Statistics and Application 统计学与应用, 2016, 5(4), 397-403
Published Online December 2016 in Hans. http://www.hanspub.org/journal/sa
http://dx.doi.org/10.12677/sa.2016.54043

Exploration on Spatial Correlation of Two Space Variables

Qiongyu Chang, Jianjun Dou*

Longqiao College of Lanzhou University of Finance and Economics, Lanzhou Gansu
Email: *726877004@qq.com

Received: Dec. 9th, 2016; accepted: Dec. 23rd, 2016; published: Dec. 29th, 2016

Abstract

This paper studies the spatial correlation of two space variables and defines the spatial correlation coefficient of them. By using Monte Carlo random simulation method, it is obtained that when there is a linear relationship between the two space variables, the new coefficient measures the space structure similarity of them.

Keywords

Space Variable, Spatial Autocorrelation, Spatial Correlation, Data Generation Process

两个空间变量空间相关性的分析

常琼玉，窦剑军*

兰州财经大学陇桥学院，甘肃 兰州
Email: *726877004@qq.com

收稿日期：2016年12月9日；录用日期：2016年12月23日；发布日期：2016年12月29日

摘　要

本文研究了两个空间变量空间相关性的问题，定义了一个测度两个空间变量空间相关性的系数。通过

*通讯作者。

文章引用：常琼玉，窦剑军. 两个空间变量空间相关性的分析[J]. 统计学与应用, 2016, 5(4): 397-403.
http://dx.doi.org/10.12677/sa.2016.54043

Monte Carlo模拟发现，当两个空间变量存在线性关系时，这个新的系数测度了两个空间变量空间结构的相似性。

关键词

空间变量，空间自相关，空间相关，数据生成过程

1. 引言

两个空间变量的空间相关关系一直是空间统计学的热门研究课题。特别是在空间计量经济学和气候研究中，例如，经济增长和能源消费之间的关系，经济增长与污染物排放之间的关系，一个地区降雨量与气温之间的关系。

有关空间变量的空间相关性分析，国内外学者都是通过 Moran’s I 检验出空间的自相关性，然后建立空间计量模型，通过估计的模型的参数检验空间的依赖性。Badi H. Baltagi、Seuck Heun Song、Byoung Cheol Jung 和 Won Koh (2003) [1]建立了面板数据回归模型，检验了序列的相关性，空间自回归和随机效应；Pesaran、Mohammad Hashem (2004) [2]建立了面板数据回归模型诊断检验横截面数据的依赖性；张晓旭和冯宗宪(2008) [3]建立了空间误差自回归模型和广义空间模型分析中国人均 GDP 的空间相关与地区收敛，得出正的空间自相关以及空间异质性的存在，而且空间上的相互作用和地理位置对于人均收入增长的作用随时间不断加强；Qian Li、Jinping Song、Enru Wang、Hao Hu、Jianhui Zhang、Yeyao Wang (2014) [4]基于中国县一级的数据研究了经济增长与污染物排放的关系，得出中国近几年工业化的进程以及重工业的转移。

但是这些研究都是基于空间具体的模型上对两个空间变量进行回归分析，建立回归方程然后在进行假设检验的统计推断。然而到目前能够刻画两个空间变量空间相关关系的定义还没有，鉴于此本文定义了一个与经典统计学测度两个随机变量相关性的 Pearson 相关系数类同的统计量，测度两个空间变量的空间相关关系，命名为新相关系数。用这个新定义相关系数来测度两个空间变量的空间关系，通过模拟和实证分析来探讨这个新的定义究竟刻画了两个空间变量怎样的空间关系。全文安排如下，第二部分研究新相关系数的定义和它的一些性质，第三部分通过 Monte Carlo 模拟得到两个具有空间自相关的空间变量之间的相关程度，第四部分为实证分析，第五部分为结论和进一步的讨论。

2. 新相关系数及其性质

Pearson’s 相关系数是测度两个随机变量之间相关性的统计量，它的定义为

$$\rho=\frac{E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}$$

类似的定义两个空间变量的相关系数为

$$\rho=\frac{E\left[(X-WX)^{\mathrm{T}}(Y-WY)\right]}{\sqrt{E\left[(X-WX)^{\mathrm{T}}(X-WX)\right]}\sqrt{E\left[(Y-WY)^{\mathrm{T}}(Y-WY)\right]}}$$

