

各国经济水平、工业水平与森林资源的评价与关系探究

摘 要

2006 年“十一五”期间,节能减排作为一项国家重大战略规划被正式提出。环境与国民经济、工业化进程有着密切的联系,探究世界各国经济水平、工业水平与森林资源之间的关系具有重要的实际意义。为了解决所给的问题,我们利用时间序列模型来预测中国的森林面积,并建立了退耕还林政策影响模型与经济、工业、环境评价模型。

在问题一中,我们通过观察中国 1990-2020 年间森林面积的时间序列图,发现数据呈稳步上升的趋势,于是我们选择 ARIMA(0,2,0) 模型进行时间序列分析。模型通过了白噪声检验并得到了中国在 2025-2035 年间的预测值。在 2035 年,中国森林面积预计将达到 2481717.30 平方公里。

在问题二中,我们从国家统计局网站的年鉴中搜集到能够反映国家工业水平与国民经济的 7 个指标,并对数据进行归一化处理来消除量纲的影响,粗略地画出实施退耕还林政策前后十年的数据变化。我们参考了《退耕还林条例》,从中选取出造林面积、补助标准等 5 个指标,对其与反映国家工业水平与国民经济的 7 个指标分别进行相关性分析,结果表明退耕还林政策对我国国民经济与工业化水平有很大的影响。为了进一步探究,我们对这些指标利用 Stata 软件各自进行了 OLS 回归分析,我们的回归结果通过了 BP 检验与多重共线性检验。结果表明,反映国家工业化的 4 个指标受到技术研究、宣传教育的影响较深,反映国民经济的 3 个指标受到宣传教育的影响较深。

在问题三中,我们通过样条插值等方法对 99 个国家的各个指标数据进行预处理。利用系统聚类法将其分为三大类、三小类,并给出每一类关于经济、工业、森林的评价。然后,从每一类中选择一个代表国家,选取三个时间序列变量进行 ADF 平稳性检验、协整检验与 Granger 因果关系检验。我们发现,经济水平、工业水平与森林资源这三者之间的关系在三类国家之间均不相同。例如对于以俄罗斯联邦为代表的第 1 类国家,森林资源是能够促进经济与工业化发展的优势。

最后,我们向联合国递交了一封报告书来阐明我们的发现,并针对不同类型的国家提出了具体可靠的建议。

关键词: 退耕还林影响模型 OLS 回归 Granger 因果分析 系统聚类

一、 问题重述

近年来,随着全球范围内的植被破坏,森林面积逐渐减少,而工业发展又增加了碳排放量,环境的保护对国民经济的发展有着重要意义。对比全球趋势,中国在环境治理方面已经做出了巨大的贡献。然而,随着工业化进程的加快,环境保护方面的问题仍然刻不容缓。

(1) 预测中国在 2025-2035 年内的森林面积。

(2) 通过分析中国的有关数据,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) 根据附件中给出的国家数据构建模型对世界各国经济水平、工业水平和森林植被面积进行评价。

(4) 向联合国递交一封报告书,描述发现,并提出一些具体措施。

二、 问题分析

2.1 问题一的分析

问题一要求我们预测森林未来十年里的面积,根据附录中给的数据,画出时间序列图可以发现,在过去的十几年里森林面积是呈上升趋势的,因此我们考虑使用时间序列模型来对其进行预测。

2.2 问题二的分析

问题二要求我们分析退耕还林政策对国民经济和工业的影响,我们通过查询相关条例发现退耕还林政策可以使农民获得国家补贴,促进产业调整,推动第一产业向第二产业转变。因此我们查阅相关资料后选取能反映政策和经济工业水平的指标,对它们进行相关分析,看之间是否有很大关联。接着,为了做进一步的探究,我们考虑使用 OLS 回归来探索指标与指标之间的联系,由此来发现影响的程度与决定性因素。

2.3 问题三的分析

问题三中要求我们对各个国家的经济、工业和森林覆盖率进行评价,在附录的《国家》中我们发现,各个国家的收入水平有所差异,它们的经济工业水平也各不相同,因此我们要先对它们进行聚类,再根据每一类的突出特点评价它们的经济工业和森林面积,并最后根据不同的类别找出它们三者的关系

三、 模型假设

1. 假设附件中所给的数据真实可靠。
2. 假设森林资源的年龄结构趋于稳定。
3. 在预测的 2025-2035 年期间,中国无大型自然灾害大规模破坏森林面积的现象出现。

4. 退耕还林政策能够持续实施, 退耕还林的农户能够领到补助或者找到工厂工作。
5. 假设十几年前收录在中国知网中的学术论文不会被下架。

四、 定义与符号说明

符号	含义
A_i	第 i 大类国家
B_j	第三大类的第 j 小类国家
VIF	方差膨胀系数
$Adj R - squared$	变量之间相关程度的值

五、 模型的建立与求解

5.1 中国森林面积预测模型

我们对附录里文件森林面积(平方公里)中的中国 1990-2020 年的数据进行预处理, 首先我们画出森林面积的时间序列图来观察中国森林面积再过去 30 年间的变化趋势。

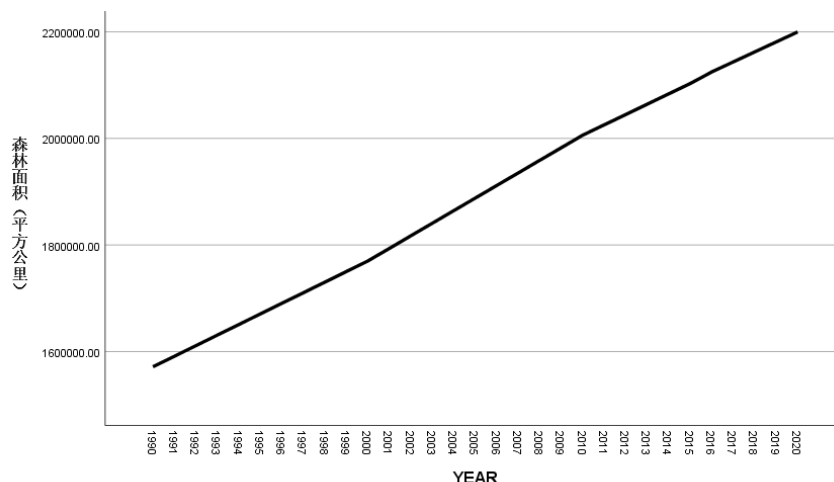


图 1 中国森林面积变化

由图可知森林面积随时间变化朝着一定方向呈现出持续稳定的上升趋势。因此, 我们考虑采用时间序列分解。即通过历史森林数据揭示现象随时间变化的规律, 将这种规律延伸到未来, 从而预测 2025-2035 年的森林面积。

时间序列分析方法, 主要用来解决具有随机性、季节性以及平稳性的时间序列问题的, 它是由博克斯-詹金斯 (BoxJenkins) 发现的。对于平稳的时间序列, 一般采用 ARMA

模型, 我们的计算步骤如下:

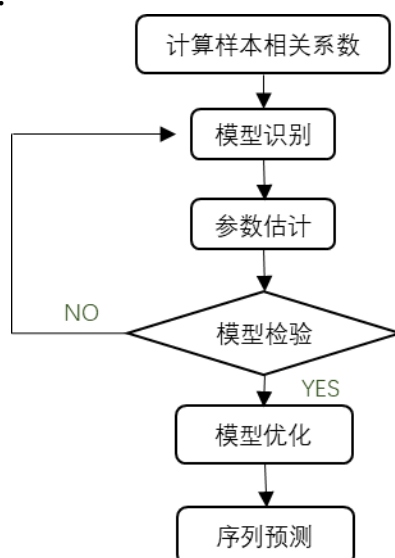


图 2 ARMA 计算步骤

然而, 根据我们画出的时间序列图可知其属于非平稳时间序列。对于非平稳的时间序列, 我们采用 ARIMA 模型, 设 $\{Y_t\}$ 是一个不平稳的时间序列, 在 d 次差分运算后, 可将其转化为平稳的时间序列模型再进行上述求解。

令 $W_t = Y_{t-1}$:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

用序列符号 Y_t 来表示:

$$Y_t - Y_{t-1} = \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) + \dots + \phi_p (Y_{t-p} - Y_{t-p-1}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

通过化简可以得到:

$$Y_t = (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + (\phi_3 - \phi_2)Y_{t-3} + \dots + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

其中, c 是常数, $\phi_1, \phi_2, \dots, \phi_p$ 是自回归模型 AR 的系数, p 是 AR 的阶数, $\theta_1, \theta_2, \dots, \theta_q$ 是滑动平均模型 MA 的系数, q 是 MA 的阶数, e_t 是均值为 0 方差为 σ^2 的白噪声序列。该模型即为差分方程形式。具体操作步骤如下:

● 模型识别

对于一个确定的时间序列, 模型识别是首要任务, 我们利用 SPSS 专家建模器进行模型选择。SPSS 专家建模器会自动查找每个相依序列的最佳拟合模型。如果指定了自变量(预测)变量, 则专家建模器为 ARIMA 模型中的内容选择那些与该相依序列具有统计显著关系的模型。适当时, 使用差分和/或平方根或自然对数转换对模型变量进行转换, 缺省情况下, 专家建模器既考虑指数平滑法模型也考虑 ARIMA 模型。

● 参数估计

通常采用极大似然估计法来估计 ARIMA(p, d, q) 模型中的未知参数, 这里专家建模器估计出来的最优模型为 ARIMA(0, 2, 0) 模型。

● 模型检验

模型对时间的拟合效果我们用平稳的 R 方来判断。模型检验既要参数的估计值进行检验, 又要对残差序列进行检验。从残差的 ACF 和 PACF 图形中可以看出, 所有滞后阶数的自相关系数和偏自相关系数均和 0 没有显著的差异。因此 ARIMA(0, 2, 0) 模型能很好地识别本题中的森林面积数据。

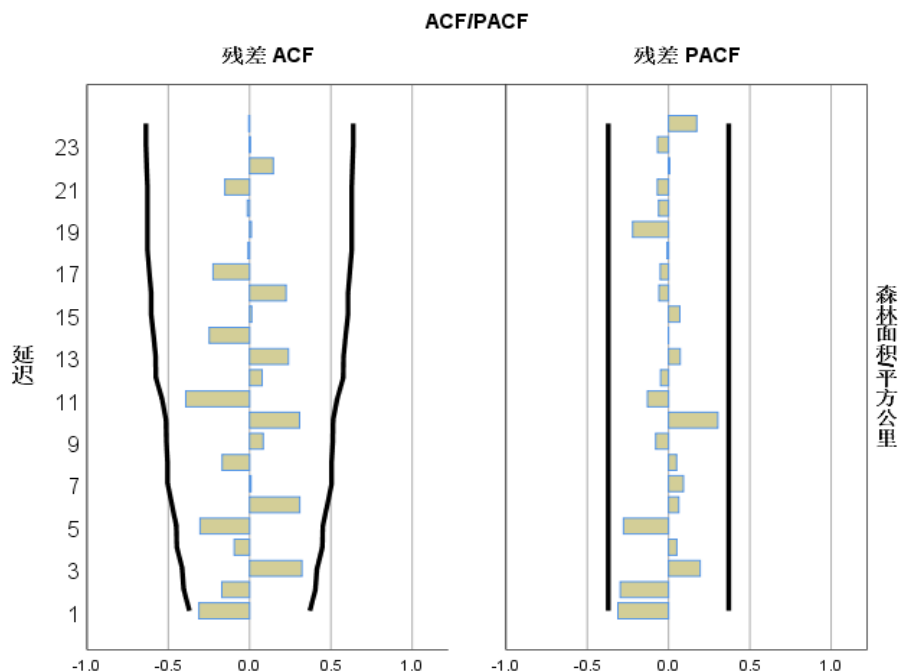


图 3 ACF/PACF 图像

● 模型预测

在 95%的置信水平下，我们继续使用该模型对未来 2025-2035 年的森林面积进行预测。

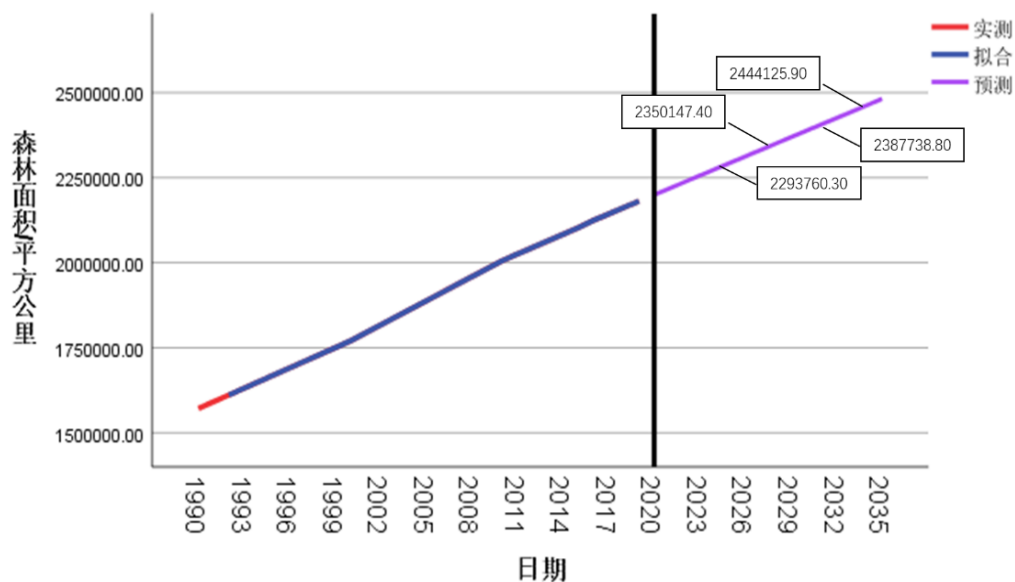


图 4 预测图

由图表可以看出，预测值总体来说是符合其变化规律及周期性的，且通过了残差的白噪声检验，具有一定的可信度。具体的预测值如下表所示：

表 1 森林面积预测值

年份	预测值	置信区间下限	置信区间上限
2025	2293760.30	2293760.30	2293760.30
2026	2312556.00	2312556.00	2312556.00
2027	2331351.70	2331351.70	2331351.70

