

多群组混合环境基尼系数及其在五大城市群绿色发展评价中的应用

张仁杰*, 束 慧, 闵莉花, 孙 静

南京邮电大学理学院, 江苏 南京

收稿日期: 2023年9月25日; 录用日期: 2023年10月18日; 发布日期: 2023年10月25日

摘 要

我国区域间发展不平衡性导致了其数据的厚尾性, 基于此背景, 在参数化基尼的框架上构建了多群组混合基尼系数及其分解形式, 并结合区域数据的厚尾特征推导了有限样本下的多群组混合基尼系数的表达式, 并运用该基尼系数对我国五大城市群2005~2020年的经济产出、资源消耗或污染物排放等方面的基尼系数进行计算, 对区域综合绿色发展情况进行评价和分析。结果表明, 相比于传统的基尼系数方法, 该基尼系数可以很好地适应厚尾数据, 其中组间差异是区域发展不平衡的主要来源, 其贡献率约为70%。相比于十一五、十二五时期, 十三五期间的经济和生态之间不平衡得到了一定的改善, 能源消耗基尼系数从0.45降到了0.35, 我国五大城市群的“清洁生产效率”同时呈现出下降趋势, 绿色发展程度提升。

关键词

多群组基尼系数分解, 厚尾特征, 有限样本量, 空间异质性

Multi-Group Environment Gini Coefficient and Its Application in the Evaluation of Economic and Ecological Balance of Five Urban Agglomerations

Renjie Zhang*, Hui Shu, Lihua Min, Jing Sun

College of Science, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Sep. 25th, 2023; accepted: Oct. 18th, 2023; published: Oct. 25th, 2023

*通讯作者。

文章引用: 张仁杰, 束慧, 闵莉花, 孙静. 多群组混合环境基尼系数及其在五大城市群绿色发展评价中的应用[J]. 应用数学进展, 2023, 12(10): 4378-4391. DOI: [10.12677/aam.2023.1210431](https://doi.org/10.12677/aam.2023.1210431)

Abstract

The regional imbalance of economy and ecology between regions leads to the thick tail of the data. Based on this background, the multi-group mixed Gini coefficient and its decomposition form are constructed in the framework of parameterized Gini to evaluate the spatial balance of economy and ecology in the region. Combined with the thick tail characteristics of regional data, the expression of the multi-group mixed Gini coefficient under limited samples is derived. The Gini coefficient is used to evaluate and analyze the economic output, resource consumption, or pollutant emissions of the five major urban agglomerations from 2005 to 2020. Compared with the traditional Gini coefficient method, the results show that the Gini coefficient can be well adapted to the thick mantissa data. Inter-group difference is the main source of regional economic and ecological spatial inequality, and its contribution rate is about 70%. Compared with the period of 11th and 12th Five-Year, the spatial balance of economy and ecology has been improved to a certain extent during the 13 Five-Year, with the Gini coefficient of energy consumption dropping from 0.45 to 0.35. The “clean production efficiency” among the five major urban agglomerations also shows a downward trend and the degree of green development increased.

Keywords

Multigroup Gini Coefficient Decomposition, Thick Tail Characteristics, Limited Sample Size, Spatial Heterogeneity

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

党的二十大对新发展阶段区域发展和空间治理作出了重大部署，着力构建优势互补、高质量发展的区域经济布局 and 国土空间体系，以城市群、都市圈为依托构建大中小城市协调发展的格局，这对解决发展不平衡不充分的问题，加快构建新发展格局，推动高质量发展具有重大而深远的意义。因此，探究城市群之间经济产出和生态效益的差异对于针对性促进城市群乃至区域整体协调发展有着重大意义[1] [2]。

基尼系数是最被广泛运用的社会经济指标之一，它是基于洛伦兹曲线定量确定平衡的指标工具。基尼系数常常用来度量收入、消费和财富的不平等，但其实基尼系数可以用来度量任何事物的分布不均情况。传统的基尼系数方法如今已被广泛的运用到区域的生态和经济发展平衡评价之中[3]。这些研究中的基尼系数计算方法多种多样，但其中大部分方法与收入分配的计算方法相同。然而，居民收入分配数据与区域经济产出、污染排放数据在分布类型上存在较大差异，收入分布数据与区域发展数据的样本单位也处于不同的数量水平，这些差异对于基尼系数的计算以及分解将产生重大影响。中国的经济生态发展带有浓厚的区域特征[4] [5] [6]，呈现出经济高度空间聚集，以及东西部地区每单位 GDP 上的污染排放高度分布不平等，导致不同地区的工业产出数据和污染排放数据在一定程度上延伸出不同的重尾特征[7]。

在非参数方法中，人们一般直接从总体实际数据出发，不做任何分布假定，直接构造基尼系数或者相关表达式。但 Bahadur 和 Savage [8]指出当数据呈现出厚尾特征时使用非参数方法估计基尼系数可能不可靠；Andrea *et al.* [9]也指出当对厚尾数据进行分析时，基尼系数的估计会存在向下的估计，这种情况下，基尼系数不能可靠地使用传统的非参数方法；但根据现有的文献的研究发现，在对区域平衡发展研究时，

使用的基尼系数算法大多是非参数方法[10]-[15]，比如 Druckman [16]构建了 AR-Gini，估计了社区之间关于特定消费品之间的不平等；Jiandong Chen *et al.* [17]等人构建了关于 CEC 的基尼系数研究了我国煤炭二氧化碳排放区域间差异的演变过程，以及排放差异偏离经济增长差异的原因。

另一种基尼系数计算方法是基于洛伦兹曲线的参数方程和连续分布函数的参数化方法。Kendall [18]在 1977 年提出了一种基于连续分布函数的基尼系数计算方法，程永宏[19] [20]利用收入分配函数和洛伦兹曲线定义了一种新的参数化基尼系数计算方法，并且在某种程度上证明的 Kendall 的表达式，同时将基尼系数从一元总体推向了多元群组，使基尼系数有了更加广泛的运用。在使用参数法计算基尼系数时，需要对总体的分布类型进行假定。陈希孺[21]认为，当对总体的分布形式有较大把握时，有利于提高估计效率。然而，这种参数化方法在区域平衡发展评价[22]中很少用到，并且收入分配数据与区域生态效益数据有一定差异，基于该类分布数据的基尼系数计算还未得到更深的研究。

在使用参数化方法评价区域平衡发展时，需要从总体基尼系数出发，考虑到区域经济和生态条件的差异。总体基尼系数在福利领域得到广泛的研究[23] [24] [25] [26]，比如艾小青和祁磊[23]实现了在收入或财富信息不完全的情况下的总体基尼系数的估算；黄和平[27]提出，不同的基尼系数计算方法有着各自的优点和适应范围，从不同角度计算的环境基尼系数也不尽相同

对总体基尼系数的另一个深入研究领域是对其分解结果的探讨，对总体基尼系数进行分解，可以有效地揭示不平等的来源。当数据被分成大小相同的组时，基尼系数将会严格降低[28]。杨桐彬[29]、Cowell [30]等人都从不同角度对总体基尼系数进行了分解，但这些分解方法大多基于非参数总体基尼系数[31] [32] [33] [34] [35]，并且几乎都来自经济研究领域。大多数学者将基尼系数引入区域经济和生态效益评价中[36] [37]时，往往都直接采用收入基尼系数的分解方法，这些应用研究几乎没有区别样本数据的分布类型，而且区域发展数据中往往以地区样本单位，样本量有限。艾小青[38]指出样本数据对于分布函数的拟合程度不是关键，真正的关键是对总体分布的假定是否准确。因此，进行区域平衡评价时，主要集中在基尼系数的适用性上，需要考虑数据类型、数据的统计特征以及基尼系数的计算方法等多方面的因素。