其中 W 为空间权重矩阵。新相关系数的估计值为

$$\hat{\rho}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x}_i)(y_i-\tilde{y}_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x}_i)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2}}$$

这里 $\tilde{x}_i$ 表示 x_i 的空间滞后，$\tilde{y}_i$ 表示 y_i 的空间滞后，其推导过程如下：
$X=(x_1,x_2,\cdots x_n)$ 为空间变量 X 的一个样本，$Y=(y_1,y_2,\cdots y_n)$ 为空间变量 Y 的一个样本

$$X-WX=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}-\begin{pmatrix}w_{11}&w_{12}\cdots&w_{1n}\\w_{21}&w_{22}\cdots&w_{2n}\\\vdots&\vdots&\vdots\\w_{n1}&w_{n2}\cdots&w_{nn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}-\begin{pmatrix}\tilde{x}_1\\\tilde{x}_2\\\vdots\\\tilde{x}_n\end{pmatrix}=\begin{pmatrix}x_1-\tilde{x}_1\\x_2-\tilde{x}_2\\\vdots\\x_n-\tilde{x}_n\end{pmatrix}$$

$$Y-WY=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}-\begin{pmatrix}w_{11}&w_{12}\cdots&w_{1n}\\w_{21}&w_{22}\cdots&w_{2n}\\\vdots&\vdots&\vdots\\w_{n1}&w_{n2}\cdots&w_{nn}\end{pmatrix}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}-\begin{pmatrix}\tilde{y}_1\\\tilde{y}_2\\\vdots\\\tilde{y}_n\end{pmatrix}=\begin{pmatrix}y_1-\tilde{y}_1\\y_2-\tilde{y}_2\\\vdots\\y_n-\tilde{y}_n\end{pmatrix}$$

所以新相关系数分子的估计值为

$$\frac{1}{n}(X-WX)^{\mathrm{T}}(Y-WY)=\frac{1}{n}(x_1-\tilde{x}_1,x_2-\tilde{x}_2,\cdots x_n-\tilde{x}_n)\begin{pmatrix}y_1-\tilde{y}_1\\y_2-\tilde{y}_2\\\vdots\\y_n-\tilde{y}_n\end{pmatrix}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x}_i)(y_i-\tilde{y}_i)$$

分母的估计值为

$$\frac{1}{n}(X-WX)^{\mathrm{T}}(X-WX)=\frac{1}{n}(x_1-\tilde{x}_1,x_2-\tilde{x}_2,\cdots x_n-\tilde{x}_n)\begin{pmatrix}x_1-\tilde{x}_1\\x_2-\tilde{x}_2\\\vdots\\x_n-\tilde{x}_n\end{pmatrix}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x}_i)^2$$

$$\frac{1}{n}(Y-WY)^{\mathrm{T}}(Y-WY)=\frac{1}{n}(y_1-\tilde{y}_1,y_2-\tilde{y}_2,\cdots y_n-\tilde{y}_n)\begin{pmatrix}y_1-\tilde{y}_1\\y_2-\tilde{y}_2\\\vdots\\y_n-\tilde{y}_n\end{pmatrix}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2$$

$$\hat{\rho}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x}_i))(y_i-\tilde{y}_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\tilde{x}_i)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2}}=\frac{\frac{1}{n}(X-WX)^{\mathrm{T}}(Y-WY)}{\sqrt{\frac{1}{n}(X-WX)^{\mathrm{T}}(X-WX)}\sqrt{\frac{1}{n}(Y-WY)^{\mathrm{T}}(Y-WY)}}$$

可以证明 ρ 满足不等式 $-1\le\rho\le1$，应用 Cauchy-Schwarz 不等式，可以重写空间变量 X,Y 满足：

$$\left\{E\left[(X-WX)(Y-WY)\right]\right\}^2\le E\left[(X-WX)^2\right]E\left[(Y-WY)^2\right]$$

等号成立当且仅当一个空间变量是另一个空间变量的倍数，这就是
$p\left[a(X-WX)=b(Y-WY)\right]=1$，$\forall a,b\in R$，$a,b$ 至少有一个不为 0。

