DOI:10.11721/cqnuj20160519

基于样条函数的时空加权回归模型变量选择*

玄海燕[1]，宋磊勃[2]，陈金淑[2]，张运虎[2]
(1. 兰州理工大学 经济管理学院；2. 兰州理工大学 理学院，兰州 730050)

摘要：为了提高时空加权回归模型的预测精度，增强时空加权回归模型的可解释性，选择对因变量具有显著影响的重要变量已成为当今统计分析中一个重要研究课题。首先对时空加权回归模型的系数使用样条函数作出逼近，其次在最小二乘的理论基础上，根据 SCAD 惩罚理论对时空加权回归模型各变量所对应的系数进行处理，并利用 BIC 准则来选择调谐参数 λ，最终通过迭代算法来选择出对时空加权回归模型有用的变量，剔除掉影响模型准确性的变量，达到精简模型、提高预测精度的目的。

关键词：时空加权回归；变量选择；SCAD；函数逼近；BIC 准则
中图分类号：O212.7　　**文献标志码**：A　　**文章编号**：1672-6693(2016)05-0054-04

由于时空数据对许多领域(如经济学、流行病等)的影响逐步加深，人们便开始寻求方法来研究它的特性并尝试把它作为重要的影响因素考虑在模型中，以便使得模型更加完善和有用，从而增强模型的可靠性，更好地供人们参考和使用，于是出现了由 Huang 等人[1]提出的时空加权回归模型，形式如下：$Y_i=\beta_0(u_i,v_i,t_i)+\sum_{l=1}^{d}\beta_l(u_i,v_i,t_i)X_{il}+\varepsilon_i, i=1,2,\cdots,n$，其中，$(Y;X_{i1}X_{i2},\cdots,X_{id})$是因变量 Y 和自变量 $X_1,X_2,\cdots,X_d$ 在观测位置 (u_i,v_i,t_i)处的观测值；$\varepsilon_i(i=1,2,\cdots,n)$为独立的随机误差项，均值为 0，方差为 σ^2；$\beta_l(u_i,v_i,t_i)$，$l=0,1,2,\cdots,d$ 是 $d+1$ 个未知的系数函数。

同时，Huang 等人[1]也对模型系数做出了估计，并且为了验证模型的优越性，他们将模型应用于影响住房价格变化的因素分析上，模拟研究表明，时空加权回归模型(GTWR)比 TWR 模型和 GWR 模型的拟合精度都要好，所得模拟结果都要优于其他两个模型。Yan 等人[2]通过一种新的二步估计方法来估计时空加权回归模型的系数，具体做法是：第一步先固定时间因素来估计系数；第二步，通过第一步估计出的具有固定时间因素的系数再继续以时间为变量估计此系数，其结果也令人满意。继而，玄海燕等人[3]给出了模型回归关系的空间平稳性检验和时间相关性检验方法。玄海燕等人[4]基于均值漂移模型，利用两步估计法对模型进行拟合，研究异常点检验问题，并构造检验统计量。还有许多学者基于此模型用其他方法对社会现象做出讨论，并且使用其他方法对此模型做出检验和完善[5]。

近些年，人们对于变系数模型在变量选择方面的研究已经较为成熟，Li 等人[6]通过使用基函数逼近和非凹惩罚函数对变系数模型进行变量选择。Zhao 等人[7]结合具有收缩估计性质的基函数逼近理论提出了一种偏差校正的变量选择方法。Fan 等人[8]提出用似然函数来处理计算的复杂性和随机误差被忽略等问题，用这种方法可以自动进行变量选择和系数估计。Wang 等人[9]专注于通过具有不同数量参数的 LAD 回归来实现变量选择。变系数模型在变量选择方面的研究为时空加权回归模型进行变量选择提供了重要的依据。

本文主要通过函数逼近的方法和 SCAD 惩罚理论，探索时空加权回归模型中变量的关系，并且可以保留有用的变量，剔除不必要的变量，这样不仅可以减少模型的复杂程度，而且还可以消除一些变量间由于相关性问题带来的不变，同时也大大提高了模型的可信度。

1 系数函数的扩展

为方便起见，把上述模型用矩阵形式表示为：

* **收稿日期**：2015-11-02　**修回日期**：2015-12-28　**网络出版时间**：2016-07-13 14:03
资助项目：国家自然科学基金(No. 11261031)
作者简介：玄海燕，女，副教授，研究方向为应用概率统计，E-mail：haiyanxuan@msn.com
网络出版地址：http://www.cnki.net/kcms/detail/50.1165.N.20160713.1403.026.html

$$\boldsymbol{Y}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}+\varepsilon, \tag{1}$$

