

基于 PCA 的中国各省市自治区能源经济发展综合评价

史喜平 (山西焦煤西山煤电太原公司, 山西 太原 030000)

摘要: 当前俄乌冲突下的百年变局叠加世纪疫情, 进一步使国际局势复杂多变, 全球经济持续衰退, 能源价格居高不下。在此影响下, 我国能源价格也持续上涨, 新时代下如何能够稳住能源经济、确保能源经济运行在合理区间就显得尤为重要。本文以中国各省市自治区为研究对象, 基于国家统计局原始数据, 选取适当指标, 利用 PCA 综合评价了各省市自治区的能源经济发展情况, 结果表明各省市自治区之间的能源经济发展情况存在较大的不平衡, 具体表现为: 东部地区省份能源经济发展相对较高, 处于第一梯队, 中部地区省份能源经济发展相对较为均衡, 西部地区省份与东北地区省份能源经济发展相对比较滞后。

关键词: 新时代; PCA; 能源经济发展

1 引言

截止 2021 年底, 我国煤炭资源储量为 2078.85 亿 t, 天然气为 63392.67 亿 m³, 石油储量为 36.89 亿 t。2021 年我国一次性能源生产总量为 43.3 亿 t, 其中原煤占比为 67%, 原油占比为 6.6%, 天然气占比为 6.1%, 由数据不难看出, 我国的能源结构现状为“富煤、少气、缺油”。

新时代随着我国工业化、城镇化及现代化建设步伐的推进, 能源需求也将持续增加, 2021 年我国能源消费总量为 52.4 亿 t, 其中原煤占比为 56%, 原油占比为 18.5%, 天然气占比为 8.9%, 总体能源自给率为 82.6%。2021 年我国的国内生产总值为 1143669.7 亿元, 按不变价计算, 较上年增长 8.1%, 居民恩格尔系数为 29.8%, 较上年下降 0.4 个百分点, 能源消费弹性系数 0.64, 较上年下降 0.36。本文就我国各省市自治区的能源经济发展进行了综合评价, 期望能够为决策者提供数据支撑。

2 研究方法及其评价指标体系的构建

2.1 研究方法

主成分分析 (Principal components analysis, PCA) 是一种通过降维技术把多个变量转化为少数几个主成分的多元统计方法, 这些主成分能够反映原始变量的大部分信息^[1]。主成分的贡献率反映了该主成分包含原始数据的信息量占全部信息量的比重, 这样的确定权数是客观、合理的, 可以克服某些评价方法中人为确定权数的不足。

薛静静等^[2]通过构建多指标综合评价体系, 运用全局主成分分析方法对我国八大经济区经济发展水平 (2000 年–2010 年) 和能源效率水平 (2000 年–2009 年) 进行系统的评价, 借助 ArcGIS 空间分析技术和协调发

展评价模型对两者发展过程中的时空差异、演变特征及协调关系进行深入分析。曾鸣等^[3]构建能源、经济、环境系统耦合协调度评价指标体系, 基于主成分分析法对这三个系统的发展水平进行测度分析, 利用构建的耦合协调度模型计算系统的协调度水平。Stanković Jelena J.etal^[4]运用 PCA 和 PROMETHEE 创建了一个综合指数, 作为衡量国家层面循环经济发展的指标, 研究结果清楚地表明, 循环经济在国家层面的发展与国家社会经济发展之间存在正相关关系, 而循环经济的进步对环境可持续性没有直接影响, 但有间接影响且滞后期过后, 影响逐步加强。

2.2 评价指标体系

指标体系是否科学合理不仅关系到系统评价是否准确可靠, 而且决定了系统发展方向的正确与否^[5]。因此, 本文遵循一定的原则, 综合分析多方观点, 拟构建反映国民经济水平、人民生活水平及能源消耗水平的指标体系 (见表 1)。其中国民经济水平包括地区生产总值 GDP^[6] (亿元) 和社会消费品零售总额 (亿元) 两个指标; 人民生活水平包括全体居民人均可支配收入^[7] (元) 和全体居民人均消费支出 (元); 能源消耗水平包括煤炭消费量 (万吨)、汽油消费量 (万吨)、天然气消费量 (亿立方米) 和电力消费量 (亿千瓦时) 等四个指标。所选指标数据均来源于国家统计局, 由于未能获取西藏自治区相关能源数据, 故未将其纳入研究对象。为了使数据能够更准确, 本文将地区生产总值 GDP、社会消费品零售总额、煤炭消费量、汽油消费量、天然气消费量、电力消费量等六个指标进行人均化处理, 即除以各省市自治区年末常住人口, 同时将逆相关的四个指标统一正向化, 即对应数据加负号处理。