2028	2350147.40	2350147.40	2350147.40
2029	2368943.10	2368943.10	2368943.10
2030	2387738.80	2387738.80	2387738.80
2031	2406534.50	2406534.50	2406534.50
2032	2425330.20	2425330.20	2425330.20
2033	2444125.90	2444125.90	2444125.90
2034	2462921.60	2462921.60	2462921.60
2035	2481717.30	2481717.30	2481717.30

5.2 退耕还林政策影响模型

5.2.1 指标选取

- 作为中国的生态重建和恢复项目，退耕还林肩负着防风固沙、退耕区经济结构调整、生态文明建设、精准扶贫等目标和重任。通过查阅《退耕还林条例》以及《各地区退耕还林情况统计》，我们选取以下能够反映退耕还林政策情况。

造林面积：造林面积是林业资源恢复和发展重要指标，通过统计各年造林面积，更直观的看到退耕还林的显著效果，是否有助于生态恢复和环境保护。

补助标准：通过统计补助标准，分析退耕还林对农村居民的经济影响，国家对退耕还林的重视程度，分析退耕还林、非农就业、农民增收等问题。

国家投资：有助于帮助分析退耕还林对经济的影响，体现国家对退耕还林的重视程度，分析退耕还林对国家经济的影响。

技术研究：分析退耕还林带来对国家工业以及科技上的影响，为生态环境建设、森林植被恢复奠定了良好研究基础。

退耕还林活动的宣传教育：通过分析相关活动的宣传教育，体现退耕还林政策在大众的普及程度。

- 通过查阅《退耕还林经济报告》，我们选取以下指标作为国民经济的影响：

农村居民人均收入：旨在评估退耕还林在农民增收、非农就业和扶贫开发等方面的政策效果，评估参与工程的农户是否能够增加收入并提高家庭的生计水平。

人均可支配收入：对比农村居民人均收入，分析在减贫和调节收入分配方面起到了积极的作用。

GDP：评估退耕还林政策是否兼顾农业产业结构的调整与优化，分析退耕还林政策变量影响，宏观上分析退耕还林工程在促进农民增收和农业产业结构调整方面作用。

- 工业影响主要体现在第一产业向第二产业的转型，可选取以下指标：

工业增加值：分析退耕还林对工业的影响，计算工业增加值的同比增速，退耕还林之后产业的转型趋势，以及后续产业的发展现状。

工业资产统计：分析退耕还林之后工业资产，对比往年的数据，评估退耕还林对工业资产的影响，解读工业资产的改变与退耕还林政策的相关程度。

平均用工数：解读工业的平均用工数，分析退耕还林之后对农民发生的变化，产业的转型趋势，当地工业与农业的分配比例，以及退耕还林的后续产业对农民的收入以及就业的影响分析。

第二产业生产总值：评估退耕还林政策对第二产业的影响，分析第二产业的生产总值的增加率，评估退耕还林后产业结构是否合理，以及后续产业存在的问题与对策。

5.2.2 相关性分析

我们利用 Spearman 相关系数来探究造林面积、补助标准、国家投资、技术研究、宣传教育与国家工业水平、国民经济之间的整体相关性。斯皮尔曼等级相关系数为：

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

其中，n 是样本的数量， d_i 为 X_i, Y_i 之间的等级差。

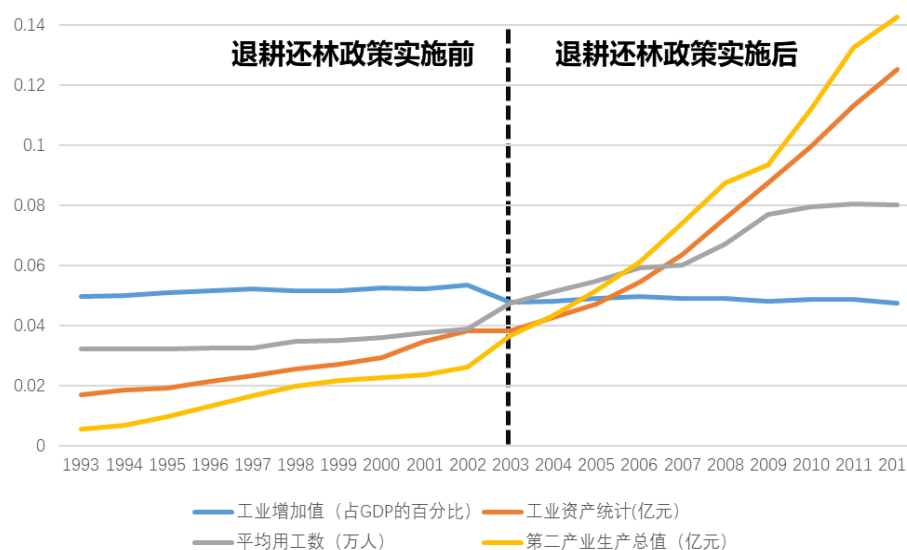
	国家工业水平	国民经济
	Spearman 相关系数	Spearman 相关系数
造林面积	-0.8112***	0.1399
补助标准	-0.0877	-0.3508
国家投资	0.7203***	-0.3427
技术研究	-0.7203***	0.7692***
宣传教育	0.2168	0.8042***

表 2 斯皮尔曼相关系数结果

由上表我们可以看出，对于国家工业水平来说，国家投资和宣传教育与工业增长值呈正相关，造林面积与之呈负相关且造林面积与之相关性最强；对于国民经济来说，技术研究、宣传教育与之呈正相关，宣传教育与之相关性更强。

5.2.3 数据预处理

通过国家统计局网站年鉴以及题目所给的附件，我们收集到了反映国家工业水平与国民经济的相关数据。具体如下表所示。为了将反映国家工业水平的四个指标的变化趋势反映在一张图上，我们通过对数据进行归一化处理来消除量纲的影响。反映国家工业水平的指标归一化结果如图所示：



从图中我们可以清晰地看出，在我国 2003 年实施退耕还林政策之前，工业资产统计、第二产业生产总值与平均用工数基本保持稳定的上升趋势但增长缓慢，在实施退耕

还林政策之后,这三项指标的增长速度有了明显的提升。而工业增加值在退耕还林政策实施之后有略微的下降趋势。由此我们可以粗略地得出结论:退耕还林政策的实施对我国工业水平有影响。

同样地,我们对国民经济的三个指标数据进行归一化处理来消除量纲影响,画出变化趋势图后同样可以得出“退耕还林政策的实施对我国国民经济有影响”的结论。

5.2.4 回归方程的建立

为了更严谨地分析退耕还林政策对我国工业水平与国民经济的影响,并揭示出影响的程度、主要影响了哪些角度,我们分别以工业的增加值、工业资产统计、平均用工数、第二产业生产总值为因变量,以反映退耕还林政策的五个指标为自变量,采用 OLS 普通最小二乘估计法进行标准化回归。