其中，$\boldsymbol{X}_i=(1,X_{i1}X_{i2},\cdots,X_{id})^{\mathrm{T}}$，$\boldsymbol{X}=(X_1,X_2,\cdots,X_d)^{\mathrm{T}}$，$\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_d)^{\mathrm{T}}$，$\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_d)^{\mathrm{T}}$，$\boldsymbol{X}$，$\boldsymbol{\beta}$ 为关于 u，v，t 的函数。为了能有效地对所需变量作出选择，笔者考虑以下做法。

假设 $\{Y_i;X_{i1},X_{i2},\cdots,X_{id}\}_{i=1}^{n}$ 是模型(1)的随机样本，基于 B 样条函数所具有的性质，在此使用 B 样条函数来逼近参数 $\boldsymbol{\beta}$，对于每个 $\beta_l(u,v,t)$，$l=0,1,2,\cdots,d$，有

$$\beta_l(u,v,t)\approx\sum_{k=1}^{K_l}\gamma_{lk}B_{lk}(u,v,t), \tag{2}$$

其中，$\{B_{lk}(u,v,t),\ k=1,2,\cdots,K_l\}$ 是样条函数所构成线性空间的基。记 $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_0^{\mathrm{T}},\cdots,\boldsymbol{\gamma}_d^{\mathrm{T}})^{\mathrm{T}}$，且 $\boldsymbol{\gamma}_l=(\boldsymbol{\gamma}_{l1},\cdots,\boldsymbol{\gamma}_{lK_l})^{\mathrm{T}}$，$l=0,\cdots,d$，其中样条函数均具有固定的度和节序，$B$ 样条函数节点的数目和 B 样条的阶决定了 K_l 的值，不同的 β_l 允许取不同的 K_l，在应用时，用 B 样条函数来逼近不同光滑度的 β_j 就提供了较强的适应性。根据(1)式和(2)式，有

$$Y_i\approx\sum_{l=0}^{d}\sum_{k=1}^{K_l}X_{il}B_{lk}(u_i,v_i,t_i)\gamma_{kl}+\varepsilon_i, \tag{3}$$

其中，$X_{i0}=1$，这里

$$\boldsymbol{B}(u,v,t)=\begin{pmatrix}B_{01} & \cdots & B_{0K_1} & 0 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & 0 & B_{0d_1} & \cdots & B_{dK_d}\end{pmatrix}$$

为 $(d+1)\times(\boldsymbol{K}_1+\cdots+\boldsymbol{K}_d)$ 矩阵。记 $\boldsymbol{Z}_i^{\mathrm{T}}=\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{B}(u_i,v_i,t_i)$，$\boldsymbol{Z}=(\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n)^{\mathrm{T}}$，基于以上的逼近，得到观测值 Y_i 与理论逼近值得残差表达式，记为 $e_i=Y_i-\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}$。对于任意给定的 K_l，解决以下最优问题：

$$\boldsymbol{L}=\sum_{i=1}^{n}\Big(Y_i-\sum_{l=0}^{d}\sum_{k=1}^{K_l}X_{il}B_{lk}(u_i,v_i,t_i)\gamma_{lk}\Big)^2K(\cdot),$$

用矩阵形式表示为：

$$\boldsymbol{L}=(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\gamma})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\gamma}), \tag{4}$$

其中，$\boldsymbol{W}$ 是关于核函数 $K(\cdot)$ 的对角矩阵，其形式为 $K(d/h)$，其中为 d 距离，h 为窗宽。对于核函数 $K(\cdot)$，根据文献[1]，在本文中令不同时刻不同地点的距离为

$$(d_{i_1i_2}^{ST})^2=\tau[(u_{i_1}-u_{i_2})^2+(v_{i_1}-v_{i_2})^2]+\mu(t_{i_1}-t_{i_2})^2,$$

则

$$\frac{(d_{i_1i_2}^{ST})^2}{\tau}=[(u_{i_1}-u_{i_2})^2+(v_{i_1}-v_{i_2})^2]+\frac{\mu}{\tau}(t_{i_1}-t_{i_2})^2。$$

一般情况下，为了减少参数，令 $\tau=1$，这样只要决定 μ 即可得时空距离。此时，可以使用交叉确认法(CV)来确定最优的 μ。

2 SCAD 惩罚理论

根据 SCAD 惩罚理论，可以得出较好的理论性质，如 Oracle 性质，对于所提出的理论，若一些变量和模型没有多大关系，则它们所对应的系数函数为零函数。

使 $\boldsymbol{R}_l=(r_{k_1k_2})_{K_lK_l}$ 表示元素为 $r_{k_1k_2}=\iiint B_{lk_1}(u,v,t)B_{lk_2}(u,v,t)\mathrm{d}u\mathrm{d}v\mathrm{d}t$ 的方阵。此时定义 $\|\boldsymbol{\gamma}_l\|_{R_l}^2=\boldsymbol{\gamma}_l^{\mathrm{T}}\boldsymbol{R}_l\boldsymbol{\gamma}_l$，则这里定义惩罚加权最小二乘准则为：

$$\boldsymbol{PL}(\boldsymbol{\gamma})=(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\gamma})^{\mathrm{T}}W(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\gamma})+\sum_{l=0}^{d}p_\lambda(\|\gamma_l\|_{R_l})。 \tag{5}$$

