

基于统计学视角的区域协调发展及环境治理效率问题研究

致理-数理 1 刘苏青* 2021013371

未央-能动 02 杨华康† 2020012990

2023 / 06 / 24

摘 要

本文基于 2006-2015 年间 31 个省份关于经济、社会和生态发展的一系列相关指标，采用熵值法、耦合协调度发展模型等建模方法，深入探讨了各省份生态-经济-社会复合系统的耦合协调发展情况，并选择长三角、京津冀和东三省三个典型区域作为研究对象，采用回归分析等统计学方法对比分析这些区域的生态环境的治理效率问题，并提出具有针对性的政策建议。有助于人们在结构调整上深下功夫，进一步优化调整产业布局，促进人与自然的和谐共生。

关键词：熵值法；耦合协调度；多重线性回归；贝叶斯信息准则 (BIC)

*liu-sq21@mails.tsinghua.edu.cn

†yanghk20@mails.tsinghua.edu.cn

目录

1	引言	3
2	文献综述	3
3	探索性数据分析	4
3.1	数据集描述	4
3.2	相关性探索分析	4
4	复合系统耦合协调发展情况	6
4.1	熵值法构建指标体系	6
4.2	构建耦合协调度发展模型	9
5	生态环境治理效率问题研究	11
6	总结与改进	15
7	附录	16
7.1	R 代码	16
7.2	参考文献	26

1 引言

在中国过去几十年的发展中，一定程度上以牺牲环境为代价换取国民经济的高速增长。这种粗放式的发展理念带来了生态环境日益恶化、产业结构不合理等一系列问题，引发了一些社会矛盾。经济发展是社会福祉得以实现的物质基础，社会效益提升是经济发展和生态保护的最终目标，而可持续的生态系统又是经济高质量发展和社会福祉提升的重要保障。在当前，实现经济发展、生态文明和社会保障等多方面的协调发展，是中国社会发展过程中的一个亟待解决的问题。所谓生态-经济-社会协调发展，就是指生态、经济、社会实现相互之间的良性循环，达到经济稳定发展、生态状况良好、社会效益提升的一种有序状态。

如今，“绿水青山就是金山银山”的思想已经深入到了人们心中，追求“人与自然和谐共生的现代化”已经成为了社会共识。因此，研究生态-经济-社会复合系统的协调发展情况具有重要的理论和实践意义，能为可持续高质量发展提供借鉴，能提升人民群众的获得感、幸福感、安全感。

本研究以 31 个省份在 2006-2015 年间的发展状况为研究对象，对生态-经济-社会系统的耦合协调发展建立评价指标体系，并采用障碍度模型分析影响耦合协调发展的障碍因素，以期为生态-经济-社会系统的高质量耦合协调发展提供建议。在此基础上，选择部分典型区域，通过回归分析来研究生态环境治理效率问题，提出更有针对性的建议。进一步，通过引入固定效应模型，处理固定效应带来的影响，对生态-经济-社会系统的耦合协调发展做更深入的探讨。

2 文献综述

对于生态-经济-社会复合系统的研究，是首先从二元系统开始的，之后逐渐向三元系统演化。

对“生态-经济”二元系统，美国著名经济学家 G.Grossman 和 A.Kureger 在 1991 年提出了 EKC 曲线 (Environmental kuznets curve)，认为生态环境与经济发展间存在倒“U”形关系，即随着经济发展水平提高，生态环境会先恶化，达到某个临界点后开始不断改善。这一观点得到了国内外学者实证分析的肯定。对“经济-社会”二元系统，有许多经济学的经典论点，如庇古在《福利经济学》中所提出的“经济增长有利于社会福利的改善”。

而对于“生态-经济-社会”三元系统的研究，李茜等从生态、经济、社会三大系统

出发，构建了综合发展指标体系，研究发现中国三大系统协调度不断增强，但存在明显的空间差异。初雪等研究发现欠发达地区的经济发展水平显著滞后于生态和社会发展水平，三者协调度较低，并进一步探索了经济欠发达地区的“生态-经济-社会”协调模式。王琦等对洞庭湖区“生态—经济—社会”耦合协调发展的时空分异进行分析，发现经济子系统和生态子系统之间的矛盾是制约洞庭湖地区整体协调发展水平的主要原因。吴义刚等对安徽农村地区的“生态—经济—社会”耦合协调发展进行研究，发现农业发展水平较低是安徽多数地区的突出问题。

3 探索性数据分析

3.1 数据集描述

本研究所使用的数据来自课程所提供的数据集。这是一份从 2006-2015 年间 31 个省份 (不包含香港、澳门和台湾) 关于经济、社会和生态发展的一系列相关指标。我们对其进行分类，国内生产总值等 8 个变量属于经济数据，年末常住人口等 14 个变量属于社会数据，建成区面积等 13 个变量属于生态数据。

研究数据来源于 2006-2015 年的《中国统计年鉴》、《中国城市统计年鉴》、《中国环境统计年鉴》、《中国环境年鉴》、《中国财政年鉴》及《中国人口统计年鉴》。

3.2 相关性探索分析

经济、社会、生态在发展的同时，是相互影响的，因此我们对经济数据、社会数据、生态数据进行两两相关性探索分析。

我们以安徽省和北京市为例，绘制经济数据、社会数据之间的相关性热力图，可以发现两组变量之间存在较为明显的相关性（图中变量从左下至右上为经济和社会指标，因此观察三角形右下角可得到两组变量间的相关性），并且可以注意到部分变量间为负相关性，这对后续的探索也有一定的启发性。

然后绘制安徽、北京经济数据与生态数据之间的相关性热力图，可以发现仍然有较强的相关性。这说明在发展过程中，经济、社会、生态子系统不是割裂的，而是存在着耦合的关系，也就意味着协调发展的水平将会是我们在综合评价某地发展时的重点。

