我们称之为退耕还林政策对国家工业水平的影响。这里的 k_1 、 k_2 有两方面的含义：一是作为偏好性加权系数，二是充当修正系数的作用，以保证就业效益和潜在效益的两个表达式的量化值在数值上近似相等。

（三）对国民经济的影响的表达式的构建

（1）经济效益

提高经济效益有利于增强相关企业市场竞争力、提高人民生活水平、增强综合国力，因此经济效益在某种程度上是一切经济活动的根本出发点，对国民经济现在、未来的发展具有重要影响。鉴于此，为了形象地描述退耕还林的经济效益，我们利用指数函数的定义对二者关系进行描述。

$$M_3 = b_1 * X_1^{b1} + b_2 * X_2^{b2} \quad (7)$$

$$b_1 + b_2 = 1 \quad (8)$$

式中， b_1 、 b_2 、 $b1$ 、 $b2$ 为系数， X_1 为农业总产值（亿元）， X_2 为林业总产值（亿元）。

（2）生物多样性

日前联合国环境规划署（UNEP）在国际期刊 Environmental Science 发文提出将生物多样性纳入国民经济核算。将生物多样性指标与国民经济账户联系起来，提供了将生物多样性纳入经济规划和监测进程的主流手段。同时，国民经济体系（SNA）涵盖所有与生产、消费和积累相关的经济活动以及所有工业部门，但是未能充分考虑到环境。SEEA EEA 旨在通

过扩展 SNA 来解决这一问题，以提供关于生态系统及其如何促进经济和福祉的综合统计数据。2021 年中共中央办公厅国务院办公厅印发《关于进一步加强生物多样性保护的意见》也显示了生物多样性对国民经济绿色发展的重要性。纵观国内外生物多样性相关科研、政策现状，我们知道生物多样性对国民经济的健康绿色发展有着重要影响。因此我们拟采用生物多样性这一指标解释其对国民经济的影响，且此种做法具有一定合理与可行性。

我们认为，对生物的多样性的影响主要由森林面积 S （占土地面积的百分比）决定，令生物多样性的值为 M_4 ，我们通过建立 S 和 M_4 的函数表达式，来对该问题进行细化量化分析。

为了便于求解，我们用极差变换：

$$x'_{ij} = \frac{x_{ij} - \min_{1 \leq i \leq n} \{x_{ij}\}}{\max_{1 \leq i \leq n} \{x_{ij}\} - \min_{1 \leq i \leq n} \{x_{ij}\}} \quad (9)$$

将不同 S 对应的 M_i 的值标准化，则 M_i 介于 $[0, 1]$ 。

令 S 为 x ， M_i 为 $y=f(x)$ ，其中， $y \in [0, 1)$ 。 y 越接近于 1，说明生物多样性越强。 y 与 x 应满足下列条件：

1. 森面面积越大，生物多样性越强，即 y 随着 x 的增加而增加
2. 森林面积为 0 时，生物多样性的初始值为一个位于区间 $(0, 1)$ 的常数，即 $f(0)=y_0 \in (0, 1)$
3. 当森面面积较小时，生物多样性的增加大致与森面面积的增加成正比
4. 当森林面积较大，渐进趋向于正无穷时，生物多样性将趋于 1

我们选取的一个符合上述条件的指数增长模型如下：

$$y = f(x) = (1 - y_0)(1 - e^{-\beta x}) + y_0, y_0 \in (0, 1) \quad (10)$$

我们取定生物多样性的初始值为 0.18，即在上述表达式中，取 $y_0=0.18$ 。结合实际情况，我们估计 β 的范围在 $-6.13 \sim -4.33$ 之间，不妨取 β 为 -5.23 （我们会在灵敏度分析中详细分析 β 取值对结果的影响）。此时得到的函数表达式如下：

