

基于面板数据的灰色指标关联模型构建及其应用

党耀国, 尚中举[†], 王俊杰, 冯 宇

(南京航空航天大学 经济与管理学院, 南京 211106)

摘要: 针对现有面板数据灰色关联模型中对象的排列顺序变换引起灰色关联序不一致的缺陷, 提出一种基于面板数据的新型灰色指标关联模型. 考虑到面板数据的时间维度和对象维度差异特征, 在时间维上利用增量表征指标的发展水平, 对象维上引入离差表征指标的分布特征, 同时将两个维度上的方向差异作为正负关联判断依据, 构建时间维度和对象维度的关联系数, 得到面板数据的灰色指标关联模型, 并对模型的唯一性、对称性和可比性等性质进行证明. 最后将该模型应用于苏南五市空气质量与主要污染物浓度的关联分析中, 验证了灰色指标关联度不受对象排列顺序的影响, 且能够反映指标间的正负关联关系.

关键词: 灰色关联模型; 面板数据; 时间维度; 对象维度; 增量; 离差; 大气污染

中图分类号: N941.5

文献标志码: A

Grey incidence model for relationship between indicators of panel data and its application

DANG Yao-guo, SHANG Zhong-ju[†], WANG Jun-jie, FENG Yu

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: An innovative grey indicator incidence model based on the panel data is presented for the inconsistency of grey relation order resulting from order changing of objects in the grey incidence model of the current panel data. For the difference between the time dimension and the object dimension of panel data, increment and deviation are introduced to respectively represent the development level of the indicator in the time dimension and distribution characteristics of the indicator in the object dimension. This paper also introduces the difference in directions between the above two dimensions to be a judgement standard of positive and negative incidence, and builds an incidence coefficient of the time dimension and object dimension. Moreover, the grey indicator incidence model of panel data is built, and the uniqueness, symmetry and comparability of the model is verified. As a result, the model is applied to the incidence analysis between air quality and concentration of main pollutant for five cities in South-Jiangsu province, verifying that the degree of grey indicator incidence is not related to the objects' order, and indicating the positive and negative incidence between indicators.

Keywords: grey incidence model; panel data; time dimension; object dimension; increment; deviation; air pollution

0 引言

随着经济社会高速发展、工业化水平不断提升、城镇化不断推进,新常态背景下中国大气污染问题日益严重,已经严重影响到人民的生活质量、交通效率和教育工作,成为全社会普遍性重大环境和社会问题. 2012年中国颁布的《环境空气质量标准》(GB3095-2012)分别调整和增设了NO₂及PM_{2.5}

浓度限制等多项内容. 由于新标准实施时间短、内容变化大,加之内外扰动的存在和人类认知水平的局限,统计结果出现了“有效数据少、信息不完全”的特点,给传统数理统计分析方法带来了诸多困扰与不便,而近年来快速发展的灰色关联分析正擅长处理这类“部分信息已知”的多因素影响不确定性系统问题.

收稿日期: 2017-11-14; 修回日期: 2018-01-23.

基金项目: 国家自然科学基金项目(71771119, 71371098, 71701101); 江苏省社会科学基金重点项目(16GLA001); 南京航空航天大学基本科研业务费重大培育项目(NP2016301); 南京航空航天大学博士学位论文创新与创优基金项目(BCXJ15-10).

责任编辑: 李登峰.

作者简介: 党耀国(1964—), 男, 教授, 博士生导师, 从事灰色系统理论、数量经济等研究; 尚中举(1993—), 男, 硕士生, 从事灰色系统理论的研究.

[†]通讯作者. E-mail: shangzhongju@163.com.

灰色关联分析既是灰色系统理论的重要内容之一,也是系统分析方法体系的一员.基于邓聚龙^[1]提出的灰色关联度模型,许多学者在新型灰色关联度模型的构造方面进行了有益的探索,如基于全局视角的广义关联度^[2]、基于接近性和相似性视角的新型关联度^[3]、基于变化态势的斜率关联度^[4]以及基于三次样条插值的样条关联度^[5]等;在模型性质方面,Xie等^[6]讨论了几类关联度模型的平行性和一致性,崔杰等^[7]讨论了几类模型的仿射性和仿射变换保序性,魏勇等^[8]指出灰色关联公理中整体性和偶对称性的不必要性,灰色关联分析在一维数据分析的发展取得了长足的进步.但诸如大气污染这类复杂系统不仅具有持续性,还与自然、地理、经济发展、能源结构和环境规制等因素相关,迫切需要运用关于面板数据、矩阵序列和高维场数据的分析方法研究解决^[9].

面板数据灰色关联分析是灰色关联分析的重要组成部分,虽然对于灰色关联模型的研究已经持续近30年,但是针对面板数据灰色关联模型的研究还处于起步阶段,现有成果主要集中在基于矩阵时空特征表示和基于三维空间几何特征表示两大类.以矩阵时空特征表示灰色关联度的代表性成果有具有时空属性的灰色矩阵关联模型^[10]、存在发展速度和增长速度的灰色矩阵相似关联模型^[11]、包含三重差异信息的灰色指标关联聚类模型^[12],此类模型旨在通过数据的潜在时空关系反映序列的关联程度,但对负关联度的构造与表示较为困难.以三维空间几何特征表示灰色关联度的代表性成果有扩展绝对关联模型^[13]和空间网络关联模型^[14].前者将由空间三角形构成的多个曲面组成曲面簇,以曲面簇间围成的曲顶柱体构造三维灰色绝对关联模型;后者将空间相邻节点连接组成网格状结构,根据网格对应独立线段的斜率构造网格关联模型,同时引入了面板数据负关联的初步表示方法.类似地,还有基于相对面积的面板数据接近关联模型^[15]、基于增量构成比和构成差的面板数据灰关联度模型^[16]、基于空间向量夹角和向量差的模的相似性和接近性关联度模型^[17]以及基于序列折线间角度的灰色动态趋势关联模型^[18]等.现有此类模型存在忽视面板数据的自身维度特征而导致几何特征(如曲面簇和网格)表示不唯一和关联序不稳定的问题,同时关联系数正负符号判别方法不严谨也给关联结果的优势分析带来了困难.

