

# SAR 模型在省域和县域农民收入中的应用研究

戴平生 陈建宝

**内容提要:**本文利用各省财政农业人均支出、农村人均用电量作为解释变量,建立了农民人均收入空间自回归(SAR)模型。模型数据分析表明:我国农民人均收入存在明显的空间自相关现象,它反映了省际农民收入存在聚集效应;农民人均收入对财政农业人均支出、农村人均用电量存在明显的空间依赖性;财政农业支出对农民增收产生积极的正面影响,农村用电量是衡量农民收入水平的重要有效指标。考虑到省际间差异可能太大,文章还对福建省 67 个县市进行了同样的问题研究,并得出了相似的结论。

**关键词:**农民收入;财政农业支出;农村用电量;空间自回归模型

**中图分类号:**C812 **文献标识码:**A **文章编号:**1002-4565(2007)09-0048-05

## Application of SAR Models on Farmer's Income of Inter-provinces and Inter-counties

Dai Pingsheng & Chen Jianbao

**Abstract:** Using the government expenditure for agriculture and rural electricity consumption as explanatory variables, this paper establishes spatial autoregressive models (SAR) for farmer's income per capita. The results indicates the following facts: the phenomena of spatial auto-correlation obviously exist in farmer's income, which shows that inter-provincial farmer's income exists clustering effect; farmer's income has significant spatial dependence on government expenditure for agriculture and rural electricity consumption; government expenditure for agriculture has positive effect on farmer's income and rural electricity consumption is an important index to reflect farmer's income level. Considering the big differences among provinces, the authors even study the same problem for 67 counties of Fujian Province and draws the similar results.

**Key words:** Farmer's income; Government expenditure for agriculture; Rural electricity consumption; Spatial autoregressive model

近年来农民收入问题一直受到政府和社会各界的高度关注,不少学者从农民的收入结构、影响因素和地区差异等方面展开了讨论,并尝试在定性分析的基础上建立时间序列模型和利用各种指数进行定量分析,取得了不少好的结果。如宋元梁、肖卫东利用协整理论并结合 Granger 因果关系研究了改革开放以来我国城镇化水平与农民收入的数量关系,得出了短期看农民收入的增长促进了城镇化水平的提高,长期看农民收入与城镇化互为因果的结论<sup>[1]</sup>。本文利用空间计量经济学(Spatial econometrics)中的空间自回归模型,通过省域和县域截面数据对农民收入的空间异质性和空间相关性问题展开讨论。空

间计量经济学是在 Cliff 和 Ord 开拓性工作的基础上建立起来的一个计量经济学分支<sup>[2]</sup>,Anselin 经过系统的研究,给出了无异方差空间自回归模型中所有参数的具体求解方法,及提供了相关的计算公式和检验统计量<sup>[3]</sup>;而 Lesage 利用 Bayesian 方法也彻底解决了有异方差空间自回归模型的参数求解问题<sup>[4]</sup>。由于他们出色的贡献,空间计量经济学方法,在许多领域都得到了广泛的应用。国内的部分学者近年来也尝试把空间计量方法引入一些经济指标的区域特征研究<sup>[5,6]</sup>,如李序颖、陈宏民通过建立居民收入与城市经济水平的空间自回归模型,分析了江、浙、沪地区居民收入的空间相关性和居民收入对城

市经济水平的空间依赖性。但这些研究基本上只有一个解释变量,没有涉及农民收入问题。

农民收入空间自回归模型中引入国家财政农业支出、农村用电量两个解释变量,使用 2004 年 31 个省、市、自治区的数据(不含港、澳、台地区)和 2005 年福建省 67 个县的数据。考虑到财政农业支出能够对农民增收起到促进作用,这是国家制定农业政策的一个基本出发点。自 2004 年起,国家财政在全国范围开始对农业实施直补,直补使财政农业支出与农民收入直接联系起来。林伯强通过对 1990 年代的数据分析,利用联立方程模型验证了公共投资是中国农村经济增长的重要源泉,说明了财政农业支出对增加农民收入的促进作用<sup>[7]</sup>。2004 年也是国家在全国范围实施农村电网改造后的第一年,考虑到农村用电量作为解释变量,主要因为乡镇企业等非农收入已成为农民收入的主要来源,电力的作用是不言而喻的。林伯强通过协整分析发现中国的电力消费是经济增长的原因<sup>[8]</sup>,希望农村用电量能够显著地反映农民的收入水平。

