

## 第 29 章 空间计量经济学

### 29.1 地理学第一定律

许多经济数据都涉及一定的空间位置。比如, 研究全国各省的 GDP、投资、贸易、R&D 等数据。

此前各章很少关注各省经济之间的互动, 通常假设各省的变量相互独立。

但各省经济有着广泛的联系, 而且越近的省份联系越密切。

根据 Tobler (1970), “所有事物都与其他事物相关联, 但较近的事物比较远的事物更关联” (Everything is related to everything else, but near things are more related than distant things)。

这被称为 “地理学第一定律” (First Law of Geography)。

各省之间的距离信息并不难获得, 比如是否相邻, 直线距离或运输距离。

将各省的变量数据, 再加上各省的位置信息(或相互距离), 即可得到 “空间数据” (spatial data 或 areal data)。

研究如何处理空间数据的计量经济学分支, 称为 “空间计量经济学” (spatial econometrics)。

空间计量经济学的最大特色在于充分考虑横截面单位之间的空间依赖性(spatial dependence)。

空间效应(spatial effects)包括空间依赖性与“空间异质性”(spatial heterogeneity)。

由于标准的计量经济学也考虑横截面单位之间的异质性(比如异方差)，故空间计量经济学的关注重点为空间依赖性。

空间计量经济学诞生于 1970 年代。近年来，空间计量经济学蓬勃发展并进入主流，可归功于两方面。

首先，由于 GIS(地理信息系统)的发展，空间数据或包含地理信息的数据(geo-referenced data)日益增多。

其次，在经济理论方面，人们越来越关注经济行为人之间的互动，而不仅仅停留于代表性厂商或个人。

比如，在考察同伴效应(peer effect)，相邻效应(neighborhood effect)，溢出效应(spillover effect)或网络效应(network effect)时，都需要明确地考虑空间因素。

## 29.2 空间权重矩阵

进行空间计量分析的前提是度量区域之间的空间距离。

记来自  $n$  个区域的空间数据为  $\{x_i\}_{i=1}^n$ ，下标  $i$  表示区域  $i$ 。记区域  $i$  与区域  $j$  之间的距离为  $w_{ij}$ ，则可定义“空间权重矩阵”(spatial weighting matrix)如下：

$$\mathbf{W} = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}$$

其中，主对角线上元素  $w_{11} = \cdots = w_{nn} = 0$  (同一区域的距离为 0)。

空间权重矩阵  $\mathbf{W}$  为对称矩阵。

最常用的距离函数为“相邻”(contiguity)，即如果区域  $i$  与区域  $j$  有共同的边界，则  $w_{ij} = 1$ ；反之，则  $w_{ij} = 0$ 。

比照(国际)象棋中棋子的行走路线，相邻关系可分为以下几种：

(1) 车相邻(rook contiguity): 两个相邻区域有共同的边。

(2) 象相邻(bishop contiguity): 两个相邻区域有共同的顶点, 但没有共同的边。

(3) 后相邻(queen contiguity): 两个相邻区域有共同的边或顶点。

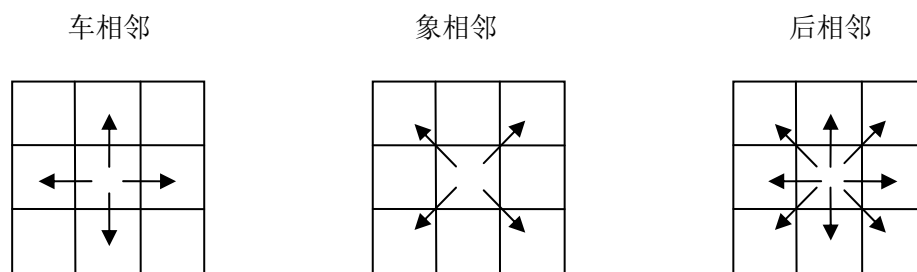


图 29.1 常用相邻关系

在实践中，为了区分“边”与“点”，须设定一个最小距离，在此距离以下为点，而在此距离以上为边。

究竟使用车、象或后相邻，取决于具体情况。

比如，区域  $i$  与区域  $j$  仅在一点相交(象相邻)，但有一条主要高速公路通过此点连接两区域，则不宜使用车相邻。

假设有如下四个区域，其变量取值分别为  $\mathbf{x} = (x_1 \ x_2 \ x_3 \ x_4)'$ 。

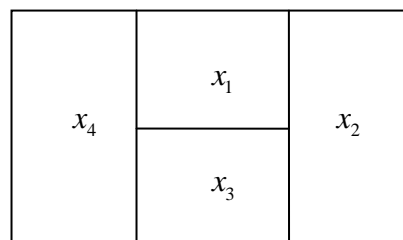


图 29.2 假想的四个区域

其空间权重矩阵为：

$$W = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

第一行表示，区域 1 与其余三个区域均相邻；第二行表示，区域 2 与区域 1、区域 3 相邻，但不与区域 4 相邻；以此类推。



空间权重矩阵考虑的是一阶邻居，还可以考虑二阶邻居，即邻居的邻居，可用矩阵 $W^2$ 来表示。

矩阵 $W^2$ 的主对角线上元素一般不再为 0，这意味着邻居的邻居也包括自己。

实践中，有时对空间权重矩阵进行“行标准化”(row standardization)，即将矩阵中的每个元素(记为 $\tilde{w}_{ij}$ )除以其所在行元素之和，以保证每行元素之和为 1：

