

基于时空窗口改进的时空加权回归分析 ——以湖北省黄石市住房价格为例

王新刚, 孔云峰

(河南大学环境与规划学院, 河南 开封 475001)

摘要: 针对地理加权回归(GWR)模型不能有效处理样本数据空间自相关性这一问题, 构造局部时空窗口统计量, 尝试改进时空加权回归(GTWR)模型。定义多时空窗口的概念, 给出其选取、计算和验证方法; 计算时空窗口包含的各样本点的被解释变量平均值, 与样本拟合点的被解释变量值的比值, 作为新的解释变量, 构建改进的时空加权回归(IGTWR)模型。以土地稀缺、多中心、资源型城市——湖北省黄石市为例, 收集2007~2012年商品住宅成交价格1.93万个数据和398个楼栋样本点, 选取小区等级、绿化率、楼栋总层数、容积率、距区域中心距离和销售年份6个解释变量, 分别利用常规线性回归(OLS)、GWR、GTWR和IGTWR方法进行回归分析。模型结果表明: 计算Moran's I 指数和分析时间序列的自相关性, 能确定时空窗口的大小和数量的选取; IGTWR模型和各变量的回归统计均通过0.05的显著性水平检验, 有关解释变量的系数估计值在空间分布上能合理解释; GWR拟合结果优于OLS, GTWR优于GWR, 而IGTWR拟合精度最好。与GTWR模型分析相比, IGTWR模型 R^2 从0.877提升到0.919, 而AICc、残差方(RSS)和均方差(MSE)分别从6 226、49 996 201和354.427下降到6 206、32 327 472和284.969。案例研究表明: IGTWR能够表达一定时空范围的时空自相关特征, 减小了估计误差, 提高了回归拟合精度。

关键词: 时空窗口; 时空加权回归(GTWR); 住房价格; 黄石市

中图分类号: P208 **文献标识码:** A **文章编号:** 1000-0690(2015)05-0615-07

城市住房价格一直是社会关注度最高的热点问题之一。分析和预测住房价格的空间分布和变化趋势是居民、投资者、开发商以及政府管理部门关注的重要问题, 也是学者研究的热点问题。城市住房价格与空间区位密切相关, 且具有空间自相关特征, 因此常采用空间自回归(SAR)或地理加权回归(GWR)模型分析住房价格各影响要素的空间分布规律^[1-4]。受房地产市场趋势和国家宏观政策影响, 时间也是影响城市住房价格的重要因素之一^[5-7]。考虑时间因素影响, 有学者建立时空自回归模型或时空加权回归(GTWR)模型^[8-12]。然而, 在时空回归模型中, 时间因素未作为解释变量。同时, 为了研究局部时空的影响, 本文引入时空窗口, 根据时空的自相关性分析, 建立改进的时空加权回归(IGTWR)模型, 从局部时空的角度, 分析、估计和解释住房价格的变化规律。

1 研究方法

1.1 时空窗口

在XYT时空坐标系中, 通常定义点 $I(x_i, y_i, t_i)$ 的时空窗口(STW)为距离 (x_i, y_i) 小于 r 、距离 t_i 小于 t 的圆柱区域^[13]。给定空间距离 r 和时间间隔 t , 本文定义点 I 的多个时空窗口 $S_m T_n W (m=1, 2, \dots; n=1, 2, \dots)$, 即为距离 (x_i, y_i) 大于 $(m-1)r$ 且小于 mr 、距离 t_i 大于 $(n-1)t$ 且小于 nt 的时空区域; m 与 n 的取值的大小表示窗口 $S_m T_n W$ 与点 I 的邻近程度。

对于时空样本点 I , 为研究不同局部时空邻域内的样本点对其时空相关性影响, 可选择其多个时空窗口 $S_m T_n W$ 作为局部时空邻域, 进行局部时空相关性分析。