### 一、空间自回归模型及相关性检验

同一时点的截面数据,可以看作是来自不同的空间位置。通常认为相邻区域比较“相似”,即所谓的“空间依赖”(Spatial dependence)。人们引入空间相邻矩阵(Spatial contiguity matrix)来反映截面数据的空间结构,通过空间加权来研究截面数据的空间相关性。对截面数据进行传统的线性回归分析,以 Gauss-Markov 假设为前提,即古典线性回归模型应满足:变量间客观上存在着线性关系、解释变量间不存在线性关系且非随机的、扰动项独立同正态分布。但类似于时间序列数据,截面数据的线性回归也同样存在扰动项空间自相关和异方差问题。

空间自回归模型的一般形式:

$$\begin{cases} Y = W_1 Y + X + \mu \\ \mu = W_2 \mu + \epsilon \\ \epsilon \sim N(0, \sigma^2 I) \end{cases} \quad (1)$$

其中 Y、X 分别为截面数据变量, W<sub>1</sub> 和 W<sub>2</sub> 为空间加权矩阵(通常进行标准化,即每行相加等于 1), μ 为扰动项, W<sub>1</sub>、W<sub>2</sub> 分别为 Y 和 μ 的空间自相关系数, I 为对角矩阵。

如果截面数据不存在空间自相关(W<sub>1</sub> = 0, W<sub>2</sub> = 0),且没有异方差(σ<sup>2</sup> = I 即对角阵为单位矩阵),上

面的模型就回到古典线性回归模型。在参数估计上空间自回归模型采用极大似然法(ML),这与古典线性回归模型的传统最小二乘法(LS)不同。对于截面数据有无异方差,可以采用 Breusch-Pagan 方法对 Y、X 的线性回归结果进行异方差检验。对无异方差空间自回归模型根据截面数据的具体适用情形,可以分为以下几类:

#### 1. 一阶空间自回归模型(FAR)

$$Y = WY + X + \mu, \quad \mu \sim N(0, \sigma^2 I) \quad (2)$$

这一模型主要用于检验截面数据是否存在空间自相关,也是研究参数估计性质的基础。

#### 2. 误差空间自回归模型(SEM)

$$\begin{cases} Y = X + \mu \\ \mu = W\mu + \epsilon \\ \epsilon \sim N(0, \sigma^2 I) \end{cases} \quad (3)$$

对普通线性回归方程的残差进行空间自相关检验,如果存在空间自相关,就可以选择该模型。检验误差空间自相关的方法通常有以下几种: Moran 统计量、Wald 值、似然比(LR)和拉格朗日乘数(LM),其中 LM 统计量的计算公式为:

$$I = \frac{e' W e}{e' e} \quad (4)$$

这里 e 表示残差, W 为已标准化的空间结构矩阵。

#### 3. 空间 Durbin 模型(SDM)

$$Y = WY + X_1 + WX_2 + \mu, \quad \mu \sim N(0, \sigma^2 I) \quad (5)$$

这一模型实际上是 SEM 通过广义差分的变化形式。

#### 4. 空间自回归模型(SAR)

$$Y = WY + X + \mu, \quad \mu \sim N(0, \sigma^2 I) \quad (6)$$

变量 Y 如果存在空间自相关(只要模型 FAR 中的系数 λ 经统计检验显著),选择这一模型就是很自然的事情。

#### 5. 一般空间自回归模型(SAC)

如果式(1)中各项系数都不等于 0(无异方差 σ<sup>2</sup> = I),就把此模型称为一般空间自回归模型(记 SAC),实际上它是模型 SAR 的一个自然发展。如果方程式中的残差也存在空间自相关,那么就考虑选择模型 SAC。通常用 LM 法检验模型 SAR 的残差是否存在空间自相关。

