

基于多元回归法的武汉市二手房价格影响因素研究

缪格^{1*} 李英冰^{1,2} 袁菲¹

(1. 武汉大学测绘学院,湖北 武汉 430079; 2. 武汉大学测绘学院时空大数据研究中心,湖北 武汉 430079)

摘要: 二手房价格受到区域因素、房屋自身因素、市场因素等多种因素综合影响,传统估价方法过多关注于房地产自身属性的差异及修正,有较强的主观性。本文以特征价格模型为基础,采用多元线性回归模型和多元对数回归模型,建立二手房价格与区位因素、房屋自身因素等影响因子之间的函数关系,并采用武汉市二手房挂牌数据、看房价网(www.kanfangjia.cn) 数据和百度热点数据,对武汉市二手房各个特征变量进行筛选、量化和整合,并对比分析不同区域各个特征变量对二手房价格的影响程度,实验结果与实际情况符合度较好。

关键词: 二手房; 特征价格; 价格评估; 影响因素; 多元回归法

1 引言

二手房价格是政府和广大人民的关注焦点,而如何给出相对客观价格,受到研究人员和市场交易的广泛关注。传统的房价评估方法常用的是市场比较法、收益还原法和成本法,但在实践中往往依赖评估师的经验,对数理模型的运用比较少^[1]。Cour 最早提出特征价格模型,当时多用于耐用消费品的定价^[2]。特征价格模型从产品的差异性出发,把二手房产品价格分解为各个特征价格,核心在于建立二手房特征与价格之间的函数关系^[2]。特征价格模型将房屋价格特征化能让消费者更能理解房价构成,购房者可以根据自己的需求与经济状况有针对性地选择自己想要的居住环境;此外,从消费者的需求角度出发,从房屋特征的角度看待房屋价格,在一定程度上反映出消费者对房屋特征变量的关心程度,使房价的评估更为简易,为政府等有关部门提供借鉴^[3]。

特征价格模型自提出以来,被国内外研究不断深入且紧密联系实际情况开展研究。1999 年, Paliwal 等基于特征价格模型,引入三角模糊数法衡量人们对土地购买的意愿^[4]; Peterson(2009) 将人工神经网络法与特征价格模型相结合^[5]; Badi H. Baltagi(2011) 用极大似然估计拟合了特征价格模型^[6]。国内学者最早是由中国人民大学蒋一军、龚江辉(1996) 利用特征价格方法提出了计算异质商品价格指数的方法,并将其应用于房地产价格指数的计算^[7]。温海珍(2003) 开启了国内特征价格模型实证研究之路,对杭州市二手房交易案例进行数据整合,

研究特征影响变量^[8]; 黄古博、李雨真(2011) 应用主成分法改进二手房的特征价格模型,排除变量间因多重共线性产生的误差^[9]; 李恒凯(2012) 结合地理信息系统(GIS) 和主成分分析(PCA) 算法研究如何提高二手房特征价格模型函数拟合效果及准确性^[10]。

二手房价格往往受到多种因素的影响,特征价格模型将二手房价格分解为各个特征价格,得出各个特征对二手房价格的影响,从而实现二手房价格的估算。本文采用特征价格模型,从消费需求出发,用多元回归分析法将房价分解为各个特征价格来研究房价的影响因素,其中通过逐步回归和变量整合以避免变量间的多重共线性,从理论上分析二者的函数关系。

2 特征价格模型的构建

2.1 特征价格模型理论及参数估计

二手房的真正效用源自于其包含的各个分散的功效特征,如面积、楼龄、朝向,如果二手房所包含的特征不同,则相应的价格也不同。消费者愿意为二手房的某个特征的每单位增加而支付的额外费用,则称为该特征的隐含价格(特征价格),它反映了各个特征对二手房价格的贡献程度。构建回归函数是特征价格模型的研究的核心, Butler 理论阐述房屋价格受区位特征(L)、建筑特征(S)、邻里环境(N) 三大特征变量影响^[11]:

$$P = f(L, S, N) \quad (1)$$

在其他条件不变前提下,将每个特征求偏导即为

* 收稿日期: 2016—08—17

作者简介: 缪格(1994—),女,硕士研究生,研究方向为时空大数据分析。

基金项目: 武汉市住房保障和房屋管理局信息化建设项目“智慧房管”主数据库设计及一期建库 WHZC-2014-075A(2210、2415、2416)