$$w_{ij} \equiv \frac{\tilde{w}_{ij}}{\sum_j \tilde{w}_{ij}}$$

如果区域  $i$  为孤岛，与其他区域均不相邻，则上式分母为 0，并不适用；可将分母改为  $\max(1, \sum_j \tilde{w}_{ij})$ 。

不包含孤岛的行标准化矩阵也称为“行随机矩阵”(row-stochastic matrix)，所有元素均介于 0 与 1 之间，且每行元素之和为 1，在形式上与离散型概率分布一样。

将前面的空间权重矩阵行标准化可得(仍记为  $W$ ):

$$W = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

行标准化的好处在于，如果将行标准化矩阵 $\mathbf{W}$ 乘以 $\mathbf{x}$ ，则可得到每个区域邻居的平均值。在上例中：

$$\mathbf{W}\mathbf{x} = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{bmatrix} (x_2 + x_3 + x_4)/3 \\ (x_1 + x_3)/2 \\ (x_1 + x_2 + x_4)/3 \\ (x_1 + x_3)/2 \end{bmatrix}$$

比如，区域 1 的邻居为区域 2, 3 和 4，而上式右边第一行元素正好为 $(x_2 + x_3 + x_4)/3$ ，即区域 1 邻居的平均值；以此类推。

比照时间序列中时间滞后(time lag)的概念,  $W\mathbf{x}$  也被称为  $\mathbf{x}$  的“空间滞后” (spatial lag), 即  $\mathbf{x}$  邻居的平均取值。

行标准化之后的空间权重矩阵一般不再是对称矩阵, 这是它的缺陷之一。

由于每行元素之和均为 1, 这意味着区域  $i$  所受其邻居的影响之和一定等于区域  $j$  所受其邻居的影响之和(任意  $i \neq j$ ); 此假定可能过强, 这是行标准化的另一局限。

定义相邻关系的另一方法基于区域间的距离。

记区域  $i$  与区域  $j$  的距离为  $d_{ij}$ ，可定义空间权重如下：

$$w_{ij} = \begin{cases} 1 & \text{若 } d_{ij} < d \\ 0 & \text{若 } d_{ij} \geq d \end{cases}$$

其中， $d$  为事先给定的距离临界值。

也可直接以距离之倒数(inverse distance)作为空间权重：

$$w_{ij} = \frac{1}{d_{ij}}$$

在上式中，距离  $d_{ij}$  既可以是地理距离，比如直线距离或大圆距离(great circle distance)；也可以是基于运输成本或旅行时间的经济距离；甚至社交网络中的距离。

例：林光平等(2005)使用基于地理相邻关系的简单权重矩阵  $W$  来研究我国 28 个省市在 1978-2002 年期间实际人均 GDP 的收敛情况。但相邻地区经济上的相互关系并不完全相同。

为此，林光平等(2005)使用地区间人均 GDP 的差额作为测度地区间“经济距离”的指标，并引入经济空间权重矩阵  $W^* = W \times E$ ，其中矩阵  $E$  的主对角线元素均为 0，非主对角线的  $(i, j)$  元素为

$$E_{ij} = \frac{1}{|\bar{Y}_i - \bar{Y}_j|} (i \neq j), \quad \bar{Y}_i \text{ 为地区 } i \text{ 样本期间的人均实际 GDP 平均值。}$$

## 29.3 空间自相关

在使用空间计量方法前，首先要考察数据是否存在空间依赖性。

如果不存在，则使用标准的计量方法即可；

如果存在，则可使用空间计量方法。

比照时间序列(time series)，空间数据有时也称为“空间序列”(spatial series)。

时间序列可视为在时间轴上分布的随机过程，而空间数据(序列)则为在空间分布的随机过程。

时间序列的一个重要特性是可能存在自相关，特别是一阶自相关。

对于空间序列，自相关的情形则更为复杂；因为时间序列只可能在一个方向上相关(过去影响现在，但现在无法影响过去)，而空间序列则可以在多个方向上相关，而且可以互相影响( $x_i$ 影响 $x_j$ ，而 $x_j$ 也影响 $x_i$ )。

