第 32 卷第 7 期
2015 年 7 月
统计研究
Statistical Research
Vol. 32, No. 7
Jul. 2015

半参数变系数回归模型的空间相关性检验*

陈建宝　乔宁宁

内容提要: 本文对半参数变系数回归模型,构造了新的空间相关性检验统计量,利用三阶矩 χ^2 逼近方法导出了其检验 p - 值的近似计算公式,蒙特卡罗模拟结果表明,该统计量在检测空间相关性方面具有较高的准确性和可靠性。同时考察了误差项服从不同分布时的检验功效,体现出该检验方法的稳健性。进一步,我们还给出了检验统计量的 Bootstrap 方法以及检验水平的模拟效果。

关键词: 半参数变系数回归模型; 空间相关性; Bootstrap 方法

中图分类号: C81　　**文献标识码:** A　　**文章编号:** 1002 - 4565(2015) 07 - 0087 - 06

Testing Spatial Correlation in Semi-parametric Varying Coefficient Regression Models

Chen Jianbao & Qiao Ningning

Abstract: Based on the semi-parametric varying coefficient regression model, a new testing statistic for spatial correlation is constructed. The corresponding approximation of p-value is derived by employing the third-moment χ^2 method. The Monte Carlo simulation studies show that the proposed testing statistic has good accuracy and reliability. Meanwhile, we investigate the simulation results of testing powers for different error terms, which show that the testing method has high robustness. Furthermore, the bootstrap method and simulation results for the testing statistic are given.

Key words: Semi-parametric Varying Coefficient Regression Model; Spatial Correlation; Bootstrap Method

一、引言

近年来,半参数变系数回归模型一直是学者们研究的重点计量模型之一,并在经济、金融等数据分析中应用广泛。该模型是在变系数回归模型的基础上发展起来的,较之变系数回归模型,这种模型允许解释变量的系数一部分为未知函数,另一部分为常数,其设定体现出更强的适应性和灵活性,不仅继承了非参数回归模型形式自由和估计稳健的特点,还保留了线性回归模型直观、容易解释的优点,且能够有效规避多元非参数回归模型中的“维数灾难”问题。Zhang 等[10] 基于局部多项式估计给出了半参数变系数回归模型的两阶段估计法,证明了参数和非参数估计量的渐近偏和方差; Zhou 和 You[11] 给出了模型参数和非参数部分的小波估计; Fan 和 Huang[3] 提出了半参数变系数回归模型的截面最小二乘估计,证明了参数部分的渐近正态性,并进一步研究了参数部分的假设检验和变系数部分的稳定性检验等问题; Li 等[7] 借鉴 Fan 和 Huang 的研究提出了半参数时变系数回归模型,同样采用截面最小二乘估计方法对此类模型进行了研究,并讨论了模型的统计推断问题。

在许多实际问题中,半参数变系数回归模型多应用于分析经济金融等时序数据。由于连续观察的数据经常是序列相关的,因此,模型误差的独立性假设有时并不能完全成立,进而对模型的直接估计将能产生序列相关问题。Fan 等[4] 对纵向数据中包含协方差结构的半参数变系数回归模型进行了相应研究; Hu 等[5] 基于经验似然方法给出了半参数变系数部分线性度量误差模型的序列相关检验; 胡雪梅和刘锋[12] 提出了两个统计量分别用来检验半参数变系数时序模型及其平行数据模型的有限阶序列相关性。事实上,对半参数变系数回归模型的运用不

* 本文为教育部人文社会科学项目“空间自回归单指数模型的理论和实践”(13YJA9100002) 的阶段性成果。

止限于时间维度，延伸到空间维度时，依然具有较大的拓展空间（Fan 和 Huang，2005；Zhang 等，2002；Xia 等，2004）。然而，当涉及到空间截面数据时，类似于时序数据的序列相关性，空间相关性的存在同样会使得模型分析变得较为复杂，这一现象不可避免，因此如何对模型中存在的空间相关性进行有效辨识和处理成为重要一环。有鉴于此，本文尝试构造新的空间相关性检验统计量实现对半参数变系数回归模型的空间相关性检验，这方面的研究在现有文献中尚未涉及。