地理时空回归分析中, 距离 r 和间隔 t 的选择要能够反映地理时空变异的尺度特征。可通过计

收稿日期: 2014-01-18; 修订日期: 2014-03-20

基金项目: 国家十二五科技支撑计划(2012BAJ051306)资助。

作者简介: 王新刚(1972-), 男, 河南郑州人, 博士研究生, 主要从事城市问题、城市住房研究。E-mail: wuhanwxg@163.com

算Moran's I 指数估计 r 值^[13]。同样,时间间隔 t ,其设定先验值^[14]或可根据时间序列相关性进行估计。

为了兼顾考虑空间和时间的相关性,有学者提出时空权重矩阵^[10]。针对某样本点的STW而言,为简便计算,可将其包含的各样本点的变量值取平均值,分析本样本点与其STW的局部时空相关性。

1.2 GTWR模型改进

GWR作为局部回归拟合方法,将地理位置坐标引入模型,处理空间异质性,探测相关变量空间变化规律^[3]。GWR利用空间变化的回归系数进行拟合估测,并未直接考虑空间依赖性^[15],模型残差 ε 具有不确定性,可能包含部分能够建模的时空相关性。GTWR是在GWR模型的基础上,将空间位置坐标和时间序列坐标构成三维坐标,同时考虑空间和时间对各解释变量的回归系数的影响^[10]。时空坐标系中,时空位置 i 的坐标为 (u_i, v_i, t_i) ,GTWR模型为:

$$Y_i = \alpha_0(u_i, v_i, t_i) + \sum_{j=1}^m \alpha_j(u_i, v_i, t_i) X_{ij} + \varepsilon_i \quad (1)$$

式中, Y_i 为样本点 i 的被解释变量值($i=1,2,\dots,n$), n 为样本点的数目, m 为解释变量的数目; t_i 为第 i 个样本点的时间坐标; $\alpha_0(u_i, v_i, t_i)$ 表示样本点 i 的时空截距项; X_{ij} 表示样本点 i 的第 j 个解释变量值; $\alpha_j(u_i, v_i, t_i)$ 表示样本点 i 处第 j 个变量的回归系数,是时空坐标的函数; ε_i 表示残差。时空权重函数 $w_i(u_i, v_i, t_i) = w_i(u_i, v_i) \otimes w_i(t_i)$,其中 $w_i(u_i, v_i)$ 是空间权重, $w_i(t_i)$ 为时间权重。计算时空距离时,可乘以参数 λ 作为时空权重的平衡因子。

因将时空三维坐标引入模型,GTWR能提高模型拟合精度,且从时空三维角度提供各解释变量对因变量的影响分析,具有比较好的解释力。但模型处理中时间因子是通过坐标影响其他解释变量的回归系数发挥作用,其自身未列入解释变量,给定某时空范围的自相关性在模型中未能体现;模型仅考虑某个核(固定核或自适应核)的局部时空邻域影响,而其他局部时空影响并未兼顾。

为有效处理时空自相关性,减少残差的不确定性,基于某样本点的时空窗口STW,计算STW包含的各样本点的被解释变量平均值(STW内无样本点时值为0),与样本拟合点的被解释变量值求比值,产生新的解释变量^[11],构造局部时空统计量,对模型进行扩展,可获取某些动态的时间影响^[8]。改进模型IGTWR公式如下:

$$Y_i = \alpha_0(u_i, v_i, t_i) + \sum_{j=1}^m \alpha_j(u_i, v_i, t_i) X_{ij} + \sum_{k=1}^p \alpha_k(u_i, v_i, t_i) X_{ik} + \varepsilon_i \quad (2)$$

式中, Y_i 、 m 、 X_{ij} 、 $\alpha_0(u_i, v_i, t_i)$ 、 $\alpha_j(u_i, v_i, t_i)$ 、 ε_i 表示的含义同式(1); $\alpha_k(u_i, v_i, t_i)$ 表示样本点 i 处第 k 个时空窗口变量的回归系数; X_{ik} 表示样本点 i 的第 k 个时空窗口变量值, p 表示时空窗口的数目。

