

Research on Spatial Regression Model with Unknown Distribution Residuals

Dexiong Teng*, Huiguo Zhang, Xijian Hu#

College of Mathematics and System Sciences, Xinjiang University, Urumqi Xinjiang
Email: #xijianhu@126.com

Received: May 26th, 2016; accepted: Jun. 13th, 2016; published: Jun. 17th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper analyzed the profile likelihood method, mainly studied the estimation of regression coefficients under the condition of unknown distribution of residuals of spatial regression model and made some simulation experiments comparing with two-stage least squares method. The simulation obtained the conclusions that the parameter estimated by profile likelihood method demonstrated more robust.

Keywords

Profile Likelihood Method, Unknown Distribution Residuals, Spatial Regression Model

关于空间回归模型残差项分布未知的研究

滕德雄*, 张辉国, 胡锡健#

新疆大学, 数学与系统科学学院, 新疆 乌鲁木齐
Email: #xijianhu@126.com

收稿日期: 2016年5月26日; 录用日期: 2016年6月13日; 发布日期: 2016年6月17日

摘要

文章分析了剖面似然方法, 在残差项分布未知的条件下研究了空间回归模型的回归系数的估计问题, 并

*第一作者。

#通讯作者。

在模拟实验中与二阶段最小二乘法做比较, 得出剖面似然方法所得到的参数估计更加稳健的结论。

关键词

剖面似然方法, 残差项分布未知, 空间回归模型

1. 引言

空间计量经济学的概念最早由 Paelinck [1] 提出, 此后经过众多学者研究和发展, 空间计量经济学的理论框架体系逐步形成。空间计量模型主要用于处理具有空间自相关性和空间异质性的局部化空间效应的数据集。Cleveland 和 Devlin [2], McMillen 和 McDonald [3] 指出有参数和非参数两种思路来设定空间计量模型的空间相关性结构, 在近年来的实证研究中, 很少采用非参数方法。Anselin [4] 在其著作中展示了空间权重矩阵外生设定方法的相关方面, 并着重分析了空间误差模型(Spatial Error Model)和空间滞后模型(Spatial Lag Model)。在实证研究中, 所建立的模型要比这两个模型更加复杂[5]。

经典的空间计量模型假定残差项服从正态分布, 且相互独立。在实证研究中, 这样的假设是很难满足的。因此, 在残差项分布未知情况下建立严谨可信的模型来研究相关问题是必要的。本文在残差分布未知的条件下来建立空间回归模型, 应用剖面似然方法的思想对模型进行估计, 通过计算机模拟实验来实现估计过程。

2. 空间回归模型参数估计的剖面似然方法

空间回归模型是空间计量经济学研究的主体内容, 其一般形式为:

$$y = \rho W y + X \beta + \varepsilon \quad (1)$$

其中因变量 y 是 $(N \times 1)$ 向量, 分量为地区的观测值, 外生变量 X 是 $(N \times p)$ 矩阵, 参数 β 为 $(p \times 1)$ 向量, ε 是独立同分布随机变量, 且服从均值为 0 方差为 σ^2 的正态分布, W 为 $(N \times N)$ 空间权重矩阵。模型(1)的似然函数为:

$$L(y; \rho, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} |A| \exp \left\{ -\frac{(Ay - X\beta)' (Ay - X\beta)}{2\sigma^2} \right\}$$

其中 $A = I - \rho W$, I 为单位矩阵。可得对数似然函数为:

$$\ln(y; \rho, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) + \ln |A| - \frac{(Ay - X\beta)' (Ay - X\beta)}{2\sigma^2} \quad (2)$$

对(2)式进行极大化就能得到模型(1)参数的极大似然估计, 由于参数 ρ 并不能直接计算得到, 应用剖面极大似然方法可以简化参数的估计过程。

首先, 我们不切实际的假设参数 ρ 已知, 得:

$$\ln(y, \beta, \sigma^2 | \rho) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) + \ln |A| - \frac{(Ay - X\beta)' (Ay - X\beta)}{2\sigma^2} \quad (3)$$

对(3)式直接进行极大化, 即可得到参数 β 和 σ^2 的估计:

$$\hat{\beta} = (X'X)^{-1} X'Ay \quad (4)$$

$$\hat{\sigma}^2 = \frac{(Ay - X\hat{\beta})' (Ay - X\hat{\beta})}{n} \quad (5)$$

将(4)式和(5)式代入(2)式得:

$$\ln(y, \rho | \hat{\beta}, \hat{\sigma}^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} - \frac{n}{2} \ln(\hat{\sigma}^2) + \ln|A| \quad (6)$$