事实上,面板数据是同时在时间和截面上取得的二维数据^[19],在一维数据中,李宏艳等^[20]指出截面数据与时间序列的本质不同,即时间序列中数的排列是

固定的,而截面数据中数的排列顺序是可以改变的,吴利丰等^[21]通过织机评价的关联决策案例验证了截面数据中对象顺序对关联序的影响,因此具有多维特征的面板数据中对象排列顺序的可变性决定了其灰色关联模型构建过程中时间维度与对象维度差别化处理的必然性.同时,社会经济等系统中常面临的单调性序列相对较少,相互联系的因素之间的发展趋势不总是呈相同方向,仅包含正关联度的灰色模型对于震荡序列的处理存在盲区,现有模型虽从多种角度对负关联度的表示进行了尝试,但有关正负关联度的比较和排序仍存在争议,有待进一步探索.

基于此,本文为了解决面板数据对象排列顺序影响关联序的问题,一方面从时间维度考察各时段不同指标的发展特征,另一方面从对象维度衡量各对象不同指标的分布特征,通过提取各维度不同指标间发展程度和方向的相对差异,对其关联大小和正负分别进行度量,构建基于指数函数的灰色指标关联模型.借鉴矢量比较的方法对关联度绝对值进行比较,以解决现有模型正负关联度的排序问题.最后将该模型应用于苏南五市空气质量与主要污染物浓度的关联分析中,验证了灰色指标关联度不受对象排列顺序的影响,且能够反映指标间的正负关联关系.

1 面板数据灰色指标关联模型的构建

1.1 面板数据的空间表示及矩阵表征

面板数据具有时间维度和对象维度的特征.图1为江苏省10个城市近10年GDP增速的变化情况.从时间维度看,各对象的观测值都组成一个时间序列;从对象维度看,是由若干个对象在某一时刻构成的截面观测值.对于面板数据中的每一个观测值,可以一一对应到矩阵的元素中.

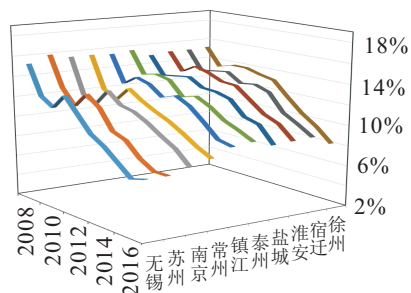


图1 2007~2016年江苏省10个城市的GDP增速变化

*注:数据来源于各地《国民经济和社会发展统计公报》

定义1 设系统中有 N 个不同指标, M 个对象,观测时刻个数为 T .令 $x_n^*(m,t)$ 为第 n 个指标($n=1,2,\dots,N$)第 m 个对象($m=1,2,\dots,M$)在第 t 时刻的观测值($t=1,2,\dots,T$),则第 n 个指标的面板数据 X_n^* 表示为

$$X_n^* = \begin{bmatrix} x_n^*(1,1) & x_n^*(1,2) & \cdots & x_n^*(1,T) \\ x_n^*(2,1) & x_n^*(2,2) & \cdots & x_n^*(2,T) \\ \vdots & \vdots & \ddots & \vdots \\ x_n^*(M,1) & x_n^*(M,2) & \cdots & x_n^*(M,T) \end{bmatrix}.$$

定义2 设第 n 个指标的面板数据为 X_n^* , 经过序列算子 $D_k (k = 1, 2)$ 作用后的序列为 X_n , 有

$$X_n = \begin{bmatrix} x_n(1,1) & x_n(1,2) & \cdots & x_n(1,T) \\ x_n(2,1) & x_n(2,2) & \cdots & x_n(2,T) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(M,1) & x_n(M,2) & \cdots & x_n(M,T) \end{bmatrix}.$$

当

$$x_n(m,t) = x_n^*(m,t)d_1 = \frac{x_n^*(m,t)}{\bar{x}},$$

$$\bar{x} = \frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T x_n^*(m,t), \quad (1)$$

称 D_1 为面板数据均值化算子^[14], $m = 1, 2, \dots, M, t = 1, 2, \dots, T$. 当

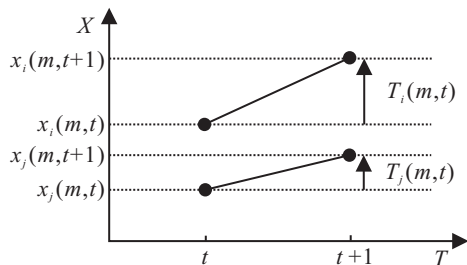
$$x_n(m,t) = x_n^*(m,t)d_2 = \ln(x_n^*(m,t)), \quad (2)$$

称 D_2 为面板数据对数化算子, $m = 1, 2, \dots, M, t = 1, 2, \dots, T$.