2 研究区域及数据

本文选取湖北省黄石市作为研究区。黄石市位于湖北省东南部,长江中游南岸,是武汉城市圈副中心城市,中国中部地区重要的原材料工业基地。黄石市是典型的资源型组团城市,地形较为复杂、土地稀缺;城区发展采用中心组团式布局结构,受地形条件、传统产业布局和2001~2020年城市规划影响,分为4个组团片区:铁山区、下陆区为工矿区;黄石港区南部为商业中心区,北部为花湖物流中心区;团城山区北部沿湖为行政中心区(图1)。



图1 研究区及住房分布

Fig.1 The study area and housing distribution

选择2007~2012年商品住宅成交价格数据1.93万个。为便于作整体分析,按楼栋计算平均住房单价,共获得398个楼栋样本点。研究区域样本住房呈非均匀分布(图1),针对样本数据进行相关分析,发现价格与销售年份、楼栋总层数、小区容积率、小区等级、绿化率呈现或强或弱的线性相关关系。运用探索性空间数据分析,住房价格的Moran's I 为0.196,具有较强的空间正自相关性;聚类 and 异常值分析,表明住房价格具有局部空间异质性;住房所在区域和区位对价格影响显著。

选择价格解释变量:反映土地稀缺和建筑成本的楼栋总层数(FL)和容积率(PR),反映社区环境的小区等级(CL)和绿化率(GR),反映城市区位属性的住房距区域中心的距离(ND),以及反映宏观经济和政策变化的销售年份(YI)。其中,小区等级划分为4个等级(四级为最高级,赋值为4);销售年份以2005年为基准,转化为年度索引值。计算各解释变量的方差膨胀因子进行检验^[15],发现其值均小于2,远小于10,表明解释变量间的多重共线性现象较弱,适合于回归分析。

3 模型分析与解释

运用常规线性回归(OLS)进行全局估计,结果见表1。模型拟合度 $R^2=0.495$,能合理解释49.5%的住房价格变化原因。由于空间的异质性,变量之间的关系是随空间位置的变化而变化的,OLS模型并未考虑这种变化特征,尚有50.5%的住房价格的变化未能解释。进一步利用GWR模型进行局部统计分析,结果见表1。GWR模型的拟合度 R^2 为0.824,与OLS模型的拟合度相比,有大幅度的提高;统计检验 F 值为2.925,相应概率 P 值为0.000,通过显著性水平为0.05的 F 检验,住房均价与影响因素之间的整体关系确实存在显著的空间非平稳性。

表1 OLS和GWR回归系数的估计结果

Table 1 The estimation results of OLS and GWR regression parameters

变量	OLS 模型				GWR 模型			
	系数	标准差	t 检验	P 值	系数	标准差	t 检验	P 值
CL	264.142	43.123	6.122	0.000	210.802	131.452	1.334	0.000
GR	23.015	4.256	5.403	0.000	14.613	8.155	1.245	0.000
FL	32.133	6.088	5.275	0.000	23.204	22.323	1.306	0.020
PR	64.312	35.265	1.823	0.030	294.105	313.566	1.594	0.000
ND	-14.324	48.892	-0.298	0.760	-172.682	-167.162	-1.016	0.060
YI	283.693	24.623	11.512	0.000	137.728	137.678	4.735	0.000

进行住房价格的Moran's I 空间自相关分析(图2a),当距离为500 m时,Moran's I 指数值为最大,所以可取500 m作为时空窗口的单位空间距离 r 。对住房价格以年份进行时间序列的自相关分析(图2b),当时间间隔为1 a时,时间自相关的系数值为最大,因此可取1 a作为时空窗口的单位时间间隔 t 。

根据时空窗口STW的 r (500 m)和 t (1a),在空间距离小于500 m时,取5个时空窗口 S_1T_1W 、 S_1T_2W 、 S_1T_3W 、 S_1T_4W 、 S_1T_5W ,其时间距离分别为小于1 a、1~2 a、2~3 a、3~4 a和4~5 a;考虑到计算量,在空间距离为500~1 000 m时,仅取2个时空窗口 S_2T_1W 、 S_2T_2W ,其时间距离分别为小于1 a、1~2 a。计算各样本点的时空窗口,产生相应的7个窗口新变量。

