

空间面板模型原理、方法与应用^①

杨威¹, 任晓航²

(1. 山西大学 管理与决策研究所, 太原 030006; 2. 中南大学 商学院, 长沙 410083)

摘要:近年来,随着计算机技术的飞速发展,空间数据的收集和存储变得更加容易,但是空间数据之间往往存在着相关性。为了更好地考虑空间相关性和异质性,空间面板模型应运而生。本文首先介绍空间面板模型的基本原理,然后对空间面板模型的估计方法和模型选择进行阐述,最后通过实例详述了在 Stata 和 R 命令中实现空间面板模型的具体步骤。

关键词:空间面板; Stata; R

中图分类号: O212.4; F224.0 **文献标识码:** A **文章编号:** (2020) 01-0114-09

0 引言

在经济领域的研究过程中,我们经常可以发现一个地区的某种经济现象或属性与邻近地区的同一经济现象或属性具有相关性。近年来,随着计算机技术的飞速发展,空间数据的收集和存储变得更加容易。尤其是在区域经济学、能源、环境科学、城市经济学等学科中,存在着大量复杂的时空数据。一般来说,空间数据大多存在着某种空间相关性,往往很难满足传统的面板数据模型中样本在空间维度独立的假设。为了解决这个问题,空间面板模型应运而生。通过在经典面板模型中加入空间权重矩阵,空间面板模型考虑了面板数据中存在的空间相关性和空间异质性问题,并能得到比经典面板模型更稳健、更有效的估计结果。近年来,由于空间面板模型的优势和其在实证领域的广泛应用,它已经成为计量经济学领域的研究热点之一。

空间面板数据模型同时考虑了空间相关性和时间依赖性,因此可以看作是经典面板模型、时间序列模型和横截面模型的一种结合。在理论方面,国内外许多学者都对空间面板模型进行深入研究。Elhorst^[1]在传统的面板模型中加入了空间误差自相关项和空间自相关项,并提供了相应的估计方法和估计量的渐近性质。Kapoor 等^[2]提出了用广义矩估计方法来估计误差项同时存在空间自相关和时间依赖性的空间面板模型。Lee 和 Yu^[3]提出了使用拟极大似然估计方法来对空间面板固定效应模型进行参数估计。在此基础上,Yu 等^[4]进一步考虑了动态空间面板模型的拟极大似然估计。

本文对空间面板模型的基本原理和估计方法进行详述,并进一步通过实证分析 Stata 和 R 命令的实现对各种空间面板模型的具体操作进行全面介绍,旨在为相关研究者提供借鉴。

1 空间面板模型的原理

1.1 空间面板模型的介绍

一般的面板固定效应模型为

$$y_{it} = x_{it}\beta + \alpha_i + u_t + \epsilon_{it}$$

式中, y_{it} 为因变量; x_{it} 为自变量; ϵ_{it} 为模型误差项; β 为模型要估计的参数; $i = 1, \dots, N$ 和 $t = 1, \dots, T$ 分别为截面和时间维度; u_t 为时间效应,表示随时间改变的因素; α_i 为个体效应,用于控制那些不随时间改变的

^① 基金项目:国家自然科学基金(71971131),山西省回国留学人员科研资助项目(2021-021)。

作者简介:杨威(1982—),男,山西大同人,博士,山西大学管理与决策研究所,副教授,博士生导师,研究方向:复杂数据建模与分析、金融工程与风险管理,Email: yangwei@sxu.edu.cn,通讯作者;任晓航(1991—),男,湖北邯郸人,博士,中南大学商学院,副教授,研究方向:金融计量与能源金融,Email: domrxh@outlook.com。

影响因素。面板固定效应模型的主要优点在于可以通过 α_i 和 u_i 来控制不可观测的个体效应或时间效应，从而得到更稳健的估计结果。

但是面板固定效应模型的解释变量仅考虑了个体自身的影响因素，并未考虑其他地区或者个体的影响因素。因此，当在面板固定效应模型中考虑被解释变量的空间滞后效应时，就会得到空间自回归模型 (spatial autoregression model)，也称作空间滞后模型：

$$y_{it} = \lambda \sum_{j=1}^N W_{ij} y_{jt} + \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + u_i + \epsilon_{it}$$