鉴于此，在已有研究基础上，本文将基于京津冀、珠三角、长江上游、成渝、长三角五大城市群 2005~2020 年城市面板数据，对区域经济和生态数据的厚尾性进行验证，并针对其厚尾特征，基于参数化基尼框架，构建了有限样本下的多群组混合环境基尼系数及其分解形式，用于度量我国区域经济和生态发展的不平衡及其组成成分。

2. 空间异质性条件下多群组环境基尼系数的构建

在针对区域性的平衡发展研究时，基尼系数主要是用来度量各发展中区域的工业产出、资源消耗以及污染物排放情况之间的社会分配平等性。利用 Kendall 在 1977 年给出的计算方法，得到总体的基尼系数为[18]：

$$G = \frac{\int_a^b F(s)ds - \int_a^b F(s)^2 ds}{u} \quad (1)$$

一个包含多个空间单元的区域污染物排放量可以看作一个连续的随机变量 S ， $S = [a, b]$ ，其中 a 和 b 表示在该区域中排放的上界和下界， F 为关于 S 的分布函数；所研究的区域的空间单元总数记为 N ；且该区域的污染物排放均值记为 u ， $u = \frac{S}{N}$ ，同时均值[4]也可以表示为 $u = b - \int_a^b F(s)ds$ 。

2.1. 具有空间异质性的统计分布函数

随着时间的推移，区域间的发展关系也会同步地发生变化。将调查时间设定为 t ， $t = 1, \dots, T$ 。调查

区域内将划分为 M 个不同的分组, 代表了不同的经济水平和生态承载力的区域各空间单元。第 m 组空间单元数记为 N_m , 第 t 年第 m 组空间单元的污染排放总量用随机变量 S_{mt} 表示, $S_{mt} \in (a_{mt}, b_{mt})$, 则对应的排放分布函数记为 $F_{mt}(s)$, $t=1, \dots, T$, $m=1, \dots, M$ 。区域的空间单元总数为 $N = \sum_{m=1}^M N_m$, 第 t 年区域污染排放总量记为 $S_t = \sum_{m=1}^M S_{mt}$, 其分布函数为 $F_t(s)$, 随机变量的值在区间 (a_t, b_t) 内, 其中 $a_t = \min_{1 \leq m \leq M} a_{mt}$, $b_t = \max_{1 \leq m \leq M} b_{mt}$ 。

为获得多群组混合的分布函数, 进而推广到多群组混合基尼系数的计算及其分解, 在此做出一个重要的前提假设, 各群组的分布函数之间是相互独立的。

$F_t(s) = P(S_t \leq s)$, 它的概率论意义表示为该区域第 t 年污染物排放不超过 s 的空间单元数占空间单元总数的比例, 用 P_m 表示第 m 组空间单元数占该区域空间单元总数的比重, $\sum_{m=1}^M P_m = 1$ 。

基于空间单元 N 比重的概率意义, 根据全概率公式, 可以清晰的证明:

$$F_t(s) = P(S_t \leq s) = P_1 P(S_{1t} \leq s) + \dots + P_m P(S_{mt} \leq s)$$

$$F_t(s) = \sum_{m=1}^M P_m F_{mt}(s) \quad (2)$$

但当考虑一个特定地区的经济产出、资源消耗、或污染物排放时, 样本个体一般选取为城镇, 样本量 N 往往是有限的, 设定以下前提条件:

所有空间单元的污染物排放量都位于区间 $[a, b]$ 内。如果各空间单位的污染物排放可以相同, 则处于最高污染物排放量 b 或最低污染物排放量 a 位置的空间单元数不唯一, 用 k_a, k_b 表示。污染排放分布函数 $F(s)$ 也在 $s=a$ 和 $s=b$ 处存在不连续。

$$\lim_{s \rightarrow a^-} F(s) = 0, \quad \lim_{s \rightarrow a^+} F(s) = \frac{k_a}{N} \quad (3)$$

$$\lim_{s \rightarrow b^-} F(s) = 1 - \frac{k_b}{N}, \quad \lim_{s \rightarrow b^+} F(s) = 1 \quad (4)$$

在收入分配的计算中, 由于样本量人口足够大, $N \gg 1$, 所以分布函数在上下界的不连续点处的左右界限处十分接近, 近似看作连续。但区域空间单元数是有限的, 区域经济产出和生态效益分布函数在上下界跳跃点处的左右界限是不相等的。除了经济和生态方面数据本身的分布特征外, 区域空间单元数也会对基尼系数的计算产生影响。

2.2. 具有空间异质性的多群组环境基尼系数的推导

在区域污染排放数据中, 总体基尼系数测算了存在模式差异时的空间不平衡。基于总体基尼系数, 可以进一步分析总体基尼系数测算的不平衡的组成成分。

设第 t 年总体的污染排放均值为 u_t , 各群组的污染排放均值为 u_{mt} , $u_t = \sum_{m=1}^M P_m u_{mt}$ 。设第 t 年各群组的污染排放占区域总排放的比重为 w_{mt} , 则

$$u_{mt} = \frac{w_{mt} * S}{P_m * N} = \frac{w_{mt}}{P_m} u_t \quad (5)$$

根据排放分布函数下基尼系数的计算公式, 总体以及各个群组的基尼系数为:

$$G_t = \frac{\int_{a_t}^{b_t} F_t(s) ds - \int_{a_t}^{b_t} F_t(s)^2 ds}{u_t} \quad (6)$$

$$G_{mt} = \frac{\int_{a_{mt}}^{b_{mt}} F_{mt}(s) ds - \int_{a_{mt}}^{b_{mt}} F_{mt}(s)^2 ds}{u_{mt}} \tag{7}$$

在样本足够大的情况下，污染排放分布函数在上下界处依然存在跳跃间断点，但可将其忽略近似看作连续。

根据艾小青[11]与程永宏[5]对于基尼系数分解的研究，可以得到：

$$\begin{aligned} & \int_{a_t}^{b_t} F_t(s)(1-F_t(s)) ds = \int_{a_t}^{b_t} \sum_{m=1}^M (P_m F_{mt} - P_m^2 F_{mt}^2) ds - \int_{a_t}^{b_t} \sum_{m=1}^M \sum_{j=1, m \neq j}^M P_m P_j F_{mt} F_{jt} ds \\ & = \int_{a_t}^{b_t} \sum_{m=1}^M (P_m F_{mt} - P_m^2 F_{mt}^2 + P_m F_{mt}^2 - P_m F_{mt}^2) ds - \int_{a_t}^{b_t} \sum_{m=1}^M \sum_{j=1, m \neq j}^M P_m P_j F_{mt} F_{jt} ds \\ & = \sum_{m=1}^M P_m^2 u_{mt} G_{mt} + \sum_{m=1}^M \sum_{j>m}^M P_m P_j \int_{a_t}^{b_t} (F_{mt} + F_{jt} - 2F_{mt} F_{jt}) ds \end{aligned} \tag{8}$$

当考虑一个特定地区的经济产出、资源消耗、或污染物排放时，样本个体一般选取为城镇，样本量 N 往往是有限的。区域经济产出和生态效益的分布以及空间单位的数量对于基尼系数中的计算有着重大的关系，不得不考虑区域空间的有限性，不能忽略数据上下界出数据分布的不连续性。

有限样本下的多群组总体基尼系数推导中的几个重要元素需要特殊处理，即对区域的空间单元根据具体情景分别分组，然后根据各组的排放数据拟合相应的分布函数，找到相应的逆函数，进而得到各子群组的上下界的估计值，即 $a_{mt} = F_{mt}^{-1}\left(\frac{k_{am}}{N_m}\right)$ 和 $b_{mt} = F_{mt}^{-1}\left(1 - \frac{k_{bm}}{N_m}\right)$ 。