在SAS9.1软件环境中,编程进行GTWR和IGTWR模型计算。考虑到本研究区域的样本点非均匀性分布和时空权重的计算,模型权函数选用Gauss自适应核函数,避免了某些样本点的邻近样本数据稀疏而造成的估计偏差^[9]。

GTWR和IGTWR模型计算时,可以根据回归拟合度 R^2 确认获取最佳的时空权重平衡参数 λ^{opt} 。GTWR模型参数 λ 选择的细节如图3a所示,当 $\lambda=0.1$ 时,回归拟合度 R^2 有最大值0.877 3,所以选取0.1为最佳的 λ 。IGTWR模型参数 λ 选择的细节如图3b,当 $\lambda=0.2$ 时,回归拟合度 R^2 有最大值0.919 3,所以选取0.2为最佳的 λ 。

GTWR和IGTWR模型分析结果,见表2。两个模型检验:统计检验 F 值分别为3.046、4.104,相应概率 P 值均为0.000,均通过显著性水平为0.05的 F 检验;各变量回归统计 t 检验的显著性概率 P 均小于0.05,均通过显著性水平为0.05的 t 检验,表明各回归系数随时空的变化均具有显著性。

IGTWR模型计算中,选用7个时空窗口变量。为验证STW选取是否合理,取10个时空窗口,分别计算出其住房价格平均值,直接作为模型的新解释变量,模型结果表明: S_1T_1W 、 S_1T_2W 、 S_1T_3W 、 S_1T_4W 、 S_1T_5W 、 S_2T_1W 、 S_2T_2W 、 S_2T_3W 、 S_2T_4W 、 S_2T_5W 的回归系数分别为-0.109、0.033、0.093、0.080、-0.150、-0.100、-0.068、-0.019、-0.012、-0.009,反映了各个时空窗口对样本点住房价格的影响程度不同;从 S_2T_2W 之后,回归系数绝对值低于2%,因此仅取前7个STW具有合理性。

相比GTWR,IGTWR模型拟合度 R^2 从87.7%提高到91.9%,AICc从6 226降低到6 206,F统计从3.046提升到4.104,残差平方和RSS从49 996 201降低到32 327 472,均方误差MSE从354.427降低到284.969。分析表明,IGTWR比GTWR能较好地改进回归拟合的精度,减少了估计的误差。

研究区的容积率与住房价格整体上呈现正相

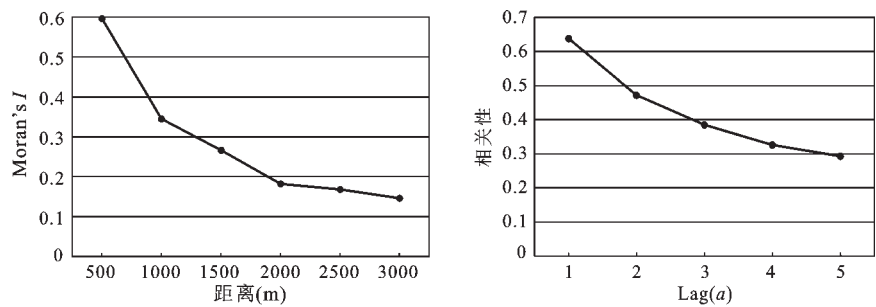


图2 时空窗口单位距离和单位时间的最优值

Fig.2 The appropriate units of distance and time lag in the spatiotemporal windows