二、半参数变系数回归模型介绍

考虑如下形式的半参数变系数回归模型：

$$Y = X\beta + M + \varepsilon \tag{1}$$

其中，$Y = (y_1, y_2, \cdots, y_n)^T$ 为被解释变量，(U, X, Z) 为协变量，为避免"维数灾难"问题，不失一般性，不妨将 U 设定为单变量，且有 $U = (u_1, u_2, \cdots, u_n)^T$，$X = (x_1, x_2, \cdots, x_n)^T$，$x_i = (x_{i1}, x_{i2}, \cdots, x_{iq})^T$，$Z = (z_1, z_2, \cdots, z_n)^T$，$z_i = (z_{i1}, z_{i2}, \cdots, z_{ip})^T$，$\beta = (\beta_1, \beta_2, \cdots, \beta_q)^T$ 为待估参数向量，并设定 $M = (\alpha^T(u_1) z_1, \alpha^T(u_2) z_2, \cdots, \alpha^T(u_n) z_n)^T$ 为模型中的非参数部分，$\alpha(u_i) = (\alpha_1(u_i), \alpha_2(u_i), \cdots, \alpha_p(u_i))^T$ 为函数形式未知的变系数向量，$\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$ 为随机误差向量，$\varepsilon \sim N(0, \sigma^2 I)$，$I$ 为 n 阶单位阵，$i = 1,2,\cdots,n$。进一步，模型(1)可以改写为：

$$y_i^* = \sum_{j=1}^{p} \alpha_j(u_i) z_{ij} + \varepsilon_i \quad i = 1,2,\cdots,n \tag{2}$$

其中，$y_i^* = y_i - \sum_{j=1}^{q} \beta_j x_{ij}$。针对模型(2)中的参数与非参数部分，本文采用 Fan 和 Huang[3] 提出的基于局部线性的截面最小二乘估计（Profile Least Square）进行拟合。即在任一点 u 处，对 $\alpha_j(u)$ 进行泰勒展开，定义：

$$\{\alpha_j(u), \alpha_j'(u)\}_{j=1}^{p} = \underset{\{\alpha_j(u), \alpha_j'(u)\}_{j=1}^{p}}{\operatorname{argmin}} \sum_{i=1}^{n} \left[y_i^* - \sum_{j=1}^{p} \{\alpha_j(u) - \alpha_j'(u)(u_i - u)\} z_{ij}\right]^2 K_h(u_i - u) \tag{3}$$

其中，$K_h(u_i - u) = h^{-1} K((u_i - u)/h)$，$K((u_i - u)/h)$ 为核函数，h 为窗宽。令：

$$W_u = diag(K_h(u_1 - u), \cdots, K_h(u_n - u)),\ D_u = \begin{pmatrix} z_1^T & h^{-1}(u_1 - u) z_1^T \\ \vdots & \vdots \\ z_n^T & h^{-1}(u_n - u) z_n^T \end{pmatrix}$$

此时，由广义最小二乘法可得：

$$[\hat{\alpha}_1(u), \cdots, \hat{\alpha}_p(u), h\hat{\alpha}'_1(u), \cdots, h\hat{\alpha}'_p(u)]^T = \{D_u^T W_u D_u\}^{-1} D_u^T W_u (Y - X\beta)$$