(1) OLS 回归

关于反映退耕还林的指标选取问题,我们主要参考了 2002 年 12 月 14 日中华人民共和国国务院令第 367 号发布,自 2003 年 1 月 20 日起施行的《退耕还林条例》。我们选取了造林面积、补助标准、国家投资、技术与研究与宣传教育五个指标。其中,我们通过该年度在中国知网上以“退耕还林”为关键词搜索显示的论文数目来衡量技术研究,以该年度“退耕还林”的百度搜索指数来衡量宣传教育。工业指标的回归结果如下:

表 3 工业指标的回归结果

		造林面积	补助标准	国家投资	技术研究	宣传教育	_cons
工业增 加值	回归系数	0.00	-0.01	0.00	0.00	0.02	42.75
	标准误	0.00	0.00	0.00	0.00	0.01	5.51
	T	-1.81	-1.91	1.46	-0.15	1.42	7.76
	标准化回归系数	-0.62	-0.39	0.38	-0.07	0.65	.
人均可 支配收入	回归系数	14.79	284.96	-0.05	143.96	-1716.82	579853.20
	标准误	9.81	119.68	0.02	90.60	373.15	165519.10
	T	1.51	2.38	-2.11	1.59	-4.60	3.50
	标准化回归系数	0.21	0.19*	-0.22*	0.29**	-0.84*	.
平均用 工数	回归系数	1.32	-5.10	0.00	-7.58	-57.28	25398.67
	标准误	0.52	6.32	0.00	4.78	19.69	8733.58
	T	2.56	-0.81	-1.41	-1.58	-2.91	2.91
	标准化回归系数	1.02*	-0.19	-0.42	-0.85*	-1.55*	.
第二产 业生产 总值	回归系数	8.82	177.50	-0.04	110.75	-1323.98	470932.80
	标准误	7.52	91.71	0.02	69.43	285.95	126836.10
	T	1.17	1.94	-2.29	1.60	-4.63	3.71
	标准化回归系数	0.16	0.15	-0.23*	0.29	0.83**	.

四个工业指标 $Adj R-squared$ 的值分别为 0.7089, 0.9536, 0.6082, 0.9558, 拟合效果较好。且各项 p 值分别对应为 0.0217, 0.0001, 0.0493, 0.0001 均小于 0.05, 四次回归结果都通过了联合显著性检验, 并且调整后的拟合优度均达到 0.9 以上。

同样的, 我们以农民人均收入、农民可支配收入、GDP 作为因变量, 进行标准化回归, 得到回归结果如下表所示:

表 4 国民经济指标回归结果

		造林面积	补助标准	国家投资	技术研究	宣传教育	_cons
农村居民人均收入	回归系数	0.38	10.72	0.00	3.49	-47.07	16684.66
	标准误	0.29	3.53	0.00	2.67	11.01	4883.61
	T	1.32	3.04	-2.59	1.31	-4.28	3.42
	标准化回归系数	0.20	0.27**	-0.29**	0.26	-0.85***	.
人均可支配收入	回归系数	0.77	18.60	0.00	6.57	-94.02	33215.70
	标准误	0.55	6.70	0.00	5.08	20.90	9272.22
	T	1.41	2.77	-2.52	1.29	-4.50	3.58
	标准化回归系数	0.20	0.24**	-0.27**	0.25	0.87***	.
GDP	回归系数	24.17	526.72	-0.10	238.59	-3132.29	1085117.00
	标准误	18.31	223.34	0.04	169.07	696.33	308867.20
	T	1.32	2.36	-2.38	1.41	-4.50	3.51
	标准化回归系数	0.19	0.20*	-0.25*	0.27	0.85***	.

三个经济指标 $Adj R-squared$ 的值分别为 0.9451, 0.9484, 0.9506 拟合效果较好。且各项 p 值分别对应为 0.0002, 0.0001, 0.0001, 均小于 0.05, 三次回归结果都通过了联合显著性检验。

(2) 异方差检验

如果扰动项存在异方差, 那么不仅 OLS 估计出来的回归系数是无偏、一致的, 而且假设检验无法使用 (构造的统计量失效了), 与此同时 OLS 估计量不再是最优线性无偏估计量 (BLUE)。因此我们选用 BP 检验来对回归结果进行异方差检验。

BP 检验的原理是既然在条件同方差下, 稳健标准误还原为普通标准误, 那么这二者之间的差别就可以用来度量条件异方差。

在同方差的原假设 $H_0: E(\varepsilon_i^2|X) = \sigma^2$ 下, 稳健协方差矩阵与普通协方差矩阵之差收敛到一个零矩阵。BP 检验结果如下:

表 5 BP 检验结果

	工业增加值	人均可支配收入	平均用工数	第二产业生产总值	农村居民人均收入	人均可支配收入	GDP
BP 检验	p=0.8231	p=0.771	p=0.1018	p=0.9571	p=0.8700	p=0.9160	p=0.9825

在 95% 的置信水平下, $p > 0.05$, 因此我们接受原假设, 即我们认为该回归结果的扰动项不存在异方差。

(3) 多重共线性检验

现将退耕还林政策的五个指标作为自变量, 假设第 m 个自变量的 $VIF_m = \frac{1}{1 - R_{1-k|m}^2}$,

$R_{1-k|m}^2$ 是将第 m 个自变量作为因变量, 对剩下的 4 个自变量回归得到的拟合优度。 VIF_m 越大 (此时 $R_{1-k|m}^2$ 越大), 说明第 m 个变量和其他变量的相关性越大。定义回归模型的 $VIF = \max \{VIF_1, VIF_2, \dots, VIF_5\}$ 检验结果如下:

表 6 回归模型检验结果

Variable	VIF	VIF	1/VIF
技术研究		8.09	0.123628
宣传教育		7.93	0.126088
造林面积		4.44	0.225069
国家投资		2.51	0.397899
补助标准		1.58	0.632953
Mean VIF		4.91	

分析结果可以发现, 变量的 VIF 均小于 10, 因此不存在多重共线性, 不影响到回归系数的显著性水平。

5.2.5 回归结果分析

在方程通过了异方差检验、多重共线性检验后, 我们来对回归结果进行分析。

首先, 退耕还林政策影响了我国的工业水平, 除工业增加值外, 人均可支配收入、平均用工数与第二产业生产总值均在不同的置信水平下, 受到了退耕还林政策中的某些方面的影响。

人均可支配收入受到了补助标准、国家投资、技术与研究及宣传教育的影响。其中, 技术研究对其影响最为显著, 表现为在 5% 的置信水平下, 技术研究每增长 1%, 人均可支配收入增长 0.29%。

平均用工数受到了造林面积、技术与研究及宣传教育的影响。在 10% 的置信水平下, 宣传教育每增长 1%, 平均用工数减少 1.55%, 造林面积每增加 1%, 平均用工数增加 1.02%, 技术研究每增加 1%, 平均用工数减少 0.85%。这与实际情况是相符的。