最大化(6)式, 得到 ρ 的估计, 然后将 ρ 的估计带入到(4)和(5)式中, 便得到 β 和 σ 的估计。

3. 残差项分布未知的空间回归模型的估计

如果残差项的真实分布已知, 则极大似然方法能够得到参数的有效估计。在实证研究中, 残差项的真实分布往往未知。此时, 建立半参数模型可以解决残差项分布未知的问题。半参数模型一般用剖面似然方法来估计。Severini 和 Wong [6]的研究表明: 剖面似然能得到半参数的有效估计, 其基本思想是基于核回归(kernel regression)方法, 用非参数核密度估计代替未知的非参数函数。本文应用 Linton, Sperlich 和 Van Keilegom [7]给出的用于估计半参数变换模型的剖面似然方法来估计残差项分布未知的空间回归模型。

对于给定 $\delta \in \Theta, \delta = \{\rho, \beta\}$, 令 $\varepsilon_i = y_i - \rho W_i y - \sum_j x_{ij} \beta_j$, (y_i, x_i) 是来自于模型(1)的独立样本, x_i 与 ε_i 相互独立, $\text{Var}(x_i)$ 有限, $\delta \in \text{interior}\Theta$, Θ 是 R^p 中的紧集, 且 $\sqrt{n}(\delta^* - \delta) \xrightarrow{d} N(0, \Sigma^*)$, 则:

$$\int f_{\varepsilon(\delta^*)}(\mu) \log\left(f_{\varepsilon(\delta^*)}(\mu)\right) d\mu < \int f_{\varepsilon(\delta^*)}(\mu) \log\left(f_{\varepsilon(\delta^*)}(\mu)\right) d\mu \quad (7)$$

对于任意 $\delta^* \neq \delta$ 。可以得到:

$$\delta^* = \arg \max_{\delta \in \Theta} \int f_{\varepsilon(\delta^*)}(\mu) \log\left(f_{\varepsilon(\delta^*)}(\mu)\right) d\mu \quad (8)$$

其中 δ^* 存在且唯一。

令 $\varepsilon_i = y_i - \rho W_i y - \sum_j x_{ij} \beta_j$, 即可得到 $\varepsilon(\delta)$ 的密度函数估计 $\hat{f}_{\varepsilon(\delta)}(\mu)$ 。 $\hat{f}_{\varepsilon(\delta)}(\mu)$ 的表达式为:

$$\hat{f}_{\varepsilon(\delta)}(\mu) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\varepsilon_i(\delta) - \mu}{h}\right)。$$

其中, h 是一个合适的窗宽参数, K 是一个核函数。由此, 可通过下面给出的式子来估计参数 δ :

$$\hat{\delta} = \arg \max_{\delta} \int \hat{f}_{\varepsilon(\delta)}(\mu) \log\left(\hat{f}_{\varepsilon(\delta)}(\mu)\right) d\mu \quad (9)$$

上述方法能得到残差项分布未知条件下空间回归模型的有效估计, 即估计的渐进方差达到估计的有效下界。

4. 随机模拟的结果和分析

在随机模拟实验中, 目的是为了探究上述估计方法在有限样本下的表现。在模拟试验中分别产生 50 组数据, 每组数据由模型(1)生成的 200 次观测值组成。 W 是 (200×200) 的 rook 邻接的空间权重矩阵, 参数 ρ 的真实值为 0.8, 参数 β 的真实值为 $(0.5, 0.7, 0.9)$ 。考虑残差项在不同分布下, 在模拟实验中计算参数估计的平均值和标准差。

首先, 应重视对核密度估计过程窗宽参数的选择问题。下图是窗宽参数从 0.5 增大至 1.5 参数估计的标准差的变化过程。

从图 1 可以看出, 窗宽参数增大至 0.8 后, 参数估计的标准差几乎不再变化, 因此在接下来的模拟实验中, 窗宽参数定为 1.2。

在模拟试验中, 分别考虑标准正态分布, 混合正态分布, 自由度为 3 的卡方分布以及标准柯西分布, 比较二阶段最小二乘方法与文中采用的剖面似然方法所得到的参数估计的差别。从表 1 可以看出, 当残