式中， W_{ij} ($1 < i, j < N$) 为 $N \times N$ 维空间权重矩阵 \mathbf{W} 中的元素； λ (一般假设 $|\lambda| < 1$) 为空间相关系数。空间自回归模型主要用来描述空间面板数据中存在的空间相依性 (spatial dependence)。

当面板模型中加入空间滞后误差项后，将得到空间误差模型 (spatial error model, SEM)：

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + u_i + \pi_{it} \\ \pi_{it} &= \sigma \sum_{j=1}^N W_{ij}^* \pi_{jt} + \varepsilon_{it} \end{aligned}$$

式中， W_{ij}^* ($1 < i, j < N$) 为 $N \times N$ 维空间权重矩阵 \mathbf{W}^* 中的元素； σ (一般假设 $|\sigma| < 1$) 为空间相关系数； π_{it} 用来表示扰动项存在空间相关。空间误差模型主要用来描述空间面板数据中存在的空间异质性 (spatial heterogeneity)。

当同时考虑自变量和因变量的空间滞后项，面板固定效应模型将变为空间杜宾模型 (spatial Durbin model, SDM)。其向量形式如下：

$$\mathbf{y}_t = \lambda \mathbf{W}_1 \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\beta} + \mathbf{W}_2 \mathbf{x}_t \boldsymbol{\theta} + \boldsymbol{\alpha} + u_t \mathbf{1}_N + \boldsymbol{\varepsilon}_t$$

式中， \mathbf{W}_1 和 \mathbf{W}_2 为 $N \times N$ 维空间权重矩阵； $\boldsymbol{\varepsilon}_t \sim N(0, \sigma_\varepsilon^2 \mathbf{I}_N)$ ， $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]'$ ， $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_k]'$ ； $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_k]'$ ， k 为向量 \mathbf{x}_{it} 中变量的个数，在本文中，'表示矩阵转置； $\mathbf{1}_N$ 为每个元素都为 1 的 ($N \times 1$) 列向量，

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix}, \mathbf{x}_t = \begin{bmatrix} 1 & x_{21t} & \cdots & x_{k1t} \\ 1 & x_{22t} & \cdots & x_{k2t} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2Nt} & \cdots & x_{kNt} \end{bmatrix}$$

当同时考虑空间滞后误差项和因变量的空间滞后项，将得到广义空间自回归模型。其形式如下：

$$\begin{aligned} y_{it} &= \lambda \sum_{j=1}^N W_{ij} y_{jt} + \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + u_i + \pi_{it} \\ \pi_{it} &= \sigma \sum_{j=1}^N W_{ij}^* \pi_{jt} + \varepsilon_{it} \end{aligned}$$

式中， \mathbf{W} 和 \mathbf{W}^* 为 $N \times N$ 维空间权重矩阵。广义空间自回归模型可以同时考虑空间面板数据中存在的空间相依性和空间异质性。本文主要考虑的是固定效应的空间计量模型，这是因为随机效应的空间计量模型可以看作是空间自回归模型的一个特殊情况。

更进一步，当同时考虑空间滞后误差项、自变量和因变量的空间滞后项时，可以得到空间杜宾误差模型 (spatial Durbin error model, SDEM)。其向量形式如下：

$$\begin{aligned} \mathbf{y}_t &= \lambda \mathbf{W}_1 \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\beta} + \mathbf{W}_2 \mathbf{x}_t \boldsymbol{\theta} + \boldsymbol{\alpha} + u_t \mathbf{1}_N + \boldsymbol{\pi}_t \\ \boldsymbol{\pi}_t &= \sigma \mathbf{W}_3 \boldsymbol{\pi}_t + \boldsymbol{\varepsilon}_t \end{aligned}$$

式中， \mathbf{W}_1 、 \mathbf{W}_2 和 \mathbf{W}_3 为 $N \times N$ 维空间权重矩阵， $\boldsymbol{\pi}_t = [\pi_{1t}, \pi_{2t}, \dots, \pi_{Nt}]'$ 。空间杜宾误差模型是空间自回归模型、空间误差模型和空间杜宾模型的结合，比广义空间自回归模型更一般化。上述的几种模型都是静态空间模型，它们只考虑了同时期的因变量和自变量。当在空间自回归模型和空间杜宾模型中考虑因变量的滞后项时，可以得到动态空间自回归模型和动态空间杜宾模型。其向量形式分别如下。

