

空间计量模型在房地产批量评估中的应用研究

——以昆明市盘龙区二手房挂牌价为例

■ 金 杰 罗婷婷

【摘 要】利用 2019 年 9 月至 2020 年 3 月昆明市盘龙区 340 个住宅小区二手房挂牌价格样本数据构建经典特征价格模型,并以此为基础构建空间滞后模型和空间误差模型进行对比研究,寻求适合昆明市盘龙区二手房住宅最佳的批量评估模型。结果表明:昆明市盘龙区二手房住宅价格存在着空间自相关性,空间误差模型和空间滞后模型对经典特征价格模型均有所改进,空间误差模型的统计检验指标和估计精度最优;在进行房地产批量评估时应用空间计量模型能够考虑房地产之间的空间相关性,降低评估误差从而提高模型估计准确性。

【关键词】批量评估 特征价格模型 空间滞后模型 空间误差模型

一、引言

随着房地产税开征预期和房地产评估的需求日益增多,在大数据时代,批量评估系统的未来可挖掘发展的潜力巨大。国际估价官协会(IAAO)将批量评估定义为利用共同数据,标准化的方法和统计检验技术评估一组财产确定日期价值的活动。批量评估产生于税价评估中对一致性和统一性的需求,克服单独个案评估的成本高、效率低、税负不公、税收流失等问题,实现低成本、高效率、大规模完成评估任务。

因此,本文将考虑房地产之间的空间相关性,分别建立以经典特征价格模型和空间计量模型为基础的批量评估模型,并进一步分析不同的模型在批量评估应用中评估结果的科学性与合理性,以期更好的运用空间计量模型进行房地产批量评估。目前批量评估技术主要有四种方法:多元回归分析,适应估价技术,时间序列分析和人工神经网络。大量学者运用不同方法分别进行了评估实证研究,然而,传统特征价格模型容易存在变量遗漏或者假设各样

本点是相互独立的等模型设定不当问题,导致估计结果出现偏差和特征指标存在内生性。因此,很多学者认为不同房地产样本之间的空间相关性不应忽视,且对不同空间计量模型进行了比较研究。基于经典特征价格模型进行房地产的批量评估,重点研究住宅异质性对房地产价格的影响,而忽略了住宅价格之间的空间依赖性,且大多利用空间计量模型做房价影响因素分析,将之在房地产批量评估方面的应用研究较少。住房价格与区位因素密切相关,因此在建立批量评估模型时不仅应将空间因素考虑进去,而且适用不同地区且最优的房地产批量评估空间计量模型更值得研究。

二、理论模型

(一) 经典特征价格模型

目前利用经典特征价格模型进行房地产批量评估的模型形式主要分为以下三种:

(1) 线性函数形式:

$$p_i = b_0 + \sum_{i=1}^k b_i X_{iu} + \varepsilon_i \quad (1)$$

(2) 半对数函数形式:

$$\ln p_i = b_0 + \sum_{i=1}^k b_i X_{it} + \varepsilon_i \quad (2)$$

(3) 对数函数形式:

$$\ln p_i = b_0 + \sum_{i=1}^k b_i \ln X_{it} + \varepsilon_i \quad (3)$$

式中: p_i 为第 t 个住宅样本的价格; X_{it} 为第 t 个住宅样本的特征变量; b_0 为常数项; $b_1, b_2, b_3, \dots, b_k$ 为不同特征变量的估计系数; ε_i 为随机误差项。

(二) 空间计量模型

在建立空间计量模型之前要定义空间权重矩阵 W 来表示样本空间数据的空间依赖性, 空间权重矩阵 W 是 $n \times n$ 的矩阵, 矩阵元素代表房地产之间的关系强弱, 其一般形式如下:

$$W = \begin{pmatrix} 0 & w_{12} & L & w_{1n} \\ w_{21} & 0 & L & w_{2n} \\ M & M & 0 & M \\ w_{n1} & w_{n2} & L & 0 \end{pmatrix} \quad (4)$$