第二产业生产总值主要是受到了宣传教育的影响, 在 5% 的置信水平下, 宣传教育每增长 1%, 第二产业生产总值增加 0.85%。这说明, 当人民了解到退耕还林活动后, 从事第一产业的农民便会逐渐转向第二、三产业, 导致第二产业生产总值增加。

其次, 退耕还林政策影响了我国的国民经济。农村居民人均收入受到了补助标准、国家投资与宣传教育的影响, 其中受宣传教育影响最大。在 1% 的置信水平下, 宣传教育每增长 1%, 农村居民人均收入减少 0.85%。人均可支配收入受到补助标准、国家投资与宣传教育的影响, 在 1% 的置信水平下, 宣传教育每增长 1%, 人均可支配收入增加 0.87%。GDP 受到补助标准、国家投资与宣传教育的影响, 其中宣传教育每增长 1%, GDP 增加 0.85%。总体来说, 国民经济受到退耕还林政策的宣传教育影响最大。

5.3 经济、工业、环境评价模型

5.3.1 数据预处理

我们对附录《国家》表格中的国家进行筛选, 再将各个指标中缺失值较多的国家删去。对于由空缺值但空缺值不多的数据我们进行插值处理。最后利用 SPSS 检验异常值并删除, 最后剩下 99 个国家。然后, 我们再选取二氧化碳排放量、工业增加值、耗电量、能源使用量、农业用地、森林面积、森林面积占比作为国家工业、经济以及森林面积的评价指标, 并整合到同一张数据表中。

5.3.2 系统聚类

为了对世界各国经济水平、工业水平和森林植被面积进行评价,我们首先根据上述指标利用系统聚类法对 99 个国家进行分类,再分别研究它们的各项水平。

系统聚类的合并算法通过计算两类数据点间的距离,对最为接近的两类数据点进行组合,并反复迭代这一过程,直到将所有数据点合成一类,并生成聚类谱系图。

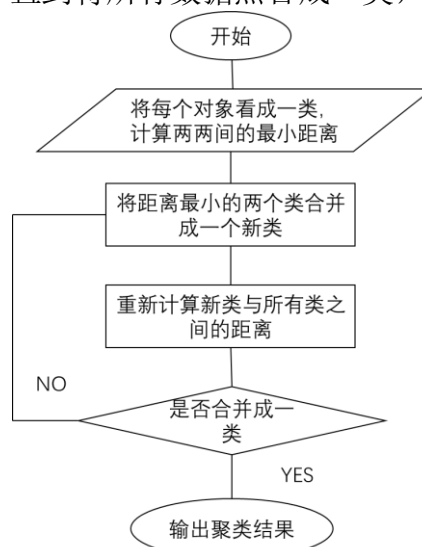


图 6 聚类谱系图

我们用欧几里得距离来计算两类数据点间的距离 $d(x, y) = [\sum_{k=1}^p |x_k - y_k|^2]^{\frac{1}{2}}$, 用类平均法来衡量类与类之间的距离 $D(G_1, G_2) = \frac{1}{n_1 n_2} \sum_{x_i \in G_1} \sum_{x_j \in G_2} d(x_i, x_j)$ 并用 SPSS 软件对 99 个国家的数据进行聚类分析。

我们使用平均联接（组间）的方法，由于生成的谱类系数图过大，我们将它放在了支撑材料中。根据谱系图可以将上述的 99 个国家分为三类，并对分类出来的国家进行特征分析：

	第一类	第二类	第三类
特征	二氧化碳排放量多、工业增加值和能源使用量较大、耗电量和农业用地少、森林面积和森林面积占比较多	二氧化碳排放量较多、工业增加值较大、耗电量较少、能源使用量较少、农业用地多、森林面积和森林面积占比少	由于第三类所涉及国家过多，其特征分析需要第二次聚类。
国家	俄罗斯联邦 巴西 加拿大	中国 美国	尼日尔 多哥 ...

表 7 第一次聚类分析结果分析

由第一次聚类分析我们可以发现，第一类国家经济比较发达，工业化水平高且它们的森林面积也很多；第二类国家工业化水平也很高，经济发展不如第一类国家且它们的森林面积覆盖不够多。而对于第三类国家，需第二次聚类来分析它们的特征。

删除第一类和第二类共计五个国家，对第三类的 94 个国家再次进行聚类分析，我

们再次使用平均联接（组间）的方法，对第三类国家聚类。

根据第二次的分类结果，我们又可将它们分为三类：

	第一类	第二类	第三类
特征	二氧化碳排放量少、工业增加值和耗电量较大、能源使用量较少、农业使用地较多、森林面积和森林面积占比较多	二氧化碳排放量较少、工业增加值较少、耗电量较少、能源使用量较少、农业用地多、森林面积和森林面积占比多	二氧化碳排放量较少、工业增加值较少、耗电量较少、能源使用量较少、农业用地多、森林面积和森林面积占比少
国家	澳大利亚 印度尼西亚	印度 南亚 德国 日本	尼日尔 多哥 乌拉圭 ...

表 8 第二次聚类结果分析

我们可以看出第三大类国家的二氧化碳的排放量都是偏少的，但是有些国家可能是由于工业化水平高的原因，可以控制碳排放，如：日本、澳大利亚等，它们的工业增加值大，说明它们的工业化水平已经很高了；而从能源消耗层面来看，它们的经济的发展不如前两大类国家，且这点可通过《国家》中的收入水平得以验证。而且第一小类与第二小类由于地形、气候的优势，它们的森林面积是高于第三小类的。

分类结果如下图所示：

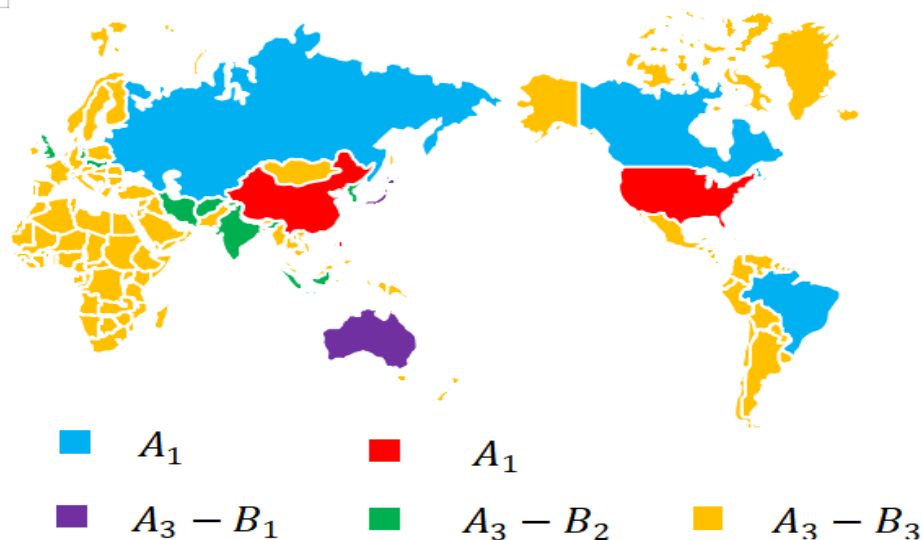


图 7 国家分类结果

5.3.3 Granger 因果关系分析

我们从三类国家中分别选出一个代表国家，对该国家长期以来在经济水平、工业水平和森林植被面积水平三个方面表现进行关系的探索。这里以第二类国家为例，我们选取中国为代表国家并且选取了三个指标作为经济、工业、森林这三个方面的代表，分别是耗电量、工业增加值和森林面积占土地面积的百分比。其中有必要说明的是耗电量与经济活力息息相关，能反应居民生活与工业生产水平，是“克强指数”的分项指标之一。