$$y = 0.82(1 - e^{-5.23x}) + 0.18 \quad (11)$$

通过 MATLAB 画出的函数图像如下：

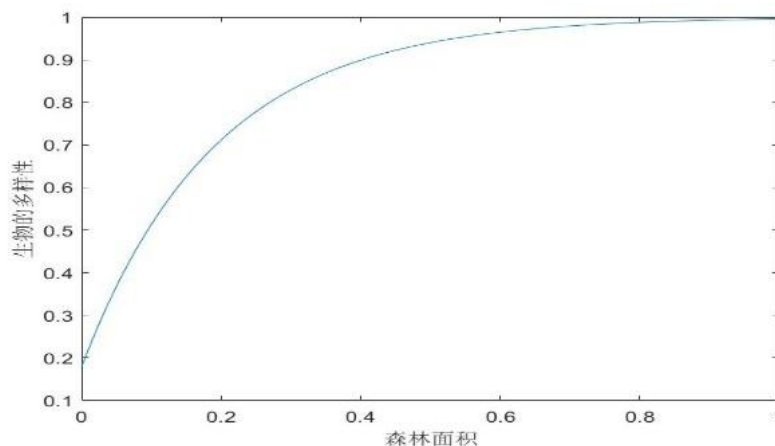


图 11 森林面积 S 和生物的多样性 M_4 的关系

用偏好性系数加权法将多目标问题转化为单目标问题，其表达式为：

$$N = k_3 M_3 + k_4 M_4 \quad (12)$$

这里表达式体现了综合体现了退耕还林政策在经济效益和生物多样性方面对国民经济的影响，我们称之为退耕还林政策对国民经济的影响。这里的 k_3, k_4 有两方面的含义：一是作为偏好性加权系数，二是充当修正系数的作用，以保证经济效益和生物多样性方面的两个表达式的量化值在数值上近似相等。

（四）模型中参数的确定

（1） d_1 、 d_2 、 d_3 和 k

d_1 、 d_2 、 d_3 分别为流向一、二、三产业的劳动力的占比，我们选取不同年份一、二、三产业的占比为 d_1 、 d_2 、 d_3 的值。

k 的取值可将实际值代入表达式来确定。因为我们最终求得的是折合成绩，为了便于计算， k 的值可取 1。

$$\text{式（3）即为 } M_1 = (d_1 * x_i + d_2 * y_i + d_3 * z_i)$$

（2） d 和 r

根据搜集到的数据，可以很快计算出 1980-2020 年工业增加值的平均值 d 为 0.4425。

r 为修正因子，当 $D=1$ 时，令 $M_2=2$ ，即可解得 $r=0.3067$ 。

$$\text{式（5）即为 } M_2 = e^{\frac{D}{0.4425/0.3067}}。$$

（3） b_1 、 b_2 、 $b1$ 、 $b2$

可取 b_1 、 b_2 为农业产值和林业产值比例，对搜集到的数据进行计算可得，农业产值的平均值为 22275.76，林业产值的平均值为 1736.192，所以， b_1 可取 0.9277、 b_2 可取 0.0723。

$b1$ 、 $b2$ 的取值可将实际值代入表达式来确定。因为我们最终求得的是折合成绩，为了便于计算， $b1$ 、 $b2$ 的值可取 1。

$$\text{式（7）即为 } M_3 = 0.9277 * x_1 + 0.0723 * x_2。$$

（4） k_1 、 k_2 、 k_3 、 k_4

k_1 、 k_2 、 k_3 、 k_4 的取值可将实际值代入表达式来确定。因为我们最终求得的是折合成绩，为了便于计算， k_1 、 k_2 、 k_3 、 k_4 要充当修正系数的作用，以保证就业效益、潜在效益、经济效益和生物多样性方面的四个表达式的量化值在数值上近似相等。

$$\text{令 } k_1=0.01、k_2=10，\text{式（6）即为 } M = 0.01M_1 + 10M_2。$$

$$\text{令 } k_3=0.001、k_4=100，\text{式（12）即为 } N = 0.001M_3 + 100M_4。$$

（五）模型求解与结果分析

（1） 对国家工业水平的影响

构造折合成绩需要确定标准值，我们选取相关数据的平均值代入表达式计算，将得到的 M 的值作为标准值。

计算可得，第一、二、三产业的占比的平均值为 17.06098，44.25366，38.6878。第一、二、三产业的贡献率的平均值为 9.097561，50.69268，40.20488。 M_1 的标准值为 3953.9883。工业增加值 D 的平均值为 0.4425，则 M_2 的标准值为 1.3589。所以 M 的标准值为 53.1289。

工业增加值 D 的平均值为 0.4425，则 M_2 的标准值为 1.3589。所以 M 的标准值为 53.1289。

将历年的数据代入表达式可得每年的得分，结果如下：

表 2 不同年份退耕还林政策对国家工业水平的影响

年份	得分	年份	得分	年份	得分	年份	得分
2020	56.3544	2009	57.17636	1999	54.81117	1989	49.60997
2019	61.09329	2008	56.67504	1998	54.39919	1988	52.14732
2018	60.34754	2007	57.83596	1997	54.7467	1987	50.65919
2017	59.77777	2006	57.08047	1996	54.41271	1986	50.22505
2016	59.48562	2005	56.46096	1995	54.55306	1985	50.7795
2015	58.40963	2004	55.3883	1994	55.1891	1984	47.62233
2014	57.66248	2003	57.024	1993	54.71927	1983	47.78607
2013	57.44942	2002	55.88714	1992	53.02972	1982	46.03944
2012	57.15113	2001	55.31343	1991	51.84853	1981	43.37337
2011	57.42999	2000	55.83926	1990	47.29386	1980	57.03144
2010	57.55343						