可以重写不等式(2.6)为

$$-1\le\rho=\frac{E\left[(X-WX)^T(Y-WY)\right]}{\sqrt{E(X-WX)^2\cdot E(Y-WY)^2}}\le1$$

因为 $E\left[\left(X-WX\right)^2\right]$ 与 $E\left[\left(Y-WY\right)^2\right]$ 严格为正。

3. Monte Carlo 模拟

考虑如下三种数据生成情况，在模拟过程中空间权重矩阵为 Rook 规则型权重矩阵，观察的样本量为 20×20 单元格上的 400 个样本，模拟次数均为 50 次。

情况 1：假设数据生成过程为 $X=\rho_1WX+\varepsilon$ 和 $Y=\rho_2WY+\varepsilon$，

这是 X 与 Y 相互独立即 $X=\left(I-\rho_1W\right)^{-1}\varepsilon$ 和 $Y=\left(I-\rho_2W\right)^{-1}\varepsilon$，当固定 ρ_1 不变时 ρ_2 和 ρ_{XY} 之间关系如表 1。

从表 1 可以看出只有当 $\rho_1=\rho_2$ 是，两个空间变量的相关系数为 1，当 $\rho_1\neq\rho_2$ 时，两个空间变量的相关系数都趋于 0。

情况 2：假设数据生成过程为

$$\begin{cases}X=\left(I-\rho_1W\right)^{-1}\varepsilon\\ Y=\rho_2X+\varepsilon\end{cases}\quad \varepsilon\sim N\left(0,\sigma^2I\right)$$

从表 2 可以看出当 $\rho_1=0.9,\ 0.7,\ 0.6,\ 0.5$ 时随着 ρ_2 的增大，ρ_{XY} 也呈现出有一定规律的增大，当 $\rho_1=0.8$ 时 ρ_{XY} 随 ρ_2 的变化比较特殊，当 $\rho_1=0.4,\ 0.3$ 时，当 $\rho_2<0$ 时，ρ_{XY} 除个别点外基本上趋于−1，当 $\rho_2>0$ 除个别点外基本上趋于 1，当 $\rho_1=0.2,\ 0.1$ 两个空间变量的相关系数基本上都大于 0 并且随 ρ_2 的增大而增大，当 $\rho_1=0$ 是，两个空间变量的相关系数恒为 1.当 $\rho_2<-1$ 时，不论 ρ_1 怎样变化 ρ_{XY} 趋于−1，当 $\rho_2>1$，不论 ρ_1 怎样变化 ρ_{XY} 趋于 1。

情况 3：假设数据生成过程为

$$\begin{cases}X=\left(I-\rho_1W\right)^{-1}\varepsilon\\ Y=\rho_2WX+\varepsilon\end{cases}\quad \varepsilon\sim N\left(0,\sigma^2I\right)$$

从表 3 可以看出，当 $\rho_1=\rho_2$ 时，它们之间的相关系数恒等于 1。当 $\rho_1\neq\rho_2$ 时，$\rho_1=0.9,\ 0.7,\ 0.6,\ 0.5$ 是 ρ_{XY} 随着 ρ_2 的增大呈现出一定规律的变化，$\rho_1=0.8$ 和第二种情况一样变化比较特殊，$\rho_1=0.4,\ 0.3$ 时，当 $\rho_2>0$ 时 ρ_{XY} 趋于 1，$\rho_2<0$ 时趋于−1，当 $\rho_1=0.2,\ 0.1,\ 0$ 时，ρ_{XY} 呈现出先上升后下降的趋势。

4. 实证分析

这一部分用新定义的相关系数来计算 GDP 和固定资产投资，GDP 和社会消费总额，GDP 和进出口额这三对空间变量之间的系数，进一步来确认这个新的定义。使用 2013 年中国 31 个省份(不含港澳台地区) GDP，固定资产投资，社会消费总额，进出口额的数据，数据来源 2014 年中国统计年鉴。

空间权重矩阵是计算两个空间变量相关系数的关键，也是地区间空间影响方式的体现。最常用的是简单二分权重矩阵，遵循的判定规则是 Rook 相邻规则，即两个地区拥有共同边界则视为相邻。矩阵 W 的

Table 1. Relationship between ρ_2 and ρ_{XY}

表 1. ρ_2 与 ρ_{XY} 之间的关系

ρ_1	0.9									
ρ_2	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
ρ_{XY}	1	0.096	−0.003	0.0075	0.0478	0.038	−0.019	0.009	0.063	0.024