## 二、省域农民收入问题探讨

### (一)数据来源和处理

本文数据取自农业部中国农业信息网公布的统计资料,集中了 2004 年 31 个省、市、自治区农民人

均纯收入、农村人口数、财政农业支出、农村用电量等四组数字。把财政农业支出及农村用电量分别除以农村人口数得到两组人均数据:财政农业人均支出和农村人均用电量,希望用这两组数据帮助解释农民人均收入。另一方面,由于区域经济发展的聚集特性,农民收入在空间上还可能存在一定的相关性。为了建立我国农民收入的空间自回归模型,首先必须考虑截面数据的异方差问题。利用 Breusch-Pagan 异方差检验(原假设  $H_0$ :无异方差),笔者发现直接利用财政农业人均支出、农村人均用电量对农民人均收入进行线性回归,存在严重的异方差问题,说明农民收入具有空间异质性特征。笔者对三组数据取自然对数,截面数据通过了异方差检验,表 1 为异方差检验的结果。

表 1 数据的 Breusch-Pagan 异方差检验

数据	LM 统计量	<sup>2</sup> 概率
人均数原值	13.53	0.0012
自然对数值	0.4595	0.7947

这说明截面数据取自然对数后消除了异方差,同时由于模型中的回归系数都成了弹性系数也有利于方程的解释。先考虑消除异方差,是因为在空间自回归模型中如果存在异方差,参数估计就会变得更为复杂,通常要使用 Bayesian 空间统计方法,其中涉及 Markov Chain Monte Carlo 模拟(简称为 MCMC)的 Gibbs Sampling 技术<sup>[4]</sup>。

## (二) 模型及参数估计结果

在模型的确定过程中,一方面从传统的线性回归入手,寻找建立空间自回归模型的途径,即检验残差是否存在空间自相关,如果存在自相关,就可以按误差空间自回归模型(SEM)建模,并获得相关的参数估计。另一方面,可以预先检验各组数据是否存在空间自相关,考虑建立相应的空间自回归模型(SAR),并通过进一步的残差空间自相关检验,决定是否建立更一般的空间模型(SAC)。

利用 OLS 方法进行回归分析可以发现:财政农业人均支出(所有变量为自然对数值,为了陈述方便略去了全称,下同)与农民人均收入存在显著的相关性,农村人均用电量与农民人均收入也存在显著的相关性,但相对而言农村人均用电量具有更强的解释能力。当农民人均收入( $Y$ )同时对财政农业人均支出( $X_1$ )和农村人均用电量( $X_2$ )进行线性回归时,财政农业人均支出的显著性略有下降,回归结果列

于表 2 的模型 1(记  $M_1$ )。说明两个解释变量存在一定的相关性,在解释上出现重叠。

对模型  $M_1$  的残差进行空间自相关检验,几种方法检验的结果发现残差存在高度显著的自相关(参见表 3),这表明不同省份的农民人均收入并非处于完全的随机状态。因此必须建立误差空间自回归模型(SEM),参数估计的结果列于表 2 的模型  $M_2$  中。在  $M_2$  中  $\rho$  值高度显著,这与表 3 回归残差空间自相关检验的结果十分吻合。比较模型  $M_1$  和  $M_2$ ,发现 SEM 的回归系数显著性水平有了提高,反映模型拟合优度的  $R^2$  从 0.7522 提高到 0.8606,表明模型有了较大的改善。这里需要说明的是,尽管两个模型估计参数的方法有所不同(前者用的是最小二乘法,后者用的是极大似然法),但仍然可以用拟合优度作为模型评价标准。模型  $M_2$  表明,如果财政农业投入增加 1 倍,农民收入就倾向于增加 14.54%;农村用电量往上翻一番,就意味着农民收入增加了 17.56%;同时农民收入不可解释的部分,与相邻省份有 71.90% 的趋同性。把模型  $M_2$ :

$$\begin{cases} Y = 6.3437 + 0.1454 X_1 + 0.1756 X_2 \\ \mu = 0.7190 W\mu \end{cases}$$

经广义差分变形为

$$Y = 0.7190 WY + X - 0.7190 WX$$

其中  $W$  表示省份间的相邻结构,通常以相邻省份取 1,其他情况为 0 的矩阵为基础进行标准化,这里取  $W_1 = W_2 = W$ 。也就是说,农民收入不仅可以用所在省份的财政农业支出和农村用电量来解释,而且还可以用相邻省份的农民收入和财政农业支出和农村用电量来解释。这说明了相邻省份间社会经济生活的密切联系,以及彼此间的相互渗透和影响。