3 PCA 的实证研究

3.1 对处理后的数据进行标准化处理

为使其具有可比性,需要对处理后的数据进行标准化处理以消除量纲的影响,通过 spss 软件计算,得到标准化矩阵 Z。

3.2 计算特征值及各个主成分的贡献率和累计贡献率

根据图 1 所示, KMO 值为 0.751 (> 0.5), Bartlett 检验的近似卡方值为 302.863, P 值 $< 0.01 < 0.05$, 表明该原始数据适合主成分分析。

KMO 和 Bartlett 检验

KMO 取样适切性量数。		.751
Bartlett 球形度检验	近似卡方	302.863
	自由度	28
	显著性	.000

图 1 KMO 值和 Bartlett 检验

3.2.1 计算特征值

解特征方程 $|\lambda I - R| = 0$ 。常用雅克比法 (Jacobi) 求出特征值, 并使其按大小顺序排列: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_8 \geq 0$, 从表 1 中可明显看出, $\lambda_1 > \lambda_2 > 1$ 。

3.2.2 计算各个主成分贡献率和累计贡献率

各个主成分贡献率计算公式如下:

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} (i=1, 2, \dots, p) \quad (1)$$

累计贡献率计算公式如下:

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i=1, 2, \dots, p) \quad (2)$$

通过 spss 软件计算, 得到各个主成分贡献率和累计贡献率 (见表 1)。一般取累计贡献率达 80% 以上的特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 所对应的第一、第二、...、第 m ($m \leq p$) 个主成分。从表 1 中可以明显看出, 前两个主成分的累计贡献率为 82.314%, 大于 80%, 因此在本文中选取前两个主成分进行研究。

表 1 初始特征值、方差贡献率和累计贡献率

序号	特征值	方差贡献率 /%	累计贡献率 /%
1	4.697	58.708	58.708
2	1.889	23.607	82.314
3	0.709	8.867	91.181
4	0.409	5.114	96.295

5	0.16	2.004	98.299
6	0.11	1.378	99.676
7	0.017	0.21	99.886
8	0.009	0.114	100

3.3 计算各省市自治区各主成分的得分

采用回归方法, 通过 spss 软件计算各主成分得分系数矩阵 Z (见表 2):

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{bmatrix}$$

表 2 主成分得分系数矩阵

序号	主成分 1	主成分 2
ZX ₁	0.208	-0.015
ZX ₂	0.195	0.078
ZX ₃	0.207	-0.008
ZX ₄	0.208	-0.007
ZX ₅	0.049	0.439
ZX ₆	-0.169	-0.01
ZX ₇	-0.121	0.271
ZX ₈	-0.022	0.506

因此, 两个主成分的评价得分函数如下:

$$F_1 = 0.208ZX_1 + 0.195ZX_2 + 0.207ZX_3 + 0.208ZX_4 + 0.049ZX_5 - 0.169ZX_6 - 0.121ZX_7 - 0.022ZX_8 \quad (3)$$

$$F_2 = -0.015ZX_1 + 0.078ZX_2 - 0.008ZX_3 - 0.007ZX_4 + 0.439ZX_5 - 0.01ZX_6 + 0.271ZX_7 + 0.506ZX_8 \quad (4)$$

将标准化后 ZX₁、ZX₂、...、ZX₈ 代入各主成分线性表达式, 通过计算, 得出中国各省市自治区的各主成分得分。

3.4 计算综合评价, 并排序

以贡献率为权数, 构建综合评价函数如下:

$$SCORE = \sum_{i=1}^n \alpha_i F_i (i=1, 2, \dots, n) \quad (5)$$

其中为累计贡献率, 具体计算公式如下:

$$SCORE = 0.58708F_1 + 0.23607F_2 \quad (6)$$

将 F₁、F₂ 代入公式 (6), 得到中国各省市自治区的综合评价得分 (见表 3), 并将其按从高到低排序, 显然指数越高代表其能源利用率越好。

表 3 综合评价得分

地区	综合评价得分	排序
北京市	1.832	1
上海市	1.648	2
天津市	0.647	3

浙江省	0.608	4
江苏省	0.528	5
福建省	0.482	6
广东省	0.417	7
湖北省	0.284	8
重庆市	0.224	9
辽宁省	0.12	10
湖南省	0.058	11
四川省	-0.024	12
安徽省	-0.03	13
海南省	-0.074	14
山东省	-0.127	15
江西省	-0.15	16
河南省	-0.199	17
黑龙江省	-0.212	18
陕西省	-0.232	19
云南省	-0.234	20
吉林省	-0.292	21
贵州省	-0.3	22
广西壮族自治区	-0.387	23
河北省	-0.393	24
甘肃省	-0.482	25
青海省	-0.507	26
新疆维吾尔自治区	-0.662	27
内蒙古自治区	-0.707	28
山西省	-0.715	29
宁夏回族自治区	-1.117	30

从排名情况来看,排在前十的有北京市、上海市、天津市、浙江省、江苏省、福建省、广东省、湖北省、重庆市、辽宁省,十一到二十的有湖南省、四川省、安徽省、海南省、山东省、江西省、河南省、黑龙江省、陕西省、云南省,后十位的是吉林省、贵州省、广西壮族自治区、河北省、甘肃省、青海省、新疆维吾尔自治区、内蒙古自治区、山西省、宁夏回族自治区。通过 PCA 分析可以发现:排名前十的,东部地区有 7 个,中部地区有 1 个,东北地区有 1 个,西部地区省份多但仅有 1 个;排名十一到二十的,东部地区有 2 个,中部地区有 4 个,东北地区有 1 个,西部地区有 3 个;排名后十位的,东部地区有 1 个,中部地区有 1 个,东北地区有 1 个,西部地区有 7 个。从结果中看,东部地区整体排名相对较高,处于第一梯队,仅有河北一省处于后十名,河北作为能源消耗大省,能源经济发展有点不尽如人意;中部地区有且仅有山西一个省份排名处于后十名,而且是倒数第二,山西作为煤

炭大省、能源基地,对于保障国家能源安全有着极其重要的作用,自新中国成立以来,为祖国的建设奉献了很多,自身的发展却相对比较滞后;西部地区整体排名相对靠后,内蒙古同样作为煤炭能源大省,目前也面临着和山西较为相似的窘境;东北地区三省份排名也一般,东三省近几年人口外流严重,严重制约了当地的能源经济发展。

中西部地区和东北地区作为“内循环”的重要承载地,在国家中西部崛起和振兴东三省战略的支持下,在“一带一路”经济的带动下,在东部地区的对口帮扶下,构建以国内大循环为主体、国内国际双循环相互促进的新发展格局,共同促进我国经济高质量发展。

4 结束语

当下,疫情已呈现常态化的趋势,俄乌冲突下的百年变局进一步使国际局势变得复杂起来,我国要想在后疫情时代保持经济高质量发展,就需要稳住能源经济、确保能源经济运行在合理区间。

参考文献:

- [1] 谢中华. Matlab 统计分析与应用:40 个案例分析 [M]. 北京:北京航空航天大学出版社,2010.
- [2] 薛静静,沈镭,刘立涛,高天明. 中国区域能源利用效率与经济水平协调发展研究 [J]. 资源科学,2013,35(04):713-721.
- [3] 曾鸣,王亚娟. 基于主成分分析法的我国能源、经济、环境系统耦合协调度研究 [J]. 华北电力大学学报(社会科学版),2013(03):1-6.
- [4] Jelena J. Stanković, Vesna Janković-Milić, Ivana Marjanović, Jasmina Janjić. An integrated approach of PCA and PROMETHEE in spatial assessment of circular economy indicators [J]. Waste Management, 2021(28):154-166.
- [5] 张洁. 企业信息化工程 [M]. 北京:电子工业出版社,2008.
- [6] 许宪春,吕峰. 改革开放 40 年来中国国内生产总值核算的建立、改革和发展研究 [J]. 经济研究,2018,53(08):4-19.
- [7] 杨天宇,陈明玉. 消费升级对产业迈向中高端的带动作用:理论逻辑和经验证据 [J]. 经济学家,2018(11):48-54.

作者简介:

史喜平(1988-),男,山西山阴人,硕士研究生,经济师,研究方向:工业工程与管理、管理科学研究。