边际隐含特征价格。国外文献出现最频繁的模型变量为面积、总层数、装修程度、有无 CBD、公交站、有无地铁、楼龄、教育设施与生活、娱乐设施等^[1]。在函数形式的设定中,最常用的函数形式为线性形式和对数形式,计算公式分别为式(2)和式(3):

$$P = \alpha_0 + \sum \alpha_i Z_i + \varepsilon \quad (2)$$

$$\ln(P) = \alpha_0 + \sum \alpha_i \ln(Z_i) + \varepsilon \quad (3)$$

其中 P 是特征价格, Z_i 是住房的特征变量, α_i 是常数项, ε 是随机干扰项。

对数模型中,自变量和因变量以对数形式进入模型,则回归系数对应着特征的价格弹性。对数模型是假设函数经过显性化处理后得到的表达,对数模型考虑到了二手房特征应当遵循边际效用递减规律,即房屋价格随某种特征的增加而增加,但增加速率会越来越慢,对数模型能较好地反映价格与各个特征之间的函数关系。

经典线性计量经济学模型最常用的最小二乘法实现参数估计法的应用。参数的最小二乘估计(OLS):选择合适的参数 α_i 使得全部样本值的残差平方和(Rss)最小,即

$$\min \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \quad (4)$$

$\sum e_i^2$ ——为残差平方和

Y_i ——为样本的真实值

\hat{Y}_i ——为样本估计值

线性、无偏性、最小方差性是运用最小二乘法所得的多元线性回归参数估计值所具有的特性^[2]。

2.2 模型变量的选取与量化

选取房价相应的解释变量是在对武汉市二手房价格分析建模前必须考虑的步骤。根据已有的数据及后

期可以获取到的数据,结合国内外特征价格模型常用的解释变量,初步选取室厅、面积、有无景点、首付、层数、有无车库、总层数、朝向、楼龄、有无地铁、装修程度、学校、公园、CBD、公交线路、医院、商场这 17 个解释变量研究。

特征变量,详细信息主要包括室厅、面积、总层数、朝向、装修程度、楼龄、学校、公园、医院、商场、中央商务区(CBD)、有无地铁等解释变量。

二手房的特征变量分定量、定性变量,为了从数理模型是研究各个特征变量对二手房价格产生的影响,需要将各个变量进行量化整合,不同类别的变量量化方式不同,如表 1 所示。

住宅解释变量的初步选取 表 1

特征分类	特征变量	可供调查的内容
区位特征	CBD(中央商务区)	5 000 m 内是否有 CBD
	公交线路	500 m 公交站数目
	地铁	2 000 m 内是否有地铁
建筑特征	朝向	房间的主要朝向
	房间数目	室厅总数
	面积	建筑面积/m ²
	房龄	竣工至今年数
	装修程度	装修程度
	楼层	所在楼层的层数
	车位	住宅小区内是否有车库
	总层数	所在楼栋的总层数
邻里特征	教育配套	有无幼儿园、小学、中学(500 m)
	生活服务配套	有无 shopping 中心、医院、湖泊、公园、景点(500 m)

(1) 定性变量的量化

定性变量用三种方法来量化,即虚拟变量量化法、李克特量表量化法和综合性指标量化法^[3],具体量化法如表 2 所示:

定性变量量化表 表 2

定性变量量化法	变量名称	量化指标	数据来源	预期影响
二元虚拟变量法	CBD	临近 CBD 赋值为 1,其他为 0	百度热点查询	+
	地铁	临近地铁赋值为 1,其他为 0	百度热点查询	+
	朝向	南北朝向赋值为 1,其他为 0	挂牌数据	+
李克特量表法	装修程度	未装修(1分)、普通装修(2分)、精装修(3分)、豪华装修(4分)	挂牌数据、看房价网查询	+
综合性指标法	生活娱乐配套	车库、商场、医院、湖泊、景点、公园,每项计一分,共 6 分	百度热点、看房价网查询	+
	教育配套	幼儿园、小学、初中、高中、大学,每项计一分,共 5 分	百度热点查询	+

表 2 中,地铁凸显了交通的便捷程度,因此二手房周围有地铁也会让房价相应提升,预期影响为正;其次,将装修程度量化为四个等级,装修程度越高,表明住宅的成本价格越高,住宅价格也越高。

当特征变量选取较多时,自变量之间可能存在较大的相关性,为了避免自变量的多重共线性问题,采用综合

性指标量化法来进行特征变量的整合以减少变量数目。

(2) 定量变量的量化

二手房价格、面积、楼龄、室厅数、层数、总层数、公交站数这 7 个定量变量直接用实际数值(对数模型中对二手房面积进行简单的自然对数变换)进行量化,如表 3 所示,变量内涵也比较简单客观^[1]。

定量变量量化表 表 3

变量名称	量化指标	数据来源	预期影响
二手房价格	一套二手房的出售挂牌总价	挂牌数据	--
面积	一套二手房的总建筑面积	挂牌数据	+
楼龄	房屋竣工后的时间长度/年	挂牌数据	-
层数	所处的楼层层数/层	挂牌数据	未知
总层数	楼栋的总层数/层	挂牌数据	未知
室厅数	室厅总数	挂牌数据	+
公交站数	周围公交站个数/个	百度热点查询	+