同时,由于 SEM 模型  $M_2$  中的  $\rho$  值显著不为 0,说明各省的农民人均收入与本省的财政农业人均支出和农村人均用电量存在明显的空间依赖性。

由于区域经济在一定空间范围内的聚集效应,相邻省份间的某些经济指标会产生较大的相似性,表现为经济数据的空间自相关。利用各省的农民人均收入、财政农业人均支出和农村人均用电量数据,笔者分别建立了三个一阶空间自回归模型(FAR),即表 2 中的模型  $M_3$ 、 $M_4$  和  $M_5$ 。从模型  $M_3$  中  $\rho = 0$ ,发现农民人均收入具有显著的空间自相关,即相邻省份的农民人均收入是高度相关的。而且  $\rho$  值大

于 0,表明相邻省份的农民收入具有同方向变化特征,即相邻省份的农民收入相互影响,具有同增同减的变化趋势。但相对而言,财政农业人均支出和农村人均用电量的空间自相关会弱一些。

表 2 模型的参数估计

		0	1	2		R <sup>2</sup>
M <sub>1</sub>	普通回归模型	6.04 (0.0000)	0.0978 (0.0980)	0.2661 (0.0000)		0.7522
M <sub>2</sub>	误差空间自回归模型	6.34 (0.0000)	0.1454 (0.0009)	0.1756 (0.0000)	0.7190 (0.0000)	0.8606
M <sub>3</sub>	Y 的一阶空间自回归模型				0.7770 (0.0000)	0.6077
M <sub>4</sub>	X <sub>1</sub> 的一阶空间自回归模型				0.4890 (0.0545)	0.1895
M <sub>5</sub>	X <sub>2</sub> 的一阶空间自回归模型				0.5220 (0.0334)	0.2719
M <sub>6</sub>	空间自回归模型	2.35 (0.0030)	0.0868 (0.0278)	0.1849 (0.0000)	0.5270 (0.0000)	0.8308
M <sub>7</sub>	一般空间模型(有截距)	1.94 (0.0318)	0.0789 (0.0351)	0.1757 (0.0000)	0.5900 (0.0000)	-0.1656 (0.6126)
M <sub>8</sub>	一般空间模型(无截距)		0.0846 (0.0040)	0.1078 (0.0000)	0.8780 (0.0000)	-0.6980 (0.0050)

注:括号内为参数估计显著性检验统计量的 p 值。

表 3 普通回归残差的空间自相关检验

	Moran 统计量	LR	Wald 值	LM
指标值	3.25	12.92	18.54	5.87
p 值	0.0011	0.0003	0.0000	0.0154

农民人均收入作为被解释变量存在着高度的自相关,因此必须考虑建立空间自回归模型(SAR),参数估计的结果列于表 2 的模型 M<sub>6</sub> 中。发现 值高度显著,且大于 0。但与 SEM 模型 M<sub>2</sub> 相比,R<sup>2</sup> 略有下降。从模型 M<sub>6</sub> (SAR) 可以看出,农民收入不仅可以用本省的财政农业支出和农村用电量来解释,而且还可以用相邻省份的农民收入来解释。对于 SAR 模型的残差是否存在空间自相关,仍然可以使用 W 作为空间加权矩阵进行检验。对模型 M<sub>6</sub> 的检验结果参见表 4。

表 4 自回归模型(SAR)残差的空间自相关 LM 检验

	LM	p 值	<sup>2</sup> (1)的 1%临界值
指标值	26.31	0.0000	6.63

LM 检验表明模型 M<sub>6</sub> 的残差存在高度显著的空间自相关。应该考虑使用一般空间模型(SAC),参数估计的结果列于表 2 的模型 M<sub>7</sub> 中,但笔者发现一个十分意外的情形,反映残差空间自相关的系数统计检验十分不显著,这与上面的检验结果不符合,同时模型 M<sub>7</sub> 中的截距也不显著。于是消除截距项重新估计参数得到了表 2 中模型 M<sub>8</sub> 的结果,这时系数 的 p 值几乎高度显著,R<sup>2</sup> 也从 M<sub>7</sub> 的