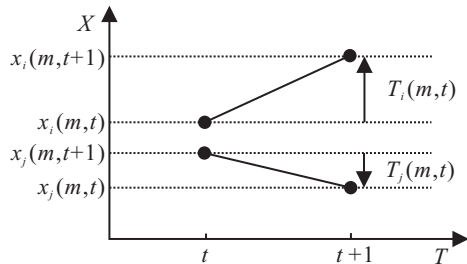
1.2 面板数据时间维度和对象维度的特征表征

1.2.1 时间维度的特征表征

在时间维度, 第 m 个对象 ($m = 1, 2, \dots, M$) 的第 i 个指标和第 j 个指标 ($i, j = 1, 2, \dots, N$ 且 $i \neq j$) 在 $[t, t + 1]$ 时段 ($t = 1, 2, \dots, T$) 的发展方向可能出现图2(a)和图2(b)两种情形.



(a) 同向



(b) 反向

图2 面板数据时间维度的增量

定义3 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 $X_j (i, j = 1, 2, \dots, N$ 且 $i \neq j)$, 称

$$T_i(m,t) = x_i(m,t+1) - x_i(m,t), \quad (3)$$

$$T_j(m,t) = x_j(m,t+1) - x_j(m,t) \quad (4)$$

分别为第 m 个对象 ($m = 1, 2, \dots, M$) 的第 i 个指标和第 j 个指标在 $[t, t + 1]$ 时段 ($t = 1, 2, \dots, T - 1$) 的增量.

定义4 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 $X_j (i, j = 1, 2, \dots, N$ 且 $i \neq j)$, 称

$$\gamma_{ij}^T(m,t) = \|T_i(m,t) - |T_j(m,t)\| \quad (5)$$

为第 m 个对象 ($m = 1, 2, \dots, M$) 的第 i 个指标和第 j 个指标在 $[t, t + 1]$ 时段 ($t = 1, 2, \dots, T - 1$) 的增量差. 称

$$\text{sgn}(T_{ij}(m,t)) = \begin{cases} 1, & T_i(m,t)T_j(m,t) > 0 \text{ or} \\ & T_i(m,t)T_j(m,t) = 0, T_i(m,t) + T_j(m,t) \geq 0; \\ -1, & T_i(m,t)T_j(m,t) < 0 \text{ or} \\ & T_i(m,t)T_j(m,t) = 0, T_i(m,t) + T_j(m,t) < 0 \end{cases} \quad (6)$$

为第 m 个对象 ($m = 1, 2, \dots, M$) 的第 i 个指标和第 j 个指标在 $[t, t + 1]$ 时段 ($t = 1, 2, \dots, T - 1$) 发展方向的差异. 其中: “1” 表示 $T_i(m,t)$ 与 $T_j(m,t)$ 发展方向相同, “-1” 表示 $T_i(m,t)$ 与 $T_j(m,t)$ 发展方向相反.

1.2.2 对象维度的特征表征

在对象维度, 为避免因对象顺序的变化对关联度造成影响, 采用基于点到点的差量进行刻画.

定义5 设第 n 个指标的面板数据为 $X_n (n = 1, 2, \dots, N)$, 称

$$\bar{x}_n(t) = \frac{1}{M} \sum_{m=1}^M x_n(m,t) \quad (7)$$

为第 n 个指标在第 t 个时刻 ($t = 1, 2, \dots, T$) 的所有对象平均水平.

定义6 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 $X_j (i, j = 1, 2, \dots, N$ 且 $i \neq j)$, 称

$$M_i(m,t) = x_i(m,t) - \bar{x}_i(t), \quad (8)$$

$$M_j(m,t) = x_j(m,t) - \bar{x}_j(t) \quad (9)$$

分别为第 t 时刻 ($t = 1, 2, \dots, T$) 的第 i 个指标和第 j 个指标关于对象 $m (m = 1, 2, \dots, M)$ 的离差.

如图3所示, 用 $M_i(m,t)$ 和 $M_j(m,t) (i, j = 1, 2, \dots, N$ 且 $i \neq j)$ 的数值大小和正负符号反映对象 $m (m = 1, 2, \dots, M)$ 对序列均值水平的“拉升”和“打压”作用.

定义7 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 $X_j (i, j = 1, 2, \dots, N$ 且 $i \neq j)$, 称

$$\gamma_{ij}^M(m,t) = \|M_i(m,t) - |M_j(m,t)\| \quad (10)$$

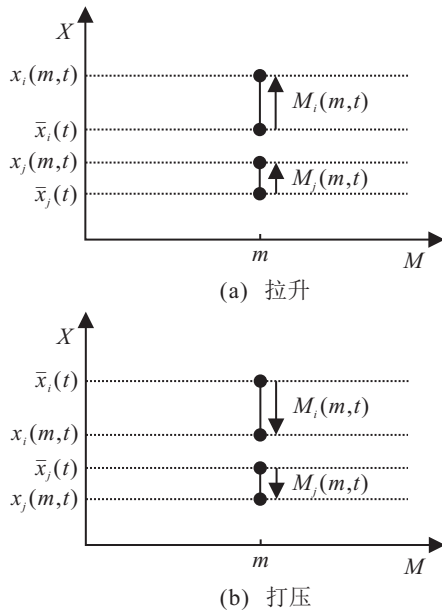


图3 面板数据对象维度的离差

为第 t 时刻 ($t = 1, 2, \dots, T$) 的第 i 个指标和第 j 个指标关于对象 m 的离差的差 ($m = 1, 2, \dots, M$). 称

$$\text{sgn}(M_{ij}(m, t)) = \begin{cases} 1, & M_i(m, t)M_j(m, t) > 0 \text{ or} \\ & M_i(m, t)M_j(m, t) = 0, \\ & M_i(m, t) + M_j(m, t) \geq 0; \\ -1, & M_i(m, t)M_j(m, t) < 0 \text{ or} \\ & M_i(m, t)M_j(m, t) = 0, \\ & M_i(m, t) + M_j(m, t) < 0 \end{cases} \quad (11)$$

为第 t 时刻 ($t = 1, 2, \dots, T$) 的第 i 个指标和第 j 个指标关于对象 m ($m = 1, 2, \dots, M$) 作用方向的差异.