动态空间自回归模型：

$$\mathbf{y}_t = \tau \mathbf{y}_{t-1} + \varphi \mathbf{W}_1 \mathbf{y}_{t-1} + \lambda \mathbf{W}_1 \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\beta} + \boldsymbol{\alpha} + u_t \mathbf{1}_N + \boldsymbol{\varepsilon}_t,$$

动态空间杜宾模型：

$$y_t = \tau y_{t-1} + \varphi W_1 y_{t-1} + \lambda W_1 y_t + x_t \beta + W_2 x_t \theta + \alpha + u_t 1_N + \varepsilon_t,$$

式中, $\varphi (|\varphi| < 1)$ 为空间滞后相关系数。

前文多次提到空间权重矩阵, 顾名思义, 空间权重矩阵是用来反映空间面板数据中个体的相互依赖关系的矩阵, 其形式如下:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{pmatrix}$$

式中, $w_{ij} (1 < i, j < N)$ 为空间中第 i 个体对空间中第 j 个体的影响程度。例如, 假设因变量 y 是我们关注的地区空气污染水平, 则 y_{it} 就是地区 i 在时间 t 的空气污染水平。我们已经知道, 由于可能存在空间相关性, 则某一地区的空气污染水平可能会受到其他地区空气污染水平的影响, 将空间相关性引入其中, 可以表述为

$$y_{it} = \lambda \sum_{j=1}^N W_{ij} y_{jt} + \epsilon_{it}$$

空间权重矩阵中不同的 W_{ij} 大小不同, 表示不同地区之间的影响权重不同。为了更好地理解空间权重, 通常的做法是将空间权重矩阵 W 进行行标准化, 得到空间权重每一行的和为 1。

1.2 空间权重矩阵原理简介

空间权重矩阵被用来描述不同地区的实物之间的相关程度, 主要分为空间邻接矩阵和空间距离矩阵^[5]。

1.2.1 邻接矩阵

空间相邻关系既可以是拥有共同边界也可以是拥有共同顶点。因此, 空间邻接矩阵主要分为三种类型: Bishop 邻接、Rock 邻接 [图 1 (a)] 和 Queen 邻接 [图 1 (b)]。其中, Bishop 邻接表示共顶点连接, Rock 邻接表示共邻边连接, Queen 邻接表示既是共顶点连接又是共邻边连接。一般在实证研究中, 大多选择 Queen 邻接构建空间相邻矩阵。