而对于各项指标的变化趋势，由于我们在下文将对经济、社会、生态子系统的发展水平做较详细的讨论，此处不做过多分析。

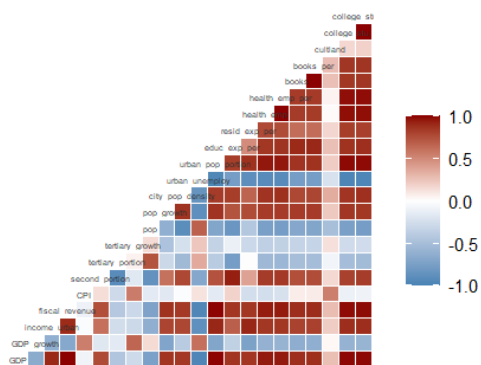


Figure 1: 安徽经济、社会数据相关性

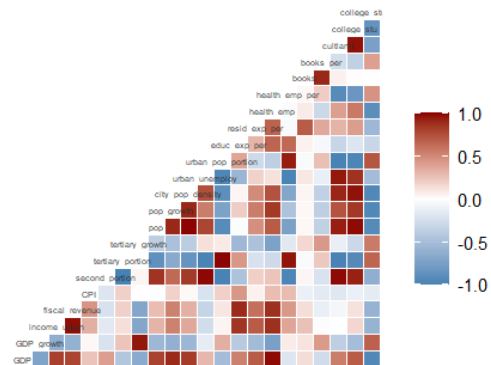


Figure 2: 北京经济、社会数据相关性

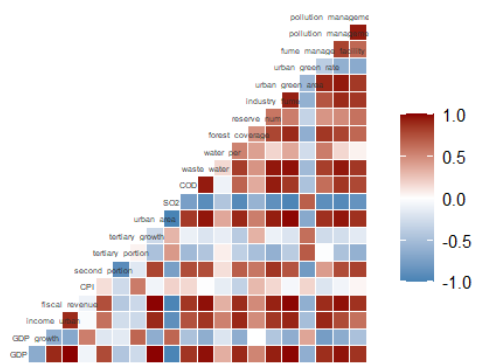


Figure 3: 安徽经济、生态数据相关性

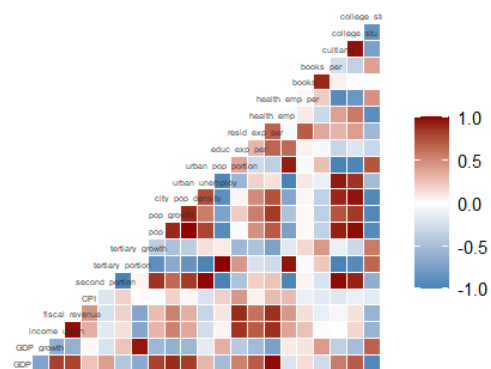


Figure 4: 北京经济、生态数据相关性

4 复合系统耦合协调发展情况

本研究从经济子系统、社会子系统、生态子系统的耦合协调关系出发，共使用 35 个指标构建生态-经济-社会系统耦合协调度评价指标体系。构建过程包括：将 35 个指标进行划分，分别属于经济子系统、社会子系统、生态子系统，然后采用熵值法计算各指标的权重。基于指标权重，计算 2006-2015 年间 31 个省份的经济子系统、社会子系统、生态子系统综合发展指数，绘制子系统综合发展指数的变化趋势图。根据各省综合发展指数的数据，结合耦合协调度模型，测算得出 2006-2015 年间 31 个省份的生态-经济-社会系统耦合协调度，并绘制变化趋势图。

4.1 熵值法构建指标体系

在信息论中，熵是对不确定性的一种度量。我们可以通过计算熵值来判断一个事件的随机性及无序程度，也可以用熵值来判断某个指标的离散程度，一个指标的离散程度越大，该指标对综合发展水平的影响权重越大。根据这一特性，我们可以采用熵值法来衡量各省份每年的生态、经济和社会的综合发展水平，步骤如下：

首先，我们将数据集划分为经济、社会和生态三个子系统。对于每一个子系统，其数据矩阵记为 $X_{n \times q}$ ，其中 X_{ij} 代表第 i 个样本的第 j 个指标数值。由于各指标的尺度、量纲以及和总体指标的正负相关性可能存在不同，我们需要采用极差法对各指标的原始数据分正向指标和负向指标分别归一化。同时，为避免归一化后的数据出现极端数据 0 或 1，我们对传统的极差法做适当的收缩¹：

- 正向指标：

$$z_{ij} = 0.002 + (0.098 - 0.002) \cdot \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}}$$

- 负向指标：

$$z_{ij} = 0.002 + (0.098 - 0.002) \cdot \frac{\max_i \{x_{ij}\} - x_{ij}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}}$$

然后，根据归一化的数据矩阵 $Z_{n \times q}$ ，我们可以计算第 i 个样本的第 j 个指标数值的占比 p_{ij} ：

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^n z_{ij}}$$

¹收缩后的归一化方法只会得到 $[0.002, 0.998]$ 之间的数据。

进而可以计算第 j 个指标数值的熵值 e_j 和权重 w_j ，各变量权重计算结果如表所示：

$$e_j = -\frac{1}{\ln q} \sum_{i=1}^n (p_{ij} \cdot \ln p_{ij})$$

$$w_j = \frac{1 - e_j}{\sum_{j=1}^q (1 - e_j)} = \frac{1 - e_j}{q - \sum_{j=1}^q e_j}$$

子系统	指标层	权重 w_i
经济子系统	国内生产总值	0.1184
	GDP 增速	0.1290
	城镇居民人均可支配收入	0.1218
	地方一般公共预算收入	0.1177
	通货膨胀指数	0.1296
	第二产业 GDP 占比	0.1294
	第三产业 GDP 占比	0.1247
	第三产业 GDP 增速	0.1294
社会子系统	年末常住人口	0.0707
	人口自然增长率	0.0740
	城市人口密度	0.0728
	城镇登记失业率	0.0748
	城镇人口比重	0.0738
	城镇居民家庭人均教育文化消费支出	0.0731
	城镇居民家庭人均居住消费支出	0.0642
	卫生机构人数	0.0709
	人均卫生机构人员数	0.0740
	公共图书馆总藏量	0.0697
	人均公共图书馆总藏量	0.0687
	耕地面积	0.0691
	高等学校在校学生数	0.0709
	人均高等学校在校学生数	0.0732
生态子系统	建成区面积	0.0845
	二氧化硫排放量	0.0837
	化学需氧量排放量	0.0840
	废水排放总量	0.0846
	人均水资源量	0.0365
	森林覆盖率	0.0793
	自然保护区个数	0.0752
	工业废气排放总量	0.0846
	城市绿化覆盖面积	0.0738
	建成区绿化覆盖率	0.0769
	废气治理设施数	0.0784
	环境污染治理投资总额	0.0759
	环境污染治理投资占 GDP 比重	0.0822