其中, w_{ij} 为空间权重矩阵的每一个元素, 其表示不同区域的距离。空间权重矩阵主要包括相邻空间权重矩阵和距离空间权重矩阵, 最常用的距离函数为“相邻”, 即如果区域 i 与区域 j 有共同的边界, 则 $w_{ij}=1$, 反之, 则 $w_{ij}=0$ 。本文在进行空间计量模型建立时使用的是 Queen 邻接矩阵。

在使用空间计量方法前要考察数据是否存在空间依赖性, “莫兰指数 I ” (Moran's I) 是度量空间自相关最常用的方法:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (5)$$

其中: $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$ 为样本方差, w_{ij} 为空间权重矩阵的 (i, j) 元素 (用来度量区域 i 与区域 j 之间的距离), 而 $\sum_{i=1}^n \sum_{j=1}^n w_{ij}$ 为所有空间权重之和。如果空间权重矩阵为行标准化, 则 $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = n$, 此时莫兰指数 I 可写作:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x})^2} \quad (6)$$

莫兰指数 I 的取值一般介于 -1 到 1 之间, 大于 0 表示正自相关, 即高值与高值相邻、低值与低值相邻; 小于 0 表示负自相关, 即高值与低值相邻。

空间计量模型中常见的两种模型是空间滞后模型和空间误差模型, 其中空间滞后模型中引入了变量的空间滞后项, 其一般化表达式为:

$$y = \lambda W y + X \beta + \varepsilon \quad (7)$$

式中, y 代表样本的被解释变量, W 为空间权重矩阵, $W y$ 表示空间滞后的被解释变量; λ 为空间回归系数; X 为 $n \times k$ 解释变量矩阵, 解释变量为 k 列; $\beta_{k \times 1}$ 表示解释变量 X 对被解释变量 y 的影响系数; ε 为随机误差向量。

空间误差模型是将空间依赖性体现在误差项中, 其一般化表达式为:

$$y = X \beta + \mu \quad (8)$$

其中误差项 μ 存在空间依赖性, 其空间自回归表达形式为:

$$\mu = \rho W \mu + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n) \quad (9)$$

在误差项 μ 的回归式中, W 为空间权重矩阵, 参数 ρ 为误差项 μ 的空间相关系数, ρW 则表示样本的空间误差项。

三、实证研究

(一) 数据来源

本文通过安居客, 房天下等昆明市房地产行业网站采集了 2019 年 12 月 -2020 年 3 月的昆明市盘龙区 380 条真实二手房挂牌数据 (剔除了缺失值和异常值), 随机筛选出 340 个样本用于构造模型, 剩下的 40 个样本数据用来检验模型的优劣。样本数据来自于 136 个住宅小区, 建筑面积限定在 50-200m² 之间, 卧室数和客厅数至少满足 1 室 1 厅的要求, 建筑年代考虑商品房发展舍去建筑年代比较久远的, 将其限定在 1998 年之后建成的小区。

(二) 特征变量的选择及量化

在建立模型特征变量选择方面, 根据昆明市盘龙区二手房的市场特点, 参考王娟娟, 毛博研究统计的特征变量文献使用次数和指标显著次数表, 从建筑特征、区位特征和邻里特征三方面选取 19 个特征变量, 特征变量及具体量化方法如表 1:

首先, 对样本数据描述性统计分析, 如表 2: 剔除无效数据及缺失值后, 选取的昆明市盘龙区 340 个小区样本。二手房挂牌总价的平均价格为 141.51

表 1 特征变量及其量化表

	特征变量	变量量化方法	预期影响
实际变量	卧室数	房屋实际卧室数 (个)	+
	客厅数	房屋实际客厅数 (个)	+
	建筑面积	房屋建筑实际面积 (m2)	+
	建筑年代	建成年份至 2020 年年数	-
	容积率	小区的容积率	-
	绿化率	小区的绿化率	+
	物业费	小区的物业费 (元)	+
	距离小区最近的公交距离	通过百度地图 GPS 测距 (km)	-
	距离小区最近的地铁距离	通过百度地图 GPS 测距 (km)	-
	距离小区最近的公园距离	通过百度地图 GPS 测距 (km)	-
	距离小区最近的三甲综合医院距离	通过百度地图 GPS 测距 (km)	-
	距离小区最近的普通医院的距离	通过百度地图 GPS 测距 (km)	-
	距离小区最近的商圈的距离	通过百度地图 GPS 测距 (km)	-
	距离小区最近的重点小学的距离	通过百度地图 GPS 测距 (km)	-
虚拟变量	房屋朝向	没有朝南方向：1；有一面朝南：2；南北通透：3	+
	装修情况	毛坯：1；简装修：2；中装修：3；精装修：4； 豪华装修：5	+
	所处楼层	低层：1；中层：2；高层：3	未知
	建筑类别	板楼：1；其他类别：0	+
	学区信息	普通学区：1；二类优质学区：2； 一类优质学区：3	+

万元，最大值为 400 万元，最小值为 63 万元。

续表

(三) 经典特征价格模型建立

通过 SPSS 软件对三种常见的传统特征价格模型函数进行分析对比，如表 3，选择调整的拟合优度最优的经典特征价格模型。

表 2 特征变量描述性统计分析表

特征变量	最小值	最大值	平均值	标准差
建筑面积	55	189	103.61	22.787
房屋总价	63	400	141.51	42.097
朝向	1	3	2.24	0.796
楼层	1	3	2.05	0.81
装修情况	1	5	3.82	0.722
建筑年代	-1	30	12.59	5.915
建筑类别	0	1	0.89	0.319
容积率	0.35	6.07	2.7641	1.284
绿化率	20	70	42.88	8.107
物业费	0.4	3.8	1.239	0.5627

特征变量	最小值	最大值	平均值	标准差
最近的公交站 距离	0.047	0.809	0.21305	0.1220
最近的地铁 距离	0.088	3.3	1.07549	0.7193
最近的公园 距离	0.056	4.2	0.76678	0.4351
最近的三甲医 院距离	0.258	7.3	2.89351	1.4155
最近的普通医 院距离	0.034	4.7	1.03721	0.9524
最近的商圈 距离	0.16	10.4	3.53332	1.9409
学区信息	1	3	1.29	0.614
最近的重点小 学距离	0.066	4.4	0.82563	0.7511
卧室数	1	4	2.83	0.632
客厅数	1	3	1.96	0.228

表 3 三种特征价格模型对比

线性模型			半对数模型			全对数模型		
特征变量	B	Sig	特征变量	B	Sig	特征变量	B	Sig
常数项	-1.534	0.942	常数项	3.934	0.000	常数项	-0.033	0.906
建筑面积	1.314	0.000	建筑面积	0.009	0.000	Ln（建筑面积）	0.912	0.000
朝向	-0.204	0.909	朝向	-0.005	0.650	朝向	-0.007	0.534
楼层	1.833	0.269	楼层	0.007	0.546	楼层	0.004	0.671
装修情况	4.719	0.013	装修情况	0.033	0.009	装修情况	0.051	0.000
建筑年代	-2.065	0.000	建筑年代	-0.017	0.000	Ln（建筑年代）	-0.125	0.000
建筑类别	8.233	0.056	建筑类别	0.051	0.073	建筑类别	0.028	0.297
容积率	-0.927	0.499	容积率	-0.011	0.246	容积率	-0.005	0.512
绿化率	0.376	0.031	绿化率	0.003	0.006	Ln（绿化率）	0.210	0.000
物业费	-5.187	0.146	物业费	-0.041	0.085	物业费	-0.024	0.306
最近的公交站	7.410	0.522	最近的公交站	0.075	0.330	Ln（最近的公交站）	0.036	0.031
最近的地铁	-6.270	0.004	最近的地铁	-0.042	0.004	Ln（最近的地铁站）	-0.042	0.002
最近的公园	-0.127	0.969	最近的公园	0.004	0.856	Ln（最近的公园）	0.010	0.506
最近的三甲医院	-2.092	0.089	最近的三甲医院	-0.011	0.167	Ln（最近的三甲医院）	-0.023	0.238
最近的医院	-3.635	0.087	最近的医院	-0.030	0.031	Ln（最近的医院）	-0.043	0.001
最近的商圈	-5.160	0.000	最近的商圈	-0.036	0.000	Ln（最近的商圈）	-0.132	0.000
学区信息	7.298	0.003	学区信息	0.041	0.012	Ln（最近的重点小学）	0.001	0.959
最近的重点小学	1.713	0.539	最近的重点小学	0.007	0.686	学区信息	0.032	0.032
卧室数	5.877	0.065	卧室数	0.052	0.014	卧室数	0.047	0.020
客厅数	1.541	0.798	客厅数	0.037	0.359	客厅数	0.054	0.157