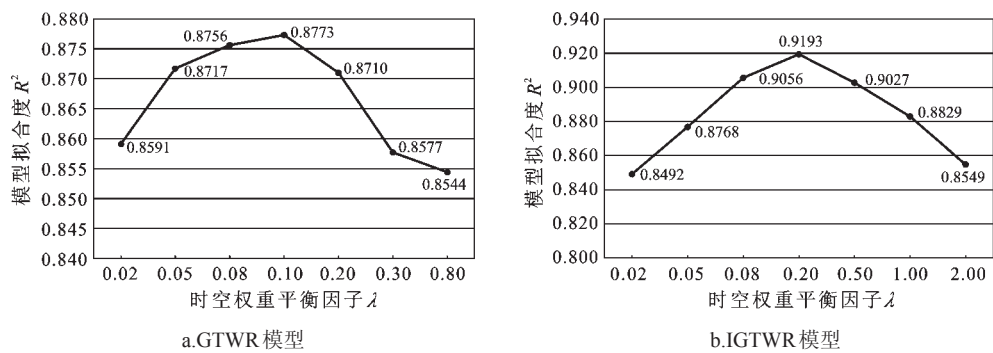


图3 GTWR 及 IGTWR 模型参数 λ 选择

Fig.3 The parameters λ selection for the GTWR model and the IGTWR model

表2 GTWR和IGTWR回归系数的估计结果

Table 2 The estimation results of GTWR and IGTWR regression parameters

变量	GTWR 模型						IGTWR 模型($r=0.5\text{ km}, t=1\text{ a}$)					
	最小值	中值	最大值	均值	t 检验	P 值	最小值	中值	最大值	均值	t 检验	P 值
CL	-2867.193	165.244	2231.183	276.852	4.232	0.000	-417.322	99.966	823.353	145.245	5.346	0.000
GR	-181.562	5.302	477.492	11.833	3.223	0.000	-36.433	10.975	129.874	10.316	3.053	0.000
FL	-96.943	33.891	95.763	32.091	2.456	0.000	-59.315	26.434	68.336	21.036	3.086	0.000
PR	-400.376	73.003	1811.225	175.270	3.125	0.000	-394.116	125.052	1822.655	168.225	4.235	0.000
ND	-3121.423	-135.422	1419.204	-303.752	2.174	0.030	-760.367	-76.726	289.257	-104.683	-2.922	0.020
S_1T_1W	-	-	-	-	-	-	-859.132	-18.015	738.148	-26.712	3.253	0.010
S_1T_2W	-	-	-	-	-	-	-1298.575	49.756	1568.713	152.156	2.634	0.020
S_1T_3W	-	-	-	-	-	-	-1088.636	345.266	1268.425	410.037	3.278	0.000
S_1T_4W	-	-	-	-	-	-	-792.299	746.387	1229.202	561.235	4.166	0.030
S_1T_5W	-	-	-	-	-	-	-1274.415	-494.655	1523.185	-201.415	2.895	0.000
S_2T_1W	-	-	-	-	-	-	-437.276	-294.443	355.759	-381.826	2.363	0.020
S_2T_2W	-	-	-	-	-	-	-745.934	-343.354	1026.043	-238.003	3.162	0.030
模型检验	$R^2=0.877, AICc=6\ 226, F=3.046, P=0.000$ $RSS=49\ 996\ 201, MSE=354.427$						$R^2=0.919, AICc=6206, F=4.104, P=0.000$ $RSS=32\ 327\ 472, MSE=284.969$					

关性,原因可能是黄石市土地资源紧缺,多数住宅设计为高层建筑,在提高容积率的同时增加了建筑成本。其回归系数在不同片区差异显著(图4a):