Table 1: 生态-经济-社会系统耦合协调度评价指标体系及指标权重

最后，根据归一化的数据矩阵 $Z_{n \times q}$ 和第 j 个指标数值的权重 w_j ，我们可以计算这个子系统第 i 个样本的综合发展水平值²：

$$S_i = \sum_{j=1}^q z_{ij} w_j$$

根据以上计算结果，我们从 31 个省份中选择长三角、京津冀、东三省等典型区域，对其区域内各省份经济、社会和生态子系统的综合发展水平值绘制折线图如下：

可以发现，在 2006-2015 年间，长三角各省份的经济和社会综合发展水平都表现出显著的上升趋势，但生态综合发展水平却只有上海呈现出上升趋势，另外三省均表现出

²这里的 S_i 只表示某一个子系统第 i 个样本的综合发展水平值，不同子系统的综合发展水平值可能不同。

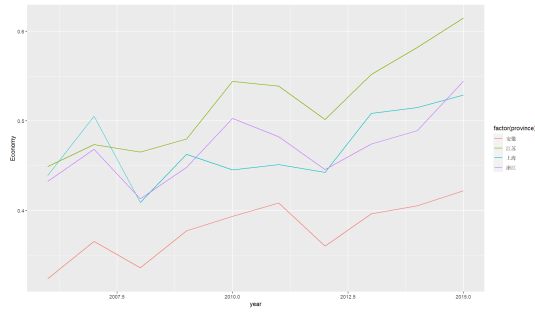


Figure 5: 长三角经济指数

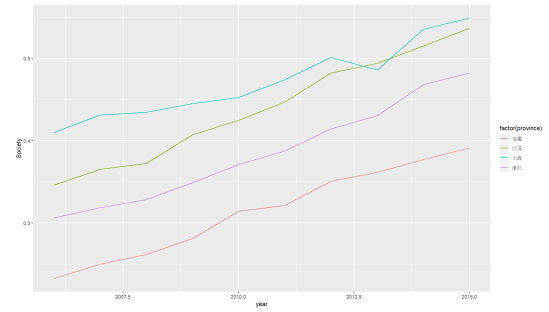


Figure 6: 长三角社会指数

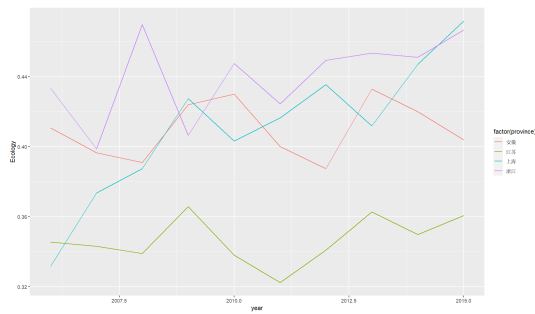


Figure 7: 长三角生态指数

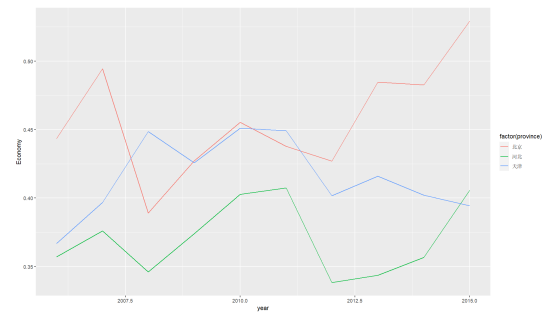


Figure 8: 京津冀经济指数

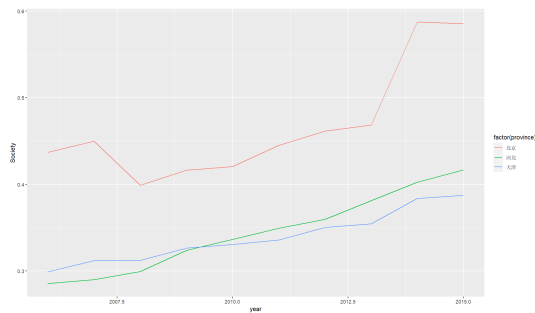


Figure 9: 京津冀社会指数

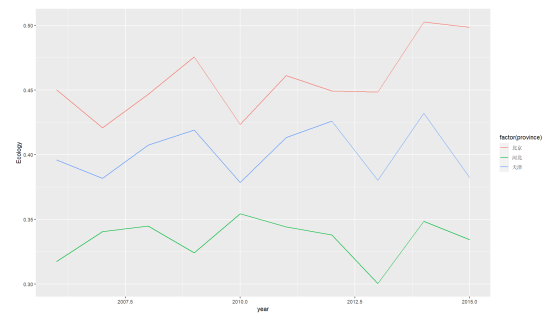


Figure 10: 京津冀生态指数

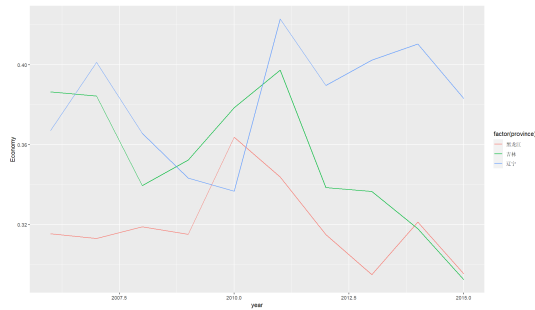


Figure 11: 东三省经济指数

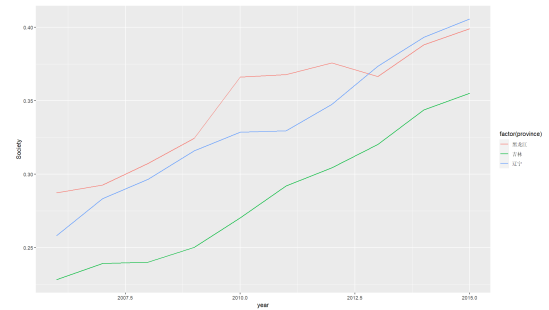


Figure 12: 东三省社会指数

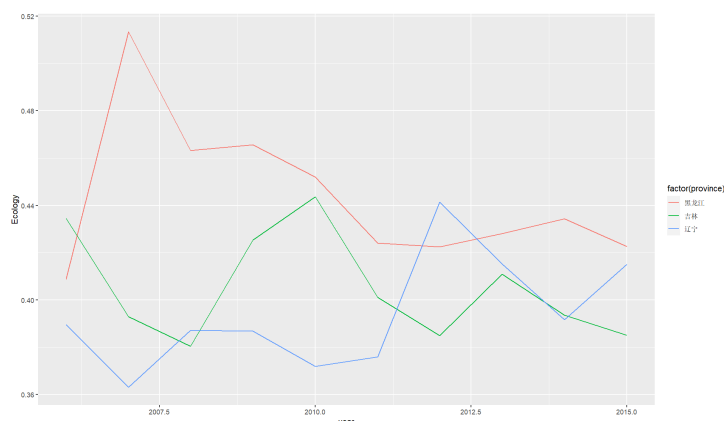


Figure 13: 东三省生态指数

一定范围内的上下波动。这说明，在这一时期，长三角地区的发展带动了经济效益和社会效益的显著提升，但生态效益未出现显著改善。因此未来长三角地区应将生态效益放到更重要的地位上，聚焦生态环境改善、加大生态环境保护力度。

而在京津冀地区，三个省份在经济和社会发展水平上均有提升，但北京的经济、社会发展水平的提升幅度显著超过天津和河北，对于生态，北京也是唯一实现显著提升的省份，天津地区在 2015 年的生态水平甚至低于 2006 年，河北地区也只是小幅提升。这说明在这一时期，北京地区的各项发展均优于河北和天津，一定程度上反应了京津冀发展不平衡的问题；而河北和天津在生态上的发展明显不足，未来需要在保证生态的条件下提升经济和社会效益。