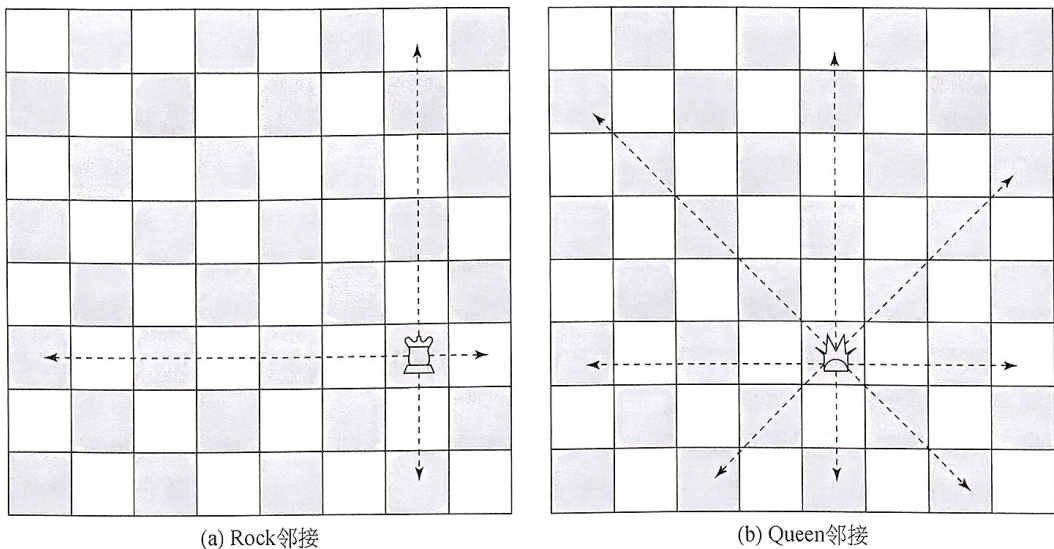


图 1 Rock 邻接和 Queen 邻接
Fig. 1 Rock adjacency and Queen adjacency

因此根据连接方式, 我们即可构建如下空间相邻矩阵 W 来反映不同事物之间的邻接关系

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{pmatrix}$$

式中,

$$w_{ij} = \begin{cases} 1, & \text{位置 } i \text{ 与位置 } j \text{ 相邻时} \\ 0, & \text{位置 } i \text{ 与位置 } j \text{ 不相邻时} \end{cases}$$

当 $i = j$ 时, $w_{ij} = 0$ 。

1.2.2 距离矩阵

地理学第一定律表示:“任何事物都是紧密相连的,只不过越相邻的事物连接更紧密”。因此,距离也是表现空间相关性的一个重要指标,距离越近两者关系越紧密,空间权重影响也越大。在空间计量经济学中,距离通常分为狭义距离和广义距离,狭义的距离通常指从地理角度测度的物理距离,而广义的距离指的是经济距离或者社会距离等多种形式的虚拟距离。在经济学实证研究中,经济距离常常被用来衡量地区经济发展的区域相关性。

(1) 物理距离权重矩阵

假设用 d_{ij} 表示地区 i 与地区 j 之间的地理距离(区域质心距离或者区域行政中心之间的距离)。那么空间距离权重矩阵中的元素 w_{ij} 可以表示为

$$w_{ij} = (d_{ij})^{-\alpha}$$

$$w_{ij} = e^{-\alpha d_{ij}}$$

式中, α 为正整数。由于空间距离满足对称性: $d_{ij} = d_{ji}$, 因此上述空间距离权重矩阵是对称矩阵。

(2) 经济距离权重矩阵

通常我们使用 GDP 或者人均 GDP 来衡量一个地区的经济发展水平, 设地区 i 与地区 j 的 GDP 分别为 y_i 和 y_j , 若定义地区 i 与地区 j 之间的经济距离为 $d_{ij} = |y_i - y_j|$, 那么空间经济距离权重矩阵中的元素 w_{ij} 可以表示为

$$w_{ij} = \frac{1}{|y_i - y_j|}$$

因此, d_{ij} 越小就表明地区 i 与地区 j 之间的经济发展水平越相似, 经济距离越近, 两者之间的空间权重系数越大。

1.2.3 空间权重矩阵的进展

由于传统空间权重矩阵是外生给定的, 只要能够给出合理的解释, 就可以人为设定矩阵形式。目前, 大多数现有文献研究都将重心放在了空间模型的参数估计和假设检验上, 关于空间权重矩阵的选择, 通常使用常见的空间邻接矩阵和空间距离矩阵。在实证研究中, 由于空间权重矩阵的选择非常敏感和关键, 因此为了确保研究的严谨性, 经常需要对比多种空间权重矩阵的估计结果, 从而得到最优选择。

近年来, 已经有部分学者开始关注空间权重矩阵的内生性问题, 即假设空间权重矩阵不再是外生给定的, 可以通过数据进行估计的^[6-8]。常用的方法是使用一些惩罚技术迫使许多权重为零, 从而可以估计出来空间权重矩阵, 如使用统计中广泛使用 LASSO 方法^[9] 和 SCAD 方法^[10]。例如, Lam 和 Souza^[11] 使用自适应 LASSO 方法同时估计空间权重矩阵和模型参数。Ahrens 和 Bhattacharjee^[12] 提出两步 LASSO 估计方法用以估计空间面板自回归模型中空间权重矩阵。

2 空间面板模型的应用案例及软件实现

为了更清晰地展示空间面板模型在实证分析中的应用, 本节将给出一个具体案例, 并提供空间面板模型的 STATA 和 R 的命令实现。

2.1 空间面板模型的应用案例

近年来, 空气污染已成为国际社会面临的最严重问题之一, 控制和减少空气污染得到了全球各国政府和人民的高度重视。作为最大的发展中国家, 我国存在严重的雾霾问题, 对居民的健康和生命造成严重的威胁。作为雾霾污染的主要组成部分, $PM_{2.5}$ 的浓度得到了政府的广泛关注。为此, 政府出台了一系列政策来降低 $PM_{2.5}$ 。为了更好地了解造成 $PM_{2.5}$ 的主要原因, 许多学者对 $PM_{2.5}$ 的影响因素进行了研究。因此, 本节将 $PM_{2.5}$ 的影响因素分析作为实际案例, 通过空间面板模型对 $PM_{2.5}$ 的影响因素进行分析。