则 M 的初次估计为：$\tilde{M} = S(Y - X\beta)$，其中，$S = \begin{pmatrix} [z_1^T\ 0] & \{D_{u_1}^T W_{u_1} D_{u_1}\}^{-1} D_{u_1}^T W_{u_1} \\ \vdots & \vdots \\ [z_n^T\ 0] & \{D_{u_n}^T W_{u_n} D_{u_n}\}^{-1} D_{u_n}^T W_{u_n} \end{pmatrix}$。将 $\tilde{M}$ 代入式(1)，整理可得线性回归模型：

$$(I - S) Y = (I - S) X\beta + \varepsilon$$

利用最小二乘法得到 β 的最终估计为：

$$\hat{\beta} = [X^T (I - S)^T (I - S) X]^{-1} X^T (I - S)^T (I - S) Y$$

变系数 $\alpha(u)$ 的最终估计为：

$$\hat{\alpha}(u) = (\hat{\alpha}_1(u), \hat{\alpha}_2(u), \cdots, \hat{\alpha}_p(u))^T = e^T \{D_u^T W_u D_u\}^{-1} D_u^T W_u (Y - X\hat{\beta})$$

其中，$e = (I_p, O_p)^T$，I_p 和 O_p 分别为 $p \times p$ 单位阵和元素全部为 0 的 $p \times p$ 矩阵。同时得到 M 的最终估计为 $\tilde{M} = S(Y - X\hat{\beta})$。从而可得 $Y = (y_1, y_2, \cdots, y_n)^T$ 的拟合值向量为：

$$\hat{Y} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n)^T = \hat{M} + X\hat{\beta} = S(Y - X\hat{\beta}) + X\hat{\beta} = (I - S) X\hat{\beta} + SY = LY$$

其中，$L = (I - S) X [X^T (I - S)^T (I - S) X]^{-1} X^T (I - S)^T (I - S) + S$。相应的残差向量为 $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n)^T = Y - \hat{Y} = (I - L) Y$。

对于上述估计方法，光滑参数 h 可用交叉确认法进行确定。即选取合适的 h，使得 $CV(h) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_{(-i)}(h))^2$ 达到最小，其中，$\hat{y}_{(-i)}(h)$ 是在给定 h 值下删掉第 i 组观测数据后得到的 y_i 的预测值。

三、空间相关性检验

众所周知，空间相关性是空间数据的主要特征之一，相关性的存在时常会导致一些标准的计量方法失效（Anselin[1]；Cordy 和 Griffith[2]；等等）。更重要的是，空间相关行为往往说明模型中忽视了重要的回归因子或采用了错误的函数形式。因此，在对空间数据进行研究时，检验回归模型的误差项是否存在空间相关性已成为一种标准做法。然而，对于半参数变系数回归模型，至今仍没有相关研究涉足空间相关性方面的检验。以此为契机，本文尝试提

出了一类新的检验空间相关性的 Moran's I 指标，并首次采用三阶矩 χ^2 逼近方法对其检验 p - 值进行逼近。这里，对 Moran's I 指标的选择主要借鉴了线性参数回归模型中的空间相关性检验思想，同时基于 Fan 和 Huang[3] 对半参数变系数回归模型的研究框架，我们能够得到模型参数估计的 $\sqrt{n}$ 一致性，这在一定程度上保障了检验统计量 Moran's I 指标的一致性。另外，对于三阶矩 χ^2 逼近方法，已有理论研究(Pearson[8]；Imhof[6])与模拟结果(梅长林和王宁[14])表明该方法不但能大大降低计算量，而且具有较高的精度。

关于空间相关性 Moran's I 指标的选择，其本身具有一定意义。由于 Moran's I 检验能够识别出任何形式的空间关系，使得这种检验在空间数据分析中能够更有效地捕捉到研究对象之间的空间关联性。这里，将原假设 H_0 设定为模型估计误差不存在空间相关性，即 $\mathrm{var}(\varepsilon)=E(\varepsilon\varepsilon')=\sigma^2 I$，备择假设 H_1 为模型估计误差存在空间相关性。由于实际中真正的误差项不可观测，因此采用回归模型的残差向量 $\hat{\varepsilon}$ 进行代替，其中，$\hat{\varepsilon}=(\hat{\varepsilon}_1,\hat{\varepsilon}_2,\cdots,\hat{\varepsilon}_n)^T$。根据第二部分半参数变系数回归模型的估计结果，可得 $\hat{Y}=LY$，$\hat{\varepsilon}=(I-L)Y=NY$，其中 $N=I-L$。给定残差向量 $\hat{\varepsilon}$ 后，构造的 Moran's I 指标为:

$$I_0=\frac{n}{s}\frac{\hat{\varepsilon}^T W^* \hat{\varepsilon}}{\hat{\varepsilon}^T\hat{\varepsilon}} \tag{4}$$