窗宽对估计的影响

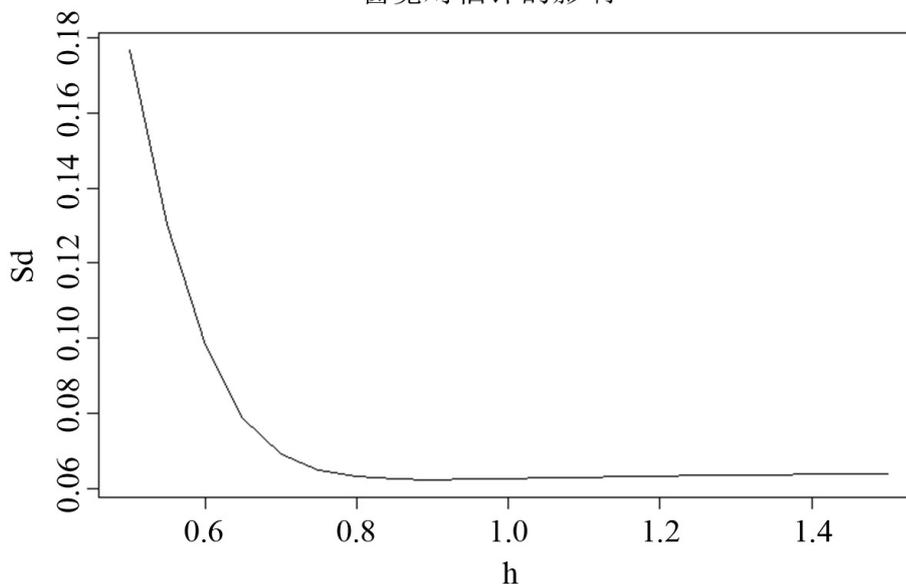


Figure 1. Change of standard deviation of parameter's estimation with bandwidth increasing
图 1. 随窗口增大参数估计的方差的变化

Table 1. Mean and standard deviation of parameter's estimation
表 1. 参数估计的平均值和标准差

| | ρ | | β_1 | | β_2 | | β_3 | |
|-------------------------|--------|--------|-----------|--------|-----------|--------|-----------|--------|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| N(0,1) | | | | | | | | |
| 2LS | 0.8013 | 0.0137 | 0.4972 | 0.0664 | 0.6897 | 0.0608 | 0.8958 | 0.0596 |
| MPL | 0.8843 | 0.0315 | 0.4789 | 0.0629 | 0.6692 | 0.0619 | 0.8546 | 0.0604 |
| 0.5N(0,0.5) + 0.5N(0,2) | | | | | | | | |
| 2LS | 0.8035 | 0.0153 | 0.4866 | 0.0953 | 0.6822 | 0.0772 | 0.8933 | 0.0780 |
| MPL | 0.8948 | 0.0324 | 0.4716 | 0.0654 | 0.6676 | 0.0600 | 0.8455 | 0.0573 |
| Chi-square with df = 3 | | | | | | | | |
| 2LS | 0.8802 | 0.0150 | 0.4900 | 0.1420 | 0.7084 | 0.1412 | 0.8576 | 0.1309 |
| MPL | 0.8965 | 0.0421 | 0.4938 | 0.0821 | 0.6808 | 0.0890 | 0.8364 | 0.0955 |
| Standard Cauchy | | | | | | | | |
| 2LS | 0.8380 | 0.1724 | 1.3667 | 5.3746 | 0.4504 | 6.9500 | 0.0935 | 7.1815 |
| MPL | 0.8085 | 0.0132 | 0.4892 | 0.1053 | 0.7187 | 0.0956 | 0.8740 | 0.1200 |

差项服从的分布与正态分布差别越大，二阶段最小二乘方法所得到的参数估计越偏离真实值，并且估计的标准差也越大；文中所采用的剖面似然方法所得到的参数估计总体来看，变化并不明显，可信度更高。

5. 结论

本文在残差项分布未知的条件下应用剖面似然方法研究空间回归模型的估计问题。剖面似然方法的基本思想是基于核回归方法用非参数核密度估计代替未知非参数函数来估计回归系数。在随机模拟实验

中，比较了二阶段最小二乘方法与本文采用的剖面似然方法，结果表明，在残差项分布未知情况下，剖面似然方法得到的估计更加稳健。

参考文献 (References)

- [1] Paclinck, J. and Klaassen, L. (1979) *Spatial Econometrics*. Saxon House, Farnborough.
- [2] Cleveland, W.S. and Devlin, S.J. (1988) Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, **83**, 596-610. <http://dx.doi.org/10.1080/01621459.1988.10478639>
- [3] McMillen, D.P. and McDonald, J.F. (2004) Locally Weighted Maximum-Likelihood Estimation: Monte Carlo Evidence and an Application. In: Anselin, L., Florax, R.J.G.M. and Rey, S.J., Eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer, Heidelberg. http://dx.doi.org/10.1007/978-3-662-05617-2_10
- [4] Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht. <http://dx.doi.org/10.1007/978-94-015-7799-1>
- [5] Parent, O. and LeSage, J.P. (2008) Using the Variance Structure of the Conditional Autoregressive Spatial Specification to Model Knowledge Spillovers. *Journal of Applied Econometrics*, **23**, 235-256. <http://dx.doi.org/10.1002/jae.981>
- [6] Severini, T.A. and Wong, W.H. (1992) Profile Likelihood and Conditionally Parametric Models. *Annals of Statistics*, **20**, 1768-1802. <http://dx.doi.org/10.1214/aos/1176348889>
- [7] Linton, O., Sperlich, S. and Keilegom, I.V. (2008) Estimation of a Semiparametric Transformation Model. *Annals of Statistics*, **36**, 686-718. <http://dx.doi.org/10.1214/009053607000000848>