对于惩罚函数，有多种形式可选择，这里引用 Fan 等人[7]提出的 SCAD 惩罚函数，定义为：

$$p_\lambda(\theta)=\begin{cases}\lambda\theta, & 0\leqslant\theta\leqslant\lambda,\\ -\dfrac{\theta^2-2a\lambda\theta+\theta^2}{2(a-1)}, & \lambda<\theta<a\lambda,\\ \dfrac{(a+1)\lambda^2}{2}, & \theta\geqslant a\lambda。\end{cases}$$

其中,a 是一个调谐参数,这里,取 $a=3.7$,在本文中,这是一个合理的取值,若使 $\boldsymbol{PL}(\gamma)$最小化的估计值表示为 $\hat{\gamma}$,那么可以得到$\hat{\beta}_l$ 的估计值 $\hat{\beta}_l=\sum_{k=1}^{K_l}\hat{\gamma}_{lk}B_{lk}(u,v,t)$。

3 计算算法

由于惩罚函数 $p_\lambda(\|\gamma_l\|_{R_l})$的不可微性,通常的梯度优化法在这里不适用,在本节中,参考 Fan 等人的研究结果[7],给出非凸惩罚函数 $p_\lambda(\|\gamma_l\|_{R_l})$的局部二次逼近形式,进而使用迭代算法选择出迭代过程中存在值为零的系数函数,为了更容易地表示 $p_\lambda(\|\gamma_l\|_{R_l})$,在给定 $\theta_0\in\mathbf{R}^+$($\mathbf{R}^+$为正实数集)的邻域内

$$[p_\lambda(\theta)]'=p'_\lambda(\theta)\mathrm{sgn}(\theta)\approx\{p'_\lambda(\theta_0)/\theta_0\}\theta,\ \theta_0\neq 0,$$

那么有 $p_\lambda(\theta)\approx p_\lambda(\theta_0)+\frac{1}{2}\{p'_\lambda(\theta_0)/\theta_0\}(\theta^2-\theta_0^2)$。

令 $\theta=\|\gamma_l\|_{R_l}$,有:

$$p_\lambda(\|\gamma_l\|_{R_l})=p_\lambda(\|\gamma_l^{(0)}\|_{R_l})+\frac{1}{2}\{p'_\lambda(\|\gamma_l^{(0)}\|_{R_l})/\|\gamma_l^{(0)}\|_{R_l}\}(\|\gamma_l\|_{R_l}^2-\|\gamma_l^{(0)}\|_{R_l}^2)。\tag{6}$$

将(6)式带入(5)式中并去掉不相关的项,则有以下表示:

$$\boldsymbol{PL}_1(\boldsymbol{\gamma})=(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\gamma})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\gamma})+\frac{1}{2}\gamma^{\mathrm{T}}\boldsymbol{\Omega}_\lambda(\gamma^{(0)})\boldsymbol{\gamma},\tag{7}$$

其中,$\boldsymbol{\Omega}_\lambda(\gamma^{(0)})$为对角阵,它的元素为 $p'_\lambda(\|\gamma_l^{(0)}\|_{R_l})/\|\gamma_l^{(0)}\|_{R_l}\boldsymbol{R}_l$,通过对(7)式求导令其为零,可以得到以下二次形式的方程:

$$\boldsymbol{ZWY}=\left(\boldsymbol{ZWZ}^{\mathrm{T}}+\frac{\boldsymbol{\Omega}_\lambda(\boldsymbol{\gamma}^{(0)})}{2}\right)\boldsymbol{\gamma}。\tag{8}$$

根据以上讨论,笔者给出以下算法:

第 1 步 初始化,$\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(0)}$;

第 2 步 给定一个 $\boldsymbol{\gamma}^{(m)}$,通过解上面的方程替换到 $\boldsymbol{\gamma}^{(m+1)}$,这里 $\boldsymbol{\gamma}^{(0)}$用 $\boldsymbol{\gamma}^{(m)}$替换;

第 3 步 重复第 2 步直至收敛。

由于使用非凸惩罚 SCAD 函数,有时算法不会收敛,并且几乎所有基于非凸的惩罚理论都会遇到这样的问题,故在第一步中,γ 的初始估计值可使用非惩罚估计,这里使用样本个数为$\frac{n(n-1)}{2}$的伪观测值$\{Z_m-Z_n,Y_m-Y_n\}_{m<n}$,通过拟合 L_1 回归解得。在第二步迭代过程中,若$\gamma_l^{(m)}{}_{R_l}$小于$\varepsilon(>0)$。这时取$\hat{\gamma}_m=0$,即筛选了不必要的变量,如此进行,最终得到有效变量。