花湖新区和团城山开发区及黄石港片区偏远位置,容积率对住房价格的影响较小;而下陆片区、黄石港片区靠近中心位置、团城山开发区靠近湖

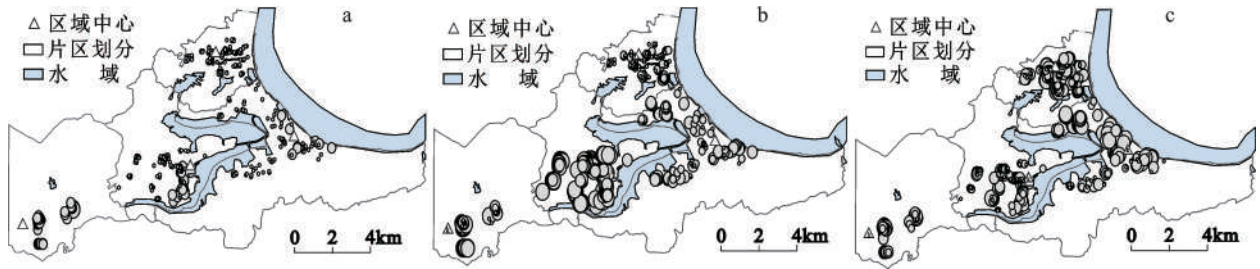


图4 小区容积率(a)、小区等级(b)及距区域中心距离(c)回归系数的分布

Fig.4 The distribution of regression coefficients on cell plot ratio, cell level and distance from the regional centre

边位置,容积率能显著的影响住房价格。其原因是局部位置小区,建设高层住宅提高容积率,造成房屋成本较大增加,促使住房价格有明显的升高。小区等级代表小区的档次定位,其对住房价格的影响具有显著的片区空间差异性(图4b):在花湖新区核心区,小区等级对住房价格影响较小;而在团城山片区,对住房价格影响比较大。原因在于团城山区居住的白领人士居多,比较注重住宅小区的的品质和档次。距区域中心距离反映了城市区位对住房价格的影响,呈具显著的负相关性(图4c)。当然,同一小区楼栋到区域中心距离相差较小,住房价格受此影响较小,图4c中差异不明显;而不同小区楼栋受距区域中心距离的影响较大。限于篇幅,楼栋总层数、绿化率未列出。

IGTWR 模型 7 个时空窗口变量的回归系数和为: $-26.712 + 152.156 + 410.037 + 561.235 + (-201.415) + (-381.826) + (-238.003) = 275.476$, 反映出每年的时空窗口的回归系数,即给定空间距离小于 1 000 m 范围内年度住房价格变化为 275.476。时空窗口 S_1T_1W 的回归系数为负值,在空间上显著的变化(图 5a),反映出住房价格的波动呈空间差异性,对偏远位置住房价格影响较大,而靠近区域中心影响较小,原因可能在于国家对房地产市场宏观调控及经济和金融等政策的影响;

S_1T_3W 、 S_1T_4W 的回归系数为较大的正值,在不同区域呈现显著的空间差异性(图 5b),在下陆片区、团城山的偏远位置及黄石港片区的工矿区,其对住房价格影响较小,原因估计在于这些位置住房发展不充分,速度缓慢,表明住房开发周期对住房价格具有较强正相关性影响; S_2T_1W 、 S_1T_2W 的回归系数为绝对值较大的负值,在空间上呈明显变化(图 5c),偏离区域中心位置,对住房价格影响较大,说明空间距离 500 m~1 000 m 范围内,随空间距离的增加,住房价格显著降低。

GWR、GTWR、IGTWR 三种模型诊断及回归系数估计,见表 3。模型的 R^2 和 F 统计值依次增大,校验 AICc 和残差平方和 RSS 的值依次减小;GTWR 比 GWR 拟合精度有较大的提高(提高 5.3%),同时估计误差大幅减少(减少 40.1%);IGTWR 与 GTWR 相比,也较好提高了拟合精度(提高 4.8%),估计误差也部分减少(35.3%)。小区等级、绿化率、楼栋总层数、容积率在 3 种模型中均呈现出与住房价格的正相关性,IGTWR 模型中各回归系数估计值小于其他 2 个模型。距区域中心距离与住房价格呈现很大的负相关性,按模型顺序相关性先增加后减弱。销售年份的回归系数,GTWR 没有体现,IGTWR 比 GWR 有较大增强。