令标准值 53.1289 对应的成绩为 100，则折合成绩如下：

表 3 折合成绩

年份	成绩	年份	成绩	年份	成绩	年份	成绩
2020	106.0711	2009	107.6182	1999	103.1664	1989	93.37662
2019	114.9907	2008	106.6746	1998	102.391	1988	98.15245
2018	113.587	2007	108.8597	1997	103.0451	1987	95.35147
2017	112.5146	2006	107.4377	1996	102.4164	1986	94.53434
2016	111.9647	2005	106.2717	1995	102.6806	1985	95.57792
2015	109.9395	2004	104.2527	1994	103.8777	1984	89.63546
2014	108.5332	2003	107.3314	1993	102.9934	1983	89.94365
2013	108.1321	2002	105.1916	1992	99.81332	1982	86.65611
2012	107.5707	2001	104.1117	1991	97.59008	1981	81.638
2011	108.0956	2000	105.1015	1990	89.01721	1980	107.3454
2010	108.3279						

(2) 结果分析

为直观显示分数的变化，绘制不同年份的折合成绩、 M_1 、 M_2 的图：

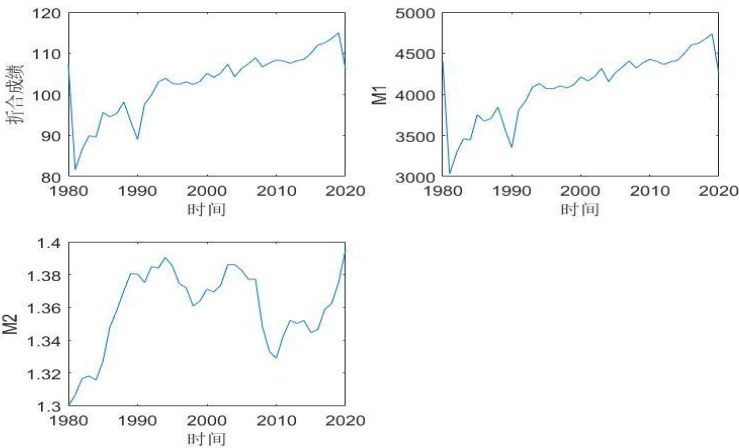


图 12 不同年份的折合成绩、 M_1 、 M_2

1、 $M1$ （就业效益）：由图可知，退耕还林政策的实施对带动第二、三产业就业水平影响总体呈现上升态势。尤其是经过 1980 年断崖式下跌之后，自 2003 年开始，退耕还林政策对第二、三产业就业发展的影响呈现稳步增长态势。而到 2020 年就业效益出现短暂下滑，原因可能是受疫情影响，人员流动限制，第二、三产业，特别是第三产业就业发展受挫。但与此同时，随着退耕还林政策逐步推进实施，截至 2020 年，就业效益较之 20 世纪数据来说，仍保持在较高水平。这也充分说明了随着退耕还林政策实施，我国就业效益（第二、第三产业）得到了较好改善，因此我们认为退耕还林政策实施对促进国家工业水平提升具有重要作用。

2、 $M2$ （潜在效益）：由图可知，退耕还林政策对国家工业水平的潜在价值影响在稳步增长中呈现一定周期性波动规律，这可能是经济周期波动造成的工业工业水平随之波动。但同时我们能在波动中看出工业发展水平较之以前，正在逐步突破最高点。因此我们认为退耕还林政策的实施在推动国家工业水平发展的同时能够一定程度增加国家工业发展水平的潜在效益，且在此基础上使国家工业水平在未来更具发展活力。

3、折合成绩：通过偏好性系数加权法我们得出退耕还林政策对国家工业水平影响综合成绩图。由图可知，在 2003 年退耕还林政策实施之前国家工业水平在上升趋势中存在一定程度波动且 1990 年出现了较为明显的下滑，而在 2003 年退耕还林政策实施之后，国家工业水平发展呈现稳步上升态势，且在 2019 年达到峰值（预估 2020 年因疫情影响出现下滑）。因此退耕还林政策的实施的确给国家工业水平发展注入了更多动力与活力。

（3）对国民经济的影响

对搜集到的数据进行计算可得，农业产值的平均值为 22275.76，林业产值的平均值为 1736.192。则 M_3 的标准值为 20790.7492。森林面积 S 的平均值为 0.2。则 M_4 的标准值为 0.7119。所以 N 的标准值为 91.9807。