接下来我们运用单位根检验、协整检验以及格兰杰因果检验分析 1990-2014 年间中

国的工业增加值（占 GDP 的百分比）、耗电量（人均千瓦时）、森林面积（占土地面积的百分比）三者之间的关系。

（1）单位根检验（ADF 平稳性检验）

格兰杰因果检验的前提条件是数据是平稳的，因此我们先对数据进行单位根检验。如果不存在单位根，则时间序列平稳。否则我们通过差分的方式处理数据，直至数据表现出平稳性。

	ADF 值	临界值	结论
工业增加值	-1.653	-3.240*	非平稳
耗电量	-0.551	-3.240*	非平稳
森林面积	-1.860	-3.240*	非平稳
D(工业增加值)	-3.744	-3.600**	平稳
D(耗电量)	-3.362	-3.600**	非平稳
D(森林面积)	-1.020	-3.240*	非平稳
S(工业增加值)	-6.673	-4.380***	平稳
S(耗电量)	-5.652	-4.388***	平稳
S(森林面积)	-4.647	-4.380***	平稳

注：***，**，*分别代表 1%,5%,10%的显著性水平。D 表示一阶差分，S 表示二阶差分。

表 9 ADF 平稳性检验

由上表可以看出，在水平序列下，各变量都是不平稳的；在一阶差分下，除了工业增加值外，其余变量也都是不平稳的；在进行二阶差分以后，各变量都趋于平稳。

（2）协整检验

协整分析是在时间序列的向量自回归分析的基础上发展起来的时空结构与时间动态相结合的建模方法与理论分析方法。这里我们主要应用协整检验来区别变量之间长期均衡关系和短期波动关系。

因变量	自变量（系数）	Adj R-squared	F 值	协整检验
工业增加值	耗电量（0.1288）	0.0166	0.39	通过
森林面积	工业增加值（0.2842）	0.0408	2.02	未通过
耗电量	森林面积（0.9578）	0.9138	255.53	通过

表 10 协整检验

通过协整检验的结果可以看出，工业增加值与耗电量、耗电量与森林面积之间构成一个协整的系统，说明中国的工业增加值与耗电量之间、耗电量与森林面积之间存在着一个长期的均衡的关系。

（3）Granger 因果关系检验

格兰杰因果检验方法是分析时间序列变量之间的因果关系。考虑到经济中通常出现的时滞效应，在对时间序列进行因果关系检验时，我们将对滞后各期的变量之间关系进行检验，其检验结果列入下表内。

原假设	滞后阶数	P 值	结论
工业增加值不是耗电量的 Granger 原因	2	0.750	不拒绝
耗电量不是工业增加值的 Granger 原因	2	0.025	拒绝
森林面积不是工业增加值的 Granger 原因	2	0.073	不拒绝
工业增加值不是森林面积的 Granger 原因	2	0.795	不拒绝

耗电量不是森林面积的 Granger 原因	2	0.000	拒绝
森林面积不是耗电量的 Granger 原因	2	0.000	拒绝

表 11 格兰杰因果关系检验

由表可知，耗电量是工业增长值的 Granger 原因，但工业增长值不是耗电量的 Granger 原因，仅存在单项因果关系。再根据协整分析，我国耗电量与工业增长值之间存在长期均衡关系，揭示了长期以来我国耗电量与工业增长值之间的变化规律，协整系数为 0.1288，即耗电量每增加 1%，工业增长值增长 0.1288%，不仅显示我国耗电量对工业增长值具有正向性关系，同时也表明我国工业的增长在一定程度上依赖于国家经济水平，国家经济水平对工业增长起到了积极的促进作用。

耗电量是森林面积的 Granger 原因，森林面积也是耗电量的 Granger 原因，二者之间存在双项因果关系。再根据协整分析，我国耗电量与森林面积之间存在长期均衡关系，揭示了长期以来我国耗电量与森林面积之间的变化规律，协整系数为 0.9578，即耗电量每增加 1%，森林面积增长 0.9578%，不仅显示我国耗电量与森林面积之间具有正向性关系，同时也表明国家经济水平与国家森林资源之间具有互相促进的关系。

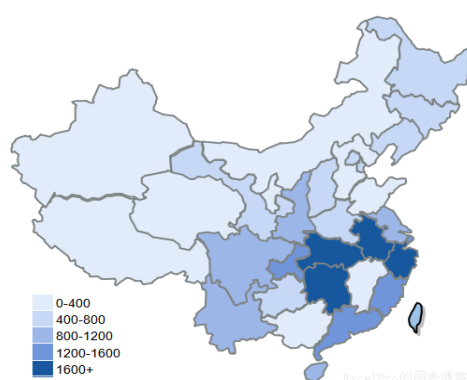


图 8 降水量分布

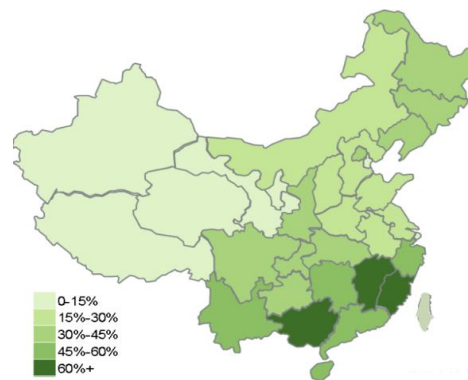


图 9 森林覆盖率

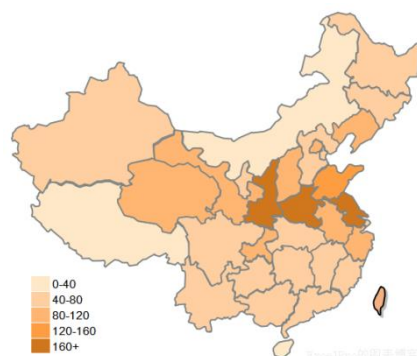


图 10 AQI 指数

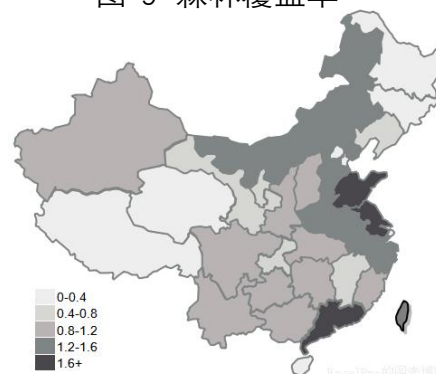


图 11 碳排放

通过搜集中国近年来在降水量、森林覆盖率、AQI 指数、碳排放的数据可知，虽然中国在经济、工业方面有明显的优势，但森林面积与森林面积占比数据不够理想，环境资源稀缺。且在森林覆盖率小的地方，AQI 指数偏高，降水量偏小，与第一大类国家环境还是有些差距。根据我们的因果分析，可以通过加大对环境治理的投资与研发高效的环境治理技术来填补这一劣势，尤其对于碳排放与空气质量较为严重的地区，更要提倡环境治理。并且，随着“绿水青山就是金山银山”这一号召的提出，以及 2003 年起我