在东三省地区，情况与另外两个地区有明显不同。三个省份的经济发展水平都没有显著提升，吉林省甚至出现了明显的下滑，另外两省也处于波动。三省的生态发展水平都处于小幅波动中，吉林省出现了小幅下滑。只有在社会发展水平上，三省均实现了明显的提升。这说明，东三省在经济发展速度上与长三角、京津冀存在较为明显的差距，而生态发展基本维持稳定，未来东三省应将重心放在经济发展上，当然也应该注意与生态保护相协调。

以上分析为从生态、经济、社会各自的综合发展水平出发进行讨论，为下文对生态-经济-社会复合系统耦合协调度的讨论奠定基础。

4.2 构建耦合协调度发展模型

运用耦合协调度发展模型，研究经济子系统、社会子系统和生态子系统的耦合协调度水平。其中，耦合度用于反映各子系统间的相互影响程度，耦合度越小，说明各子系统间的相互影响程度越低，反之则说明各子系统间相互影响程度越高。耦合度的计算方

法为：记 S_A 、 S_B 、 S_C 分别代表经济、社会和生态子系统的综合发展水平值，我们可以以此构建耦合协调发展模型，并计算第 i 个样本的耦合度 C_i ：

$$C_i = 3 \left[\frac{S_{Ai} \times S_{Bi} \times S_{Ci}}{(S_{Ai} + S_{Bi} + S_{Ci})^3} \right]^{\frac{1}{3}}$$

进而可以计算第 i 个样本的耦合协调度 D_i ：

$$D_i = \sqrt{C_i \times T_i}$$

其中 $T_i = \alpha \times S_{Ai} + \beta \times S_{Bi} + \gamma \times S_{Ci}$ ， α 、 β 和 γ 均为待定系数。在本研究中，我们假设经济子系统、社会子系统和生态子系统对城市的综合发展具有同等重要的程度，因此 $\alpha = \beta = \gamma = \frac{1}{3}$ 。易知 $D_i \in (0, 1]$ ， D_i 越接近 1，表明经济子系统、社会子系统和生态子系统之间的良性耦合水平越高，三者的发展处于和谐共进的状态。

将上述理论应用到本研究中，计算得到 31 个省份在 2006-2015 年期间每一年的耦合协调度，并选择长三角、京津冀、东三省这三个区域，绘制区域内各省份的耦合协调度变化折线图：

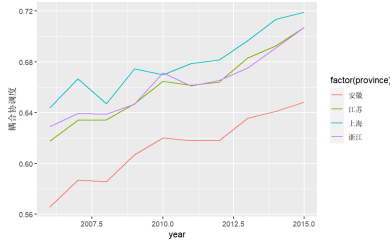


Figure 14: 长三角耦合协调度

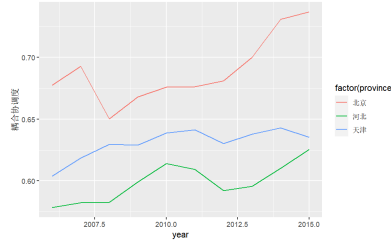


Figure 15: 京津冀耦合协调度

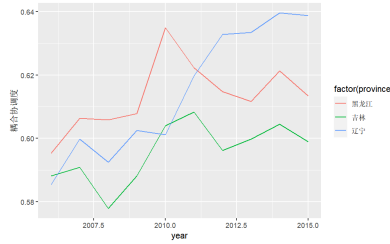


Figure 16: 东三省耦合协调度

参考汪永生等的研究，将耦合协调度分为 10 个等级：

[0,0.1]：极度失调，(0.1,0.2]：严重失调，(0.2,0.3]：中度失调，(0.3,0.4]：轻度失调，(0.4,0.5]：濒临失调。

(0.5,0.6]：勉强协调，(0.6,0.7]：初级协调，(0.7,0.8]：中度协调，(0.8,0.9]：良好协调，(0.9,1.0]：优质协调

可以发现, 在 2006-2015 年间, 长三角省份中的江浙沪三省的耦合协调度基本维持在一个较高的水平小幅波动, 均处于初级协调 (0.6-0.7) 的状态; 而安徽则从 0.5659 提升到 0.6484, 已经跨入初级协调的阶段。这一时期的长三角在生态-经济-社会协调发展上处于全国的领先地位, 优秀的协调发展水平与长三角地区雄厚的综合实力是分不开的。

而京津冀地区仍然体现出了不平衡的特点——北京的耦合协调度在 0.65 到 0.73 之间波动, 基本处于初级协调的阶段, 而天津的耦合协调度在 0.60 到 0.65 的范围内波动, 河北的耦合协调度在 0.57 到 0.63 的范围内波动, 两省都属于勉强协调的阶段, 但并未表现出提升的趋势。未来河北和天津应重视生态-经济-社会协调发展, 力争向北京学习, 三地应协力形成京津冀地区更加协同的发展。

与子系统综合发展水平的分析结果类似, 东三省地区的情况与另外两个地区有略有不同。辽宁的耦合协调度在 0.58 到 0.64 的范围内波动, 黑龙江的耦合协调度在 0.59 到 0.63 的范围内波动, 吉林的耦合协调度在 0.57 到 0.60 的范围内波动, 虽然其起步值与辽宁和黑龙江基本相同, 但从 2011 年开始相对另外二省呈现出一定的差距。这一时期的东三省在生态-经济-社会协调发展上处于全国的落后地位, 与长三角、北京等先进地区有较为明显的差距, 并且吉林省提升的趋势较缓。我们认为这与东北三省较为疲软的经济有关, 未来东北三省实现协调发展的关键在于提振经济发展的活力。

5 生态环境治理效率问题研究

我们继续选取长三角、京津冀以及东三省地区作为研究区域, 研究这些区域的生态环境的治理效率问题。考虑到生态环境的治理效率可以用各种污染物指数来衡量, 我们可以从生态子系统的各指标中选取各种污染物的排放情况作为因变量 (共 4 个), 并将其余各指标作为自变量 (共 9 个), 分别进行多重线性回归。

我们考虑过在每个区域内部根据省份对数据分别进行回归, 然而数据集中每个省份的数据只有 10 个, 自变量个数却有 9 个。在 $n = p = 10$ 的条件下, 后续的数据分析过程可能出现问题, 这导致我们不得不采用正则化方法来进行处理, 例如引入 L_2 惩罚项来进行岭回归或者引入其他惩罚项。为了避免选择正则化参数 λ 带来的复杂度, 我们可以引入简单的领域知识: 每个区域内的经济-社会-生态发展可以视为协同的。因此, 在后续的数据分析过程中, 我们将每个区域内部的各个省份进行合并处理, 只研究各区域整体的生态环境的治理效率问题。

首先, 我们使用 `shapiro_test` 函数对各指标进行检验, 观察各指标是否服从正态分

布。如果存在指标未通过 shapiro_test 检验，可以使用 density 函数检查该指标的分布情况，若不存在明显的偏态分布，则直接线性回归的结果仍然稳健。经验证，对于长三角、京津冀和东三省地区，均不存在明显的偏态分布，因此无需对各指标进行变换。