在样本量有限的情况下，总体的基尼系数为：

$$G = \frac{bF(b^-)^2 - aF(a^+)^2 - bF(b^-) + aF(a^+) + \int_a^b F(s) ds - \int_a^b F(s)^2 ds}{bF(b^-) - aF(a^+) - \int_a^b F(s) ds} \tag{9}$$

令 $T = bF(b^-)^2 - aF(a^+)^2 - bF(b^-) + aF(a^+)$,

则各子群组内 $T_{mt} = b_{mt}F(b_{mt}^-)^2 - a_{mt}F(a_{mt}^+)^2 - b_{mt}F(b_{mt}^-) + a_{mt}F(a_{mt}^+)$,

各子群组的均值为 $\hat{u}_{mt} = b_{mt}F(b_{mt}^-) - a_{mt}F(a_{mt}^+) - \int_{a_{mt}}^{b_{mt}} F(s) ds$ 。

在将式(9)进行多群组分解时，使用有限样本下的各子群组均值 $\sum_{m=1}^M P_m \hat{u}_{mt}$ 来表示总体的均值 \hat{u}_t (\hat{u}_t 是各元素在有限样本下的表达形式)，近似代替 $bF(b^-) - aF(a^+) - \int_a^b F(s) ds$ ，且根据式(8)可将上式变化为，

$$G = \frac{T - \sum_{m=1}^M P_m^2 T_{mt} + \sum_{m=1}^M P_m^2 T_{mt} + \int_a^b F(s) ds - \int_a^b F(s)^2 ds}{\sum_{m=1}^M P_m \hat{u}_{mt}} \tag{10}$$

因此该区域的污染排放总量为：

$$\begin{aligned} \hat{S}_t & = \sum_{m=1}^M N_m \hat{u}_{mt} = \sum_{m=1}^M N_m \left[b_{mt} F_{mt}(b_{mt}^-) - a_{mt} F_{mt}(a_{mt}^+) - \int_{a_{mt}}^{b_{mt}} F_{mt}(s) ds \right] \\ & = \sum_{m=1}^M \left[N_m b_{mt} - k_b b_{mt} - k_a a_{mt} - N_m \int_{a_{mt}}^{b_{mt}} F_{mt}(s) ds \right] \end{aligned} \tag{11}$$

根据定积分的积分性质，关于分布函数 F 的积分可变形为，

$$\begin{aligned} \int_{a_t^+}^{b_t^-} P_m F_{m_t} ds &= \int_{a_t^+}^{a_{m_t}} P_m F_{m_t} ds + \int_{a_{m_t}}^{b_{m_t}} P_m F_{m_t} ds + \int_{b_{m_t}}^{b_t^-} P_m F_{m_t} ds \\ &= \int_{a_{m_t}}^{b_{m_t}} P_m F_{m_t} ds + P_m (b - b_m) \end{aligned} \quad (11)$$

同理得，

$$\int_{a_t^+}^{b_t^-} P_m P_j F_{m_t} F_{j_t} ds = \int_{a_{m_t}}^{b_{m_t}} P_m P_j F_{m_t} F_{j_t}(s) ds + P_m P_j (b - b_m) \quad (12)$$

$$\int_{a_t^+}^{b_t^-} P_m^2 F_{m_t}^2 ds = \int_{a_{m_t}}^{b_{m_t}} P_m^2 F_{m_t}^2 ds + P_m^2 (b - b_m) \quad (13)$$

于是，将总体基尼系数的区间由 (a_t, b_t) 化作 (a_{m_t}, b_{m_t}) 后，得到的表达式余项为：

$$\begin{aligned} &\sum_{m=1}^M P_m (b - b_m) - \sum_{m=1}^M P_m^2 (b - b_m) - \sum_{m=1}^M \sum_{j=1, m \neq j}^M P_m P_j (b - b_m) \\ &= \sum_{m=1}^M P_m (1 - P_m) (b - b_m) - \sum_{m=1}^M \sum_{j=1, m \neq j}^M P_m P_j (b - b_m) = 0 \end{aligned} \quad (14)$$

根据式(14)可以看出，在计算基尼系数时，对于某些特定表达式的积分，区间的变化，并不会改变积分的值。根据式(9)以及式(8)的推算过程可得：

$$\begin{aligned} \hat{G}_t &= \frac{T - \sum_{m=1}^M P_m^2 T_{m_t} + \sum_{m=1}^M P_m T_{m_t} + \int_{a_{m_t}}^{b_{m_t}} \sum_{m=1}^M P_m F_{m_t} - \left(\sum_{m=1}^M P_m F_{m_t} \right)^2 ds}{\hat{u}_t} \\ &= \frac{T - \sum_{m=1}^M P_m^2 T_{m_t}}{\hat{u}_t} + \sum_{m=1}^M P_m \hat{w}_{m_t} \hat{G}_{m_t} + \sum_{m=1}^M \sum_{j>m}^M P_m P_j \hat{d}_{m_j} \end{aligned} \quad (15)$$

$$\text{这里 } \hat{d}_{m_j} = \frac{\int_{a_{m_t}}^{b_{m_t}} (F_{m_t} + F_{j_t} - 2F_{m_t} F_{j_t}) ds}{\hat{u}_t}, \quad \hat{G}_L = \frac{T - \sum_{m=1}^M P_m^2 T_{m_t}}{\hat{u}_t}$$

即将 \hat{d}_{m_j} 定义为两个子群组之间污染排放分布差异的相对指标。 \hat{G}_L 表示在样本有限的情况下，由于总体上下界处的跳跃点的不连续带来的误差。那么，总体基尼系数在假设四下的最终分解形式为：

$$\hat{G}_t = \hat{G}_L + \sum_{m=1}^M P_m \hat{w}_{m_t} \hat{G}_{m_t} + \sum_{m=1}^M \sum_{j>m}^M P_m P_j \hat{d}_{m_j} \quad (16)$$

可以看出，总体的基尼系数分解成了三个部分，组内差异、组间差异以及样本误差项。有限样本混合环境基尼系数相对于大样本收入基尼系数的主要区别在于考虑了跳跃点处界限的影响，多了一个样本误差项 \hat{G}_L ，误差项是对基尼系数的修正，并无实际的解释意义。并且该公式与区域各群组排放数据的上下界 a_{m_t} ， b_{m_t} 相关联，对各群组内的排放情况的表示更加清晰。在混合基尼系数的表达形式上看，混合基尼系数是组内基尼系数、组间基尼系数和样本误差项的线性组合。

3. 环境基尼系数在区域平衡发展评价中的应用

本节采用上述方法对我国的工业经济和生态空间平衡进行了评价和分析。从城市以及城市群的角度出发，按照我国的城市群的地理位置，选取了 5 个核心城市群共计 92 个城市作为研究对象，分别是京津冀、珠三角、长三角、长江中游和成渝城市群。本节整理了 2005~2020 年各城市的 GDP (亿元)、社会用电量(亿千瓦时)、工业废水排放量(万吨)和工业 SO_2 排放量(吨)的数据，样本数据来自《中国统计年鉴》、《中国城市统计年鉴》以及各省份统计年鉴。区域 GDP 被选为工业产出指标，并且以 2005 为基期，对

各城市群 2006~2020 年的 GDP 进行平减处理；以市社会用电量作为能源消耗的指标；以 SO_2 排放代表大气污染，以工业废水排放代表水资源污染。

从图 1 可以看出，长江上游和长三角城市群占 GDP、能源消耗和环境污染排放的主要比例。工业能源消耗、污染排放和区域工业产出的空间特征表现出一定程度的复杂性，各城市群之间存在较大差异，各城市群内部城市之间亦存在较大差异。