国正式实行退耕还林政策，我们对中国能够改善这一局面抱有足够的信心。

同样地，根据上面的方法，在第一类国家中，我们选取俄罗斯联邦作为代表国家，对其在 1990-2014 年间的工业增加值（占 GDP 的百分比）、耗电量（人均千瓦时）、森林面积（占土地面积的百分比）数据进行单位根检验、协整检验以及格兰杰因果检验分析。

原假设	滞后阶数	P 值	结论	协整检验
工业增加值不是耗电量的 Granger 原因	2	0.210	不拒绝	未通过
耗电量不是工业增加值的 Granger 原因	2	0.796	不拒绝	未通过
森林面积不是工业增加值的 Granger 原因	2	0.005	拒绝	通过
工业增加值不是森林面积的 Granger 原因	2	0.281	不拒绝	通过
耗电量不是森林面积的 Granger 原因	2	0.000	拒绝	通过
森林面积不是耗电量的 Granger 原因	2	0.000	拒绝	通过

表 12 检验分析结果

由表可知，俄罗斯联邦在 1990-2014 年间，森林面积是工业增加值、耗电量的 Granger 原因。其中，森林面积与工业增加值是单项因果关系，即工业增加值不是森林面积的 Granger 原因，尽管它们之间有着长期的均衡关系。森林面积与耗电量之间是双向因果关系，耗电量是森林面积的 Granger 原因，二者之间具有相互促进的关系，

结合以俄罗斯联邦为代表的第二类国家特征，我们可以发现，这类国家的森林面积和森林面积占总地面的比例较大，具有良好的森林资源优势且环境治理水平较高。于是，可以通过大量的森林来达到固碳、涵养水源、节能减排的作用，从而减少在环境治理方面的投资，促进工业发展。

同样地，在第三类国家中，我们选取尼日尔作为代表国家，对其在 2000-2014 年间的工业增加值（占 GDP 的百分比）、耗电量（人均千瓦时）、森林面积（占土地面积的百分比）数据进行单位根检验、协整检验以及格兰杰因果检验分析。

原假设	滞后阶数	P 值	结论	协整检验
工业增加值不是耗电量的 Granger 原因	2	0.270	不拒绝	未通过
耗电量不是工业增加值的 Granger 原因	2	0.566	不拒绝	未通过
森林面积不是工业增加值的 Granger 原因	2	0.345	不拒绝	未通过
工业增加值不是森林面积的 Granger 原因	2	0.281	不拒绝	未通过
耗电量不是森林面积的 Granger 原因	2	0.350	不拒绝	未通过
森林面积不是耗电量的 Granger 原因	2	0.730	不拒绝	未通过

表 13 第三类国家分析

由表可知，尼日尔在 2000-2014 年间，耗电量、森林面积与工业增加值两两之间不存在任何的因果关系，且均不存在长期的均衡关系。我们结合尼日尔的国家经济水平、工业水平和森林资源水平分析可知，这很有可能是因为尼日尔的国家综合实力不够强大，在国际环境中，三个方面的水平均无明显地优势，三者之间彼此牵制，无法得到显著的发展。

5.4 关于各国经济、工业、森林的因果关系探究的调查报告

5.4.1.研究背景

习近平总书记说,“绿水青山就是金山银山”。环境的保护对国民经济的发展有着重要意义。因此探究中国经济、工业、森林的因果关系显得尤为重要。

5.4.2.研究分析

通过分析查找相关数据,我们构建数学模型,通过相应指标进行第一次类聚分析,分成三类国家:

第一类:二氧化碳排放量多工业增加值和能源使用量较大、耗电量和农业用地少、森林面积和森林面积占比较多,如俄罗斯联邦、巴西、加拿大。

第二类:二氧化碳排放量较多、工业增加值较大、耗电量较大、能源使用量较少、农业用地多、森林面积和森林面积占比少,如中国、英国。

第三类:由于第三类所涉及国家过多,其特征分析需要第二次聚类。如:尼日尔、多哥。

由第一次聚类分析我们可以发现,第一类国家经济比较发达,工业化水平高且它们的森林面积也很多;第二类国家工业化水平也很高,经济发展不如第一类国家且它们的森林面积覆盖不够多。而对于第三类国家,需第二次聚类来分析它们的特征。

第一类 B1:二氧化碳排放量少、工业增加值和耗电量大、能源使用量较少、农业使用地较多、森林面积和森林面积占比较多。如澳大利亚、印度尼西亚。

第二类 B2:二氧化碳排放量较少、工业增加值较少、耗电量较少、能源使用量较少、农业用地多、森林面积和森林面积占比多。如印度、南亚、德国、日本。

第三类 B3:二氧化碳排放量较少、工业增加值较少、耗电量较少、能源使用量较少、农业用地多森林面积和森林面积占比少。如尼日尔、多哥、乌拉圭。

我们可以看出第三大类国家的二氧化碳的排放量都是偏少的,但是有些国家可能是由于工业化水平高的原因,可以控制碳排放,如:日本、澳大利亚等,它们的工业增加值大,说明它们的工业化水平已经很高了;而从能源消耗层面来看,它们的经济发展不如前两大类国家,且这点可通过《国家》中的收入水平得以验证。而且第一小类与第二小类由于地形、气候的优势,它们的森林面积是高于第三小类的。

5.4.3.研究进程

因此,我们从三类国家中分别选出一个代表国家,对该国家长期以来在经济水平、工业水平和森林植被面积水平三个方面表现进行关系的探索。我们运用各国工业增加值、耗电量、森林面积三者之间的关系,进行单位根检验、协整检验以及格兰杰因果检验分析

第一类:以俄罗斯为例。俄罗斯联邦在 1990-2014 年间,森林面积是工业增加值、耗电的 Granger 原因。其中,森林面积与工业增加值是单项因果关系,即工业增加值不是森林面积的 Granger 原因,尽管它们之间有着长期的均衡关系。森林面积与耗电量之间是双向因果关系,耗电量是森林面积的 Granger 原因,二者之间具有相互促进的关系。

第二类:以中国为例。耗电量是工业增长值的 Granger 原因,但工业增长值不是耗

电量的 Granger 原因, 仅存在单项因果关系。再根据协整分析, 我国耗电量与工业增长值之间存在长期均衡关系, 揭示了长期以来我国耗电量与工业增长值之间的变化规律, 协整系数为 0.1288, 即耗电量每增加 1%, 工业增长值增长 0.1288%, 不仅显示我国耗电量对工业增长值具有正向性关系, 同时也表明我国工业的增长在一定程度上依赖于国家经济水平, 国家经济水平对工业增长起到了积极的促进作用。