然后，我们使用 BIC 方法来筛选有效变量。相对于 $\text{adjust-}R^2$ 方法，BIC 方法的 $\log n \cdot p$ 项可以有效控制 p 的数量，因而可以在找到有效变量的同时保证 p 较小。考虑到二氧化硫排放量是非常典型且重要的一种污染指标，下面以二氧化硫排放量为因变量做分析。以长三角、京津冀和东三省地区的二氧化硫排放量作为因变量筛选的有效变量如下：

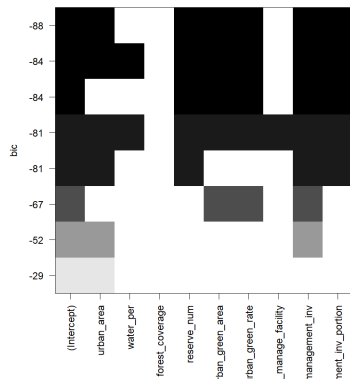


Figure 17: 长三角 BIC

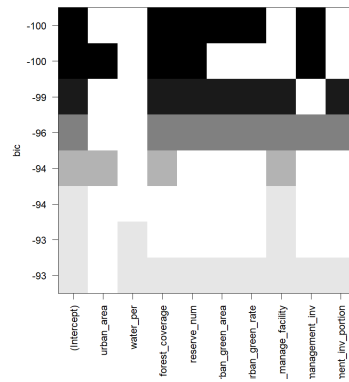


Figure 18: 京津冀 BIC

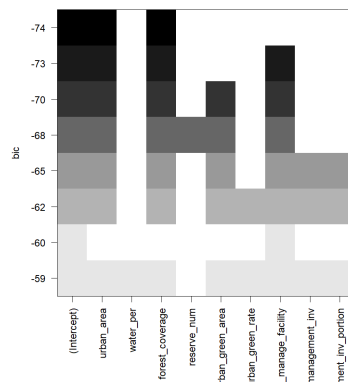


Figure 19: 东三省 BIC

最后，我们取因变量和其对应的有效变量进行回归。以长三角、京津冀和东三省地区的二氧化硫排放量作为因变量的回归结果如下：

Table 2: 长三角二氧化硫排放量回归分析

	Dependent variable:
	SO2
urban_area	0.027** (0.010)
reserve_num	-0.254*** (0.056)
urban_green_area	0.039*** (0.011)
urban_green_rate	-54.300*** (14.448)
pollution_management_inv	-0.179*** (0.018)
pollution_management_inv_portion	35.505*** (5.603)
Constant	21.122 (15.823)
Observations	40
R ²	0.941
Adjusted R ²	0.931
Residual Std. Error	7.600 (df = 33)
F Statistic	88.494*** (df = 6; 33)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: 京津冀二氧化硫排放量回归分析

	Dependent variable:
	SO2
forest_coverage	-2.465*** (0.318)
reserve_num	4.541*** (0.279)
Constant	12.604 (8.085)
Observations	30
R ²	0.908
Adjusted R ²	0.901
Residual Std. Error	17.605 (df = 27)
F Statistic	132.946*** (df = 2; 27)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: 东三省二氧化硫排放量回归分析

	Dependent variable:
	SO2
urban_area	0.068*** (0.008)
water_per	-0.006*** (0.002)
forest_coverage	-6.698*** (0.806)
reserve_num	0.047 (0.029)
urban_green_area	-0.033** (0.013)
urban_green_rate	53.696** (21.117)
pollution_management_inv	-0.061** (0.026)
pollution_management_inv_portion	17.832*** (6.006)
Constant	214.835*** (39.738)
Observations	30
R ²	0.989
Adjusted R ²	0.985
Residual Std. Error	3.958 (df = 21)
F Statistic	232.481*** (df = 8; 21)
Note:	*p<0.1; **p<0.05; ***p<0.01

我们基于上述结果进行讨论与分析。与长三角地区二氧化硫排放量关联性最显著的是环境污染治理投资总额，注意到自然保护区个数的关联性较弱。与京津冀地区二氧化硫排放量关联性最显著的自然保护区个数，而与东三省二氧化硫排放量关联性最显著的是则森林覆盖率，注意到自然保护区个数的关联性较弱，另外废水排放总量是关联性第二强的自变量。

我们结合环境治理的现实情况进行讨论。长三角地区经济发达、科技先进，在环境治理上投入较多，这也是提升长三角生态环境治理效率的主要动力，即用严格监管、技术进步来推动环境治理，因此环境污染治理投资总额与二氧化硫排放量的关联性很强；而长三角地区多为城市，自然保护区数量很少，对环境治理的贡献较小，因此关联性较弱。事实上，自然保护区对治理二氧化硫的贡献主要来自于森林对二氧化硫的吸收作用。而京津冀地区在 2006-2015 年间空气污染情况较为严重，一大原因便是二氧化硫控制不力（在数据中也能得到验证，例如河北省的二氧化硫排放量大大超过长三角各省份），我们推测京津冀地区在这一时期并未有很大的环境治理投资，因此自然保护区对二氧化硫治理的贡献便凸显出来。东三省地区有着丰富的森林资源，而森林在治理二氧化硫方面有着重要作用，因此森林覆盖率成为与二氧化硫排放量关联性最强的变量。但同时我们

注意到，环境治理投资的关联性较弱，说明东三省在对环境治理的投资还不够大，与长三角地区有着明显差距（从数据中也可以得到直接验证）。另外，废水排放总量与二氧化硫排放量表现出较强的关联性且系数为正，这说明东三省的污水排放控制做的不好，这间接导致了其他污染情况的治理。

综上所述，长三角地区应继续保持对环境治理的高投入，加强监管、加大科技投入，持续巩固环境治理成果；京津冀地区应加强三地协同发展，共同提高对环境治理的重视，以期改善环境治理的现状；东三省在拥有较好森林资源的情况下，应加大对环境治理的投入，并且重视对多种污染的同时控制，避免出现污染之间相互加剧的情况。

6 总结与改进

本研究首先通过构建生态-经济-社会系统耦合协调度评价指标体系，探讨长三角、京津冀、东三省三个典型地区的经济、社会生态子系统综合发展水平，进而运用耦合协调度发展模型，研究三个地区在 2006-2015 年期间经济子系统、社会子系统和生态子系统的耦合协调度水平。得到以下结论：

- 长三角在生态-经济-社会协调发展上处于全国的领先地位，其中江浙沪三省的耦合协调水平均处于中度协调（0.7 0.8）的状态，安徽虽然相比江浙沪尚有差距，但表现出显著的上升趋势，即将跨入中度协调的阶段。优秀的协调发展水平与长三角地区雄厚的综合实力是分不开的。
- 京津冀地区的协调发展体现出不平衡的特点，北京的耦合协调度在较高水平波动，但天津和河北则在初级协调的阶段波动，且并未体现出提升趋势。未来河北和天津应重视生态-经济-社会协调发展，三地应协力形成京津冀地区更加协同的发展。
- 东北三省都属于初级协调的阶段，与长三角、北京等地区存在明显差距，并且未表现出提升的趋势，未来东北三省实现协调发展的关键在于提振经济发展的活力。