将历年的数据代入表达式可得每年的得分，结果如下：

表 4 不同年份退耕还林政策对国民经济的影响

年份	得分	年份	得分	年份	得分
2020	142.8002	2009	100.6923	1999	82.10281
2019	137.2622	2008	98.18466	1998	81.88211
2018	132.7008	2007	94.79764	1997	81.17057
2017	129.2616	2006	91.69688	1996	80.52587
2016	126.7486	2005	89.53824	1995	78.62976
2015	125.0742	2004	87.78347	1994	75.74415
2014	122.6022	2003	84.36092	1993	72.99344
2013	119.6016	2002	84.01359	1992	71.67711
2012	115.486	2001	83.17743	1991	70.89184
2011	110.9981	2000	82.23098	1990	70.33511
2010	106.5625				

令标准值 91.9807 对应的成绩为 100，则折合成绩如下：

表 5 折合成绩

年份	成绩	年份	成绩	年份	成绩
2020	155.2502	2009	109.4711	1999	89.26092
2019	149.2293	2008	106.7448	1998	89.02097
2018	144.2703	2007	103.0625	1997	88.24739

2017	140.5312	2006	99.69143	1996	87.54649
2016	137.7991	2005	97.34459	1995	85.48507
2015	135.9788	2004	95.43683	1994	82.34787
2014	133.2912	2003	91.7159	1993	79.35734
2013	130.029	2002	91.33828	1992	77.92625
2012	125.5546	2001	90.42922	1991	77.07252
2011	120.6754	2000	89.40025	1990	76.46725
2010	115.8531				

(4) 结果分析

为直观显示分数的变化，绘制不同年份的折合成绩、 M_3 、 M_4 的图：

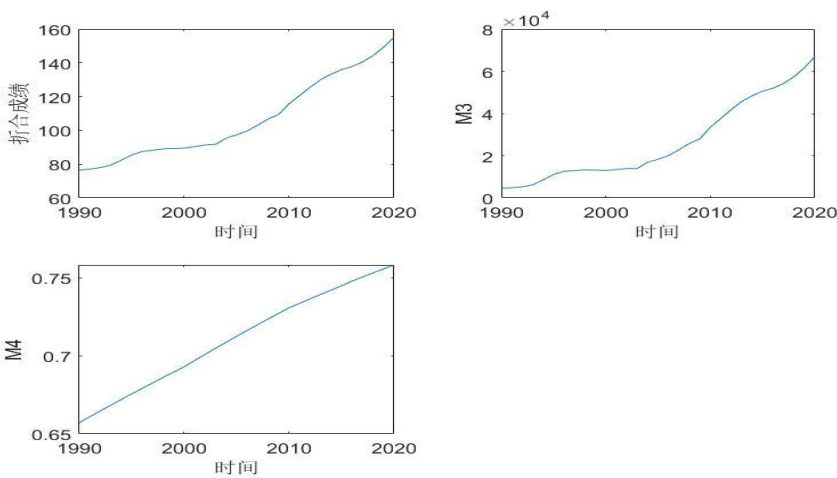


图 13 不同年份的折合成绩、 M_3 、 M_4

1、 M_3 （经济效益）：由图可知，退耕还林政策实施有力提升了经济效益。在退耕还林政策实施（2003 年）之前，农业与林业总产值以较慢速度增长，且在 1996 年增长速度明显放缓，一直停滞发展至 2003 年。在 2003 年国务院常务会议通过《退耕还林》条例之后，经济效益放缓增长态势逐步消失，且随着退耕还林政策逐步实施，经济效益愈发呈现出强劲发展活力与动力。

2、生物多样性：由图 11 可知，森林面积增长与生物多样性增长呈现边际效应递减。即随着森林面积增加，生物多样性增长速度放缓，随着时间推移每增加一单位森林面积，相应生物多样性增加量减少。因此由图 3 可以解释，自 2003 年退耕还林政策实施之后，生物多样性没有出现明显加快增长态势而是保持平稳增长，很可能是因为我国森林面积已保持在较高水平，因此存在着生物多样性边际递减效应。

3、折合成绩：由图 13 可知，自 1990 年以来，由经济效益与生物多样性加权得出的国民经济综合得分呈现稳定增长态势。但自 1997 年增长速度明显放缓，而自 2003 年退耕还林政策实施后，国民经济发展态势愈发向好。

第三部分：问题 3 的模型

(一) 主成分分析法

主成分分析就是将多项指标转化为少数几项综合指标，用综合指标来解释多变量的方差—协方差结构。综合指标即为主成分。所得出的少数几个主成分，要尽可能多地保留原始变

量的信息，且彼此不相关。

主成分分析法的步骤如下：

1. 对原始数据进行标准化处理
2. 计算相关系数矩阵 R
3. 计算特征值和特征向量
4. 选择主成分，计算综合评价值

（二）模型的求解

选定题目所给的部分数据进行主成分分析，数据分析及结果如下所示：

利用 MATLAB 求得相关系数矩阵的前 4 个特征根及其贡献率如表 6 所列

表 6 主成分分析结果

序号	特征根	贡献率	累计贡献率	序号	特征根	贡献率	累计贡献率
1	2.7730	69.3246	69.3246	3	0.2118	5.2945	99.5003
2	0.9952	24.8812	94.2058	4	0.0200	0.4998	100.0000