“空间自相关” (spatial autocorrelation)可理解为位置相近的区域具有相似的变量取值。

如果高值与高值聚集在一起，低值与低值聚集在一起，则为“正空间自相关” (positive spatial autocorrelation)。



如果高值与低值相邻，则为“负空间自相关”(negative spatial autocorrelation)；较少见。

如果高值与低值完全随机地分布，则不存在空间自相关。

考虑空间序列 $\{x_i\}_{i=1}^n$ 。

文献中提出了一系列度量空间自相关的方法，最为流行的是“莫兰指数  $I$ ”(Moran's  $I$ ):

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

其中,  $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$  为样本方差,  $w_{ij}$  为空间权重矩阵的  $(i, j)$  元素(用来度量区域  $i$  与区域  $j$  之间的距离), 而  $\sum_{i=1}^n \sum_{j=1}^n w_{ij}$  为所有空间权重之和。

如果空间权重矩阵为行标准化, 则  $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = n$ , 莫兰指数  $I$  为:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

莫兰指数  $I$  的取值一般介于-1 到 1 之间，大于 0 表示正自相关，即高值与高值相邻、低值与低值相邻；小于 0 表示负自相关，即高值与低值相邻。

如果莫兰指数  $I$  接近于 0，则表明空间分布是随机的，不存在空间自相关。

莫兰指数  $I$  可视为观测值与其空间滞后(spatial lag)的相关系数。

如果将观测值与其空间滞后画成散点图，称为“莫兰散点图”(Moran scatterplot)，则莫兰指数  $I$  就是该散点图回归线的斜率。

考虑原假设“ $H_0 : \text{Cov}(x_i, x_j) = 0, \forall i \neq j$ ”(即不存在空间自相关)。

在此原假设下，莫兰指数  $I$  的期望值为

$$E(I) = \frac{-1}{n-1}$$

莫兰指数  $I$  的方差表达式更为复杂，记为  $\text{Var}(I)$ 。

标准化的莫兰指数  $I$  服从渐近标准正态分布：

$$I^* \equiv \frac{I - E(I)}{\sqrt{\text{Var}(I)}} \xrightarrow{d} N(0, 1)$$

在使用莫兰指数  $I$  检验空间自相关时，须注意两个问题。

问题之一，莫兰指数  $I$  取决于空间矩阵  $W$ ，如果空间矩阵设定不正确，则可能导致错误的结果。

问题之二，莫兰指数  $I$  的核心成分为  $(x_i - \bar{x})(x_j - \bar{x})$ ，其隐含假设是  $\{x_i\}_{i=1}^n$  的期望值为常数(constant mean)，不存在任何趋势(trend)。如果存在趋势，则可能导致检验结果出现偏差。

为了解决问题一，须仔细选择合适的空间矩阵，或使用不同的空间矩阵以考察结果的稳健性。

为了解决问题二，可引入协变量，通过回归的方法去掉趋势，然后对残差项进行莫兰指数  $I$  检验。

以上的莫兰指数  $I$  也被称为“全局莫兰指数  $I$ ”(global Moran's  $I$ ), 考察整个空间序列  $\{x_i\}_{i=1}^n$  的空间集聚情况。

如果想知道某区域  $i$  附近的空间集聚情况, 可使用“局部莫兰指数  $I$ ”(local Moran's  $I$ ):

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x})$$

局部莫兰指数  $I$  的含义与全局莫兰指数  $I$  相似。

莫兰指数  $I$  并非唯一的空间自相关指标, 另一常用指标为“吉尔里指数  $C$ ”(Geary's  $C$ )(Geary, 1954), 也称为“吉尔里相邻比率”(Geary's Contiguity Ratio):

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{2 \left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right]}$$

吉尔里指数  $C$  的核心成分为  $(x_i - x_j)^2$ 。

吉尔里指数  $C$  的取值一般介于 0 到 2 之间(2 不是严格上界), 大于 1 表示负相关, 等于 1 表示不相关, 而小于 1 表示正相关。

吉尔里指数  $C$  与莫兰指数  $I$  呈反向变动; 前者比后者对于局部空间自相关更为敏感。

在不存在空间自相关的原假设下，吉尔里指数  $C$  的期望值为 1，而方差的表达式较复杂，记为  $\text{Var}(C)$ 。

标准化的吉尔里指数  $C$  服从渐近标准正态分布：

$$C^* \equiv \frac{C-1}{\sqrt{\text{Var}(C)}} \xrightarrow{d} N(0, 1)$$

莫兰指数  $I$  与吉尔里指数  $C$  的共同缺点在于，即无法分别“热点” (hot spot) 与“冷点” (cold spot) 区域。

所谓热点区域，即高值与高值聚集的区域；而冷点区域则是低值与低值聚集的区域。热点区域与冷点区域都表现为正自相关。



Getis and Ord (1992)提出了以下 “Getis-Ord 指数  $G$ ”:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j \neq i}^n x_i x_j}$$