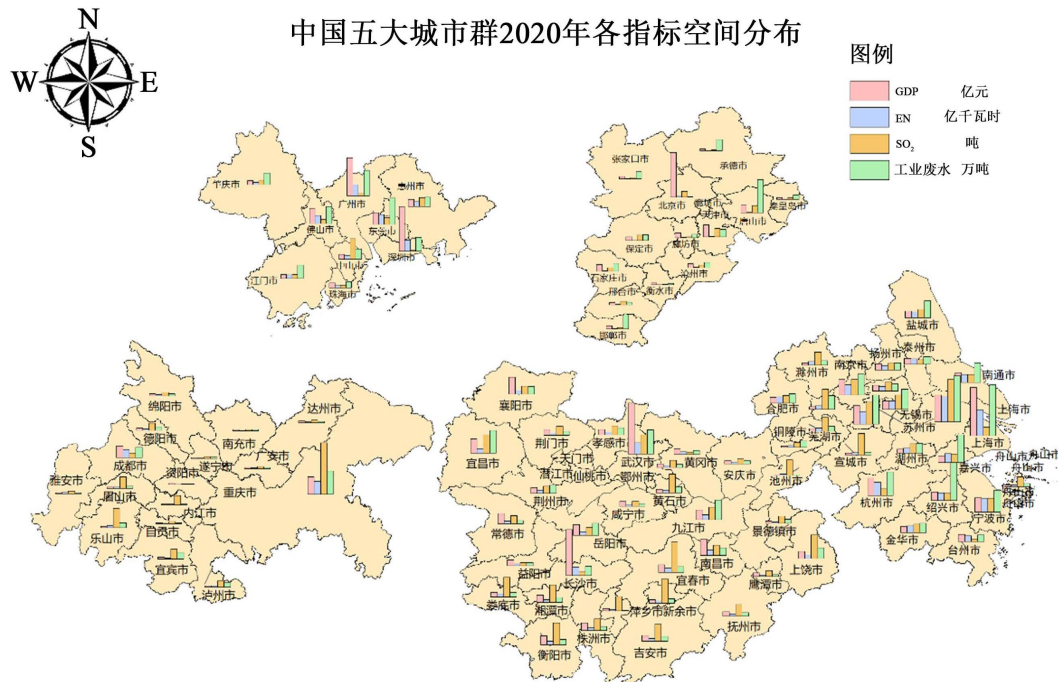


Figure 1. Spatial distribution of industrial output, energy consumption and pollution emissions in China's five major urban agglomerations

图 1. 中国五大城市群工业产出、能源消耗、污染排放空间分布图

3.1. 用非分组数据和分组数据拟合工业产出、能源消耗和污染排放的分布

大部分样本数据对于厚尾分布比如对数正态、对数逻辑、Gamma 以及 Weibull 分布的拟合效果更好，而对于正态分布、指数分布或极值分布等的拟合优度 P 值往往小于 0.05，因此在样本量有限时，厚尾分布对于样本数据有着更强的适应性。

3.1.1. 非分组总体数据的分布拟合

将 2005~2020 年的各指标数据按国民经济的 5 年规划期为一个时间段进行分组，以各分组均值为基础进行分布函数拟合。而在厚尾分布中，对数逻辑和对数正态对于总体非分组数据的拟合效果更佳。

样本数据的主要特征主要体现在“厚尾”上，对数正态分布和对数逻辑分布均表现出明显的“峰值”和“厚尾”特征，且这两种分布类型有着相近的分布特点，而且在三参数形式下的各分布均有着相近的分布特点。在十一五到十三五期间，GDP 数据和能源消耗数据的峰值特征在逐渐下降，但厚尾特征依然存在；而污染排放数据的峰值特征在逐渐上升，且厚尾特征依然存在(图 2)。

由于我国疆域辽阔且分布复杂，地区发展机会不均等，区域发展分化严重，差距明显，一线核心城市发展程度远高于其他城市。随着区域协调发展战略以及可持续发展战略的贯彻实施，各地区发展逐渐提升且差距缩小，与一线城市间差距依旧显著，数据厚尾特征减小但依旧存在。

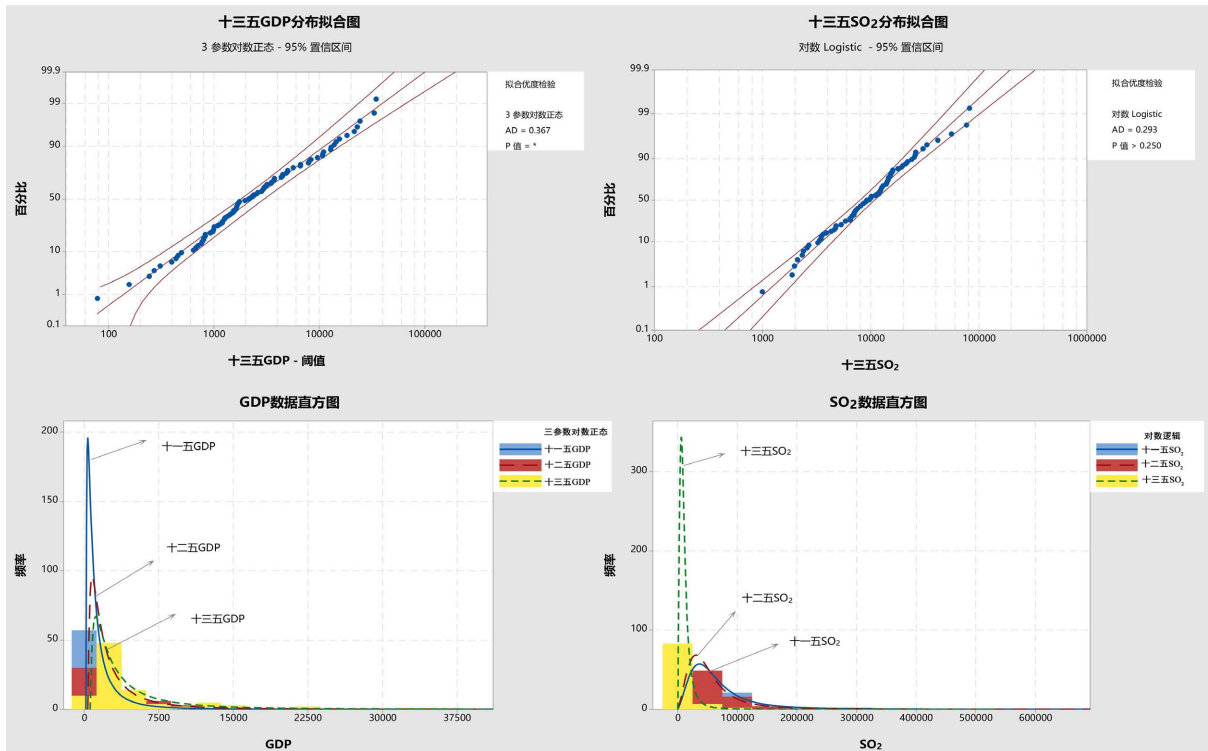


Figure 2. Fitting plot of GDP and SO₂ distribution during the 13th Five-Year Plan period

图 2. 十三五时期 GDP 和 SO₂ 分布拟合图

3.1.2. 以城市群为分组的分布拟合

分组数据每组采用四类经济和生态效益指标，分布拟合方法和统计检验方法与非分组数据相同。与非分组数据相比，各城市群的样本数据对于四种厚尾分布的拟合和效果均表现较为良好(表 1)，因此这里各统计指标均采用标准形式的厚尾分布。各城市群数据对于不同类型分布的拟合检验中，基本上最佳分布都为对数正态分布或对数逻辑分布。对每种指标选取一种分布进行可视化后如图 3 所示，可以看出区域数据表现出明显的厚尾特征。