表 4 三种特征价格模型统计检验对比表

函数形式	R ²	调整的 R ²	标准估算误差	方差分析检验	D-W	显著变量个数
线性形式	0.699	0.681	23.777	0.000	1.224	7
半对数形式	0.715	0.699	0.1574	0.000	1.213	9
全对数形式	0.745	0.729	0.1493	0.000	1.259	10

通过对比三种函数形式的系数以及检验结果，如表 4，得出全对数形式的特征价格模型的拟合优度最高，标准估算误差最小，能更好地应用在盘龙区

二手房批量评估模型中，因此在下文建立空间计量模型时将以全对数模型为基础。

在建立回归模型时，自变量选择的基本指导思想是少而精，易于评估，因此将上述全对数模型进行逐步回归得到最优回归子集，最终得出运用在建立空间计量模型中的特征变量（10 个）分别为：Ln（建筑面积）、Ln（最近的商圈距离）、Ln（建筑年代）、装修状况、Ln（绿化率）、Ln（最近的地铁站）、卧室数、Ln（最近的普通医院距离）、学区信息、Ln（最近的公交站）。

（四）空间计量模型建立

1. 空间自相关性检验

表 5 Moran's I 检验结果表

variables	I	E (I)	sd (I)	z	p-value*
price	0.124	-0.003	0.019	6.547	0.000

通过分析表 5 可得：Moran's I 为 0.124 大于 0，在显著性水平为 0.05 时拒绝原假设，说明盘龙区二手房价格存在正的空间自相关性，盘龙区二手房价格在空间上表现为高 - 高集聚或者低 - 低集聚，在进行批量评估时有必要采取空间计量模型对传统价格模型进行修正。

表 6 LM 检验

TEST	MI/DF	VALUE	PROB
Lagrange Multiplier (lag)	1	43.623	0.00000
Robust LM (lag)	1	6.0714	0.01374
Lagrange Multiplier (error)	1	55.6321	0.00000
Robust LM (error)	1	18.0506	0.00002

由表 6 可知，空间滞后模型和空间误差模型均通过了极大似然 LM-lag 和 LM-error 检验，表明昆明市盘龙区二手房样本数据存在空间相关性，在显著性水平为 0.01 时，空间滞后模型未通过 Robust LM-lag 检验，因此，建立空间误差模型能更好得解释昆明市盘龙区二手房房价。

2. 模型检验与对比分析

根据前文分析建立房屋总价与 10 个特征变量的经典特征价格模型，在此基础上利用 Geoda 软件建立空间滞后模型与空间计量模型。通过对比选择 Queen 邻接权重矩阵作为空间权重矩阵。三种回归模型对比分析如表 7。

分析表 7 可知，三种模型中除距离小区最近的公交站距离未通过显著性检验且系数不符合经济意义预期外，其他系数估计均通过显著性检验，一般来说，小区距离公交站越近，便于居民出行，由此，小区房价会高，而在模型中，小区距离公交站的距离对房屋价格影响不显著。分析评估案例原因有以下两点：一是案例中更多选择私家车或者地铁到达目的地，公交车站到小区的距离对房价没有显著性

影响；二是住宅价格考虑主要影响为安静的环境，大部分小区公交的距离非房价主要考虑因素，这点从下述最近的商圈距离因素也得到印证。

在三个模型中，建筑面积对二手房总价的影响均为正向，且均通过显著性检验，系数的估计与预期相同。而空间误差模型的 Log likelihood 值大于空间滞后模型和经典特征价格模型，而其 AIC=-351.186 为三个模型中最小的，由此判断出空间误差模型较另外两个模型对昆明盘龙区二手房价格具有更好的解释性。