上式可进一步简化为 $I_0=\dfrac{\hat{\varepsilon}^T W^*\hat{\varepsilon}}{\hat{\varepsilon}^T\hat{\varepsilon}}$。其中，$s=\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}$，$W^*=(W^T+W)/2$，$W$ 为相邻单元之间构成的标准化空间权重矩阵。上面的 Moran's I 指标即为原假设和备择假设下空间相关性的检验统计量，令 r 为 I_0 的观测值，则 I_0 的检验 p - 值表示为:

$$p=P(I_0\geqslant r)\ \text{或}\ p=P(I_0\leqslant r) \tag{5}$$

其中，$I_0\geqslant r$ 衡量误差项存在正空间相关性；反之，$I_0\leqslant r$ 衡量误差项存在负空间相关性。对于给定的显著性水平 α，若 $p<\alpha$，则拒绝原假设 H_0，误差项之间存在空间相关性。

在原假设 H_0 下，我们有 $\varepsilon\sim N(0,\sigma^2 I)$。这里，不妨设定 $E(\hat{\varepsilon})=E(Y-\hat{Y})=0$(假设窗宽的选择使得回归偏差可近似忽略)，即有 $\hat{\varepsilon}=\hat{\varepsilon}-E(\hat{\varepsilon})=N[Y-E(Y)]=N\varepsilon$，因此有:

$$I_0=\frac{\varepsilon^T N^T W^* N\varepsilon}{\varepsilon^T N^T N\varepsilon} \tag{6}$$

能够看出，I_0 关于 σ^2 具有不变性。不失一般性，可以假定 $\sigma^2=1$，即在原假设 H_0 下，$\varepsilon\sim N(0,I)$。此时，式(5)可变为:

$$p=P(I_0\leqslant r)=P\left(\frac{\varepsilon^T N^T W^* N\varepsilon}{\varepsilon^T N^T N\varepsilon}\leqslant r\right)=P[\varepsilon^T N^T(W^*-rI)N\varepsilon\leqslant 0] \tag{7}$$

下面我们将采用三阶矩 χ^2 逼近(Imhof，1961)来计算 $P(I_0\leqslant r)$①。计算结果如下:

令 $Q=\varepsilon^T N^T(W^*-rI)N\varepsilon$，那么利用三阶矩 χ^2 逼近可求得其检验 p-值，主要结果分为两种情况:

当 $E[Q-E(Q)]^3>0$ 时，

$$p=P(I_0\leqslant r)=P(Q\leqslant 0)\approx P\{\chi_d^2\leqslant d-\frac{1}{2}E(Q)\mathrm{var}(Q)/tr[N^T(W^*-rI)N]3\} \tag{8}$$

当 $E[Q-E(Q)]^3<0$ 时，

$$p=P(I_0\leqslant r)\approx 1-P\{\chi_d^2\leqslant d-\frac{1}{2}E(Q)\mathrm{var}(Q)/tr[N^T(W^*-rI)N]^3\} \tag{9}$$

其中，

$E(Q)=tr[N^T(W^*-rI)N]$，$\mathrm{var}(Q)=2tr[N^T(W^*-rI)N]^2$，

$E[Q-E(Q)]^3=8tr[N^T(W^*-rI)N]^3$，$d=\dfrac{8[\mathrm{var}(Q)]^3}{\{E[Q-E(Q)]^3\}^2}=\dfrac{\{tr[N^T(W^*-rI)N]^2\}^3}{\{tr[N^T(W^*-rI)N]^3\}^2}$。