其中,  $x_i > 0, \forall i$ ; 而  $w_{ij}$  来自非标准化的对称空间权重矩阵, 且所有元素均为 0 或 1。

如果样本中高值聚集在一起, 则  $G$  较大; 如果低值聚集在一起, 则  $G$  较小。

在无空间自相关的原假设下， $E(G) = \frac{\sum_{i=1}^n \sum_{j \neq i}^n w_{ij}}{n(n-1)}$ 。

如果  $G$  值大于此期望值，则表示存在热点区域；

如果  $G$  值小于此期望值，则表示存在冷点区域。

标准化的  $G$  服从渐近标准正态分布：

$$G^* \equiv \frac{G - E(G)}{\sqrt{\text{Var}(G)}} \xrightarrow{d} N(0, 1)$$

如果  $G^* > 1.96$ ，则可在 5% 水平上拒绝无空间自相关的原假设，认为存在空间正自相关，且存在热点区域。

如果  $G^* < -1.96$ ，则可在 5% 水平上拒绝无空间自相关的原假设，认为存在空间正自相关，且存在冷点区域。

如果要考察某区域  $i$  是否为热点或冷点，则可使用“局部 Getis-Ord 指数  $G$ ”：

$$G_i = \frac{\sum_{j \neq i} w_{ij} x_j}{\sum_{j \neq i} x_j}$$

以上各种空间自相关指标仅提供是否存在空间效应的初步检验，深入检验有赖于建立正式的空间计量模型。

## 29.4 空间自回归模型

对于时间序列，最常见的建模方式为一阶自回归 AR(1)。

空间自回归的形式更为复杂，因为空间滞后可以来自不同的方向，而且可以双向。

对于空间序列 $\{y_i\}_{i=1}^n$  (样本容量为  $n$ )，即使假设空间自相关的形式为线性，待估计的参数在理论上最多可达 $(n^2 - n)$ 个(每个区域最多可受 $n-1$ 个区域影响，共有  $n$  个区域)，大大超出样本容量。

必须假设空间依赖性服从某种共同的模式，才能简化参数。

首先回顾时间序列的一阶自相关模型：

$$y_t = \beta y_{t-1} + \varepsilon_t \quad (t = 2, \dots, n)$$

假设  $n = 4$ ，而  $y_1 = \varepsilon_1$ 。将这 4 个方程叠放在一起可得：

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \beta \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\mathbf{W}} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \equiv \beta \mathbf{W} \mathbf{y} + \boldsymbol{\varepsilon}$$

其中，矩阵  $\mathbf{W}$  的次主对角线元素均为 1，而其他元素都为 0。

矩阵 $W$ 可视为“时间滞后矩阵”，它之所以呈现出次对角线元素为 1 而其他元素为 0 的规律性排列，正是由于时间滞后在方向上的单一性。

对于空间序列，矩阵 $W$ 则为空间权重矩阵，其元素排列方式表现出多样性(因为空间自相关可以多方向)。

引入以下“空间自回归模型”(Spatial Autoregression, 简记 SAR):

$$\mathbf{y} = \lambda W \mathbf{y} + \boldsymbol{\varepsilon}$$

其中， $W$ 为已知的空间权重矩阵(非随机)，而空间依赖性仅由单一参数 $\lambda$ 。

$\lambda$ 度量空间滞后 $Wy$ 对 $y$ 的影响，称为“空间自回归系数”(spatial autoregressive parameter)。

此模型也称为“空间滞后模型”(Spatial Lag Model)。

相邻地区的被解释变量(比如犯罪率)可能相互依赖，并最终形成一个均衡的结果。

又比如，假设以地区税收为被解释变量，则不同地区的政府出于相互竞争或博弈的考虑(比如，竞相吸引 FDI)，在制定本地区税收时，会考虑周边地区的税收水平。

假设 $W$ 为前例的权重矩阵，则方程可写为：

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \lambda \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = \lambda \begin{pmatrix} y_2 + y_3 + y_4 \\ y_1 + y_3 \\ y_1 + y_2 + y_4 \\ y_1 + y_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

如果对此联立方程系统进行 OLS 估计，将导致联立方程偏差 (simultaneity bias)。

由于空间依赖性，导致变量  $\{y_i\}_{i=1}^n$  之间互相影响，产生内生性。

更一般地，可在方程中加入自变量：

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$



其中， $\mathbf{X}$  为  $n \times k$  数据矩阵，包括  $k$  列解释变量； $\boldsymbol{\beta}_{k \times 1}$  为相应系数。

此方程也称为 SAR 模型。

如果  $\lambda = 0$ ，则简化为一般的线性回归模型。

可通过检验原假设 “ $H_0: \lambda = 0$ ” 来考察是否存在空间效应。

对于空间自回归模型，常使用 MLE 估计。

首先，假设扰动项  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ 。

其次，方程可写为

$$\mathbf{A}\mathbf{y} \equiv (\mathbf{I} - \lambda\mathbf{W})\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