本文选取了2004~2016年的中国269个城市作为研究对象,选取了城市人均GDP(GDP),总用电量(EL)、城镇建设用地占比(PUC)、FDI占GDP的比例(FDI)、人口密度(PD)、公共汽车数量(BUS)作为解释变量。数据均来自《中国城市统计年鉴》。在实证估计时,我们选取物理距离权重矩阵,对其进行“标准化”处理,并且对所有变量取自然对数。

首先,我们使用莫兰散点图(图1)来检验 $PM_{2.5}$ 是否存在空间相关性。从图2中我们可以看出, $PM_{2.5}$ 存在较强的空间相关性。为了更好地选择空间计量模型,我们可以应用Hausman检验和LeSage and Pace^[13]提出的检验方法对三种模型进行检验。表1中的结果表明3个检验都在1%的显著性水平下拒绝原假设,因此应该选取具有固定效应的SDM模型。表2展示了分别采用极大似然估计SAR、SEM、SDM的估计结果。从结果中,我们可以发现中国城市 $PM_{2.5}$ 污染存在显著的空间溢出效应。

表1 模型选择检验
Table 1 Model selection test

| | χ^2 | p 值 |
|----------------|-----------|-------|
| Hausman 检验 | 73.16 *** | 0.000 |
| SAR versus SDM | 52.85 *** | 0.000 |
| SEM versus SDM | 55.54 *** | 0.000 |

注:***为在1%水平统计显著。

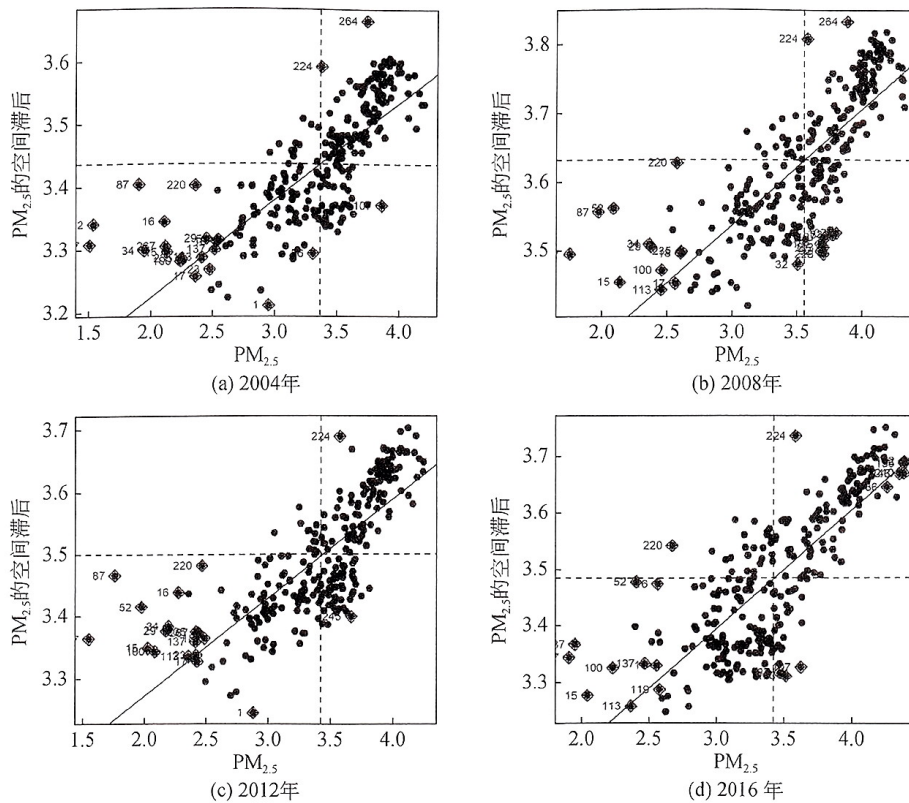


图2 2004年、2008年、2012年和2016年 $PM_{2.5}$ 污染物的莫兰散点图

Fig. 2 Moran scatter plot of $PM_{2.5}$ pollutants in 2004, 2008, 2012 and 2016

表2 估计结果
Table 2 Estimated result

| 变量 | (1) | (2) | (3) |
|-----|-------------|-----------|-----------|
| | SAR | SEM | SDM |
| GDP | 0.007 66 ** | -0.005 48 | -0.005 73 |
| | (0.003 33) | (0.010 0) | (0.010 1) |