1.3 基于指数函数的面板数据灰色关联度

通过对面板数据时间维度和对象维度的特征分析,提取两维度的发展程度和方向信息,将指标之间“增量差”和“离差的差”的绝对值转化为度量曲线形状的相似程度,并将其方向差异作为正负关联的判断依据. 本文在刘震等^[15]接近性模型思想的启发下,以下降速度较快的指数函数为基础并进行扩展.

定义8 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 X_j ($i, j = 1, 2, \dots, N$ 且 $i \neq j$), 称

$$\Upsilon_{ij}^T(m, t) = \text{sgn}(T_{ij}(m, t))e^{-\gamma_{ij}^T(m, t)} \quad (12)$$

为第 m 个对象 ($m = 1, 2, \dots, M$) 的第 i 个指标和第 j 个指标在 $[t, t+1]$ 时段 ($t = 1, 2, \dots, T-1$) 的关联系数. 同理有

$$\Upsilon_{ij}^M(m, t) = \text{sgn}(M_{ij}(m, t))e^{-\gamma_{ij}^M(m, t)} \quad (13)$$

为第 t 时刻 ($t = 1, 2, \dots, T$) 的第 i 个指标和第 j 个指标关于对象 m 的关联系数.

定义9 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 X_j ($i, j = 1, 2, \dots, N$ 且 $i \neq j$), 称

$$\Upsilon_{ij}^T = \frac{\sum_{m=1}^M \sum_{t=1}^{T-1} \Upsilon_{ij}^T(m, t)}{M(T-1)}, \quad (14)$$

$$\Upsilon_{ij}^M = \frac{\sum_{m=1}^M \sum_{t=1}^T \Upsilon_{ij}^M(m, t)}{MT} \quad (15)$$

分别为第 i 个指标面板数据和第 j 个指标面板数据在时间维度和对象维度的关联度.

Υ_{ij}^T 和 Υ_{ij}^M 的绝对值越大表示关联程度越大, 反之, 绝对值越小表示关联程度越小. Υ_{ij}^T 和 Υ_{ij}^M 的正负符号表示关联方向, 正号表示存在同向关联, 负号表示存在反向关联.

定义10 设第 i 个指标和第 j 个指标的面板数据为 X_i 和 X_j ($i, j = 1, 2, \dots, N$ 且 $i \neq j$), 称

$$\Upsilon_{ij} = \omega_1 \Upsilon_{ij}^T + \omega_2 \Upsilon_{ij}^M \quad (16)$$

为第 i 个指标面板数据和第 j 个指标面板数据的综合关联度, 其中 ω_1 和 ω_2 分别为面板数据 X_i 与 X_j 在时间维度和对象维度关联度的权重, $\omega_1, \omega_2 \in [0, 1]$ 且 $\omega_1 + \omega_2 = 1$. 在实际应用中, 按重要性程度对时间维度和对象维度赋予相应的权重.

1.4 面板数据灰色指标关联模型的性质

性质1 若第 i 个指标和第 j 个指标的面板数据为 X_i 和 X_j ($i, j = 1, 2, \dots, N$ 且 $i \neq j$), 则关联系数 $\Upsilon_{ij}^T(m, t) \in [-1, 0) \cup (0, 1], \Upsilon_{ij}^M(m, t) \in [-1, 0) \cup (0, 1]$.

证明 在时间维度, 根据定义 (8), $\Upsilon_{ij}^T(m, t) = \text{sgn}(T_{ij}(m, t))e^{-\gamma_{ij}^T(m, t)}$. 因为 $\gamma_{ij}^T(m, t) \in [0, \infty)$, 当 $\text{sgn}(T_{ij}(m, t)) = 1$ 时, 有 $\Upsilon_{ij}^T(m, t) \in (0, 1]$; 当 $\text{sgn}(T_{ij}(m, t)) = -1$ 时, 有 $\Upsilon_{ij}^T(m, t) \in [-1, 0)$. 对象维度 $\Upsilon_{ij}^M(m, t)$ 同理可证. \square

性质2 若第 i 个指标和第 j 个指标的面板数据为 X_i 与 X_j ($i, j = 1, 2, \dots, N$ 且 $i \neq j$), 则关联度 $\Upsilon_{ij}^T \in [-1, 1], \Upsilon_{ij}^M \in [-1, 1]$, 且 $\Upsilon_{ij} = 1 \iff X_i = X_j, \Upsilon_{ij} = -1 \iff X_i = -X_j$.

证明 在时间维度, 根据性质 1 $\Upsilon_{ij}^T(m, t) \in [-1, 0) \cup (0, 1]$, 又由 $\Upsilon_{ij}^T = \frac{\sum_{m=1}^M \sum_{t=1}^{T-1} \Upsilon_{ij}^T(m, t)}{M(T-1)}, \Upsilon_{ij}^T \in [-1, 1]$ 易证, 对象维度 $\Upsilon_{ij}^M \in [-1, 1]$ 同理可证. 根据 $\Upsilon_{ij} = \omega_1 \Upsilon_{ij}^T + \omega_2 \Upsilon_{ij}^M, \omega_1, \omega_2 \in [0, 1]$ 且 $\omega_1 + \omega_2 = 1$, 那么 $\Upsilon_{ij} = \omega_1(\Upsilon_{ij}^T - \Upsilon_{ij}^M) + \Upsilon_{ij}^M$, 若 $\Upsilon_{ij}^T > \Upsilon_{ij}^M$, 则 $\omega_1 = 1$ 时 Υ_{ij} 取得最大值, $\omega_1 = 0$ 时 Υ_{ij} 取得最小值. 又因为 $\Upsilon_{ij}^T, \Upsilon_{ij}^M \in [-1, 1]$, 所以 $\Upsilon_{ij} \in [-1, 1]$. 当 $\Upsilon_{ij}^T < \Upsilon_{ij}^M$ 时