0.8791 提高到 0.8992,而且明显高于模型 M<sub>2</sub> (SEM) 和 M<sub>6</sub> (SAR),更为重要的是模型中所有参数都是显著的。利用广义差分可以把 M<sub>8</sub> :

$$\begin{cases} Y = 0.8780 WY + X \\ \mu = -0.6980 W\mu \end{cases}$$

变形为

$$Y = X + 0.6980 WX + 0.1800 WY + 0.6128 W^2 Y$$

其中 W<sup>2</sup> 表示相邻省份的相邻省份,是一个二阶相邻概念,即虽然两个省份不直接相邻,但通过与彼此相邻的省份联系在一起。上式实际上相当于时间序列分析中的二阶自回归,因此可以称之为 Y 的二阶空间自回归(混合)模型。由于变形后各项的系数全都大于 0,表明了各种作用完全综合之后,等式右侧所有变量与左侧变量的变化方向是一致的,产生正向影响。即由于区域经济的聚集效应,一个省份农民收入的提高会受到周边省份的推动,同时相邻省份的财政农业支出、农村用电量的增加具有强化提高整体农民收入的倾向。

表 2 中除去模型 M<sub>3</sub>、M<sub>4</sub> 和 M<sub>5</sub>,其他五个模型都能在不同程度上反映农民人均收入与财政农业支出和农村用电量的数量关系,但考虑到参数估计的显著性以及模型对数据的拟合效果,选择无截距一般空间模型 M<sub>8</sub> 作为最终建模的结果。从表 2 的分析结果可以发现,由于存在空间自相关,普通回归模型 M<sub>1</sub> 的 R<sup>2</sup> 为 0.7522,在这 5 个反映农民收入的数量关系中最小的一个。

对比普通回归模型 M<sub>1</sub> 与无截距一般空间模型 M<sub>8</sub> 发现,由于忽视了截面数据的自相关,模型 M<sub>1</sub> 在省际间农民人均收入的分析中过分夸大了农村人均用电量的解释作用(系数为 0.2661,而后者仅为 0.1078),淡化了财政农业支出对农民收入影响(在模型引入空间自相关后,系数的显著性水平迅速上升)。同时,模型 M<sub>8</sub> 能够较大程度地提高回归方程的解释能力(R<sup>2</sup> 从 0.7522 提高到 0.8992)。知道模型的 R<sup>2</sup> 是解释变量的回归变差(RSS)与被解释变量总变差(TSS)的比值,这里的总变差实际上反映的是省际间农民收入的差异性。在这里结合考虑数据的空间自相关,提高了模型解释总变差的能力,也就是说借助省际间农民收入的自相关,极大地增强了解释彼此间收入差异的能力。

### 三、县域农民收入问题探讨

笔者对 2005 年福建省的县域数据(67 个县市)

进行了类似分析。其中财政农业支出取自《2005年全国地市县财政统计资料》(北京:中国财政出版社),农民人均纯收入、农村用电量及农村人口数取自《2005年福建经济与社会统计年鉴——农村篇》(福州:福建人民出版社)。

数据直接通过了 Breusch Pagan 异方差检验, LM 统计量为 0.0769, 卡方概率等于 0.9623, 说明不存在异方差问题。与省域数据相比, 这 67 个县市的县域数据差异要小得多(省域数据通过自然对数变换才通过异方差检验)。模型估计的结果列于表 5。从误差自回归模型  $M_{10}$  来看: 财政农业支出每增加 1 元, 农民收入可增加 4.49 元, 说明了财政支出的杠杆作用十分明显; 农村用电量每增加 1 千瓦时, 农民收入可增加 0.2801 元, 说明了农村用电量在农民增收方面的重要作用, 这也进一步验证了有关学者的研究结果<sup>[8]</sup>。在 8 个方程中, 仍是无截距空间自回归模型  $M_{16}$  最优, 拟合优度从普通回归方程<sub>9</sub> 的 0.4099 上升到 0.8692, 这表明模型有了十分明显的, 这充分说明在建模过程中, 如果数据存在空间相关性, 选择空间模型建模更能准确地反映变量间的数量关系。