续表

| 变量 | (1) | (2) | (3) |
|---------------------|--------------|------------------------|-------------|
| | SAR | SEM | SDM |
| EL | 0.009 05 * | 0.008 76 * | 0.009 12 * |
| | (0.004 80) | (0.004 84) | (0.004 81) |
| PUC | -0.001 47 | 0.001 56 | -0.000 470 |
| | (0.003 89) | (0.003 90) | (0.003 88) |
| FDI | 0.000 538 | -1.12×10 ⁻⁵ | -0.001 44 |
| | (0.001 89) | (0.002 00) | (0.002 02) |
| PD | -0.029 5 *** | -0.029 0 *** | -0.026 1 ** |
| | (0.010 2) | (0.010 5) | (0.010 4) |
| BUS | -0.002 84 | -0.003 22 | -0.002 56 |
| | (0.004 91) | (0.004 97) | (0.004 95) |
| W×GDP | | | 0.094 0 *** |
| | | | (0.022 9) |
| W×EL | | | 0.069 8 |
| | | | (0.045 0) |
| W×PUC | | | -0.181 *** |
| | | | (0.028 9) |
| W×FDI | | | 0.034 9 *** |
| | | | (0.010 4) |
| W×PD | | | -0.006 11 |
| | | | (0.067 2) |
| W×BUS | | | -0.061 8 |
| | | | (0.041 5) |
| W×PM _{2.5} | 0.985 *** | | 0.983 *** |
| | (0.004 16) | | (0.004 68) |
| σ | | 0.985 *** | |
| | | (0.004 19) | |
| 观测数量 | 3 497 | 3 497 | 3 497 |
| 城市数量 | 269 | 269 | 269 |

注：括号内数值为标准误，***、**、*分别为在1%水平、5%水平、10%水平统计显著。

2.2 STATA 命令实现步骤

本节主要介绍 Federico Belotti 等开发的 `xsmle` 命令。`xsmle` 的使用需要安装命令，在联网情况下选择 `help→search`，在弹出的命令窗口中输入 `xsmle`，即可搜索并安装所需命令。

* 空间矩阵生成

```
import excel " D: \location.xlsx", sheet (" Sheet1") firstrow
spmat idistance W longitude latitude, id (id) normalize (row)
spmat save W using W.spmat
spmat export W using W.txt, noid
```

* 面板数据导入 `import excel " D: \data.xlsx", sheet (" Sheet1") firstrow`
`xtset id t`

```

* 空间矩阵导入
spmat import W using W.txt, noid replace

* SAR 模型命令
xsmle y x1 x2 x3, wmatrix (W) model (sar) fe type (ind) nolog hausman
* SEM 模型命令
xsmle y x1 x2 x3, ematrix (W) model (sem) fe
* SDM 模型命令
xsmle y x1 x2 x3, wmat (W) model (sdm) fe type (ind) nolog

* 模型检验命令
test [Wx] x1 = [Wx] x2 = [Wx] x3 = 0
testnl ( [Wx] x1 = - [Spatial] rho * [Main] x1) ( [Wx] x2 = - [Spatial] rho *
[Main] x2)