第三类: 以尼日尔为例。尼日尔在 2000-2014 年间, 耗电量、森林面积与工业增加值两两之间不存在任何的因果关系, 且均不存在长期的均衡关系。

5.4.4. 研究结果及建议

我们根据上述分析结果给出了一些可行性建议:

(1) 以俄罗斯联邦为代表的第一类国家特征, 我们可以发现, 这类国家的森林面积和森林面积占总地面的比例较大, 具有良好的森林资源优势且环境治理水平较高。于是, 可以通过大量的森林来达到固碳、涵养水源、节能减排的作用, 从而减少在环境治理方面的投资, 促进工业发展。

(2) 根据中国国情和上述调查数据我们可以发现, 中国在经济、工业方面有明显的优势, 但森林面积与森林面积占比数据不够理想, 环境资源稀缺。根据我们的因果分析, 可以通过加大对环境治理的投资与研发高效的环境治理技术来填补这一劣势。并且, 随着“绿水青山就是金山银山”这一号召的提出, 以及 2003 年起我国正式实行退耕还林政策, 我们对能够改善这一局面抱有足够的信心。

(3) 我们结合尼日尔的国家经济水平、工业水平和森林资源水平分析可知, 这很有可能是因为尼日尔的国家综合实力不够强大, 在国际环境中, 三个方面的水平均无明显地优势, 三者之间彼此牵制, 无法得到显著的发展。因此, 对于像尼日尔这一类地区提升综合国力是最为重要的。如果单注重提高经济与工业方面反而会出现顾此失彼的画面。它们应制定能够综合发展的政策, 才能让它们得以更好的发展

5.4.5. 总结

以上是我们基于各国经济、工业、森林的因果探究。环境的保护对国民经济的发展有着重要意义。随着工业化进程的加快, 环境保护方面的问题逐渐加重。我们可以通过构建模型对相关变量进行分析, 及时解决全球的环境问题, 是我们刻不容缓的责任。

六、 模型评价与推广

6.1 模型优点

- 时间序列模型可找出变量变化的特征、趋势以及发展规律, 通过模型的检验与优化对变量的未来变化进行有效地预测。
- OLS 普通最小二乘估计法标准化回归效果好, 拟合优度高。
- 运用单位根检验、协整检验以及格兰杰因果检验分析工业、经济和森林面积之间的关系。

6.2 模型缺点

在应用时间序列分析法应注意未来发展变化规律和发展水平,不一定与其历史和现在的发展变化规律完全一致,当遇到外界发生较大变化,往往会有较大偏差。

6.3 模型的优化与推广

在我们本题所采用的时间序列预测模型中,是没有考虑 2025-2035 年的极端的自然情况的,一旦出现,那我们的预测值就会不准确。我们可以将时间序列模型优化,让它与贝叶斯统计分析相结合,变静态模型为动态模型的时间序列预测方法 其基本思想是将人们的经验信息作为已知条件结合到实际模型中,即利用模型信息、数据信息及先验信息来进行预测。那么我们就可以与气象局、地质局的数据预测,添加主观条件,使得预测更为准确。

七、 参考文献

- [1] 杜昕,李钢铁,马骏骥.基于 GM(1,1)模型群内蒙古森林资源的动态预测[J].内蒙古林业科技,2015,41(01):17-22.
- [2] 陶可.森林资源与经济发展水平的关系分析.企业导报 2012 年第 11 期
- [3] 邢祥娟.退耕还林对农户收入的影响研究[D].北京林业大学,2014.
- [4] 各地区退耕还林情况.国家统计局 2003-2014
- [5] 刘豹,胡代平.神经网络在预测中的一些应用研究[J].系统工程学报,1999(04):338-344.
- [6] 丁守奎,王洁贞,冀友瑞,胡平.时间序列预测模型的贝叶斯统计分析[J].中国公共卫生,2002(09):119-120.
- [7] 申亚娟.森林资源保护与区域经济协调发展的关系探析[J].农业灾害研究,2021,11(10):139-140.
- [8] 黄森.浅谈我国森林资源保护与区域经济协调发展[J].科技创新与应用,2017(16):283.

附录

附录一 回归 (Stata)

```
import excel "C:\Users\93945\Desktop\回归数据.xlsx", sheet("经济")
firstrow

reg y1 x1 x2 x3 x4 x5,b
estat hettest,rhs iid
reg y2 x1 x2 x3 x4 x5,b
estat hettest,rhs iid
reg y3 x1 x2 x3 x4 x5,b
estat hettest,rhs iid

estat vif

import excel "C:\Users\93945\Desktop\回归数据.xlsx", sheet("工业")
firstrow clear
reg y1 x1 x2 x3 x4 x5,b
estat hettest,rhs iid
reg y2 x1 x2 x3 x4 x5,b
estat hettest,rhs iid
reg y3 x1 x2 x3 x4 x5,b
estat hettest,rhs iid
reg y4 x1 x2 x3 x4 x5,b
estat hettest,rhs iid

estat vif
```

附录二 Granger 因果分析 (Stata)

```
import excel "C:\Users\93945\Desktop\Q3.xlsx", sheet("中国") firstrow
clear
tsset 年份

dfuller 工业增加值占 GDP 的百分比,trend
dfuller 耗电量人均千瓦时 ,trend
dfuller 森林面积占土地面积的百分比 ,trend

gen d1 工业增加值占 GDP 的百分比=d.工业增加值占 GDP 的百分比
gen d1 耗电量人均千瓦时 =d.耗电量人均千瓦时
gen d1 森林面积占土地面积的百分比 =d.森林面积占土地面积的百分比
dfuller d.工业增加值占 GDP 的百分比,trend
dfuller d.耗电量人均千瓦时 ,trend
dfuller d.森林面积占土地面积的百分比 ,trend
```

```

gen d2 工业增加值占 GDP 的百分比=d2.工业增加值占 GDP 的百分比
gen d2 耗电量人均千瓦时 =d2.耗电量人均千瓦时
gen d2 森林面积占土地面积的百分比 =d2.森林面积占土地面积的百分比
dfuller d2.工业增加值占 GDP 的百分比,trend
dfuller d2.耗电量人均千瓦时 ,trend
dfuller d2.森林面积占土地面积的百分比 ,trend

reg d2.工业增加值占 GDP 的百分比 d2.耗电量人均千瓦时
var 工业增加值占 GDP 的百分比 耗电量人均千瓦时,lag(2)

reg d2.工业增加值占 GDP 的百分比 d2.森林面积占土地面积的百分比
reg d2.森林面积占土地面积的百分比 d2.工业增加值占 GDP 的百分比
var 工业增加值占 GDP 的百分比 森林面积占土地面积的百分比 ,lag(2)

reg d2.耗电量人均千瓦时 d2.森林面积占土地面积的百分比
var 耗电量人均千瓦时 森林面积占土地面积的百分比 ,lag(2)

```

附录三 支撑材料索引

1. Granger.dta
2. 回归数据.xlsx
3. 聚类系谱图 1.png
4. 聚类系谱图 2.png
5. 问题三数据.xlsx
6. 中国森林面积预测.xlsx