小区周围公交站个数越多,居民出行越方便,住宅价格越高。楼龄是住宅竣工日期到 2016 年的时间段长度,计量单位为年,例如 2015 年竣工的住宅楼龄为 1。楼龄反映建筑本身的新旧与使用程度,楼龄与住宅价格呈负相关关系。

变量经过整合和量化后,初步确定进入模型的特征变量有 12 个。

2.3 模型变量的剔除与筛选

由于初步选择的特征变量较多,特征变量之间可能存在比较严重的线性关系,上述采用综合性指标方法量化部分变量克服自变量之间的多重共线性,本文还采用逐步回归法,逐步回归是向前回归法和向后回归法的结合,首先按自变量对因变量的贡献率进行排序,按照从大到小的顺序选择进入模型的变量。每将一个变量加入或删除模型(步进)时,要进行偏 F 检验,剔除不显著的变量,然后再对留在模型中的变量进行检验,直到没有变量可以纳入,也没有变量可以剔除为止,从而尽量排除变量多重共线的可能性。步进准则中, F 检验的概率 ≤ 0.050 表示自变量对因变量的影响显著,判断结果为进入(enter)模型; F 检验的概率 ≥ 0.100 表示自变量对因变量的影响为不显著,判断结果为移出(remove)模型。

模型构建最终确定的特征变量 表 4

模型	输入的变量	移去的变量	方法
1	面积,朝向,总层数,装修程度,有无 CBD,公交站,有无地铁,楼龄,教育设施,生活、娱乐设施	层数,室厅,朝向	步进(准则: F -to-enter 的概率 ≤ 0.050 , F -to-remove 的概率 ≥ 0.100)。

a. 因变量: 二手房价格

由表 4 可见,被移去的变量有层数,室厅和朝向,可能是因为研究的数据中包含一定量的商铺,使得一些自变量与因变量的函数关系不明显。最终进入回归模型的特征变量共 9 个,分别是面积,总层数,装修程度,有无 CBD,公交站,有无地铁,楼龄,教育设施与生活、娱乐设施。

量化整合修正后共有 9 个特征变量,将其逐步导入模型,当 9 个特征变量全部导入模型时,模型的拟合效果最好。所以,最终确定的解释变量分别是面积,总层数,装修程度,有无 CBD,公交站,有无地铁,楼龄,教育设施与生活、娱乐设施。

3 实证研究

3.1 实验数据说明

武汉市近几年房价成稳定增长趋势,且武汉市各区域房价差异较大,本文选取了 2015 年武汉市 10 个区二手房交易挂牌数据 2 000 个进行研究。

通过对数据的统计分析,删除不符合常理的数据记录后,分别作各个自变量与因变量的箱线图,如图 1 所示:横坐标表示二手房的建成年份(年),反映出二手房的楼龄,纵坐标表示所研究的二手房价格(万元/ m^2)。

箱线图利用数据中的 5 个统计量:最小值、第一四分位数、中位数、第三四分位数与最大值来描述、判断数据异常值,从中鉴别出离群值和极端值。描绘个案偏离箱体边缘(上端、下端)的距离是箱体的几倍,其中“。”表示离群点(1.5 倍~3 倍),“*”表示极端离群点(超过 3 倍),然后进行异常值剔除后,最后采用的样本数为 1 494 个。

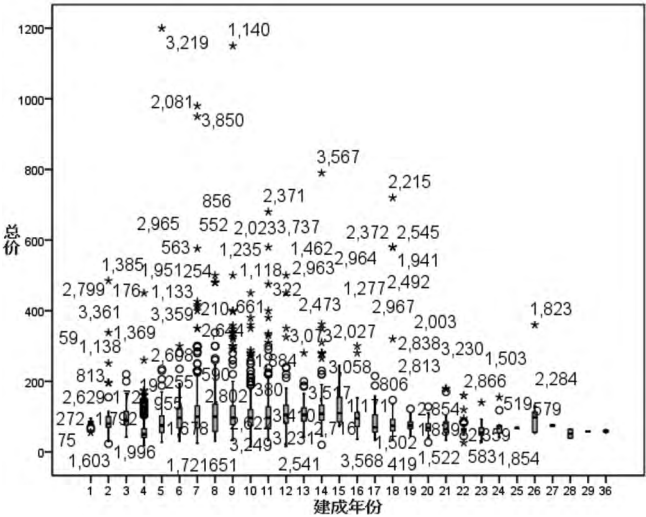


图 1 二手房总价与二手房建成年份箱线图

3.2 拟合分析与自相关 DW 检验

当选取的 9 个特征变量全部进入模型时拟合效果最好,多元回归分析结果如表 5 所示, R 代表复相关系数,在多元回归中,为了避免拟合效果单调随变量 X 的增加而变大, $R(a)$ 更能体现拟合和方程的显著性检验。 R 的平方即 R^2 表示回归方程对样本观测值的拟合程度,越接近于 1,方程的拟合度越高。线性模型与