与非分组数据相比，各城市群显示出显著的区域和时间差异，经济和生态效益指标间也表现出一定的差异。在 2005~2020 期间，GDP 和能源排放指标的峰值特征有下降趋势，而污染排放指标的峰值特征则有上升趋势；所有区域的 GDP 的厚尾特征依然存在且没有太大的变化，但 SO₂ 排放的厚尾特征逐渐下降；长三角和京津冀城市群的时间差异较为突出，其能源消耗和工业废水排放的厚尾特征随时间推移而逐渐下降，而其他城市群则对此不敏感(图 4)。

Table 1. The fitting results of different distribution function types were performed on the data of each urban agglomeration
表 1. 对各城市群数据进行不同分布函数类型的拟合结果

指标	分布类型	京津冀	长三角	珠三角	长江上游	成渝
GDP	对数正态	0.484 (0.188)	0.629 (0.091)	0.459 (0.198)	0.212 (0.838)	0.5 (0.151)
	Weibull	0.931 (0.015)	1.757 (<0.010)	0.624 (0.088)	0.476 (0.231)	0.582 (0.115)
	Gamma	1.064 (0.012)	1.624 (<0.005)	0.681 (0.089)	0.46 (>0.250)	0.618 (0.128)
	对数逻辑	0.365 (>0.250)	0.363 (>0.250)	0.454 (0.205)	0.228 (>0.250)	0.533 (0.116)

Continued

Energy	对数正态	0.309 (0.514)	0.311 (0.532)	0.388 (0.306)	0.164 (0.934)	0.946 (0.012)
	Weibull	0.522 (0.18)	1.23 (<0.01)	0.476 (0.219)	0.431 (>0.250)	1.6 (<0.01)
	Gamma	0.48 (0.246)	0.692 (0.078)	0.464 (>0.250)	0.328 (>0.250)	1.841 (<0.005)
	对数逻辑	0.315 (>0.250)	0.166 (>0.250)	0.421 (0.243)	0.161 (>0.250)	0.664 (0.047)
工业废水	对数正态	0.312 (0.526)	0.583 (0.117)	0.18 (0.885)	0.264 (0.669)	0.417 (0.292)
	Weibull	0.338 (>0.250)	1.026 (0.025)	0.271 (>0.250)	0.298 (>0.250)	0.726 (0.049)
	Gamma	0.348 (>0.250)	0.837 (0.037)	0.242 (>0.250)	0.29 (>0.250)	0.847 (0.038)
	对数逻辑	0.344 (>0.250)	0.337 (>0.250)	0.215 (>0.250)	0.304 (>0.250)	0.415 (>0.250)
SO ₂	对数正态	0.229 (0.76)	0.347 (0.456)	0.207 (0.806)	0.317 (0.52)	0.199 (0.86)
	Weibull	0.284 (>0.250)	0.363 (>0.250)	0.348 (>0.250)	0.821 (0.13)	0.484 (0.219)
	Gamma	0.255 (>0.250)	0.335 (>0.250)	0.324 (>0.250)	0.542 (0.189)	0.573 (0.174)
	对数逻辑	0.174 (>0.250)	0.384 (>0.250)	0.177 (>0.250)	0.305 (>0.250)	0.181 (>0.250)

注：括号内为拟合优度检验的 P 值，括号上面是 A-D 检验值。

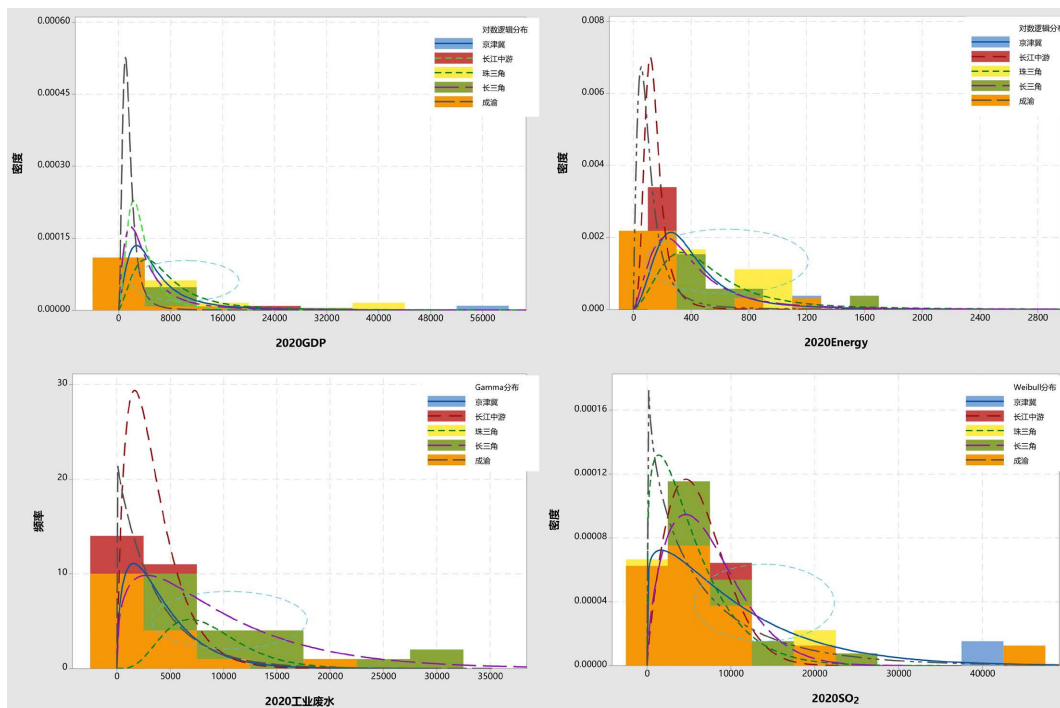


Figure 3. Histogram of grouped data by indicator in 2020

图 3. 2020 年各指标分组数据直方图

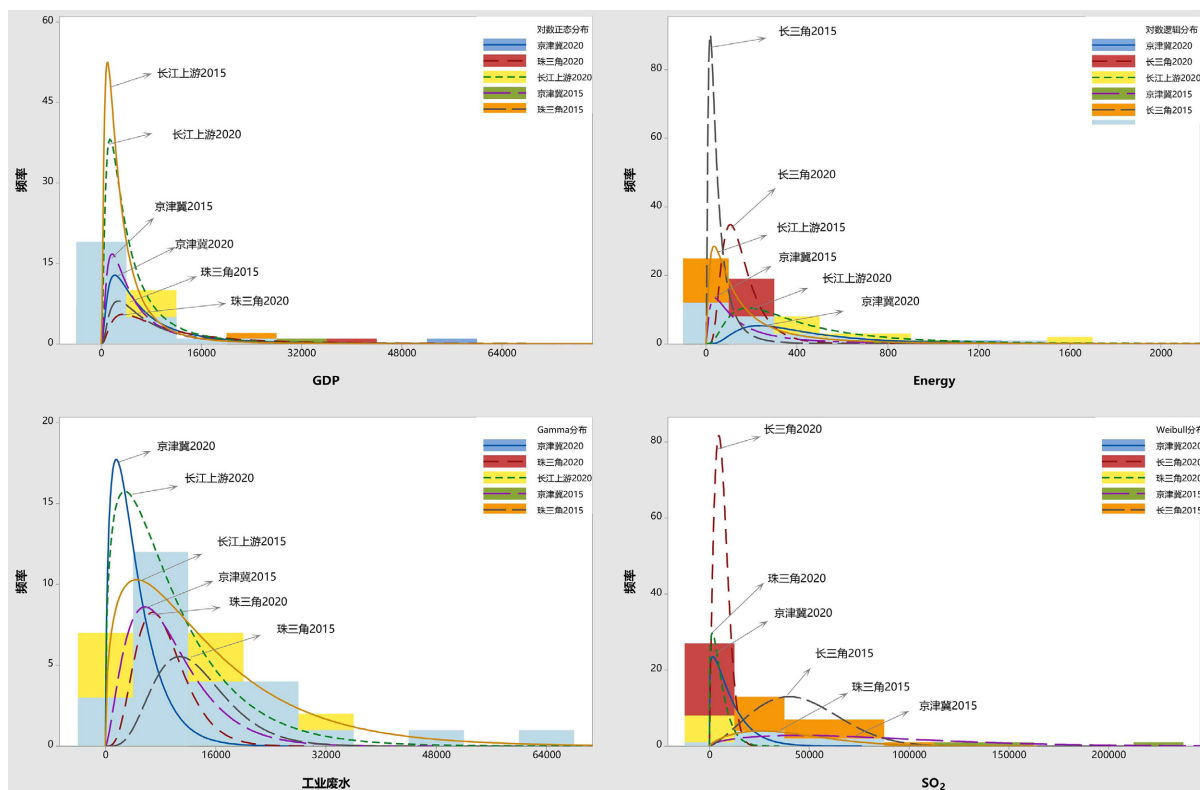


Figure 4. Histogram of grouped data for each metric under a specific distribution