实际中，$E(Q-E(Q))^3=0$ 的情况很少出现，这时有 $tr[N^T(W^*-rI)N]^3=0$。若出现这一情况，我们可采用精确方法(Imhof[6])求解，或者令 $p=\Phi[E(Q)/\sqrt{\mathrm{var}(Q)}]$，其中，$\Phi(\cdot)$ 为 $N(0,1)$ 的分布函数。

另外，借鉴半参数变系数回归模型中的空间相关性检验方法，我们同样给出了变系数回归模型的空间相关性检验结果(蒙特卡罗模拟时用到)。关于变系数回归模型的估计过程可参见 Fan 和 Zhang②，这里没有详细列出。

上述检验方法依赖于模型的误差项服从正态分布这一假定。虽然正态分布的假定在实际分析中应用广泛，但复杂的现实数据往往与此假定并不吻合，

① $P(I_0\geqslant r)$ 的计算过程类似，限于篇幅没有列出。

② Fan J，W Zhang. Statistical methods with varying coefficient models[J]. Statistics and its Interface，2008，1:179-195.

加之非参数估计偏差的影响，使得运用三阶矩 χ^2 逼近法计算检验 p - 值时，可能产生较大误差。为此，我们同时提出了另一种模拟检验 p - 值的 Bootstrap 方法。具体步骤如下：

步骤 1：对于某一给定的光滑参数 h（其值可由交叉确认法进行确定），基于观测数据 $(y_i;z_{i1},\cdots,z_{ip},x_{i1},\cdots,x_{iq},u_i)$，$i=1,2,\cdots,n$，采用截面最小二乘估计拟合半参数变系数回归模型（1），求得残差向量 $\hat{\varepsilon}_1=(\hat{\varepsilon}_{11},\hat{\varepsilon}_{21},\cdots,\hat{\varepsilon}_{n1})^T$，根据式（6）计算 Moran's I 统计量 I_0 的观测值 r；

步骤 2：对残差向量 $\hat{\varepsilon}_1$ 中心化后的向量表示为 $\hat{\varepsilon}=(\hat{\varepsilon}_{11}-\bar{\hat{\varepsilon}}_1,\hat{\varepsilon}_{21}-\bar{\hat{\varepsilon}}_1,\cdots,\hat{\varepsilon}_{n1}-\bar{\hat{\varepsilon}}_1)^T$，其中 $\bar{\hat{\varepsilon}}_1=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i1}$。对 ε 进行 Bootstrap 抽样，可采用标准 Bootstrap 或非对称 Wild Bootstrap 方法。其中，在构造 Wild Bootstrap 残差 $\{\hat{\varepsilon}_i^*\}_{i=1}^n$ 时，$\varepsilon_i^*=\hat{\varepsilon}_i\eta_i$，并且随机变量 η_i 的分布形式如下：

$$\eta_i=\begin{cases}-(\sqrt{5}-1)/2 & p=(\sqrt{5}+1)/2\sqrt{5}\\(\sqrt{5}+1)/2 & p=(\sqrt{5}-1)/2\sqrt{5}\end{cases}$$

步骤 3：采用步骤 1 中求出的拟合值向量及 Bootstrap 残差 $\hat{\varepsilon}^*$，得到新的 Bootstrap 数据集 $(y_i^*;z_{i1},\cdots,z_{ip},x_{i1},\cdots,x_{iq})$；

步骤 4：基于步骤 3 得到的 Bootstrap 样本，同样采用截面最小二乘法估计半参数变系数回归模型，得到新的残差向量 $\hat{\varepsilon}_1^*$，计算得到新的统计量值 I_{01}^*；

步骤 5：重复步骤 2 ~ 4 共 m 次，得到检验统计量 I_0 的 m 个 Bootstrap 观测值 $I_{01}^*,I_{02}^*,\cdots,I_{0m}^*$，那么，检验 p - 值的 Bootstrap 逼近为：

$$\hat{p}\approx\min\left\{\frac{1}{m}\sum_{i=1}^{m}I(I_{0i}^*\leqslant r),\frac{1}{m}\sum_{i=1}^{m}I(I_{0i}^*>r)\right\}\quad(10)$$