可见，前两个特征根的累计贡献率达到了 94%以上，主成分分析效果较好。下面取前两个主成分进行综合评价。前两个特征根对应的特征向量见表 7。

表 7 标准化变量的前 2 个主成分对应的特征向量

	x_1	x_2	x_3	x_4
第 1 特征向量	0.5965	0.5654	0.4966	0.0698
第 2 特征向量	-0.0187	-0.0960	0.9952	-0.0071

由此可得两个主成分分别为：

$$Y_1 = 0.5965x_1 + 0.5654x_2 + 0.4966x_3 + 0.0698x_4 \quad (13)$$

$$Y_2 = -0.0187x_1 - 0.0960x_2 + 0.9952x_3 - 0.0071x_4 \quad (14)$$

分别以两个主成分的贡献率为权重，构建主成分综合评价模型

$$Z = 2.7730Y_1 + 0.9952Y_2 \quad (15)$$

把数据代入上式，可以得到综合评价结果，如表 8 所列

表 8 综合评价结果

国家	结果	国家	结果
世界	8.121699	英国	-0.37926
中低收入国家	4.266676	奥地利	-0.38092
中等收入国家	4.020844	洪都拉斯	-0.38591
高收入国家	3.519982	斯里兰卡	-0.38847
人口红利之后	3.274313	瑞士	-0.39342
中高等收入国家	2.938948	喀麦隆	-0.39392
东亚与太平洋地区	2.099484	立陶宛	-0.39964
欧洲与中亚地区	1.838423	匈牙利	-0.41941
东亚与太平洋地区（不包括高收入）	1.445254	法国	-0.42266
早人口红利	1.175371	保加利亚	-0.43064
中国	1.038407	亚美尼亚	-0.43338
美国	0.892664	海地	-0.43414

中低等收入国家	0.814429	爱沙尼亚	-0.44992
利比亚	0.793108	加纳	-0.45255
欧洲与中亚地区（不包括高收入）	0.741206	孟加拉国	-0.4628
欧洲联盟	0.584236	吉尔吉斯斯坦	-0.46328
赤道几内亚	0.57213	约旦	-0.46452
文莱达鲁萨兰国	0.451421	老挝	-0.47046
科威特	0.442843	哥斯达黎加	-0.48384
伊拉克	0.417861	危地马拉	-0.48681
欧洲货币联盟	0.382092	圭亚那	-0.48947
安哥拉	0.321352	柬埔寨	-0.49066
中东与北非地区	0.288743	比利时	-0.49584
阿拉伯联盟国家	0.272734	中非共和国	-0.4965
刚果（布）	0.238764	波斯尼亚和黑塞哥维那	-0.50172
预人口红利	0.196855	圣基茨和尼维斯	-0.50357
加蓬	0.178363	布基纳法索	-0.5071
阿拉伯联合酋长国	0.155443	克罗地亚	-0.50797
印度尼西亚	0.104043	古巴	-0.51261
阿塞拜疆	0.070509	丹麦	-0.51897
日本	0.055856	以色列	-0.52818
巴西	-0.02023	阿尔巴尼亚	-0.52997
巴林	-0.02229	拉脱维亚	-0.53286
伊朗伊斯兰共和国	-0.02232	格鲁吉亚	-0.53466
加拿大	-0.05144	佛得角	-0.53557
博茨瓦纳	-0.07167	冰岛	-0.53762
列支敦士登	-0.08704	牙买加	-0.54051
大韩民国	-0.12681	科特迪瓦	-0.60612
印度	-0.13632	厄立特里亚	-0.61047
德国	-0.1441	斐济	-0.61473
阿尔及利亚	-0.14935	黎巴嫩	-0.62938
不丹	-0.16144	希腊	-0.63113
刚果（金）	-0.17618	肯尼亚	-0.64352
白俄罗斯	-0.19383	塞浦路斯	-0.64414
哈萨克斯坦	-0.19682	伯利兹	-0.64684
智利	-0.22397	贝宁	-0.65324
捷克共和国	-0.23008	安提瓜和巴布达	-0.68449
阿拉伯埃及共和国	-0.23292	巴巴多斯	-0.68503
澳大利亚	-0.27829	格林纳达	-0.68624
莱索托	-0.28129	布隆迪	-0.69037
厄瓜多尔	-0.28355	冈比亚	-0.70723
低收入国家	-0.29009	卢森堡	-0.71913
哥伦比亚	-0.29646	埃塞俄比亚	-0.73159
加勒比小国	-0.3021	多米尼克	-0.73214
爱尔兰	-0.30756	吉布提	-0.73535
几内亚	-0.32002	几内亚比绍共和国	-0.73654