分析表明,IGTWR 构造局部时空窗口统计量,

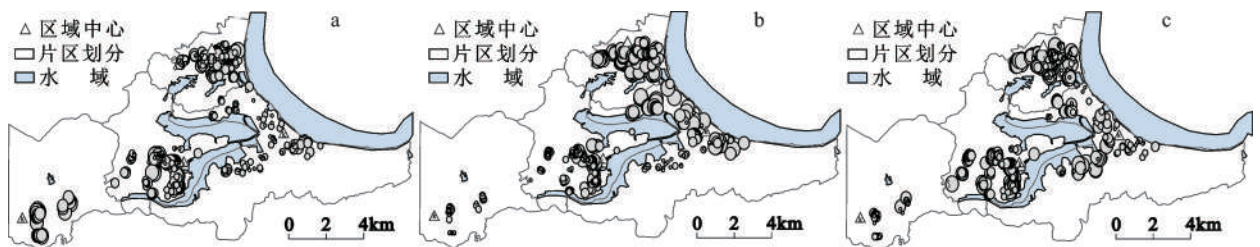
图5 S_1T_1W (a)、 S_1T_3W (b)、 S_1T_2W (c)回归系数的分布Fig.5 The distribution of regression coefficients on S_1T_1W , S_1T_3W and S_1T_2W

表3 GWR、GTWR和IGTWR的模型诊断及回归系数估计

Table 3 The statistical diagnosis and regression parameters estimation of GWR, GTWR and IGTWR models

模型	R^2	AIC_c	F	RSS	CL	GR	FL	PR	ND	YI
GWR	82.4%	6 262	2.925	71 638 362	210.802	14.613	23.204	294.105	-172.682	137.728
GTWR	87.7%	6 226	3.046	49 996 201	276.854	11.832	32.096	175.272	-303.758	-
IGTWR	91.9%	6 206	4.104	32 327 472	145.245	10.315	21.035	168.225	-104.683	275.475

弥补了 GTWR 的不足,改进了模型中时空因素作用下其他解释变量回归系数的估计,从局部时空的角度较精确地反映出时空相关性的影响。

4 结 论

针对地理加权回归(GWR)模型不能有效处理样本数据空间自相关性这一问题,计算时空窗口包含的各样本点的被解释变量平均值,与样本拟合点的被解释变量值的比值,构造局部统计量,较好地改进了时空加权回归(GTWR)模型。以湖北省黄石市的住房价格为例,分别采用常规线性回归(OLS)、GWR、GTWR、改进时空加权回归(IGTWR)作拟合分析。结果表明:IGTWR 模型和各变量的回归统计均通过 0.05 的显著性水平检验,有关解释变量的系数估计值在空间分布上能合理解释;GWR 回归拟合度优于 OLS,GTWR 优于 GWR,而 IGTWR 拟合精度最好。与 GTWR 相比,IGTWR 改进了模型中时空因素作用下其他解释变量回归系数的估计,从局部时空的角度较精确地反映出时空相关性的影响。本文案例研究表明:IGTWR 能够表达一定时空范围的时空自相关特征,减小了估计误差,提高了回归拟合精度。

参考文献:

[1] Can A.Specification and estimation of hedonic housing price models[J].Regional Science and Urban Economics, 1992,22(3): 453-474.

[2] Dubin R,Pace R K,Thibodeau T G.Spatial autoregression techniques for real estate Data[J].Journal of Real Estate Literature, 1999,7(1):79-96.

[3] Fotheringham A S,Brunsdon C,Charlton M.Geographically weighted regression[M].Chichester,UK: John Wiley and Sons, 2002.

[4] Bitter C,Mulligan G F,Dall’erba S.Incorporating spatial variation in housing attribute prices[J].Journal of Geographical Sys-

tems,2007,9(1):7-27.

[5] Demsar U,Fotheringham A S,Charlton M E.Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics[J]. Information Visualization,2008,34(7):1-17.

[6] Yu D L,Lv B Y.Measurement of Provincial Total Factor Production: Application of Geographically Weighted Regression from a Spatial Temporal Perspective[J].Journal of Chinese Soft Science,2009,16(11):160-170.

[7] Barry R R K,Clapp J M,Rodriguez,M.Spatiotemporal Autoregressive Models of Neighborhood Effects[J].Journal of Real Estate Finance and Economics,1998,17(1):15-33.