其中，$I(\cdot)$ 为指示函数。

当残差分布未知或存在异方差时，可以按照以上步骤实现对半参数变系数回归模型的 Bootstrap 检验，得到相应的空间相关性统计量 I_0 和 $I_{01}^*,I_{02}^*,\cdots,I_{0k}^*$，以期实现对空间相关性的有效检测。

四、蒙特卡洛数值模拟结果

下面通过蒙特卡洛数值模拟考察所提出检验方法的表现，主要包括检验方法的有效性，误差分布类型对检验效果的影响以及 Bootstrap 方法模拟的功效差异等。构造具有空间相关性的数据结构，我们考虑如下几种模型：

模型 M1：$Y=\rho WY+\alpha_1^T(U)z_1+\alpha_2^T(U)z_2+\varepsilon$

其中，$\alpha_1(U)=\sin(2\pi U)$，$\alpha_2(U)=3.5[\exp(-(4U-1)^2)+\exp(-(4U-3)^2)]-1.5$；

模型 M2：$Y=\rho WY+\alpha_1^T(U)z_1+\alpha_2^T(U)z_2+\beta_1x_1+\beta_2x_2+\varepsilon$

其中，$\beta_1=1.0$，$\beta_2=1.5$，$\alpha_1(U)=\sin(2\pi U)$，$\alpha_2(U)=3.5[\exp(-(4U-1)^2)+\exp(-(4U-3)^2)]-1.5$；

模型 M3：$Y=\alpha_1^T(U)z_1+\alpha_2^T(U)z_2+\eta$，$\eta=\lambda W\eta+\varepsilon$

其中，$\alpha_1(U)=\sin(2\pi U)$，$\alpha_2(U)=3.5[\exp(-(4U-1)^2)+\exp(-(4U-3)^2)]-1.5$；

模型 M4：$Y=\alpha_1^T(U)z_1+\alpha_2^T(U)z_2+\beta_1x_1+\beta_2x_2+\eta$，$\eta=\lambda W\eta+\varepsilon$

其中，$\beta_1=1.0$，$\beta_2=1.5$，$\alpha_2(U)=3.5[\exp(-(4U-1)^2)+\exp(-(4U-3)^2)]-1.5$，$\alpha_1(U)=\sin(2\pi U)$。

显然，M1 和 M3 为具有空间相关结构的变系数回归模型，M2 和 M4 为具有空间相关结构的半参数变系数回归模型，并且空间相关性结构主要分为空间滞后和误差滞后两种类型。这里，具体做出如下设定：

（1）(U,x_1,x_2,z_1,z_2) 为协变量，其中，U 为一维随机变量，产生于均匀分布 $U(0,1)$，协变量 (x_1,x_2,z_1,z_2) 产生于均值为 0、方差为 1 的多元正态分布，并且 4 个随机变量之间的相关系数都为 2/3，随机误差项服从正态分布 $N(0,1)$；

（2）空间权重矩阵 W 设定为 Rook 空间权重矩阵，分别取样本数 $n=100$ 和 $n=225$，并设定空间滞后相关系数 ρ 和空间误差相关系数 λ 在 $[-0.9,0.9]$ 内以步长 0.2 均匀取值；

（3）为了考察误差项 ε 的分布可能对检验统计量 I_0 产生的影响，我们同时给出了以下几种误差分布情况下的模拟结果。ε 的分布类型包括 $\varepsilon\sim U(-\sqrt{3},\sqrt{3})$、$\varepsilon\sim\frac{1}{\sqrt{2}}t(4)$。其中各分布满足均值为 0，方差为 1，这样可使得模拟结果具有一定的可比性。

（4）模拟中采用常见的 Epanechnikov 核函数：$K(u)=(3/4\sqrt{5})(1-1/5u^2)I(u^2\leqslant5)$。最优窗宽的选择采取交叉确认法。