同理可证. 当 $X_i = X_j$ 时, 显然时间维度关联度 $\Upsilon_{ij}^T = 1$, 对象维度关联度 $\Upsilon_{ij}^M = 1$, 在不考虑维度权重 ω_1 与 ω_2 差别的情况下, $\Upsilon_{ij} = 1$; 同理, 当 $X_i = -X_j$ 时, 有 $\Upsilon_{ij}^T(m, t) = -1, \Upsilon_{ij}^M(m, t) = -1$, 因此 $\Upsilon_{ij} = -1$. \square

性质3 在时间维度, 第 i 个指标和第 j 个指标的增量差与关联系数之间满足

$$\lim_{\gamma_{ij}^T(m,t) \rightarrow 0} \Upsilon_{ij}^T = \begin{cases} 1, & \text{sgn}(T_{ij}(m, t)) = 1; \\ -1, & \text{sgn}(T_{ij}(m, t)) = -1. \end{cases}$$

在对象维度, 第 i 个指标和第 j 个指标的离差的差与关联系数之间满足

$$\lim_{\gamma_{ij}^M(m,t) \rightarrow 0} \Upsilon_{ij}^M = \begin{cases} 1, & \text{sgn}(M_{ij}(m, t)) = 1; \\ -1, & \text{sgn}(M_{ij}(m, t)) = -1. \end{cases}$$

证明 在时间维度, 由式 (6), 当 $T_i(m, t)$ 与 $T_j(m, t)$ 发展方向相同时, $\text{sgn}(T_{ij}(m, t)) = 1$, 由式 (8), $\Upsilon_{ij}^T(m, t) = \text{sgn}(T_{ij}(m, t)a^{-\gamma_{ij}^T(m,t)})$, 若 $\gamma_{ij}^T(m, t) \rightarrow 0$, 那么 $a^{-\gamma_{ij}^T(m,t)} \rightarrow 1$, 则 $\Upsilon_{ij}^T \rightarrow 1$; 当 $T_i(m, t)$ 与 $T_j(m, t)$ 发展方向相反时, $\text{sgn}(T_{ij}(m, t)) = -1, \Upsilon_{ij}^T \rightarrow -1$. 证明成立. \square

性质4(数乘变换保序性) 对于面板数据 X_i, X_j 和 $X_k(i, j, k = 1, 2, \dots, N$ 且 $i \neq j \neq k)$, 若 $\forall x_k^*(m, t) = ax_j^*, a = \text{const}$, 则 $\Upsilon_{ij} = \Upsilon_{ik}$.

证明 根据面板数据预处理算子的作用, 由 $y = \ln x$ 的性质, 在时间维度 $[t, t + 1]$ 时段 ($t = 1, 2, \dots, T - 1$), 其增量满足

$$\begin{aligned} \ln x_k^*(m, t + 1) - \ln x_k^*(m, t) &= \\ \ln ax_j^*(m, t + 1) - \ln ax_j^*(m, t) &= \\ (\ln a + \ln x_j^*(m, t + 1)) - (\ln a + \ln x_j^*(m, t)) &= \\ \ln x_j^*(m, t + 1) - \ln x_j^*(m, t). \end{aligned}$$

在对象维度关于第 m 个对象 ($m = 1, 2, \dots, M$), 其离差满足

$$\begin{aligned} \ln x_k^*(m, t) - \ln \bar{x}_k^*(t) &= \\ \ln ax_j^*(m, t) - \ln a\bar{x}_j^*(t) &= \\ (\ln a + \ln x_j^*(m, t)) - (\ln a + \ln \bar{x}_j^*(t)) &= \\ \ln x_j^*(m, t) - \ln \bar{x}_j^*(t). \end{aligned}$$

因此 $\Upsilon_{ij} = \Upsilon_{ik}$, 均值化算子同理可证. \square

注意到对数化算子因对数函数本身的性质, 对原始数据有以下两点要求: 1) 序列中观测值均为正值; 2) 观测值一般是绝对量, 而不是相对量.

性质5 灰色指标关联模型满足唯一性、对称性和可比性.

证明 由于给定面板数据的特征值选取和计算

步骤唯一, 且相对特征的大小和符号处理方式唯一, 结果具有唯一性. 由于模型选取两个指标的增量差和离差的差刻画序列关系, 指标的“相对”差异使模型满足对称性. 因为模型满足唯一性, 根据性质1和性质2, 模型正负关联度对称且采用关联度符号和数值分离的方式进行关联排序, 所以模型满足可比性. \square

2 实例分析

2.1 样本数据来源

近年来, 我国高频次、大范围的大气环境污染日益严重, 已经成为制约我国区域经济社会发展的重要因素, 研究区域大气污染影响因素成为当今面临的极为紧迫和严峻的现实问题. Wang等^[18]指出了研究江苏省大气污染及其相关因素的必要性和迫切性. 2012年中国颁布的《环境空气质量标准》(GB3095-2012)是根据国家经济社会发展状况和环境保护要求进行的第三次修订, 此次修订调整和设置了包括细颗粒物(PM_{2.5})、可吸入颗粒物(PM₁₀)、二氧化硫(SO₂)、二氧化氮(NO₂)、一氧化碳(CO)和臭氧(O₃)共6种环境空气污染物基本项目, 并对前4种污染物“年平均”浓度数据的有效性进行了一致规定. 本文以最新《环境空气质量标准》为指导, 选取