图 4. 特定分布下的各指标分组数据直方图

3.2. 城市群间经济生态区域平衡发展的混合环境基尼系数评价

本节将使用多种方法计算四种指标的基尼系数，包括非参数方法、参数化方法、以及多群组基尼分解方法。首先，在非分组数据下，根据样本数据的特点，分析基尼系数计算方法对基尼系数结果的影响；其次，根据多群组混合基尼系数的分解公式，分析了空间异质性对于区域平衡发展的影响以及其动态发展趋势，参照收入基尼系数的标准对环境基尼系数进行解读，以 0.4 为界，小于 0.4 认为可以接受的区间。

如图 5 结果表明，非参数方法与参数方法计算得到的基尼系数结果趋势大致吻合。对于拟合精度较低的分佈，非参数基尼系数¹与连续状态下参数基尼系数结果相近，且基尼系数结果偏高，基本都在 0.4 以上，能源排放甚至超过了 0.6，放大了现实中的区域间发展不平衡。不连续下的基尼系数结果相比原来降低了 0.04 左右，说明对于以空间单元为样本单位的数据，不连续下的基尼系数有着更好的解释效果。为探索区域发展不平衡的组成成分，将总体划分若干群组，构建多群组下的混合基尼系数进行区域平衡发展评价更具有现实性意义。

多群组混合基尼系数相比于非分组参数基尼系数结果要略低，并且根据式(16)可知，多群组混合基尼系数将发展失衡细分为组间差异和组内差异，并且组间差异是区域发展不平衡的主要来源，其贡献率在约在 0.7 左右(表 2)，因此其较之总体基尼系数更有解释性，更加贴近真实的发展状态。除工业废水排放数据的不连续混合基尼系数结果相比连续混合基尼略有降低，其余指标的不连续混合基尼系数则比连续混合基尼系数要高，但其结果的动态趋势大致相同，说明基于有限样本前提的混合基尼系数推导是合理的，并且解释性更佳。

¹ 非参数基尼系数的计算公式为：
$$G = \frac{\sum_{j=1}^N \sum_{i=1}^N |s_j - s_i|}{2N^2 u}$$

Table 2. Contribution rate of between-group and within-group differences in multi-group mixed Gini coefficients
表 2. 多群组混合基尼系数的组间和组内差异的贡献率

指标	发展失衡	2005	2010	2015	2020
GDP	组间差异	0.704	0.734	0.741	0.729
	组内差异	0.296	0.266	0.259	0.271
Energy	组间差异	0.745	0.761	0.717	0.681
	组内差异	0.255	0.239	0.283	0.319
工业废水	组间差异	0.668	0.641	0.702	0.671
	组内差异	0.332	0.359	0.298	0.329
SO ₂	组间差异	0.707	0.68	0.706	0.703
	组内差异	0.293	0.32	0.294	0.297

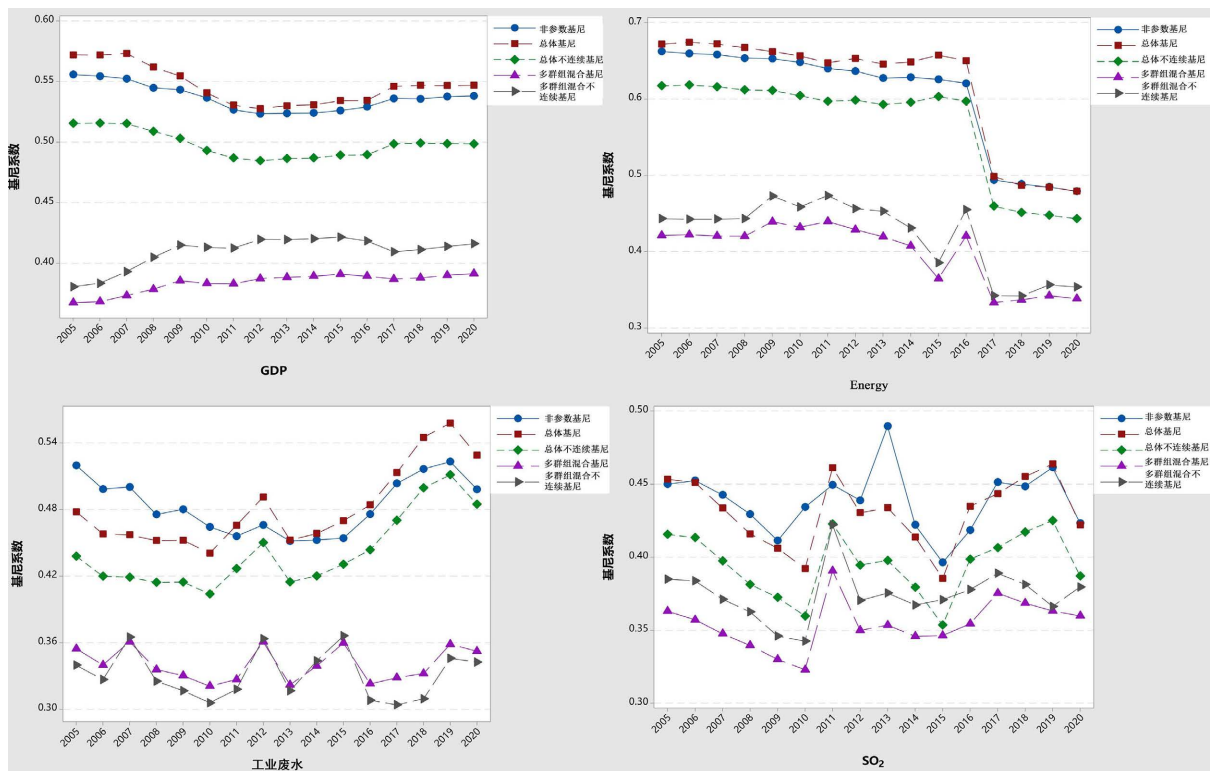


Figure 5. Gini coefficients for GDP, Energy, industrial wastewater and SO₂ under different calculation methods
图 5. 不同计算方法下的 GDP、Energy、工业废水和 SO₂ 的基尼系数

根据图 6 中的计算结果来看，在 08~20 年期间 GDP 基尼系数仍大于 0.4，且无下降趋势，表明我国区域经济效益的发展不平衡并未得到很好的改善。在考察期内，虽然区域污染排放基尼系数呈现出波动状态，但总体上较为平衡，基尼系数基本小于 0.4；而能源消耗失衡显著，但其不平衡状态大体上呈现出下降趋势。我们以能源消耗基尼系数为横轴，以经济产出和污染排放系数为纵轴，显示了“清洁生产效率”的演化趋势，结果如图 6 所示。各区域的清洁生产效率的差异是分阶段性的。