表 5 模型的参数估计

		0	1	2			R <sup>2</sup>
$M_9$	普通回归模型	3992.93 (0.0000)	3.47 (0.0682)	0.6263 (0.0000)			0.4099
$M_{10}$	误差空间自回归模型	4188.48 (0.0000)	4.49 (0.0009)	0.2801 (0.0005)		0.7350 (0.0000)	0.6475
$M_{11}$	$Y$ 的一阶空间自回归模型				0.7350 (0.0000)		0.5008
$M_{12}$	$X_1$ 的一阶空间自回归模型				0.0350 (0.8907)		0.0009
$M_{13}$	$X_2$ 的一阶空间自回归模型				0.0520 (0.2728)		0.0515
$M_{14}$	空间自回归模型	1057.95 (0.0082)	3.99 (0.0025)	0.4067 (0.0000)	0.6680 (0.0000)		0.5828
$M_{15}$	一般空间模型(有截距)	59.47 (0.5768)	1.65 (0.0110)	0.3383 (0.0000)	0.9360 (0.0000)	-1.95 (0.0000)	0.8668
$M_{16}$	一般空间模型(无截距)		1.70 (0.0058)	0.1078 (0.0000)	0.8780 (0.0000)	-0.6980 (0.0050)	0.8692

注: 括号内为参数估计显著性检验统计量的 p 值。

## 四、结论

由于区域经济发展的不平衡, 农民收入存在明显的空间差异。数据分析表明, 我国省域农民收入与国家地方财政农业支出、农村用电量存在明显的空间依赖性, 同时农民收入本身也呈现出以空间自

相关为特征的空间聚集。利用空间计量模型, 可以更好地解释省际间农民收入的差异性和空间聚集特点。考虑到省际间差异可能太大, 还对福建省 67 个县市进行了同样的问题研究, 并得出了相似的结论。推测相似的结论也可能对全国县市成立, 只所以没在文章中进行模型分析, 原因在于一方面其他省份县市的有关数据获得比较困难, 另一方面计算量太大(涉及上千阶空间相邻矩阵的确定以及相关计算)。从农民收入的空间自回归模型可以看出, 财政农业支出对农民增收产生积极的正面影响, 农村用电量是衡量农民收入水平的重要有效指标。

在研究实际经济问题的过程中, 许多具有空间聚集效应的经济指标其截面数据通常存在空间自相关关系, 只有通过建立空间自回归模型进行定量分析, 才能更科学地解释经济现象在空间演变中的变化机制和规律。

## 参考文献

- [1] 宋元梁, 肖卫东. 中国城镇化发展与农民收入增长关系的动态计量经济分析[J]. 数量经济技术经济研究, 2005(9): 30—39.
- [2] Cliff, A. and J. Ord. Spatial Processes, Models and Applications[M]. London: Pion, 1981.
- [3] Anselin L. Spatial Econometrics: Methods and Models[M]. Dordrecht: Kluwer Academic Publishers, 1988.
- [4] LeSage J. P. Bayesian estimation of spatial autoregressive models[J]. International Regional Science Review, 1997, 20: 113—129.
- [5] 邹艳芬, 陆宇海. 基于空间自回归模型的中国能源利用效率区域特征分析[J]. 统计研究, 2005(10): 67—71.
- [6] 李序颖, 陈宏民. 居民收入与城市经济水平的空间自回归模型[J]. 系统工程理论方法应用, 2005(5): 395—399.
- [7] 林伯强. 中国的政策公共支出与减贫政策[J]. 经济研究, 2005(1): 11—11.
- [8] 林伯强. 电力消费与中国经济增长: 基于生产函数的研究[J]. 管理世界, 2003(11): 27—37.

## 作者简介

戴平生, 男, 广东兴宁人。2004年毕业于厦门大学经济学院, 获经济学博士, 现为厦门大学计划统计系讲师, 研究方向: 经济统计分析。

陈建宝, 男, 云南曲靖人。澳大利亚科庭理工大学理学博士, 厦门大学计划统计系教授、博士生导师, 中国统计教育常务理事。研究方向: 空间统计学, 计量经济学。

(责任编辑: 李峻浩)