表1 苏南五市2013年~2016年空气质量达到国家二级标准天数与主要影响因素观测值

指标	城市	2013	2014	2015	2016
DAQ/ ($\mu\text{g}/\text{m}^3$)	南京市	202	190	235	242
	苏州市	217	232	244	252
	无锡市	199	208	234	244
	常州市	214	231	258	270
	镇江市	223	241	263	268
PM _{2.5} / ($\mu\text{g}/\text{m}^3$)	南京市	77.00	73.80	57.00	47.90
	苏州市	69.00	66.00	58.00	46.00
	无锡市	75.00	68.00	61.00	53.00
	常州市	72.00	67.00	58.00	49.00
	镇江市	72.00	68.00	59.00	50.00
PM ₁₀ / ($\mu\text{g}/\text{m}^3$)	南京市	137.00	123.00	96.00	85.20
	苏州市	95.00	86.00	80.00	72.00
	无锡市	112.00	105.00	94.00	82.00
	常州市	102.00	104.00	92.50	81.00
	镇江市	124.00	107.00	82.00	80.00
SO ₂ / ($\mu\text{g}/\text{m}^3$)	南京市	37.00	25.00	19.00	18.20
	苏州市	31.00	24.00	21.00	17.00
	无锡市	45.00	29.00	26.00	18.00
	常州市	41.00	36.00	27.00	19.00
	镇江市	30.00	24.00	25.00	24.00
NO ₂ / ($\mu\text{g}/\text{m}^3$)	南京市	55.00	54.00	50.00	44.30
	苏州市	53.00	53.00	54.00	51.00
	无锡市	47.00	45.00	41.00	47.00
	常州市	48.00	40.00	40.00	37.00
	镇江市	42.00	46.00	42.00	38.00

注: 常州市2015年PM₁₀、SO₂和NO₂根据环境状况公报推算得出

PM_{2.5}、PM₁₀、SO₂、NO₂的年平均浓度4个因素指标,对苏南五市(南京、苏州、无锡、常州、镇江)空气质量达到国家二级标准天数(以下简称DAQ)及其主要影响因素进行分析,样本区间为2013年~2016年,数据来源于各城市《环境状况公报》,详见表1。

表1中,数据具有典型的“少数据、贫信息”特征,主要体现在:1)新标准实施时间短,数据总量小;2)常州市2015年存在数据缺失,序列不完整。为保证建模分析,根据常州市2014年~2016年《环境状况公报》内容,对缺失的数据进行推算或运用灰色均值生成算

子计算填充。

2.2 模型计算及分析

为了分析苏南五市空气质量与主要污染物年均浓度之间的关系,对系统中选取的5组面板数据,利用本文提出的面板数据灰色指标关联模型分别计算DAQ与PM_{2.5}、PM₁₀、SO₂和NO₂在时间维度和对象维度的关联系数以及综合关联系数,从而得出综合关联度。为方便计算,采用本文提出的对数化算子进行数据预处理,式(16)中 ω_1 与 ω_2 各取0.5,计算结果见表2。

表2 苏南五市空气质量达到国家二级标准天数与主要因素的关联系数及关联度

指标	时间维度					γ_{ij}^T	截面维度				γ_{ij}^M	综合关联度
	南京市	苏州市	无锡市	常州市	镇江市		2013	2014	2015	2016		
DAQ-PM _{2.5}	-0.2798	-0.9071	-0.9482	-0.9488	-0.9301	-0.8028	-0.9777	-0.5700	0.1940	0.1948	-0.2897	-0.5463
DAQ-PM ₁₀	-0.3083	-0.9586	-0.9608	-0.3217	-0.9210	-0.6941	-0.1451	-0.5647	-0.2081	0.1701	-0.1870	-0.4405
DAQ-SO ₂	-0.4027	-0.8613	-0.7956	-0.8406	-0.2961	-0.6393	-0.5364	-0.2299	0.5432	0.5538	0.0827	-0.2783
DAQ-NO ₂	-0.2759	0.3096	-0.3551	-0.3241	-0.3105	-0.1912	-0.1816	-0.1750	-0.5507	-0.9248	-0.4580	-0.3246

由表2可见,时间维度和对象维度的各关联系数和关联度基本为负值,反映了PM_{2.5}、PM₁₀、SO₂和NO₂四个主要因素与空气质量达到国家二级标准天数呈反向关联关系,与预期结果一致。

1) 由综合关联度可知,苏南五市这4年与空气质量达到国家二级标准天数的主要影响因素关联序是PM_{2.5} > PM₁₀ > NO₂ > SO₂,且均呈负向关联关系。关联度最高的是DAQ和PM_{2.5},反映了PM_{2.5}对DAQ的影响最大,且随着PM_{2.5}的增加DAQ减小;关联度最低的是DAQ和SO₂,反映了SO₂对DAQ的影响最小。事实上,各地统计结果显示,苏南五市在这4年中所选指标对DAQ影响最严重的是PM_{2.5},虽然各城市的PM_{2.5}年均浓度逐年依次减少,但该组面板数据中20个观测值无一达到国家二级标准,从而验证了模型的有效性。

2) 从时间维度来看,与空气质量达到国家二级标准天数综合关联系数最高的是PM_{2.5},其次是PM₁₀、SO₂、NO₂,其中各对象的NO₂与DAQ关联系数除苏州外均为负值。如图4所示,2013年~2016年苏州NO₂的年均浓度变化幅度较小,观测值呈“平-增-降”的动态走势,而DAQ呈稳定上升趋势,NO₂的上下波动直接影响关联系数正负的变化。NO₂浓度在高位波动致使苏州的NO₂与DAQ出现正向关联,从而使该维度二者的关联程度较其他影响因素偏低。究其原因,NO₂主要来源于燃煤烟气和机动车尾