```

2.3 R 命令实现步骤

本节主要介绍 R 的 `splm` 包。首先需要安装 `splm` 包，在联网情况下输入 `install.package ("splm")`，即可安装 `splm` 包。

```

* 载入数据
Library ("splm")
data ("Produc", package = "Ecdat")
* 空间矩阵生成
data ("usaww")
library ("spdep")
usalw <- mat2listw (usaww) #生成 listw 格式矩阵

* SAR 模型 ML 估计命令
随机效应模型
sararremod <- spml (formula = fm, data = Produc, index = NULL, listw = usalw,
model = "random", lag = TRUE, spatial.error = "b")
summary (sararremod)
固定效应模型
sararfemod <- spml (formula = fm, data = Produc, index = NULL, listw = usalw,
lag = TRUE, spatial.error = "b", model = "within", effect = "individual",
method = "eigen", na.action = na.fail, quiet = TRUE, zero.policy = NULL,
interval = NULL, tol.solve = 1e-10, control = list (), legacy = FALSE)
summary (sararfemod)

* SAR 模型 GM 估计命令
随机效应模型
GM_error <- spgm (formula = fm, data = Produc, listw = usaww, moments = "full-
weights", model = "random", spatial.error = TRUE)
summary (GM_error)

```

固定效应模型

```
GM_error <- spgm (formula = fm, data = Produc, lag = TRUE, listw = usaww, model
= " within", spatial.error = TRUE)
summary (GM_error)
Hausman 检验
test1 <- sphtest (x = fm, data = Produc, listw = mat2listw (usaww), spatial.
model = " error", method = " GM")
test1
```

3 结论和展望

近年来，空间面板模型在理论方法方面取得了巨大的发展，在实证分析方面也得到了广泛的应用，形成了丰富的研究成果。现有研究表明空间面板模型是经济、区域、城市、环境、能源等领域分析和解决问题的有力工具。在实证研究领域，随着空间计量模型的不断改进，其解决经济社会实际问题的能力也将不断提高，经济社会中存在的复杂空间关系也将会被进一步揭示。随着大数据时代的到来，时空数据量级不断增加，为了使空间计量模型能够更好地估计复杂数据，计算技术需要得到进一步改进。随着空间计量理论和研究方法的进一步成熟，空间计量模型势必会被更广泛地应用到各个研究领域。

参考文献：

- [1] Elhorst J P. Specification and estimation of spatial panel data models [J]. *International Regional Science Review*, 2003, 26 (3): 244-268.
- [2] Kapoor M, Kelejian H H, Prucha I R. Panel data models with spatially correlated error components [J]. *Journal of Econometrics*, 2007, 140 (1): 97-130.
- [3] Lee L, Yu J. Estimation of spatial autoregressive panel data models with fixed effects [J]. *Journal of Econometrics*, 2010, 154 (2): 165-185.
- [4] Yu J, De Jong R, Lee L. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large [J]. *Journal of Econometrics*, 2008, 146 (1): 118-134.
- [5] LeSage J, Pace R K. *Introduction to Spatial Econometrics* [M]. Florida: Chapman and Hall, 2009.
- [6] Bhattacharjee A, Holly S. Understanding interactions in social networks and committees [J]. *Spatial Economic Analysis*, 2013, 8 (1): 23-53.
- [7] Kelejian H H, Piras G. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes [J]. *Regional Science and Urban Economics*, 2014, 46: 140-149.
- [8] Qu X, Lee L f. Estimating a spatial autoregressive model with an endogenous spatial weight matrix [J]. *Journal of Econometrics*, 2015, 184 (2): 209-232.
- [9] Tibshirani R. Regression shrinkage and selection via the lasso [J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 58 (1): 267-288.
- [10] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *Journal of the American statistical Association*, 2001, 96 (456): 1348-1360.
- [11] Lam C, Souza P C L. Regularization for spatial panel time series using the adaptive lasso [M]. LSE: STICERD, 2014.
- [12] Ahrens A, Bhattacharjee A. Two-step lasso estimation of the spatial weights matrix [J]. *Econometrics*, 2015, 3 (1): 128-155.

Principles, Methodology and Applications of Spatialpanel Models

*Yang Wei*¹, *Ren Xiaohang*²

1. Institute of Management and Decision, Shanxi University, Taiyuan 030006, China

2. Business School, Central South University, Changsha 410083, China

Abstract: In recent years, with the rapid development of computer technology, the collection and storage of spatial data have become more accessible, but the correlation between spatial data cannot be ignored. In order to better consider spatial correlation and heterogeneity, spatial panel models came into being. This paper first introduces the basic principles of spatial panel models and then describes the appropriate estimation methods and model selection. Finally, we detail the specific steps of how to implement the spatial panel models in Stata and R through examples.

Key words: Spatial Panel Models; Stata; R