因此，根据空间误差模型可知：盘龙区房屋面积越大房价越高，Ln（建筑面积）每增加一个单位，Ln（房屋总价）平均增加 0.861 个单位；建筑年代与房屋价格呈现负相关影响，建筑年代越久远，房价相对较低，Ln（建筑年代）每增加一个单位，Ln（房屋总价）平均下降 0.106 个单位；商圈与小区的距离对房屋价格产生负向影响，Ln（商圈距离）每增加一个单位，Ln（房屋总价）将会平均下降 1.262 个单位，一定程度是由于商圈产生的嘈杂环境；二手房的装修情况等级与房屋价格呈正相关关系，一个住宅装修程度越完善相应房价更高，房屋的装修等级每提高一个层次，Ln（房屋总价）平均升高 0.056 个单位；伴随着时代的发展，人们越来越要求交通便利，在其他条件不变的情况下，Ln（地铁距离小区距离）每增加 1 个单位，Ln（房屋总价）平均减少 0.041 个单位；另外，小区的学区每提高一个级别，Ln（房屋总价）平均增加 0.053，说明现在人们更加注重子女教育，在购房时更倾向于学区房。

（五）批量评估模型结果对比分析

将剩下的 40 个盘龙区住宅小区的相关特征变量数据分别代入经典特征价格模型和空间误差模型中分别得出住宅评估价格并与其价格进行对比。通过对比发现：利用空间计量模型进行批量评估的平均误差为 9.81%，标准差为 0.143，其中最小的评估误差为 0.2%；利用经典特征价格模型进行批量评估时的平均误差为 15.1%，标准差为 0.163，其中最小的评估误差为 2%。

根据《IAAO 房产批量评估准则》中常见的指标：比率中位数（M）、离散系数（COD）、价格相关差（PRD）对两个评估模型得出的评估模型的合理性进行对比分析，根据表 8 得出：空间计量模型评估的

表 7 OLS 模型、SLM 模型、SEM 模型结果对比

特征变量	经典特征价格模型 (OLS)		空间滞后模型 (SLM)		空间误差模型 (SEM)	
	B	Sig	B	Sig	B	Sig
常数项	-0.0360	0.8890	-1.8522	0.0000	0.1631	0.5086
Ln (建筑面积)	0.9210	0.0000	0.8858	0.0000	0.8608	0.0000
Ln (最近的商圈距离)	-0.1360	0.0000	-0.1123	0.0000	-0.1262	0.0000
Ln (建筑年代)	-0.1040	0.0000	-0.0967	0.0000	-0.1061	0.0000
装修情况	0.0520	0.0000	0.0515	0.0000	0.0559	0.0000
Ln (绿化率)	0.2010	0.0000	0.1911	0.0000	0.1998	0.0000
Ln (最近的地铁站距离)	-0.0410	0.0020	-0.0271	0.0301	-0.0408	0.0340
卧室数	0.0470	0.0160	0.0543	0.0030	0.0584	0.0010
Ln (普通医院)	-0.0390	0.0010	-0.0354	0.0012	-0.0446	0.0020
学区信息	0.0400	0.0050	0.0489	0.0002	0.0527	0.0004
Ln (最近的公交站距离)	0.0390	0.0160	0.0267	0.0802	0.0311	0.0576
ρ			0.3921	0.0000		
λ					0.5870	0.0000
R-squared	0.7369		0.7616		0.7688	
Log likelihood	169.84		184.559		186.593	
Akaike info criterion	-317.681		-345.118		-351.186	

结果均在比率检验标准合理范围内，经典特征价格模型评估的结果比率中的比率中位数为 1.13，超出了比率检验标准的合理范围。因此，建立的空间误差模型在进行房地产批量评估时更适用，估计误差更小。