[8] Pace R K,Barry R,Gilley O W,Sirmans C F.A method for spatial-temporal forecasting with an application to real estate prices [J]. International Journal of Forecasting,2000,16(2):229-246.

[9] Sun H,Tu Y,Yu S.A spatio-temporal autoregressive model for multi-unit residential market analysis[J].The Journal of Real Estate Finance and Economics,2005,31(2):155-187.

[10] Huang B,Huang W,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.

[11] Dubé J, Legros D.A spatio-temporal measure of spatial dependence: An example using real estate data [J]. Papers in Regional Science, 2013, 92(1): 19-30.

[12] Ingrid Nappi-Choulet,Tristan-Pierre Maury.A Spatial and Temporal Autoregressive Local Estimation for the Paris Housing Market[J].Journal of Regional Science,2011,51(4):732-750.

[13] Dubé J,Legros D.Dealing with spatial data pooled over time in statistical models[J].Letters in Spatial and Resource Sciences, 2013,6(1):1-18.

[14] Smith T E,Wu P.A spatio-temporal model of housing prices based on individual sales transactions overtime[J]. Journal of Geographical Systems,2012,11(4):333-355.

[15] Kumar S,Lal R,Liu D.A geographically weighted regression kriging approach for mapping soil organic carbon stock[J]. Geoderma,2012,189(190):627-634.

[16] Robinson C,Schumacker R E.Interaction effects:Centering,variance inflation factor,and interpretation issues[J]. Multiple Linear Regression Viewpoints,2009,35(1):6-11.

An Improved Spatiotemporally Weighted Regression Analysis Based on Spatiotemporal Windows: A Case Study of Housing Price of Huangshi City, Hubei Province

WANG Xin-gang, KONG Yun-feng

(College of Environment and Planning, Henan University, Kaifeng, Henan 475001, China)

Abstract: Geographically weighted regression (GWR) is a useful technique for exploring spatial nonstationarity by calibrating a regression model which allows different relationships to exist at different points in space. However, spatial autocorrelation can invalidate the model assumption and sometimes may result in residual dependency. This article aims to improve the spatiotemporal weighted regression (GTWR) by introducing additional variables based on spatiotemporal windows. The size parameters for defining spatiotemporal windows are estimated by spatial and temporal statistics of all the sample data. The new window variables are calculated by averaging the explained variables which are located in its spatio-temporal window. The new variables are added in GTWR as an improved regression (IGTWR) model. Huangshi City, a resource-dependent, land-scarce and multi-center city in Hubei Province, is selected as the study area. 19 300 commercial housing units and 398 buildings in 2007-2012 are collected as sample data. Based on general and spatial statistics, the number of building floors, the plot ratio, the greening ratio, the level of property management, the distance to region center, and the year of sale are selected as explanatory variables. The sample data are analyzed by four regression models respectively: ordinary linear regression (OLS), GWR, GTWR and IGTWR. The optimum size and number of spatiotemporal window are estimated by the Moran's I index and the correlation coefficients between temporal sequences. Modeling results indicate that both the IGTWR model and its variables pass the statistical test at the significant level 0.05. The spatial distribution of the variable coefficients can be explained reasonably. The comparison of all modeling results shows that GWR is better than OLS, GTWR is better than GWR, and IGTWR is better than GTWR, in terms of the measure of goodness of fit (R^2), the Akaike information criterion (AICc), the residual sum of squares (RSS) and the mean squared error (MSE). In the case study, compared with GTWR, the R^2 , AICc, RSS and MSE from IGTWR are improved from 0.877 to 0.919, 6 226 to 6 206, 49 996 201 to 32 327 472 and 354.427 to 284.969 respectively. The case study indicates that the IGTWR model is effective for temporal and spatial analysis of urban housing price. By introducing window based indicators in GTWR model as new variables, the IGTWR model may estimate the impact of spatial and temporal autocorrelation between geographic data, and thus is able to reduce the model error and increase the model accuracy.

Key words: spatio-temporal windows; spatiotemporally weighted regression (GTWR); housing price; Huangshi City