其中，  $\mathbf{A} \equiv (\mathbf{I} - \lambda \mathbf{W})$ 。

由于雅可比行列式  $J \equiv \left| \frac{\partial \boldsymbol{\varepsilon}}{\partial \mathbf{y}} \right| = \left| \frac{\partial (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\partial \mathbf{y}} \right| = \left| \frac{\partial \mathbf{A}\mathbf{y}}{\partial \mathbf{y}} \right| = |\mathbf{A}'| = |\mathbf{A}|$ ，根据多维正态密度公式，可写出样本的似然函数：

$$L(\mathbf{y} | \lambda, \sigma^2, \boldsymbol{\beta}) = (2\pi\sigma^2)^{-n/2} (\text{abs}|\mathbf{A}|) \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

其中，  $\text{abs}|\mathbf{A}|$  表示行列式  $|\mathbf{A}|$  的绝对值。

对数似然函数为：

$$\ln L(\mathbf{y} | \lambda, \sigma^2, \boldsymbol{\beta}) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln(\text{abs}|A|) - \frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

此最大化问题分两步。

第一步，在给定 $\lambda$ 的情况下，选择最优的 $\boldsymbol{\beta}, \sigma^2$ 。

第二步，代入第一步的最优 $\boldsymbol{\beta}, \sigma^2$ ，选择最优的 $\lambda$ 。

在第一步，由于 $\boldsymbol{\beta}$ 只出现于最后一项，故等价于使 $(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ 最小，即 $\mathbf{A}\mathbf{y}$ 对 $\mathbf{X}$ 进行回归：

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1} X'Ay = (X'X)^{-1} X'(I - \lambda W)y \\ &= (X'X)^{-1} X'y - \lambda (X'X)^{-1} X'Wy \equiv \hat{\beta}_0 - \lambda \hat{\beta}_L\end{aligned}$$

其中， $\hat{\beta}_0 \equiv (X'X)^{-1} X'y$ （ $y$  对  $X$  的回归系数），而  $\hat{\beta}_L \equiv (X'X)^{-1} X'Wy$ （ $Wy$  对  $X$  的回归系数）。

只要知道  $\lambda$ ，即可计算  $\hat{\beta}$ 。对  $\sigma^2$  求偏导可得  $\sigma^2$  的 MLE 估计：

$$\hat{\sigma}^2 = \frac{e'e}{n} = \frac{(M_X Ay)'(M_X Ay)}{n}$$

其中， $e$  为  $Ay$  对  $X$  回归的残差向量， $M_X \equiv I_n - X(X'X)^{-1}X'$  为消灭矩阵。

由于  $A\mathbf{y} = (I - \lambda W)\mathbf{y}$ , 故

$$\mathbf{e} = M_X A\mathbf{y} = M_X (I - \lambda W)\mathbf{y} = M_X \mathbf{y} - \lambda M_X W\mathbf{y} \equiv \mathbf{e}_0 - \lambda \mathbf{e}_L$$

其中,  $\mathbf{e}_0 \equiv M_X \mathbf{y}$  ( $\mathbf{y}$  对  $X$  的回归残差), 而  $\mathbf{e}_L \equiv M_X W\mathbf{y}$  ( $W\mathbf{y}$  对  $X$  的回归残差)。代入可得:

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n} = \frac{(\mathbf{e}_0 - \lambda \mathbf{e}_L)'(\mathbf{e}_0 - \lambda \mathbf{e}_L)}{n}$$

只要知道  $\lambda$ , 即可计算  $\hat{\sigma}^2$ 。

在第二步, 将  $\hat{\boldsymbol{\beta}}(\lambda), \hat{\sigma}^2(\lambda)$  代入对数似然函数, 可得 “集中对数

似然函数”(concentrated log likelihood function), 它只是 $\lambda$ 的函数。

由于 $\lambda$ 出现在行列式 $|\mathbf{A}| = |\mathbf{I} - \lambda \mathbf{W}|$ , 给计算带来不便。

可利用等式 $|\mathbf{A}| = \prod_{i=1}^n (1 - \lambda v_i)$ 来计算, 其中 $v_1, \dots, v_n$ 为矩阵 $\mathbf{A}$ 特征值(eigenvalues)。

为保证扰动项协方差矩阵为正定, 还须限制 $\lambda$ 的取值为 $\frac{1}{v_{\min}} < \lambda < \frac{1}{v_{\max}}$ , 其中 $v_{\min}$ 与 $v_{\max}$ 分别为矩阵 $\mathbf{A}$ 的最小与最大特征值, 而 $v_{\min}$ 一定为负数。

对于 MLE 估计量的渐近协方差矩阵，通过信息矩阵来估计：

$$\mathbf{I}(\boldsymbol{\theta})^{-1} \equiv -\left\{ \mathbf{E} \left[ \frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\}^{-1}$$