为了进一步验证评估的准确性，研究收集了同期昆明市盘龙区三套二手房买卖交易成交价格（市场比较法案例），与经典特征价格模型以及空间误差模型进行估价对比分析，案例评估结果如表 9 所示，通过对比可以看出，不仅空间误差模型相较于单宗房地产价值评估效率更高，而且评估的误差相对另外两种方法更小。

表 8 OLS 模型与 SEM 模型比率检验指标对比一览表

	比率中位数	离散系数	价格相关差
比率分析检验标准	0.9-1.1	5-15	0.98-1.03
传统特征价格模型	1.13	11.1	1.019
空间误差模型	1.08	9.7	1.018

表 9 案例评估对比表

单位：元

	市场法 成交价格	经典特征价 格模型 批量评估价 格	空间误差 模型 批量评估 价格	挂牌价
案例 1 评估值	885 000	95 707.25	905 995.19	920 000
案例 2 评估值	1 728 000	1 859 219.07	1773694.12	1 800 000
案例 3 评估值	1 180 000	1 272 317.83	122290.25	1 230 000

四、结论

本文结合房地产评估的特征价格理论和空间计量模型对昆明市盘龙区二手房住宅的价格进行了实证研究和评估。首先通过 Moran 指数和 LM 检验证明昆明市盘龙区二手房房价存在误差项的空间自相关性并且二手房价格在空间上表现为高 - 高集聚或

者低-低集聚;其次将经典特征价格模型,空间滞后模型,空间误差模型进行对比,根据AIC准则以及Log likelihood的结果表明空间误差模型在对所选取地区样本具有更好的解释性,且在批量评估结果误差分析中,空间误差模型得出的评估值与真实值之间的误差相比更小,且比率检验都在合理范围内。由此,本文认为在房地产批量评估方面,空间误差模型一方面克服了传统特征价格模型未考虑样本之间空间自相关的特性,另一方面能够提高房地产批量评估的效率与准确性。

【参考文献】

- [1] 纪益成,傅成锐.从价税的税基评估方法[J].中国资产评估,2005(11):5-9
- [2] 朱晓,蒋文军.特征价格法在涉税房地产批量评估中的应用[A].中国房地产估价与经纪2012(1).
- [3] 王莎莎.基于回归分析的长沙市房产税税基批量评估实证研究[D].西南财经大学,2013.
- [4] 肖雅.基于特征价格模型的房地产评估——以昆明市为例[D].云南大学,2018.
- [5] 张奎洋.基于空间特征法的房地产批量评估研究[D].江西财经大学,2019.
- [6] 姚丽,谷国峰,王建康.基于空间计量模型的郑州城市新建住宅价格模型影响因素分析[J].经济地理,2014(34):70-74
- [7] 廖志远.基于空间计量模型的我国商品房价格影响因素研究[D].湖南大学,2015.
- [8] 王少剑,王洋,蔺雪芹,张虹鸥.中国县域住宅价格的空间差异特征与影响机制[J].地理学报,2016(71):1229-1334.
- [9] 刘玲妙,刘建华,乔红芳.我国房地产价格影响因素的区域差异——基于空间计量[J].闽南师范大学学报,2018(4):67-74.
- [10] 王健,景霖霖,彭山桂,张振,吴群.基于空间计量模型的商品住宅开发商定价互动影响及其空间分异研究[J].中国土地科学,2019,33(1):34-31.
- [11] 纪益成,马铮等.基于空间误差模型的房地产批量评估研究[J].建筑经济,2015,36(02):62-67.
- [12] 李颖丽,刘勇,刘秀华.重庆市主城区住房价格影响因子的空间异质性[J].资源科学,2017,39(2):335-345.
- [13] 王娟娟,毛博.基于特征价格模型(HPM)的房地产评估研究综述[J].行政事业资产与财务,2016(01):84-86.
- 英文文献略
- (本文是云南省哲学社科规划项目(YB2019013)“地价房价联动视角的云南房地产市场平稳健康发展研究”,云南财经大学科学研究基金(80059900266)“基于GIS空间优化的昆明市保障性住房建设长效机制研究”的阶段性研究成果)
- (作者单位:云南财经大学城市与环境学院)