在十一五规划期间，能源消耗不平衡性显著，其基尼系数在 0.45 左右，而区域污染排放平衡状态(工业废水和 SO₂)均呈现上升趋势，区域清洁生产效率存在较大的不平衡；在十二五规划期间，区域经济

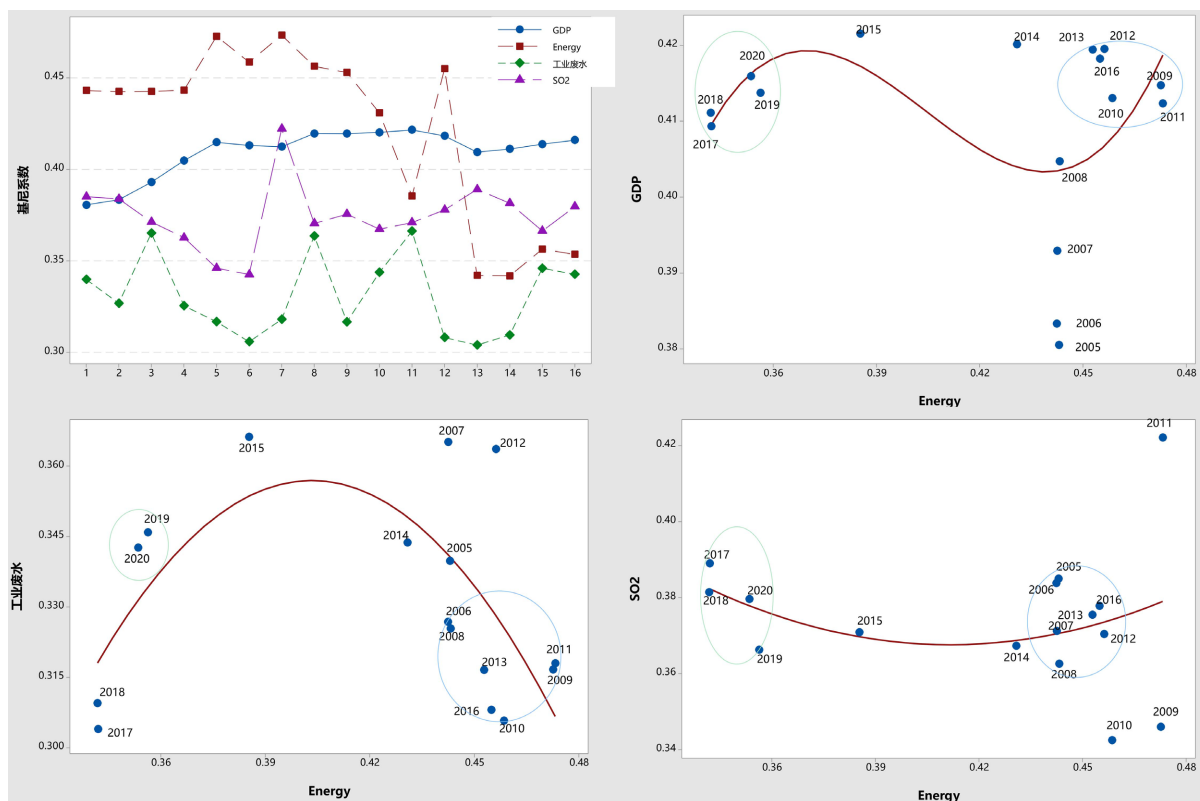


Figure 6. Time series of multi-group mixed environmental Gini coefficients for GDP, Energy, industrial wastewater and SO_2 and their relationships

图 6. GDP、Energy、工业废水和 SO_2 的多群组混合环境基尼系数的时间序列及其之间的关系

产出不平衡仍然较高，能源消耗发展失衡状态逐渐降低，污染排放则呈现出波动状态。十三五规划期间，区域能源消耗趋于平衡，基尼系数为保持在 0.35 左右，而工业废水污染排放的平衡呈现下降趋势，其清洁生产率的平衡无法维持前一阶段的趋势，呈下降趋势。这表明中国区间发展平衡主要依靠“投入端”和“产出端”之间的协调发展。

4. 结论

本文基于参数化基尼的框架将传统的收入分配基尼系数拓展到区域环境基尼系数，并在此基础上，构造出一种有限样本下的多群组总体基尼系数，并参照传统的基尼系数分解方法推导出总体基尼系数的分解形式，然后用于度量中国五大城市群经济和生态发展的区域差异及差异组成。通过对 2005~2020 年五大城市群的面板数据的分布拟合可知，我国的区域经济和生态数据呈现出明显的厚尾特征，基于厚尾分布的基尼系数可以很好地适应区域经济和生态效益指标数据。当使用多群组混合环境基尼系数来评价区域发展是否平衡时，可以充分考虑区域发展条件的差异，对区域发展趋势表达也更为精确，更符合区域平衡发展的内涵。通过对总体基尼系数的分解以及区域经济和生态指标的综合评价，从生产力水平和区域空间异质性两个方面，可以揭示我国区域发展不平衡的根源。基尼系数的计算结果显示有限样本下的多群组基尼系数相比于传统多群组基尼系数结果略高，这同时论证了我国区域数据的厚尾性，即厚尾性的存在使得真实结果要高于传统基尼方法的结果。实证研究结果表明，我国区域发展不平衡仍较为明显，城市群内部经济和生态方面的发展也并不相适应。我国区域发展不平等的主要来源于组间差异即空间异质性，其贡献率约为 70%。区域能源消耗和经济产出的不平衡以及污染排放的平衡表明，中国各

地区之间存在着“清洁生产效率”的差异。到了十三五期间，能源消耗的发展逐渐趋于平衡且清洁生产效率也呈现出下降趋势。一个区域的高质量发展应该是区域经济、生态等各个方面的协调发展，整体推进；牺牲一方面而致力于另一方面的单极化发展并不符合新发展格局的要求。因此我们在追求投入端——经济的提升的同时，也要着力于产出端——污染排放的减少，达到经济 - 生态双向整体平衡发展。

基金项目

本文获国家自然科学基金项目(项目编号: 71701105)、教育部人文社会科学研究规划项目(项目编号: 22YJA630098)、江苏省社会科学项目(项目编号: 22GLB022)、国家社会科学基金重大项目(项目编号: 17ZDA092)等项目的资助。