其中， $\boldsymbol{\theta} = (\lambda, \sigma^2, \boldsymbol{\beta})$ 。

对于空间自回归模型  $\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ，解释变量  $\mathbf{X}$  对  $\mathbf{y}$  的边际效应并非  $\boldsymbol{\beta}$ ，因为  $\mathbf{X}$  对  $\mathbf{y}$  产生作用后， $\mathbf{y}$  之间还会相互作用，直至达到一个新的均衡。

## 29.5 空间杜宾模型

空间效应的另一建模方式是, 假设区域  $i$  的被解释变量  $y_i$  依赖于其邻居的自变量:

$$y = X\beta + WX\delta + \varepsilon$$

其中,  $WX\delta$  表示来自邻居自变量的影响, 而  $\delta$  为相应的系数向量。

例如, 局域  $i$  的犯罪率不仅依赖于本区域的警力, 还可能依赖于相邻区域的警力。

此模型称为“空间杜宾模型”(Spatial Durbin Model), 简记 SDM。

此方程不存在内生性, 故可直接进行 OLS 估计; 只是解释变量  $X$  与  $WX$  之间可能存在多重共线性。



如果 $\delta = \mathbf{0}$ ，则简化为一般的线性回归模型。

将空间杜宾模型与空间自回归模型相结合：

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\delta + \boldsymbol{\varepsilon}$$

## 29.6 空间误差模型

空间依赖性还可能通过误差项来体现。考虑以下“空间误差模型”(Spatial Errors Model, 简记 SEM):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

其中，扰动项 $\mathbf{u}$ 的生成过程为

$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

其中， $\mathbf{M}$ 为空间权重矩阵。

扰动项 $\mathbf{u}$ 存在空间依赖性，不包含在 $\mathbf{X}$ 中但对 $\mathbf{y}$ 有影响的遗漏变量存在空间相关性，或者不可观测的随机冲击存在空间相关性。

如果 $\rho = 0$ ，则简化为一般的线性回归模型。

代入可得：

$$\mathbf{B}\mathbf{u} \equiv (\mathbf{I} - \rho\mathbf{M})\mathbf{u} = \boldsymbol{\varepsilon}$$

其中， $\mathbf{B} \equiv \mathbf{I} - \rho\mathbf{M}$ 。

对于 SEM 模型，尽管扰动项存在自相关，OLS 仍一致，但缺乏效率。最有效率的方法为 MLE。

样本的对数似然函数可写为

$$\ln L(\mathbf{y} | \rho, \sigma^2, \boldsymbol{\beta}) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln(\text{abs}|\mathbf{B}|) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

最大化可分两步进行。

第一步，通过 GLS 计算  $\hat{\boldsymbol{\beta}}$ ，即最小化上式最后一项  $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ ；然后通过一阶条件得到  $\hat{\sigma}^2$ 。

第二步，将  $\hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho)$  代入对数似然函数，得到集中对数似然函数(仅为  $\rho$  的函数)，求解最优的  $\hat{\rho}$ 。

在第一步中，由于目标函数涉及  $\mathbf{B} = \mathbf{I} - \rho \mathbf{M}$ ，故得不到  $\hat{\boldsymbol{\beta}}(\rho)$  与  $\hat{\sigma}^2(\rho)$  的解析表达式。

需要在第一步与第二步之间进行迭代，即给定  $\hat{\rho}^{(1)}$ ，通过第一步得到  $\hat{\boldsymbol{\beta}}(\hat{\rho}^{(1)})$ ,  $\hat{\sigma}^2(\hat{\rho}^{(1)})$  的具体取值，再通过第二步得到  $\hat{\rho}^{(2)}$ ，以此类推，直至  $\hat{\rho}$  收敛。

初始值  $\hat{\rho}^{(1)}$  可通过在第一步进行 OLS 估计，然后代入第二步而得到。

## 29.7 一般的空间计量模型

更一般的空间计量模型将空间自回归模型(SAR)与空间误差模型(SEM)结合起来:

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$$

其中, 扰动项 $\mathbf{u}$ 的生成过程为

$$\mathbf{u} = \rho \mathbf{M} \mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

其中,  $\mathbf{W}$  与  $\mathbf{M}$  分别为被解释变量 $\mathbf{y}$ 与扰动项 $\mathbf{u}$ 的空间权重矩阵, 二者可以相等。

此模型称为“带空间自回归误差项的空间自回归模型”(Spatial Autoregressive Model with Spatial Autoregressive Disturbances, 简记 SARAR)。

SAR 模型, SEM 模型以及 SARAR 模型, 也统称为“Cliff-Ord 模型”(Cliff and Ord, 1973, 1981; Ord, 1975)。

SAR 模型与 SEM 模型都是 SARAR 模型的特例, 分别对应于  $\rho = 0$  与  $\lambda = 0$  的情形。

对于 SARAR 模型, 可进行 MLE 估计。

样本的对数似然函数为

$$\ln L(\mathbf{y} \mid \lambda, \rho, \sigma^2, \boldsymbol{\beta}) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln(\text{abs}|\mathbf{A}|) + \ln(\text{abs}|\mathbf{B}|) \\ - \frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \mathbf{B} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