多米尼加共和国	-0.32028	巴哈马	-0.76094
意大利	-0.35334	圣卢西亚	-0.77609
玻利维亚	-0.35889	科摩罗	-0.7791
阿根廷	-0.36158	利比里亚	-0.79711
芬兰	-0.37106	基里巴斯	-0.80474
西班牙	-0.37536	密克罗尼西亚联邦	-0.90195

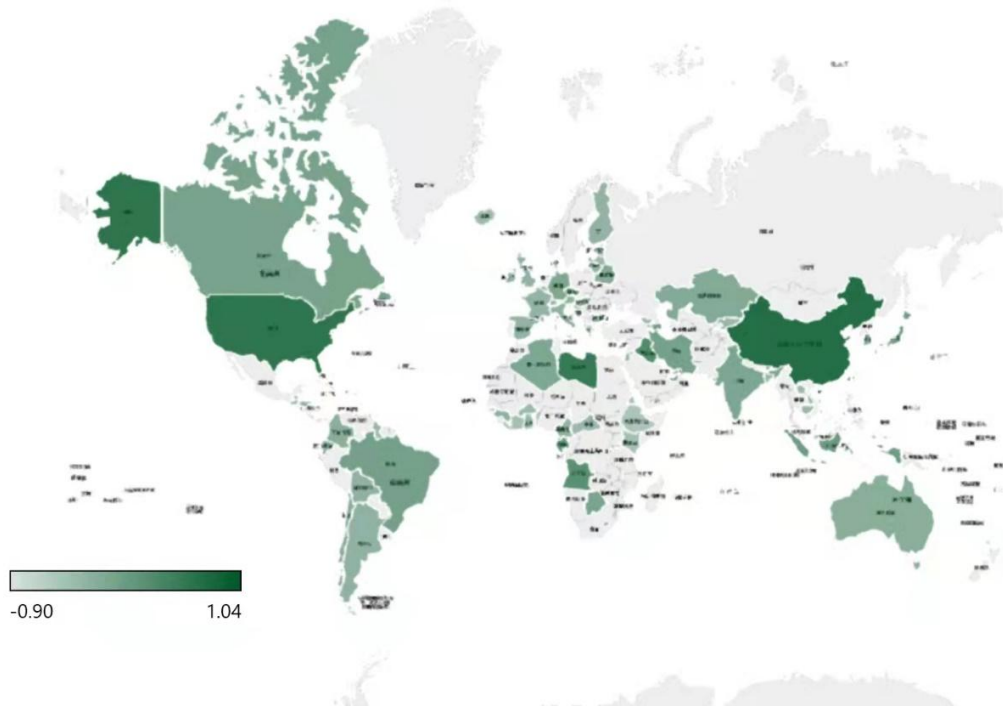


图 14 世界（部分）国家综合评价结果图

（三）结果分析

绘制综合评价结果的饼图和子母饼图

综合评价结果



- 世界
- 中低收入国家
- 中等收入国家
- 高收入国家
- 人口红利之後
- 中高等收入国家
- 东亚与太平洋地区

图 15 饼图

综合评价结果



- 世界
- 中低收入国家
- 中等收入国家
- 高收入国家
- 人口红利之後
- 中高等收入国家
- 东亚与太平洋地区

图 16 子母饼图

1. 综合评价结果中排名第一、第二位国家分别为中国、美国。目前美国作为世界第一大经济体，经过我们的模型评估其综合评价结果位居中国之后，这可能是美国某些激进的发展方式

所导致。同时，美国新总统就职后当天即签署重返《巴黎协定》的法案，也显示出美国雄心勃勃的绿色发展战略希望尽快可以回到引领全球绿色发展的地位。而作为目前引领全球绿色发展的中国则一直坚定以绿色发展为指导思想，“我们既要绿水青山，也要金山银山。宁要绿水青山，不要金山银山，而且绿水青山就是金山银山”的形象比喻也更形象表达了中国推进生态文明建设的鲜明态度和坚定决心。

2. 中低等收入国家综合评价结果普遍低于中高等国家。这可能是因为低收入国家目前亟需解决温饱、医疗等基本社会问题，其 GDP 还未达到理想发展水平。而绿色经济发展需要政府的有效政策、资金、技术投入，这些是低收入国家较难达到的，即便联合国给予相关资金支持，但就目前现实来看也很有限。如此以来，低收入国家由于主要精力在解决基本温饱、医疗等问题，且由于其经济发展需要外资注入资本力量，外国部分污染工厂的入驻也进一步加剧了其环境恶化问题，如此一来则一定程度形成恶性循环。相比较中国，由于其有较为雄厚的 GDP 基础、资金、技术及相应政策支持，因此其绿色发展则更有结合实际国情的特色及可行性。