在建构指标体系的基础上，选取了长三角、京津冀以及东三省地区的环境治理效率作为研究对象，选取污染物的排放情况作为因变量，其余各指标作为自变量，分别进行多重线性回归。基于回归分析的结果，对长三角、京津冀以及东三省地区的环境治理效率进行讨论与分析。得到以下结论，并提出具有针对性的建议：

- 与长三角地区二氧化硫排放量关联性最显著的是环境污染治理投资总额。长三角地区经济发达、科技先进，在环境治理上投入较多，这是提升长三角生态环境治理效率的主要动力，未来应继续保持对环境治理的高投入，持续巩固环境治理成果。

- 与京津冀地区二氧化硫排放量关联性最显著的自然保护区个数。2006-2015 年间，京津冀地区的二氧化硫排放量比较高，未实现很好的治理。未来京津冀地区应加强三地协同发展，共同提高对环境治理的重视、投入，以改善环境治理的现状
- 与东三省二氧化硫排放量关联性最显著的是森林覆盖率。东北三省拥有较好的森林资源，自然禀赋很好，但目前环境治理投入还存在不足。未来应着力加大对环境治理的投入，并且重视对多种污染的同时控制，避免出现污染之间相互加剧的情况。

本研究具有一定的现实意义，未来可通过更新数据、采用新方法、优化模型等方法进行改进，比如将近年来的数据纳入计算，采用障碍因子模型来量化各个指标对耦合协调水平的障碍度等，这些方法可以使结论更加明确，指导性更强。

7 附录

7.1 R 代码

```

1  ## ----- 熵值法 -----
2  ## packages
3  library(openxlsx)
4  library(factoextra)
5  library(ggplot2)
6  library(gridExtra)
7
8  data <- read.xlsx('C:/Users/liusu/Desktop/data.xlsx', sheet = 1)
9
10 #归一化
11 data[, c(3:6, 8:13, 15:24, 29:31, 33:37)] <- apply(data[, c(3:6, 8:13,
12     15:24, 29:31, 33:37)], 2, function(x, data) {
13     0.002 + 0.996 * (as.numeric(x) - min(as.numeric(x))) / (max(as.numeric
14         (x)) - min(as.numeric(x)))}, data = data)
15 data[, c(7, 14, 25:28, 32)] <- apply(data[, c(7, 14, 25:28, 32)], 2,
16     function(x, data) {
17         0.002 + 0.996 * (max(as.numeric(x)) - as.numeric(x)) / (max(as.numeric
18             (x)) - min(as.numeric(x)))}, data = data)
19
20 z <- data #z代表归一化后的数据

```



```

17 w <- data #w代表权重 (这里只是建一个同维度的dataframe)
18
19 #
20 w[, 3:37] <- apply(w[, 3:37], 2, function(x) {as.numeric(x)/sum(as.numeric
    (x))})
21 w[, 3:10] <- apply(w[, 3:10], 2, function(x) {rep(-sum(as.numeric(x)*log(
    as.numeric(x)))/log(8), 310)})
22 w[, 11:24] <- apply(w[, 11:24], 2, function(x) {rep(-sum(as.numeric(x)*log
    (as.numeric(x)))/log(14), 310)})
23 w[, 25:37] <- apply(w[, 25:37], 2, function(x) {rep(-sum(as.numeric(x)*log
    (as.numeric(x)))/log(13), 310)})
24
25
26 for (i in 1:310) {
27   w[i, 3:10] <- apply(w[i, 3:10], 2, function(x) {(1-as.numeric(x))/(8-sum
    (as.numeric(w[i, 3:10])))})
28   w[i, 11:24] <- apply(w[i, 11:24], 2, function(x) {(1-as.numeric(x))/(14-
    sum(as.numeric(w[i, 11:24])))})
29   w[i, 25:37] <- apply(w[i, 25:37], 2, function(x) {(1-as.numeric(x))/(13-
    sum(as.numeric(w[i, 25:37])))})
30 }
31
32 index <- data[, 1:5]
33 colnames(index) <- c('year', 'province', 'economy', 'society', 'ecology')
34
35 for (i in 1:310) {
36   index[i, 3] <- sum(as.numeric(z[i, 3:10])*as.numeric(w[i, 3:10]))
37   index[i, 4] <- sum(as.numeric(z[i, 11:24])*as.numeric(w[i, 11:24]))
38   index[i, 5] <- sum(as.numeric(z[i, 25:37])*as.numeric(w[i, 25:37]))
39 }
40
41 index[, 3:5] <- apply(index[, 3:5], 2, as.numeric)
42 index$D <- sqrt(3*(index$economy*index$society*index$ecology)^(1/3)/(index
    $economy+index$society+index$ecology)*(index$economy*1/3+index$society
    *1/3+index$ecology*1/3))
43
44 ##不要重复跑这段, 会覆盖
45

```

```

46 wb <- createWorkbook()
47 addWorksheet(wb, "Sheet1")
48 writeData(wb, sheet = "Sheet1", x = index)
49 saveWorkbook(wb, file = "C:/Users/liusu/Desktop/index2.xlsx", overwrite =
    TRUE)
50
51 #score <- read.xlsx('C:/Users/liusu/Desktop/index2.xlsx', sheet = 1)
52 #ggplot(score, aes(x=year, y=economy, color=factor(province)))+geom_line()
    +ylab('Economy')
53 #ggplot(score, aes(x=year, y=society, color=factor(province)))+geom_line()
    +ylab('Society')
54 #ggplot(score, aes(x=year, y=ecology, color=factor(province)))+geom_line()
    +ylab('Ecology')
55
56 ## 直接在excel里筛选数据, 在index2里
57
58 ## 长三角
59 score.yangtze <- read.xlsx('C:/Users/liusu/Desktop/index2.xlsx', sheet =
    2)
60 ggplot(score.yangtze, aes(x=year, y=economy, color=factor(province)))+geom
    _line()+ylab('Economy')
61 ggplot(score.yangtze, aes(x=year, y=society, color=factor(province)))+geom
    _line()+ylab('Society')
62 ggplot(score.yangtze, aes(x=year, y=ecology, color=factor(province)))+geom
    _line()+ylab('Ecology')
63
64 ## 京津冀
65 score.jingjinji <- read.xlsx('C:/Users/liusu/Desktop/index2.xlsx', sheet =
    3)
66 ggplot(score.jingjinji, aes(x=year, y=economy, color=factor(province)))+
    geom_line()+ylab('Economy')
67 ggplot(score.jingjinji, aes(x=year, y=society, color=factor(province)))+
    geom_line()+ylab('Society')
68 ggplot(score.jingjinji, aes(x=year, y=ecology, color=factor(province)))+
    geom_line()+ylab('Ecology')
69
70 ## 东三省
71 score.dongbei <- read.xlsx('C:/Users/liusu/Desktop/index2.xlsx', sheet =