Table 2. Relationship between ρ_1, ρ_2 and ρ_{XY}

表 2. ρ_1, ρ_2 与 ρ_{XY} 之间的关系

ρ_2 \ ρ_1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
−0.9	−0.9	−0.30	−0.87	−0.89	−0.95	−1	−1	−0.28	0.54	1
−0.8	−0.88	−0.26	−0.84	−0.86	−0.94	−1	−1	−0.09	0.73	1
−0.7	−0.85	−0.16	−0.80	−0.82	−0.92	−1	−1	0.10	0.82	1
−0.6	−0.81	−0.08	−0.74	−0.78	−0.90	−1	−1	0.30	0.88	1
−0.5	−0.75	−0.01	−0.67	−0.71	−0.86	−1	−1	0.45	0.91	1
−0.4	−0.66	0.06	−0.57	−0.61	−0.80	−1	−1	0.57	0.94	1
−0.3	−0.54	0.13	−0.44	−0.48	−0.70	−1	−1	0.67	0.95	1
−0.2	−0.36	0.20	−0.27	−0.30	−0.54	−1	−0.91	0.74	0.96	1
−0.1	−0.15	0.27	−0.07	−0.09	−0.30	−0.9	−0.72	0.79	0.97	1
0	0.01	0.33	0.14	0.14	0.03	−0.09	0.15	0.83	0.97	1
0.1	0.32	0.39	0.33	0.35	0.34	0.9	0.8	0.86	0.98	1
0.2	0.50	0.45	0.49	0.51	0.57	1	0.93	0.88	0.98	1
0.3	0.63	0.50	0.61	0.64	0.72	1	0.96	0.90	0.98	1
0.4	0.73	0.54	0.70	0.73	0.81	1	0.98	0.92	0.99	1
0.5	0.80	0.59	0.76	0.79	0.86	1	0.99	0.93	0.99	1
0.6	0.84	0.62	0.81	0.83	0.90	1	0.99	0.94	0.99	1
0.7	0.87	0.66	0.85	0.87	0.92	1	0.99	0.94	0.99	1
0.8	0.90	0.69	0.88	0.89	0.94	1	0.99	0.95	0.99	1
0.9	0.92	0.71	0.90	0.90	0.95	1	1	0.96	0.99	1

Table 3. Relationship between ρ_1, ρ_2 and ρ_{XY}

表 3. ρ_1, ρ_2 与 ρ_{XY} 之间的关系

ρ_2 \ ρ_1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
−0.9	−0.75	−0.03	−0.72	−0.78	−0.91	−1	−1	−0.13	0.55	0.81
−0.8	−0.72	−0.01	−0.70	−0.76	−0.90	−1	−1	−0.09	0.57	0.83
−0.7	−0.7	0.01	−0.67	−0.74	−0.89	−1	−1	−0.05	0.60	0.84
−0.6	−0.66	0.04	−0.64	−0.71	−0.88	−1	−1	0.004	0.64	0.86
−0.5	−0.61	0.07	−0.59	−0.68	−0.86	−1	−1	0.075	0.68	0.88
−0.4	−0.55	0.11	−0.53	−0.61	−0.83	−1	−1	0.17	0.73	0.9
−0.3	−0.46	0.15	−0.44	−0.53	−0.78	−1	−1	0.29	0.78	0.93
−0.2	−0.33	0.20	−0.31	−0.4	−0.68	−1	−1	0.45	0.85	0.96
−0.1	−0.18	0.26	−0.12	−0.18	−0.47	−1	−1	0.64	0.91	0.99
0	0.09	0.33	0.14	0.14	0.03	−0.09	0.15	0.83	0.97	1
0.1	0.37	0.42	0.44	0.51	0.66	1	1	0.96	1	0.97
0.2	0.62	0.51	0.70	0.79	0.92	1	1	1	0.97	0.86
0.3	0.79	0.62	0.86	0.92	0.98	1	1	0.98	0.87	0.65
0.4	0.90	0.74	0.95	0.98	0.997	1	1	0.93	0.74	0.39
0.5	0.95	0.84	0.98	0.996	1	1	1	0.87	0.6	0.17
0.6	0.98	0.93	0.99	1	0.998	1	1	0.83	0.49	0.005
0.7	0.99	0.98	1	0.998	0.996	1	1	0.79	0.39	−0.114
0.8	1	1	0.998	0.99	0.994	1	1	0.76	0.31	−0.2
0.9	1	0.98	0.994	0.99	0.99	1	1	0.73	0.25	−0.27