利用前面提到的三阶矩 χ^2 逼近得到的近似公式,计算每一种设定情况下检验统计量的 p - 值,重复上述试验各1000次,以1000次重复下的 p - 值小于 α(即拒绝 H_0)的频率模拟检验功效,不妨取 α = 0.05。另外,对Bootstrap检验中的每一次模拟,统计量 p - 值均基于500次Bootstrap模拟计算得到。

(一)误差项服从标准正态分布

当误差项服从标准正态分布时,从模拟结果可以看出:

首先,检验统计量Moran's I指标的检验水平与名义显著性水平相差不大,表明该统计量具有合理的检验水平性质。当原假设 H_0 为真时,拒绝 H_0 的频率比较接近于显著性水平0.05。也就是说,从检验水平角度看,空间相关性检验统计量具有较好的有限样本性质,同时也体现了三阶矩 χ^2 逼近方法的有效性。

其次,研究不同的空间相关性结构对Moran's I指标检验功效的影响。当数据生成过程为空间滞后回归时,对于半参数变系数回归模型和变系数回归模型,Moran's I指标的检验功效均呈现对称性特征,并且随着正负相关系数的增加迅速向1靠近;而当数据生成过程转变为误差滞后回归时,显现出与空间滞后回归模型检验相类似的特征,这些都表明检验统计量能够较好地捕捉到模型中存在的空间相关性。

第三,对于原假设 H_0 不成立的情况,随着样本容量的增加,Moran's I指标的检验功效逐渐趋近于1,这也意味着本文提出的统计量具有优良的检验功效性质。

(二)不同误差项分布类型

不同分布类型下的蒙特卡洛模拟结果显示,虽然检验 p - 值的计算公式是在误差项服从正态分布的假定下得到,但对于均匀分布和t分布,检验方法的表现并没有出现太大差异,其检验水平和检验功效与标准正态分布的情况较为接近。当模型中不存在空间相关性时,错误拒绝原假设的频率大多接近于显著性水平0.05;而当模型中存在空间相关性时,伴随空间滞后相关系数 ρ 或者空间误差相关系数 λ 的增加,检验功效迅速接近1,并且在样本容量增大的情况下,统计量的检验功效也越来越大。这些结果说明本文所给出的检验方法对于误差项分布的变化具有一定的稳健性。

(三)Bootstrap方法

在不同模型下,采用Bootstrap方法得到的检验统计量的检验水平与名义显著性水平相差不大,并且检验功效随着空间相关性的增加迅速向1靠近,同时,检验功效曲线随着样本量的扩大变得更为陡峭,也就是说,样本量越大,功效越大。这些现象综合反映了Bootstrap方法的有效性和合理性。显然,基于三阶矩 χ^2 逼近以及Bootstrap方法所得到的检验统计量均有着良好的检验效果,两种方法都可用于回归模型的空间相关性检验。不过,在实际蒙特卡洛模拟中,Bootstrap方法往往涉及到较大的计算量,需要消耗较长运算时间,这也是其固有的不足之处,特别在样本容量较大时体现得更为明显。

五、结论

半参数变系数回归模型是近期发展起来的一类具有广泛应用前景的回归模型,将其运用到空间数据分析时,对模型中可能存在的空间相关性进行检验具有重要意义。本文尝试构建了一类新的空间相关性Moran's I指标实现对此类模型的空间相关性检验,采取的是一种计算检验 p - 值的三阶矩 χ^2 逼近方法,蒙特卡洛模拟结果表明这种方法具有较高的准确性和可靠性。同时,本文还给出了误差项服从不同分布时的模拟结果,从侧面反映出该检验方法的稳健性。不过,这种检验方法需要依赖误差项服从正态分布的假定依然是难以避免的局限之一,借鉴很多学者的实际处理方法(龙志和等,2009;欧变玲等,2010),我们同时给出了关于该检验的Bootstrap方法以备参考。整体来看,Bootstrap方法突破了残差分布为已知的条件限制,可应用于小样本下的模拟,这些是该方法的长处。然而,受限于计算耗时等因素的影响,在大样本应用中,我们仍应以三阶矩 χ^2 逼近方法计算得到的检验统计量为基础。需要说明的是,虽然本文在解决半参数变系数回归模型的空间相关性检验问题上起到一定作用,但该检验从本质上仅能判断出研究对象是否存在空间相关性,而无法有效识别空间关系的具体表现形式,这也是其局限所在。因而,如何进一步辨识空间滞后和空间误差的相关性结构同样具有重要的理论和现实意义,值得深入研究。