气排放,特别是近几年来苏州市机动车保有量迅速增加,形成了NO₂污染的巨大压力。

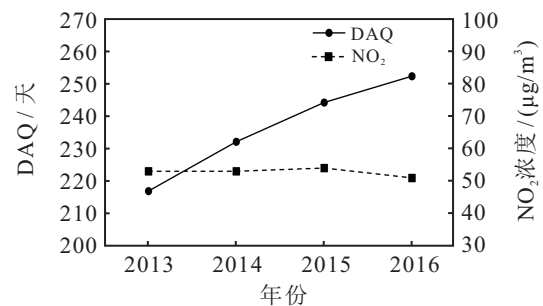


图4 苏州市年份空气质量达到二级标准天数与二氧化氮年均浓度

3) 从对象维度来看,与空气质量达到国家二级标准天数综合关联系数数值最高的是NO₂,其次是PM_{2.5}、PM₁₀,呈负向关联,SO₂的综合关联系数出现了正号但数值极小,反映了SO₂浓度对DAQ极其微弱的影响。近些年苏南城市的工厂规模外迁,硫化物排放减少,二氧化硫年均浓度逐年达标,但空气质量受PM_{2.5}、PM₁₀等多种因素的综合影响,DAQ高的主要原因是其他影响因素浓度的降低。此外,社会经济系统中多数指标都存在包含时间因素的特性,如政府某年以某空气污染物下降百分比为目标进行治理与控制,在剔除了时间因素的作用后,单纯地考虑了对象间的差异,在一定程度上抵消了时间因素对关联度的影响。两维度关联度的综合使指标间的综合关联关系趋于均衡,使模型更具有稳定性和全面性。

2.3 模型检验及对比

为检验模型关联度排序的稳定性,分别交换表1中南京市与苏州市、无锡市与常州市数据的顺序,采用相同的数据预处理方式,将本文求解结果和与本文相似的部分文献的求解结果进行对比,结果见表3.从模型改进方面来看,部分文献对时间维度和对象维度简单一致的处理方式会对关联结果造成影响,验证了构建面板数据维度差异化处理模型的必要性

和有效性,而采用其他方式回避该问题的现有文献并未给出有关负关联的表示方式和思路.从模型结果方面来看,本文模型的结果更符合现实情况,与实际统计结果有高度的匹配性和一致性,有较高的参考价值,而部分文献因对象顺序的改变导致关联大小无法判断.综合而言,本文所构建的模型在一定程度上解决了关联度稳定和正负表示的问题,具有较好的应用效果.

表3 不同模型有关苏南五市空气质量达到国家二级标准天数与主要因素关联度的对比

指标	文献[11]		文献[13]		文献[14]		文献[16]		本文	
	原顺序	交换顺序	原顺序	交换顺序	原顺序	交换顺序	原顺序	交换顺序	原顺序	交换顺序
DAQ-PM _{2.5}	0.8820	0.8820	0.8795	0.8811	-0.6472	-0.6183	-0.4284	-0.4027	-0.5463	-0.5463
DAQ-PM ₁₀	0.9134	0.9134	0.9178	0.9219	-0.6341	-0.6971	-0.4231	-0.5087	-0.4405	-0.4405
DAQ-SO ₂	0.8230	0.8230	0.8001	0.8011	-0.5535	-0.2768	-0.4101	-0.0914	-0.2783	-0.2783
DAQ-NO ₂	0.8721	0.8721	0.8522	0.8527	-0.0379	-0.2737	-0.0226	-0.2259	-0.3246	-0.3246
关联正负表示与否	否		否		是		是		是	
关联大小稳定与否	是		否		否		否		是	

3 结论

本文在灰色关联模型现有研究的启发下,以维度差异为出发点,构建了时间维度基于增量的关联模型以及对象维度基于离差的关联模型,解决了面板数据因对象顺序不同影响关联序的问题,进一步完善和处理了正负关联度的表示、对称关系和排序方法,提高了关联模型的分辨效果.然后从理论方面证明了该模型具有偶对对称性、可比性、唯一性和数乘变换保序性等性质.最后运用该模型对2013年~2016年苏南五市空气质量的影响因素进行了分析,通过实例与模型对比检验了模型的效果和应用价值.

尽管灰色指标关联模型克服了其他方法的许多缺点,但仍有许多限制需要改进.首先,正负关联系数加和求平均会使关联度的灵敏度较高,尤其是对于序列波动水平不大的情况;其次,权重的获取是模糊的,结合实际背景、专家分析和定量优化方法获取的权重应该更加合理;第三,模型仅适用于等间距且等长度序列,如何构造特定应用环境下变权、非等间距、鲁棒性高的关联分析模型是未来值得进一步研究的问题.

参考文献(References)

[1] 邓聚龙. 灰理论基础[M]. 武汉: 华中科技大学出版社, 2002: 63-170.
(Deng J L. The elements of grey system[M]. Wuhan: Press of Huazhong University of Science and Technology, 2002: 63-170.)