最大化仍可分两步进行，并进行迭代，与SEM模型的情形类似。

但MLE估计量存在三个缺陷。

首先，MLE估计量可能计算不方便，特别当空间权重矩阵 $\mathbf{W}$ 与 $\mathbf{M}$ 的维度较高时。

其次，空间计量模型MLE估计量的大样本理论尚不健全。

再次，如果扰动项 $\epsilon$ 不服从独立同分布的正态分布，则QMLE估计量可能不一致。

Arraiz et al (2010)通过蒙特卡罗模拟证明，在异方差的情况下，QMLE估计量不一致。

Kelejian and Prucha (1999, 1998, 2004, 2010)提出利用工具变量，通过GMM来估计SARAR模型，称为“广义空间二段最小二乘法”(Generalized Spatial Two-stage Least Square, 简记GS2SLS)。此IV估计量在异方差情况下也一致。



GS2SLS的计算方便，不会因为样本容量太大而产生困难。

GS2SLS所使用的工具变量为：

$\{X, WX, \dots, W^q X, MX, MWX \dots, MW^q X\}$  中线性独立的列向量，  
记此工具变量矩阵(matrix of instrumental variables)为  $H$ 。

蒙特卡罗模拟显示，在很多情况下取  $q = 2$  即可。

由于  $X$  与  $y$  相关，而且  $X$  为外生变量；故以上  $X$  的线性函数为有效工具变量。

GS2SLS法的步骤:

第一步、使用工具变量矩阵  $\mathbf{H}$  进行2SLS估计。所得估计量  $(\tilde{\lambda}, \tilde{\beta})$  虽然一致, 但并非最有效率, 因为没有考虑扰动项  $\mathbf{u}$  存在空间自回归。记所得残差为  $\tilde{\mathbf{u}}$ 。

第二步、将第一步的2SLS残差  $\tilde{\mathbf{u}}$  作为  $\mathbf{u}$  的估计量代入方程, 然后进行GMM估计, 得到估计量  $\tilde{\rho}$ 。

第三步、使用  $\tilde{\rho}$  对方程进行“空间 Cochrane-Orcutt 变换” (spatial Cochrane-Orcutt transformation), 以去掉扰动项的空间自相关。

对于时间序列  $\{y_t\}$ ，如果其扰动项存在一阶自回归，比如  $u_t = \rho u_{t-1} + \varepsilon_t$ ，则可通过做准差分（即 Cochrane-Orcutt 变换） $(y_t - \rho y_{t-1})$  来消除此自相关。

对于空间序列，可做类似的变换。

将方程两边同时左乘  $(I - \rho M)$  可得：

$$(I - \rho M)y = \lambda(I - \rho M)Wy + (I - \rho M)X\beta + \underbrace{(I - \rho M)u}_{=\varepsilon}$$

新扰动项  $\varepsilon = (I - \rho M)u$  不再存在空间自相关。

将第二步估计量  $\tilde{\rho}$  代入此方程，再次使用工具变量  $H$  对变换后的方程进行 2SLS 估计，记所得估计量为  $(\hat{\lambda}, \hat{\beta})$ ，其相应残差为  $\hat{u}$ 。

第四步、将第三步的GS2SLS残差 $\hat{u}$ 作为 $u$ 的估计量代入方程，然后进行GMM估计，得到估计量 $\hat{\rho}$ 。

由于此估计量将2SLS用于空间数据，且使用了广义最小二乘法(Cochrane-Orcutt变换是GLS的特例)，故称为“广义空间二段最小二乘法”。

GS2SLS法的优点是计算方便且结果稳健(异方差情况下也成立);缺点是必须有外生变量 $\mathbf{X}$ ，故不适用于没有外生变量的纯SAR模型。

## 29.8 含内生解释变量的 **SARAR** 模型

GS2SLS 法可推广到包含内生解释变量的情形：

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$

其中， $\mathbf{Z}$  由内生解释变量所组成，而扰动项  $\mathbf{u}$  的生成过程为

$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

假设在  $\mathbf{X}$  之外，尚有不在方程中的外生变量  $\mathbf{X}_e$ 。

定义  $\mathbf{X}_f \equiv (\mathbf{X}, \mathbf{X}_e)$ ；如果没有不在方程中的外生变量  $\mathbf{X}_e$ ，则  $\mathbf{X}_f \equiv \mathbf{X}$ 。

定义工具变量矩阵  $\mathbf{H}$  为:

$\{\mathbf{X}_f, \mathbf{W}\mathbf{X}_f, \dots, \mathbf{W}^q\mathbf{X}_f, \mathbf{M}\mathbf{X}_f, \mathbf{M}\mathbf{W}\mathbf{X}_f, \dots, \mathbf{M}\mathbf{W}^q\mathbf{X}_f\}$  中线性独立的列向量, 即可进行 GS2SLS 估计。