```

```

4)
72 ggplot(score.dongbei, aes(x=year, y=economy, color=factor(province)))+geom
   _line()+ylab('Economy')
73 ggplot(score.dongbei, aes(x=year, y=society, color=factor(province)))+geom
   _line()+ylab('Society')
74 ggplot(score.dongbei, aes(x=year, y=ecology, color=factor(province)))+geom
   _line()+ylab('Ecology')
75
76 ## ----- 耦合度 -----
77 D <- read_excel('C:/Users/liusu/Desktop/index2.xlsx')
78
79 D_chang <- D[D$province=="安徽" | D$province=="上海" | D$province=="江苏"
   | D$province=="浙江", c("year", "province", "D")]
80 D_jing <- D[D$province=="北京" | D$province=="河北" | D$province=="天津", c
   ("year", "province", "D")]
81 D_dong <- D[D$province=="黑龙江" | D$province=="吉林" | D$province=="辽宁"
   , c("year", "province", "D")]
82
83 D1 <- ggplot(D_chang, aes(x=year, y=D, color=factor(province)))+geom_line
   ()+ylab('耦合协调度')
84 ggsave("chang_D.png", D1)
85
86 D2 <- ggplot(D_jing, aes(x=year, y=D, color=factor(province)))+geom_line()
   +ylab('耦合协调度')
87 ggsave("jing_D.png", D2)
88
89 D3 <- ggplot(D_dong, aes(x=year, y=D, color=factor(province)))+geom_line()
   +ylab('耦合协调度')
90 ggsave("dong_D.png", D3)
91
92 ## ----- 长三角 -----
93 ## packages
94 library(openxlsx)
95 library(factoextra)
96 library(ggplot2)
97 library(gridExtra)
98 library(leaps)
99 library(car)

```

```

100 library(GGally)
101 library(rstatix)
102 library(stargazer)
103
104
105 ## 长三角
106 data <- read.xlsx('C:/Users/liusu/Desktop/附件1.xlsx', sheet = 1)
107 colnames(data) <- data[1, ]
108 data <- data[-1, ]
109 data <- data[data$province %in% c('安徽', '江苏', '上海', '浙江'), ]
110 province <- unique(data$province)
111 economy <- apply(data[, 3:10], 2, as.numeric)
112 society <- apply(data[, 11:24], 2, as.numeric)
113 ecology <- as.data.frame(apply(data[, 25:37], 2, as.numeric))
114
115 ## 因变量和自变量
116 Y <- as.data.frame(ecology[, c(2:4, 8)])
117 X <- as.data.frame(ecology[, c(1, 5:7, 9:13)])
118
119
120 ## 正态性检验
121 mshapiro_test(ecology)
122 assign("m", names(ecology))
123 m
124 ecology %>% shapiro_test(m)
125
126 ## 未通过正态性检验的指标
127 plot(density(Y$industry_fume))
128
129 ## scatter plot
130 pl <- ggpairs(ecology, columns = 1:13, # 选择需要绘图的变量所在列号
131               upper = list(continuous = "cor",
132                             na = "na"),
133               lower = list(continuous = "smooth",
134                             na = "na"),
135               diag = list(continuous = "densityDiag",
136                            na = "naDiag"))+
137 theme_bw()

```

```

138
139 pdf("p1.pdf",height = 15,width = 15,family = "Times")
140 print(p1)
141 dev.off()
142
143
144 ## adj-R2和BIC
145 fit <- regsubsets(Y$SO2 ~ ., data = X[, ]) #SO2可替换成别的指标
146 best_summary <- summary(fit)
147
148 par(mfrow=c(1,2))
149 plot(best_summary$adjr2,ylab = "adjr2",xlab = "number of features")
150 plot(fit,scale = "adjr2")
151
152 plot(best_summary$bic,ylab = "bic",xlab = "number of features")
153 plot(fit,scale = "bic")
154
155
156 ## 长三角
157 #SO2
158 reg <- lm(Y$SO2 ~ X$urban_area + X$reserve_num + X$urban_green_area + X$
    urban_green_rate + X$pollution_management_inv + X$pollution_management
    _inv_portion)
159 summary(reg)
160 stargazer(reg)
161
162 #COD
163 reg <- lm(Y$COD ~ X$urban_area + X$reserve_num)
164 summary(reg)
165
166 #waste_water
167 reg <- lm(Y$waste_water ~ X$urban_area + X$water_per + X$reserve_num + X$
    urban_green_area + X$fume_manage_facility + X$pollution_management_inv
    + X$pollution_management_inv_portion)
168 summary(reg)
169
170 #industry
171 reg <- lm(Y$industry_fume ~ X$urban_area + X$water_per + X$reserve_num + X

```

```

    $pollution_management_inv + X$pollution_management_inv_portion)
172 summary(reg)
173
174
175 ### ----- 京津冀 -----
176 ### packages
177 library(openxlsx)
178 library(factoextra)
179 library(ggplot2)
180 library(gridExtra)
181 library(leaps)
182 library(car)
183 library(GGally)
184 library(rstatix)
185 library(stargazer)
186
187
188 ### 京津冀
189 data <- read.xlsx('C:/Users/liusu/Desktop/附件1.xlsx', sheet = 1)
190 colnames(data) <- data[1, ]
191 data <- data[-1, ]
192 data <- data[data$province %in% c('北京', '河北', '天津'), ]
193 province <- unique(data$province)
194 economy <- apply(data[, 3:10], 2, as.numeric)
195 society <- apply(data[, 11:24], 2, as.numeric)
196 ecology <- as.data.frame(apply(data[, 25:37], 2, as.numeric))
197
198 ### 因变量和自变量
199 Y <- as.data.frame(ecology[, c(2:4, 8)])
200 X <- as.data.frame(ecology[, c(1, 5:7, 9:13)])
201
202 ### 正态性检验
203 mshapiro_test(ecology)
204 assign("m", names(ecology))
205 m
206 ecology %>% shapiro_test(m)
207
208 ### 未通过正态性检验的指标