3. 欧洲大部分国家都拥有较高的综合评价结果。我们知道欧洲是推进绿色新政的先导者。欧盟实施的是内涵最广的绿色经济模式，即将治理污染、发展环保产业、促进新能源开发利用、节能减排等都纳入绿色经济范畴加以扶持。其在推进过程中，强调多领域的协调、平衡和整合。目前中国虽为全球绿色发展引领者，但也应不断学习吸收他国发展经验，并基于自己的发展现状，走出未来具有中国特色的绿色发展道路。

六、灵敏度分析

由于在本问题中涉及的因素比较多，并且随机性比较大，所以我们对模型中的相关参数进行灵敏度分析。模型对这些参数的敏感性反映了各种因素影响结果的显著程度；反之，通过对模型参数的敏感性分析，又可以反应和检验模型的实际符合程度。

我们将灵敏度分析定在问题二中森林面积 S 和生物的多样性 M_4 的关系这一问题上。在此模型中，我们建立的森林面积和生物的多样性的关系式如下：

$$y = f(x) = (1 - y_0)(1 - e^{-\beta x}) + y_0, y_0 \in (0,1) \quad (6.1)$$

我们研究此式中参数 β 的变化对结果的影响。我们对 β 的值以增加 0.1 为单位，进行逐步变化，对其他参数保持不变，得到各自函数的图像如图所示

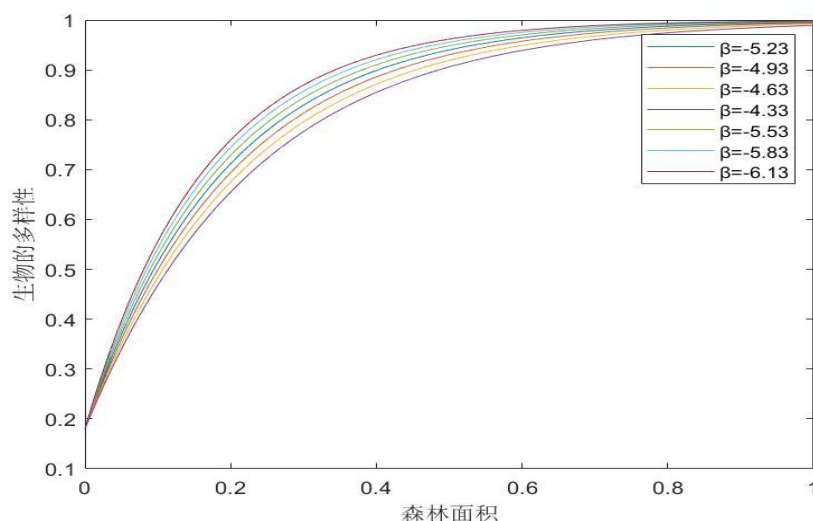


图 17 灵敏度分析图

分别计算每个函数中当 $x=0.3$ 时人们的防疫意识与措施的值的大小,并计算其相对于 $\beta = -5.23$ (模型中所取) 时的结果的变化率。计算结果如表所示

表 9 β 取值与变化率的关系

取值	森林面积为 0.3 时生物多样性的值	变化率
-4.33	0.7763	0.9362
-4.63	0.7956	0.9595
-4.93	0.8132	0.9807
-5.23	0.8292	1.0000
-5.53	0.8439	1.0177
-5.83	0.8574	1.0340
-6.13	0.8696	1.0487

从表 10 结果来看,当 β 的值每改变 0.3 (相当于 5.7%左右),其结果变化小于 3%,因此体现出模型较稳定。

七、报告书

问题 4 要求提交一份报告书给联合国,因此我们书写英文版报告书如下:

Report on Environment, Economy and Industry Comprehensive Evaluation

In this fast-changing world, with the intention of developing economy and technology, every country is producing greenhouse gases every day. What's worse, deforestation is serious all over the world. Our team is invited to analyse the current situation. Here's what we find.

China and America rank first and second respectively in our evaluation results. Currently, America is the world's biggest economy, but after modeling and analyzing we can see China is followed by America. The factor accounts for this is probably that America is taking some radical action with the intention of developing economy and technology. However just after Joe Biden took office, he announced that US returned to Paris Agreement, which revealed the green development strategy and the willing to return to the leading position of global green development as soon as possible. Meanwhile China insists on taking green development as the guiding ideology. President Xi once said that Clear waters and green mountains are as good as mountains of gold and silver, illustrating China's firm determination to promote the construction of ecological civilization.