## 29.9 空间面板模型

考察面板的空间自回归模型:

$$y_{it} = \rho \mathbf{w}'_i \mathbf{y}_t + \mathbf{x}'_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} \quad (i = 1, \dots, n; t = 1, \dots, T)$$

其中,  $\mathbf{w}'_i$  为空间权重矩阵  $\mathbf{W}$  的第  $i$  行,  $\mathbf{w}'_i \mathbf{y}_t = \sum_{j=1}^n w_{ij} y_{jt}$ ,  $w_{ij}$  为空间权重矩阵  $\mathbf{W}$  的  $(i, j)$  元素; 而  $u_i$  为区域  $i$  的个体效应。

如果不考虑空间滞后项  $\rho \mathbf{w}_i' \mathbf{y}_t$ ，则为标准的静态面板模型。

进一步，如果  $u_i$  与  $\mathbf{x}_{it}$  相关，则为固定效应模型；反之，则为随机效应模型。

对于固定效应的空间自回归模型，可先做组内离差变换，去掉个体效应  $u_i$ ；然后再用类似于横截面空间自回归模型的 MLE 估计。

对于随机效应的空间自回归模型，可先做广义离差变换，然后进行 MLE 估计。

在决定使用固定效应还是随机效应模型时，可进行通常的豪斯曼检验。

对于空间面板模型的 MLE 估计，可通过非官方命令 `xsmle` 来实现。此命令可估计很一般的空间面板模型：

$$\begin{cases} y_{it} = \tau y_{i,t-1} + \rho \mathbf{w}_i' \mathbf{y}_t + \mathbf{x}_{it}' \boldsymbol{\beta} + \mathbf{d}_i' \mathbf{X}_t \boldsymbol{\delta} + u_i + \gamma_t + \varepsilon_{it} \\ \varepsilon_{it} = \lambda \mathbf{m}_i' \boldsymbol{\varepsilon}_t + v_{it} \end{cases}$$

其中， $y_{i,t-1}$  为被解释变量  $y_{it}$  的一阶滞后(即动态面板；如果不是动态面板，可令  $\tau = 0$ )； $\mathbf{d}_i' \mathbf{X}_t \boldsymbol{\delta}$  表示解释变量的空间滞后， $\mathbf{d}_i'$  为相应空间权重矩阵  $\mathbf{D}$  的第  $i$  行； $\gamma_t$  为时间效应；而  $\mathbf{m}_i'$  为扰动项空间权重矩阵  $\mathbf{M}$  的第  $i$  行。

由于此模型太一般，故通常考虑以下特殊情形(每种情形均以命令 `xsmle` 所使用的术语来命名)：。



(1) 如果  $\lambda = 0$ ，则为“空间杜宾模型” (Spatial Durbin Model, 简记 SDM)。

(2) 如果  $\lambda = 0$  且  $\delta = \mathbf{0}$ ，则为“空间自回归模型” (Spatial Autoregression Model, 简记 SAR)。

(3) 如果  $\tau = 0$  且  $\delta = \mathbf{0}$ ，则为“空间自相关模型” (Spatial Autocorrelation Model, 简记 SAC); 即上文的 SARAR 模型。

(4) 如果  $\tau = \rho = 0$  且  $\delta = \mathbf{0}$ ，则为“空间误差模型” (Spatial Error Model, 简记 SEM)。

## 29.10 空间计量方法的局限性

空间计量的主要估计方法之一为 MLE，但此 MLE 的大样本理论尚待完善。

空间计量的另一局限是，它需要研究者主观设定一个非随机的空间权重矩阵(而非根据数据来估计此矩阵)，而此空间权重矩阵很可能无法完全反映不同地区之间复杂的相互关系。

处理区域间互动关系的另一方法是直接将地区间的距离作为变量引入回归模型。

比如，在国际贸易领域，常使用以下模型来考察两国间贸易量的决定因素：

$$Export_{ij} = \frac{\alpha Y_i^\beta Y_j^\gamma}{D_{ij}^\delta}$$

其中， $Export_{ij}$ 表示  $i$  国对  $j$  国的出口额， $Y_i, Y_j$  分别为  $i$  国与  $j$  国的 GDP，而  $D_{ij}$  为  $i$  国与  $j$  国之间的地理距离(通常以首都间的距离来衡量)。

此式很像牛顿的万有引力公式，故称为“引力模型”(gravity model)。

对方程两边取对数，并加上误差项：

$$\ln Export_{ij} = \ln \alpha + \beta \ln Y_i + \gamma \ln Y_j - \delta \ln D_{ij} + \varepsilon$$

在对引力模型的拓展研究中，经济学家又加入了影响贸易的其他因素，比如地理因素(如岛国、高山)、关税与非关税壁垒、边界效应、文化、宗教、语言、甚至政治制度(比如民主程度)，参见余淼杰(2013)。