最后,本文提出的空间相关性检验方法可推广到其他非参数/半参数计量模型,是对现有非参数/

半参数空间计量模型理论体系的有益补充，进而可延伸出一些新的、有意义的研究方向。例如，基于半参数变系数回归模型的空间相关性检验，可以进一步探讨如何处理该模型中的空间相关性问题，提出新的理论模型以及构建新的估计技术等等。

参考文献

[1] Anselin L. Spatial econometrics: methods and models [M]. Kluwer Academic, Dordrecht, 1988.

[2] Cordy C B, D A Griffith. Efficiency of least squares estimators in the presence of spatial autocorrelation [J]. Communications in Statistics: Simulation and Computation, 1993, 22(4): 1161-1179.

[3] Fan J, T Huang. Profile likelihood inferences on semiparametric varying-coefficient partially linear models [J]. Bernoulli, 2005, 11(6): 1031-1057.

[4] Fan J, THuang, R Z Li. Analysis of longitudinal data with semi-parametric estimation of covariance function [J]. Journal of the American Statistical Association, 2007, 102(478): 632-641.

[5] Hu X M, Wang Z Z, Liu F. Zero finite-order serial correlation test in a semi-parametric varying-coefficient partially linear errors-in-variables model [J]. Statistics and Probability Letters, 2008, 78(12): 1560-1569.

[6] Imhof J P. Computing the distribution of quadratic forms in normal variables [J]. Biometrika, 1961, 3-4(48): 419-426.

[7] Li D, J Chen, Z Lin. Statistical inference in partially time-varying coefficient models [J]. Journal of Statistical Planning and Inference, 2011, 141: 995-1013.

[8] Pearson E S. Note on an approximation to the distribution of non-central χ^2 [J]. Biometrika, 1959, 46: 364.

[9] Xia Y, W Zhang, H Tong. Efficient estimation for semivarying-coefficient models [J]. Biometrika, 2004, 91(3): 661-681.

[10] Zhang W, S Y Lee, X Song. Local polynomial fitting in semivarying coefficient models [J]. Journal of Multivariate Analysis, 2002, 82(1): 166-188.

[11] Zhou X, J You. Wavelet estimation in varying-coefficient partially linear regression models [J]. Statistics and Probability Letters, 2004, 68(1): 91-104.

[12] 胡雪梅，刘锋. 半参数时变系数模型的序列相关检验[J]. 应用数学学报，2011，34(6)：1103-1117.

[13] 龙志和，欧变玲，龙广平. 空间经济计量模型Bootstrap检验的水平扭曲[J]. 数量经济技术经济研究，2009(1)：151-160.

[14] 梅长林，王宁. 回归模型误差相关性的统一检验方法[J]. 高校应用数学学报A辑，2003(3)：318-326.

[15] 欧变玲，龙志和，林光平. 空间经济计量滞后模型Bootstrap Moran检验功效的模拟分析[J]. 统计研究，2010(9)：91-96.

作者简介

陈建宝，男，1965年生，云南曲靖人，2003年毕业于澳大利亚科庭理工大学统计学专业，获哲学博士学位，现为厦门大学经济学院副院长，统计系教授、博士生导师，厦门大学宏观经济研究中心和福建省统计科学重点实验室（厦门大学）兼职研究员。研究方向为统计理论与方法、经济计量。

乔宁宁，男，1986年生，山西临汾人，2014年毕业于厦门大学统计学专业，获经济学博士学位，现供职于中国人民银行太原中心支行。研究方向为经济计量。

（责任编辑：曹　麦）