```

```

209 plot(density(X$pollution_management_inv_portion))
210
211 ## scatter plot
212 p1 <- ggpairs(ecology, columns = 1:13, # 选择需要绘图的变量所在列号
213             upper = list(continuous = "cor",
214                          na = "na"),
215             lower = list(continuous = "smooth",
216                          na = "na"),
217             diag = list(continuous = "densityDiag",
218                        na = "naDiag")) +
219   theme_bw()
220
221 pdf("p1.pdf", height = 15, width = 15, family = "Times")
222 print(p1)
223 dev.off()
224
225
226 ## adj-R2和BIC
227 fit <- regsubsets(Y$SO2 ~ ., data = X[, ]) #SO2可替换成别的指标
228 best_summary <- summary(fit)
229
230 par(mfrow=c(1,2))
231 plot(best_summary$adjr2, ylab = "adjr2", xlab = "number of features")
232 plot(fit, scale = "adjr2")
233
234 plot(best_summary$bic, ylab = "bic", xlab = "number of features")
235 plot(fit, scale = "bic")
236
237
238 ## 京津冀
239 #SO2
240 reg <- lm(Y$SO2 ~ X$forest_coverage + X$reserve_num)
241 summary(reg)
242 stargazer(reg)
243
244 #COD
245 reg <- lm(Y$COD ~ X$forest_coverage + X$reserve_num + X$pollution_
           management_inv)

```

```

246 summary(reg)
247
248 #waste_water
249 reg <- lm(Y$waste_water ~ X$urban_area + X$forest_coverage + X$reserve_num
          + X$urban_green_rate + X$pollution_management_inv + X$pollution_
          management_inv_portion)
250 summary(reg)
251
252 #industry
253 reg <- lm(Y$industry_fume ~ X$forest_coverage + X$reserve_num + X$urban_
          green_area + X$urban_green_rate + X$pollution_management_inv)
254 summary(reg)
255
256
257 ## ----- 东三省 -----
258 ## packages
259 library(openxlsx)
260 library(factoextra)
261 library(ggplot2)
262 library(gridExtra)
263 library(leaps)
264 library(car)
265 library(GGally)
266 library(rstatix)
267 library(stargazer)
268
269
270 ## 东三省
271 data <- read.xlsx('C:/Users/liusu/Desktop/附件1.xlsx', sheet = 1)
272 colnames(data) <- data[1, ]
273 data <- data[-1, ]
274 data <- data[data$province %in% c('黑龙江', '吉林', '辽宁'), ]
275 province <- unique(data$province)
276 economy <- apply(data[, 3:10], 2, as.numeric)
277 society <- apply(data[, 11:24], 2, as.numeric)
278 ecology <- as.data.frame(apply(data[, 25:37], 2, as.numeric))
279
280 ## 因变量和自变量

```



```

281 Y <- as.data.frame(ecology[, c(2:4, 8)])
282 X <- as.data.frame(ecology[, c(1, 5:7, 9:13)])
283
284 ## 正态性检验
285 mshapiro_test(ecology)
286 assign("m", names(ecology))
287 m
288 ecology %>% shapiro_test(m)
289
290 ## 未通过正态性检验的指标
291 plot(density(X$urban_area))
292
293 ## scatter plot
294 p1 <- ggpairs(ecology, columns = 1:13, # 选择需要绘图的变量所在列号
295               upper = list(continuous = "cor",
296                             na = "na"),
297               lower = list(continuous = "smooth",
298                             na = "na"),
299               diag = list(continuous = "densityDiag",
300                             na = "naDiag"))+
301   theme_bw()
302
303 pdf("p1.pdf", height = 15, width = 15, family = "Times")
304 print(p1)
305 dev.off()
306
307
308 ## adj-R2和BIC
309 fit <- regsubsets(Y$SO2 ~ ., data = X[, ]) #SO2可替换成别的指标
310 best_summary <- summary(fit)
311
312 par(mfrow=c(1,2))
313 plot(best_summary$adjr2, ylab = "adjr2", xlab = "number of features")
314 plot(fit, scale = "adjr2")
315
316 plot(best_summary$bic, ylab = "bic", xlab = "number of features")
317 plot(fit, scale = "bic")
318

```

```

319
320 ## 东三省
321 #SO2
322 reg <- lm(Y$SO2 ~ X$urban_area + X$water_per + X$forest_coverage + X$
      reserve_num + X$urban_green_area + X$urban_green_rate + X$pollution_
      management_inv + X$pollution_management_inv_portion)
323 summary(reg)
324 stargazer(reg)
325
326 #COD
327 reg <- lm(Y$COD ~ X$reserve_num + X$urban_green_area + X$pollution_
      management_inv + X$pollution_management_inv_portion)
328 summary(reg)
329
330 #waste_water
331 reg <- lm(Y$waste_water ~ X$urban_area + X$forest_coverage + X$urban_green
      _rate + X$fume_manage_facility)
332 summary(reg)
333
334 #industry
335 reg <- lm(Y$industry_fume ~ X$urban_area + X$forest_coverage)
336 summary(reg)

```

7.2 参考文献

- [1] 李茜, 胡昊, 李名升, 等. 中国生态文明综合评价及环境, 经济与社会协调发展研究 [J]. 资源科学, 2015, 37(7): 1444-1454.
- [2] 初雪, 陈兴鹏, 贾卓, 李晨曦. 欠发达地区经济、社会和生态系统的协调发展研究——以甘肃省崇信县为例 [J]. 干旱区资源与环境, 2017, 31(10): 13-18. DOI:10.13448/j.cnki.jalre.2017.307.
- [3] Grossman G M, Krueger A B. Environmental impacts of a North American free trade agreement[J]. 1991.
- [4] Pigou A C. The economics of welfare[M]. Macmillan, 1924.
- [5] 王琦, 汤放华. 洞庭湖区生态—经济—社会系统耦合协调发展的时空分异 [J]. 经济地理, 2015, 35(12): 161-167+202. DOI:10.15957/j.cnki.jjdl.2015.12.023.
- [6] 吴义刚, 杨晓庆. 基于经济—生态—社会耦合关系的安徽乡村协调发展研究 [J]. 安徽农业大学学报 (社会科学版), 2023, 32(01): 23-31. DOI:10.19747/j.cnki.1009-2463.2023.01.004.

[7]Li J, Sun W, Li M, et al. Coupling coordination degree of production, living and ecological spaces and its influencing factors in the Yellow River Basin[J]. Journal of Cleaner Production, 2021, 298: 126803.

[8] 汪永生, 李宇航, 揭晓蒙, 李玉龙, 李桂君, 王文涛. 中国海洋科技-经济-环境系统耦合协调的时空演化 [J]. 中国人口·资源与环境, 2020, 30(08):168-176.

[9] 李丽, 孟得星. 资源型城市生态-经济-社会协调发展水平的测度——以陕西省延安市为例 [J]. 湖北农业科学, 2023, 62(05):196-202+208. DOI:10.14088/j.cnki.issn0439-8114.2023.05.034.

[10] 陈颖. 青海省人口、资源、环境与经济、社会的协调发展研究 [J]. 西北人口, 2007(05):25-30.

[11] 刘霞, 刘辉, 苏丽娟, 朱晓峻, 程桦. 基于 AHP—熵值综合赋权法的煤炭资源型城市生态—经济—社会协调发展分析——以鄂尔多斯市为例 [J]. 生态科学, 2023, 42(03):213-224. DOI:10.14108/j.cnki.1008-8873.2023.03.025.