参考文献

- [1] 张可云, 张江. 城市群多中心性与绿色发展效率——基于异质性的城镇化空间布局分析[J]. 中国人口·资源与环境, 2022, 32(2): 107-117.
- [2] 于伟, 张鹏, 姬志恒. 中国城市群生态效率的区域差异、分布动态和收敛性研究[J]. 数量经济技术经济研究, 2021, 38(1): 23-42.
- [3] 吴健生, 门新纳, 梁景天, 赵宇豪. 基于基尼系数的生态系统服务供需均衡研究——以广东省为例[J]. 生态学报, 2020, 40(19): 6812-6820.
- [4] 李博闻, 黄正东, 蒯希, 于溪. 基于空间公平理论的公共交通服务评价——以深圳市为例[J]. 地理科学进展, 2021, 40(6): 958-966.
- [5] Wang, F., Fan, W., Liu, J., et al. (2020) The Effect of Urbanization and Spatial Agglomeration on Carbon Emissions in Urban Agglomeration. *Environmental Science and Pollution Research*, **27**, 24329-24341. <https://doi.org/10.1007/s11356-020-08597-4>
- [6] Panzera, D. and Postiglione, P. (2020) Measuring the Spatial Dimension of Regional Inequality: An Approach Based on the Gini Correlation Measure. *Social Indicators Research*, **148**, 379-394. <https://doi.org/10.1007/s11205-019-02208-7>
- [7] Liang, H., Dong, L., Luo, X., et al. (2016) Balancing Regional Industrial Development: Analysis on Regional Disparity of China's Industrial Emissions and Policy Implications. *Journal of Cleaner Production*, **126**, 223-235. <https://doi.org/10.1016/j.jclepro.2016.02.145>
- [8] Bahadur, R.R. and Savage, L.J. (1956) The Nonexistence of Certain Statistical Procedures in Nonparametric Problems. *The Annals of Mathematical Statistics*, **27**, 1115-1122. <https://doi.org/10.1214/aoms/1177728077>
- [9] Fontanari, A., Taleb, N.N. and Cirillo, P. (2018) Gini Estimation under Infinite Variance. *Physica A: Statistical Mechanics and Its Applications*, **502**, 256-269. <https://doi.org/10.1016/j.physa.2018.02.102>
- [10] 杨泓川, 陈松林. 生态压力视角下福建省区域经济协调发展研究[J]. 生态科学, 2020, 39(1): 198-206.
- [11] Kong, Y., Zhao, T., Yuan, R., et al. (2019) Allocation of Carbon Emission Quotas in Chinese Provinces Based on Equality and Efficiency Principles. *Journal of Cleaner Production*, **211**, 222-232. <https://doi.org/10.1016/j.jclepro.2018.11.178>
- [12] Liu, S., Yang, C. and Liu, L. (2022) Identifying Spatial Relations of Industrial Carbon Emissions among Provinces of China: Evidence from Unsupervised Clustering Algorithms. *Environmental Science and Pollution Research*, **29**, 77958-77972. <https://doi.org/10.1007/s11356-022-20784-z>
- [13] Peng, G., Zhang, X., Liu, F., et al. (2021) Spatial-Temporal Evolution and Regional Difference Decomposition of Urban Environmental Governance Efficiency in China. *Environment, Development and Sustainability*, **23**, 8974-8990. <https://doi.org/10.1007/s10668-020-01007-2>
- [14] Wu, J., Wei, Y.D., Chen, W., et al. (2019) Environmental Regulations and Redistribution of Polluting Industries in Transitional China: Understanding Regional and Industrial Differences. *Journal of Cleaner Production*, **206**, 142-155. <https://doi.org/10.1016/j.jclepro.2018.09.042>
- [15] Zhong, Z., Peng, B. and Elahi, E. (2021) Spatial and Temporal Pattern Evolution and Influencing Factors of Energy-Environmental Efficiency: A Case Study of Yangtze River Urban Agglomeration in China. *Energy & Environment*, **32**, 242-261. <https://doi.org/10.1177/0958305X20923114>
- [16] Druckman, A. and Jackson, T. (2008) Measuring Resource Inequalities: The Concepts and Methodology for an Area-Based Gini Coefficient. *Ecological Economics*, **65**, 242-252. <https://doi.org/10.1016/j.ecolecon.2007.12.013>

- [17] Chen, J., Cheng, S., Song, M., *et al.* (2016) Interregional Differences of Coal Carbon Dioxide Emissions in China. *Energy Policy*, **96**, 1-13. <https://doi.org/10.1016/j.enpol.2016.05.015>
- [18] Kendall, M. and Stuart, A. (1977) *The Advanced Theory of Statistics*. Vol. 1: Distribution Theory. Griffin, London.
- [19] 程永宏. 二元经济中城乡混合基尼系数的计算与分解[J]. *经济研究*, 2006(1): 109-120.
- [20] 程永宏. 基尼系数组群分解新方法研究: 从城乡二亚组到多亚组[J]. *经济研究*, 2008, 43(8): 124-135.
- [21] 陈希孺. 基尼系数及其估计[J]. *统计研究*, 2004(8): 58-60.
- [22] Catalano, M.T., Leise, T.L. and Pfaff, T.J. (2009) Measuring Resource Inequality: The Gini Coefficient. *Numeracy*, **2**, Article No. 4. <https://doi.org/10.5038/1936-4660.2.2.4>
- [23] 艾小青, 祁磊. 信息不完全下收入或财富基尼系数的估算[J]. *数量经济技术经济研究*, 2021, 38(6): 146-165.
- [24] 艾小青. 城乡混合基尼系数分解方法研究[J]. *统计研究*, 2015, 32(9): 91-96.
- [25] Lambert, P.J. and Aronson, J.R. (1993) Inequality Decomposition Analysis and the Gini Coefficient Revisited. *The Economic Journal*, **103**, 1221-1227. <https://doi.org/10.2307/2234247>
- [26] Eliazar, I. and Cohen, M.H. (2014) On Social Inequality: Analyzing the Rich-Poor Disparity. *Physica A: Statistical Mechanics and Its Applications*, **401**, 148-158. <https://doi.org/10.1016/j.physa.2014.01.033>
- [27] 黄和平. 基于多角度基尼系数的江西省资源环境公平性研究[J]. *生态学报*, 2012, 32(20): 6431-6439.
- [28] Warrens, M.J. (2018) On the Negative Bias of the Gini Coefficient Due to Grouping. *Journal of Classification*, **35**, 580-586. <https://doi.org/10.1007/s00357-018-9267-9>
- [29] 杨欢, 吕承超. “新医改”十年: 中国医疗卫生服务效率的区域差异、动态演进及影响因素研究[J]. *中国管理科学*, 2023, 31(2): 162-172. <https://doi.org/10.16381/j.cnki.issn1003-207x.2021.0539>
- [30] Cowell, F.A. (2000) Measurement of Inequality. In: *Handbook of Income Distribution*, Vol. 1, Elsevier, Amsterdam, 87-166. [https://doi.org/10.1016/S1574-0056\(00\)80005-6](https://doi.org/10.1016/S1574-0056(00)80005-6)
- [31] 田凤平, 秦瑾龙, 杨科. 中国三大城市群经济发展的区域差异及收敛性研究[J]. *系统工程理论与实践*, 2021, 41(7): 1709-1721.
- [32] 何刚, 王莹莹, 阮君, 赵杨秋. 基于 Dagum 基尼系数的生态安全区域差异及收敛度研究——以淮河生态经济带安徽段为例[J]. *水土保持研究*, 2021, 28(6): 348-355.
- [33] Cheng, S., Fan, W., Zhang, J., *et al.* (2021) Multi-Sectoral Determinants of Carbon Emission Inequality in Chinese Clustering Cities. *Energy*, **214**, Article ID: 118944. <https://doi.org/10.1016/j.energy.2020.118944>
- [34] Miao, Z., Chen, X. and Baležentis, T. (2021) Improving Energy Use and Mitigating Pollutant Emissions across “Three Regions and Ten Urban Agglomerations”: A City-Level Productivity Growth Decomposition. *Applied Energy*, **283**, Article ID: 116296. <https://doi.org/10.1016/j.apenergy.2020.116296>
- [35] Wang, S., Zhang, Y. and Wen, H. (2021) Comprehensive Measurement and Regional Imbalance of China’s Green Development Performance. *Sustainability*, **13**, Article No. 1409. <https://doi.org/10.3390/su13031409>
- [36] Chen, J., Chen, J., Miao, Y., *et al.* (2016) Unbalanced Development of Inter-Provincial High-Grade Highway in China: Decomposing the Gini Coefficient. *Transportation Research Part D: Transport and Environment*, **48**, 499-510. <https://doi.org/10.1016/j.trd.2015.06.008>
- [37] Lau, P.L., Koo, T.T.R. and Dwyer, L. (2017) Metrics to Measure the Geographic Characteristics of Tourism Markets: An Integrated Approach Based on Gini Index Decomposition. *Tourism Management*, **59**, 171-181. <https://doi.org/10.1016/j.tourman.2016.07.019>
- [38] 艾小青. 有限总体基尼系数的统计推断[J]. *统计与信息论坛*, 2016, 31(3): 3-8.