Table 4. China's 31 provinces, cities and geographical neighbor information
表 4. 中国 31 个省市地理相邻信息

序号	地区	相邻信息	序号	地区	相邻信息
1	北京	2、3	17	湖北	12、14、16、18、22、27
2	天津	1、3、15	18	湖南	14、17、19、20、22、24
3	河北	1、2、4、5、6、15、16	19	广东	13、14、18、20、21
4	山西	3、5、16、27	20	广西	18、19、24、25
5	内蒙古	3、4、6、7、8、27、28、30	21	海南	19
6	辽宁	3、5、7	22	重庆	17、18、23、24、27
7	吉林	5、6、8	23	四川	22、24、25、26、27、28、29
8	黑龙江	5、7	24	贵州	18、20、22、23、25
9	上海	10、11	25	云南	20、23、24、26
10	江苏	9、11、12、15	26	西藏	23、25、29、31
11	浙江	9、10、12、13、14	27	陕西	4、5、16、17、22、23、28、30
12	安徽	10、11、14、15、16、17	28	甘肃	5、23、27、29、30、31
13	福建	11、14、19	29	青海	23、26、28、31
14	江西	11、12、13、17、18、19	30	宁夏	5、27、28
15	山东	2、3、10、12、16	31	新疆	26、28、29
16	河南	3、4、12、15、17、27			

资料来源：《中华人民共和国地图》。

设定方式如下：主对角线上的元素为 0，如果 i 地区与 j 地区相邻，则 W_{ij} 为 1，否则为 0。W_{ij} 经过行标准化处理，用每个元素同时除以所在行元素之和，使得每行元素之和为 1，表 4 列出了中国 31 个省市地理相邻信息。这种设置方式简单，计算简便，故使用广泛。

计算得出 Y(GDP)和固定资产投资($X1$)之间的相关系数为：0.876

Y 的空间自回归模型为：$Y = 0.639WY + e$；$X1$ 的空间自回归模型为：$X1 = 0.760WX1 + e$

计算得出 Y(GDP)和社会消费总额($X2$)之间的相关系数为：0.991

$X2$ 的空间自回归模型为：$X2 = 0.599WX2 + e$；

计算得出 Y(GDP)和进出口总额($X3$)之间的相关系数为：0.839

$X3$ 的空间自回归模型为：$X3 = 0.293WX3 + e$。

5. 结论

在上面三种情况下，模拟了两个空间变量的相关系数，出现了一定的规律性，但这种规律比较弱，它们之间的定量关系还得进一步探讨。但是通过实证分析得出，当两个空间变量的空间结构相近且这两个变量之间存在一定的关系时，它们之间新定义的这种相关系数就大，当两个空间变量的空间结构相差比较大时，它们之间新定义的这种相关系数就小，这样当两个空间变量存在一定关系时，这种新的定义刻画的是两个空间变量空间结构的相似性，故把这个新的定义称为两个空间变量空间结构的相似系数。

参考文献 (References)

[1] Baltagi, B.H., Song, S.H., Jung, B.C. and Koh, W. (2003) Testing for Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. *Journal of Econometrics*, **140**, 5-51. https://doi.org/10.1016/j.jeconom.2006.09.001

[2] Pesaran, M.H. and Hashem Pesaran, M. (2004) General Diagnostic Tests for Cross Section Dependence in Panels. CESifo Working Papers, No. 1229.

[3] 张晓旭, 冯宗宪. 中国人均 GDP 的空间关系与地区收敛: 1987-2003 [J]. 经济学(季刊), 2008(1): 399-414.

[4] Li, Q., Song, J.P., Wang, E.R., Hu, H., Zhang, J.H. and Wang, Y.Y. (2014) Economic Growth and Pollutant Emissions in China: A Spatial Econometric Analysis. *Stochastic Environmental Research and Risk Assessment*, **28**, 429-442. https://doi.org/10.1007/s00477-013-0762-6