4 调谐参数的选取

为了完成上面所述的工作,需要对调谐参数 λ 作出合适的选择,在诸多调谐参数选取的方法中,如 GCV,AIC,BIC 等,由于 BIC 准则能一致地选择在 SCAD 下的真实模型,因此,在这里应用 BIC 来选择调谐参数 λ。这里,调谐参数选择准则形式如下:

$$BIC_\lambda=\log(S_\lambda)+\frac{\bar{k}\log(n)}{n},\tag{9}$$

其中,$\bar{k}=\sum_l K_l$ 是待估参数的个数,$S_\lambda=(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\gamma}})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\gamma}})$是残差平方和的最小值。这时,最优的调谐参数可通过最小化(9)式得到,即$\hat{\lambda}_{\mathrm{BIC}}=\arg\min_{\lambda\in\mathbf{R}^+}BIC(\lambda)$,并且把相对应的模型看作是最优模型。

5 总结

本文针对时空加权回归模型,提出了变量选择方法。该方法使用样条函数逼近系数函数,并通过惩罚函数进行有效的变量选择,同时也给出了窗宽参数 h 和调谐参数 λ 的选取办法,最终,提高了模型的准确性和便利性。此方法将有助于计量经济学、生物统计学、流行病等各个领域在考虑时空加回归权模型时对其重要变量作出选择,使模型达到精简、提高预测精度的目的。

参考文献：

[1] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices[J]. International Journal of Geographical Information Science, 2010, 24(3): 383-401.

[2] Yan N, Mei C L. A two-step local smoothing approach for exploring spatio-temporal patterns with application to the analysis of precipitation in the mainland of China during 1986—2005[J]. Environmental and Ecological Statistics, 2014, 21(2): 373- 390.

[3] 玄海燕，李帅锋. 时空加权回归模型的非平稳性检验[J]. 甘肃科学报，2012，24(2)：1-4.
Xuan H Y, Li S F. The nonstationarity tests of geographically and temporally weighted regression model[J]. Journal of Gansu Sciences, 2012, 24(2): 1-4.

[4] 玄海燕，李帅峰，张运虎. 时空加权回归模型的影响分析[J]. 兰州理工大学学报，2013，39(5)：135-138.
Xuan H Y, Li S F, Zhang Y H. Influence analysis of geographically and temporally weighted regression model[J]. Journal of Lanzhou University of Technology, 2013, 39(5): 135-138.

[5] 刘锋，谭祥勇，何卓. 函数性线性回归模型分析方法及其应用[J]. 重庆理工大学学报：自然科学版，2015(11)：135-138.
Liu F, Tan X Y, He Z. Methods of functional linear regression model and its applications[J]. Journal of Chongqing University of Technology: Natural Science, 2015(11): 135-138.

[6] Li G R, Lian H, Lai P, et al. Variable selection for fixed effects varying cofficient models[J]. Acta Mathematica Sinica, English Series, 2015, 31(1): 91-110.

[7] Zhao P X, Xue L G. Variable selection for varying coefficient models with measurement errors[J]. Metrika, 2011, 74(2): 231-245.

[8] Fan J Q, Li R Z. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Journal of the American Statistical Association, 2001, 96(456): 1348- 1360.

[9] Wang M Q, Song L X, Tian G L. SCAD-penalized least absolute deviation regression in high-dimensional models[J]. Communication in Statistics: Theory and Methods, 2015, 44(12): 2452-2472.

Variable Selection of Geographically and Temporally Weighted Regression Model Based on Spline Function

XUAN Haiyan[1], SONG Leibo[2], CHEN Jinshu[2], ZHANG Yunhu[2]

(1. School of Economics and Management, Lanzhou University of Technology;
2. School of Sciences, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: In order to improve the prediction accuracy of geographically and temporally weighted regression model and to enhance the interpretation of geographically and temporally weighted regression model, selecting the important variables that has significant influence on the dependent variables has become an important research topic in the statistical analysis. In this paper, firstly the model's coefficients are approximated by using the spline function. Secondly, based on the least square theory, the corresponding coefficients of variables are processed by using the SCAD theory. And then the BIC criterion is used to select the tuning parameter λ. Finally, the useful variables for geographically and temporally weighted regression model are selected by the iterative algorithm. Eliminating the variables that affect the model's accuracy and achieving the purpose of improving the accuracy of the model.

Key words: geographically and temporally weighted regression; variable selection; SCAD; function approximation; BIC criterion

(责任编辑 游中胜)