So what about the situation of the low- and middle-income countries? We find their evaluation results are generally lower than those of high- and middle-income countries. Perhaps the reason for this is low- and middle-income countries urgently need to solve the basic social problems like food and clothing and medical treatment. Their GDP hasn't reached the level of ideal development. However, green economical development is based on the support of governments' investment of policies, funds and techniques, which is difficult to these countries. So if The United Nations give them relevant financial support to some extent, their poor conditions probably will be improved. Since the low- and middle-income countries mostly focus on solving the problems of food and clothing and medical treatment and their development of economy rely on foreign investment to a certain extent, some foreign manufactories aggravate the already serious environment problems, which is a vicious circle.

The conditions in Europe are much better. Europe is the leader of promoting Green New Deal and the European countries implement the most extensive green economy model, the economy of which consists of pollution treatment, the development of industry of environmental protection, the advancement and utilization of new energy resources, energy conservation and pollution reduction. During their promotion of the green economy model, coordination, balance and integration in multiple fields are emphasized.

To tackle the severe environmental problems, every country is supposed to continue to learn the better experience of other countries and create its own unique green development policy based on the concrete national conditions.

八、模型评价与推广

（一）优点

1. 基于时间序列模型 ARIMA，在经过数据二阶差分处理满足平稳性的基础上，得到了准确度较高的预测值。
2. 基于预测值，采用区间估计得到了 95%的 2025-2035 中国森林面积（平方公里）预测区间，较为准确全面地解决了问题一中的预测不确定性问题。且在考虑进数据误差与模型误差的基础上使得预测结果更为合理。
3. 模型简单，只需要借助内生变量而不需要借助其他外生变量。
4. 该模型基于时间序列数据采用时间序列模型，使问题描述更加清晰。
5. 使用主成分分析法，利用降维技术用少数几个综合变量来代替原始多个变量，这些综合变量集中了原始变量的大部分信息。

（二）缺点

1. 要求数据是稳定的，或者是通过差分化后是稳定的。
2. 本质上只能捕捉线性关系，而不能捕捉非线性关系。
3. 没有对模型中较多的参数进行详细的讨论。

（三）模型的推广

1. 问题二没有过多考虑产值与成本，可以进一步考虑在构建函数表达式时加入成本作为限制因素。
2. 可在问题二模型的基础上讨论退耕还林区域的影响范围和影响程度，方法如下：
 - 1) 以问题二中的折合成绩为基础，构造退耕还林区域的影响半径 R
 - 2) 以退耕还林区域为圆心，退耕半径 R 为半径画圆
 - 3) 落入圆内的区域即为被影响的区域，可根据被影响区域距离圆心的距离来衡量受影响的程度

九、参考文献

[1]施怡. 基于 R 语言 ARIMA 模型的全国硕士研究生报名人数预测研究[J]. 电脑知识与技术: 学术版, 2018(3Z):3.

[2]任宣宇, 朱冉, 姚琪, 等. 基于多元线性回归的钢水“脱氧合金化”收得率问题研究[J]. 电脑编程技巧与维护, 2020(3):4.

[3]King S, Vardon M, Grantham H S, et al. Linking biodiversity into national economic accounting. Environmental Science and Policy, 2021, 116: 20-29.

[4]苏欣欣, 肖洋, 胡晓航, 马亚怀, 李彦丽. 基于灰色关联度分析和主成分分析法评估糖用甜菜品种的适应性[J]. 中国农学通报, 2021, 37(30):39-46.

[5]赖亚飞, 朱清科. 黄土高原丘陵沟壑区退耕还林(草)工程实施综合效益评价——以陕西省吴起县为例[J]. 西北林学院学报, 2009, 24(03):219-223.

[6]李新蕊. 主成分分析、因子分析、聚类分析的比较与应用[J]. 山东教育学院学报, 2007(06):23-26.

十、附录

表 10 描述统计量

	N	极小值	极大值	均值	标准差	平均值标准误差
第一产业占比	42	7	32.8	17.06098	8.406713	1.297184
第二产业占比	42	37.8	48.1	44.25366	2.732813	0.421682
第三产业占比	42	22.3	54.5	38.6878	9.274587	1.431100
第一产业贡献率	42	-4.8	40.5	9.097561	10.009314	1.544471
第二产业贡献率	42	17.7	85.6	50.69268	12.217141	1.885146
第三产业贡献率	42	19.2	63.5	40.20488	10.833872	1.671703
农业总产值	41	1454.1	71748.23	22275.76	21124.378620	3299.073677
林业总产值	41	81.4	5961.58	1736.192	1798.885748	280.938754