[2] Liu S F, Fang Z G, Lin Y. Study on a new definition of degree of grey incidence[J]. J of Grey System, 2006, 9(2): 115-122.
[3] 刘思峰, 谢乃明, Forrest Jeffrey. 基于相似性和接近性视角的新型灰色关联分析模型[J]. 系统工程理论与实践, 2010, 30(5): 881-887.
(Liu S F, Xie N M, Jeffrey Forrest. On new models of grey incidence analysis based on visual angel of similarity and nearness[J]. Systems Engineering — Theory & Practice, 2010, 30(5): 881-887.)
[4] 党耀国, 刘思峰, 刘斌, 等. 灰色斜率关联度的改进[J]. 中国工程科学, 2004, 6(3): 41-44.
(Dang Y G, Liu S F, Liu B, et al. Improvement on degree of grey slope incidence[J]. Engineering Science, 2004, 6(3): 41-44.)
[5] 陈勇明, 张明. 灰色样条绝对关联度模型[J]. 系统工程理论与实践, 2015, 35(5): 1304-1310.
(Chen Y M, Zhang M. Cubic spline based grey absolute relational grade model[J]. Systems Engineering — Theory & Practice, 2015, 35(5): 1304-1310.)
[6] Xie N M, Liu S F. Research on the order-keeping property of several grey relational models[J]. J of Grey System, 2008, 11(3): 157-164.
[7] 崔杰, 党耀国, 刘思峰. 几类关联分析模型的新性质[J]. 系统工程, 2009, 27(4): 65-70.
(Cui J, Dang Y G, Liu S F. Novel properties of some grey relational analysis models[J]. Systems Engineering, 2009, 27(4): 65-70.)
[8] 魏勇, 曾柯方. 关联度公理的简化与特殊关联度的公理化定义[J]. 系统工程理论与实践, 2015, 35(6):

- 1528-1534.
(Wei Y, Zeng K F. The simplified relational axioms and the axiomatic definition of special incidence degrees[J]. Systems Engineering — Theory & Practice, 2015, 35(6): 1528-1534.)
- [9] 刘思峰, 蔡华, 杨英杰, 等. 灰色关联分析模型研究进展[J]. 系统工程理论与实践, 2013, 33(8): 2041-2046.
(Liu S F, Cai H, Yang Y J, et al. Advance in grey incidence analysis modelling[J]. Systems Engineering — Theory & Practice, 2013, 33(8): 2041-2046.)
- [10] 钱吴永, 王育红, 党耀国, 等. 基于多指标面板数据的灰色矩阵关联模型及其应用[J]. 系统工程, 2013, 31(10): 70-74.
(Qian W Y, Wang Y H, Dang Y G. Grey matrix relational modeling and its application based on multivariate panel data[J]. Systems Engineering, 2013, 31(10): 70-74.)
- [11] 崔立志, 刘思峰. 面板数据的灰色矩阵相似关联模型及其应用[J]. 中国管理科学, 2015, 23(11): 171-176.
(Cui L Z, Liu S F. Grey matrix similar incidence model for panel data and its application[J]. J of Management Science in China, 2015, 23(11): 171-176.)
- [12] Li X M, Keith W H, Dang Y G. An improved grey relational analysis approach for panel data clustering[J]. Expert Systems with Applications, 2015, 42(23): 9105-9116.
- [13] 张可, 刘思峰. 灰色关联聚类在面板数据中的扩展及应用[J]. 系统工程理论与实践, 2010, 30(7): 1253-1259.
(Zhang K, Liu S F. Extended clusters of grey incidence for panel data and its application[J]. Systems Engineering — Theory & Practice, 2010, 30(7): 1253-1259.)
- [14] 刘震, 党耀国, 钱吴永, 等. 基于面板数据的灰色网格关联度模型[J]. 系统工程理论与实践, 2014, 34(4): 991-996.
(Liu Z, Dang Y G, Qian W Y, et al. Grey grid incidence model based on panel data[J]. Systems Engineering — Theory & Practice, 2014, 34(4): 991-996.)
- [15] 刘震, 党耀国, 周伟杰, 等. 新型灰色接近关联模型及其拓展[J]. 控制与决策, 2014, 29(6): 1071-1075.
(Liu Z, Dang Y G, Zhou W J, et al. New grey nearness incidence model and its extension[J]. Control and Decision, 2014, 29(6): 1071-1075.)
- [16] 党耀国, 朱晓月, 丁松, 等. 基于灰关联度的面板数据聚类方法及在空气污染分析中的应用[J]. 控制与决策, 2017, 32(12): 2227-2232.
(Dang Y G, Zhu X Y, Ding S, et al. Panel data clustering method based on grey correlation degree and its application in airpollution analysis[J]. Control and Decision, 2017, 32(12): 2227-2232.)
- [17] 吴鸿华, 穆勇, 屈忠锋, 等. 基于面板数据的接近性和相似性关联度模型[J]. 控制与决策, 2016, 31(3): 555-558.
(Wu H H, Mu Y, Qu Z F et al. Similarity and nearness relational degree based on panel data[J]. Control and Decision, 2016, 31(3): 555-558.)
- [18] Wang J J, Keith W H, Dang Y G. An improved grey dynamic trend incidence model with application to factors causing smog weather[J]. Expert Systems with Applications, 2017, 87(11): 240-251.
- [19] 白仲林. 面板数据的计量经济分析[M]. 天津: 南开大学出版社, 2008: 6-10.
(Bai Z L. Econometric analysis of panel data[M]. Tianjin: Nankai University Press, 2008: 6-10.)
- [20] 李宏艳. 关于灰色关联度计算方法的研究[J]. 系统工程与电子技术, 2004, 26(9): 1231-1233.
(Li H Y. Study on the calculation method of grey relationship degree[J]. Systems Engineering and Electronics, 2004, 26(9): 1231-1233.)
- [21] 吴利丰, 闫书丽, 马利华, 等. 面向横截面数据的灰色关联度[J]. 数学的实践与认识, 2016, 46(9): 179-184.
(Wu L F, Yan S L, Ma L H, et al. Grey distance degree for cross-sectional data[J]. Mathematics in Practice and Theory, 2016, 46(9): 179-184.)

(责任编辑: 郑晓蕾)