对数线性模型的复相关系数分别为 0.927、0.935。当给模型增加自变量时,复决定系数也随之逐步增大,当自变量足够多时会得到模型拟合良好,而实际却可能并非如此,于是考虑对 R^2 进行调整,称调整 R^2 ,分别为 0.858、0.873;模型所能解释因变量差异的百分比都大于 85%,自变量和因变量间表现出较良好的线性关系,有较好的模型解释能力。但线性模型的标准估计误差远远高于对数模型,且对数模型的解释力度更强。经对比可见,对数线性模型具有更好的解释力。

估计标准误差是说明实际值与其估计值之间相对偏离程度的指标,主要用来衡量回归方程的代表性,对数模型的误差值远小于线性模型,对数模型方程代表

性更优。

更改后的 F 检验值分别为 905.605、1144.248, sig 显著性水平都为 0.000,小于 0.05,则 F 检验通过,表示统计量显著,方程中自变量因变量间线性关系能够成立,模型对样本数据的拟合在统计上有意义。

多元线性回归分析的一个假定是误差项的相互独立性,要验证在误差之间是否存在自相关联系效应,避免导致线性回归分析出现严重的偏误,所以用较为精确的自相关检验方法——自相关 DW 检验来考察,两个回归方程的 DW 检验值,分别为 1.820 和 1.744,都接近于 2,可认为两个模型中的误差项基本上是独立的,基本不存在异方差问题,如表 5 所示。

多元回归分析结果的拟合优度表

表 5

模型	R	R 方	调整 R 方	标准估计的误差	更改统计量			Durbin-Watson
					R 方更改	F 更改	Sig. F 更改	
线性模型	.927a	.859	.858	11.388	.859	905.605	.000	1.820
对数模型	.935a	.874	.873	.1237	.874	1144.248	.000	1.744

3.3 多重共线性检验与回归方程分析

多元系数回归如表 6 所示。 B 值是指回归系数和截距,常数则代表截距,变量对应的 B 值代表回归系数。

标准误差用来衡量抽样误差,越小表明样本对总体越有代表性,用样本推断总体参数的可靠度越大。经对比可见,对数模型中的标准误差远小于线性模型,说明对数模型二手房价格回归方程房价预测可靠性较好。

标准化回归系数是指去除了因变量 Y 和自变量 X_1, X_2, \dots, X_n 量纲影响之后的回归系数,其绝对值的大小直接反映了 X_i 对 Y 的影响程度。

VIF 为方差膨胀因子,用于检验自变量的多重共线性,标准为 10,超过 10 即表明自变量间存在较严重的多重共线性,表中 VIF 最大为 2.152,远远小于 10,可认为自变量间不存在显著的多重共线性,如表 6 所示。

多元回归系数表

表 6

线性模型					对数模型				
变量	B	标准误差	标准化回归系数	VIF	变量	B	标准误差	标准化回归系数	VIF
(常量)	-54.859	1.921	-	-	(常量)	-0.724	0.064	-	-
总层数	0.485	0.038	0.159	1.652	总层数	0.004	0.000	0.120	1.655
装修程度	4.992	0.349	0.145	1.087	装修程度	0.055	0.004	0.138	1.088
楼龄	-0.635	0.086	-0.099	1.891	楼龄	-0.007	0.001	-0.090	1.899
教育设施	5.714	0.444	0.184	2.152	教育设施	0.066	0.005	0.184	2.152
公交站	0.196	0.034	0.066	1.424	公交站	0.002	0.000	0.068	1.424
有无 CBD	5.537	0.695	0.085	1.193	有无 CBD	0.088	0.008	0.118	1.192
有无地铁	13.837	1.071	0.161	1.646	有无地铁	0.220	0.012	0.224	1.645
生活娱乐设施	3.638	0.286	0.173	1.936	生活娱乐设施	0.041	0.003	0.167	1.937
面积	0.827	0.011	0.727	1.051	ln 面积	0.964	0.013	0.689	1.054

由上述多元线性回归分析回归系数表可知,最终的回归方程为:

(2) 线性模型多元回归方程:

$$P = -54.859 + 0.485X_1 + 4.922X_2 - 0.635X_3 + 5.741X_4 + 0.196X_5 + 5.537X_6 + 13.837X_7 + 3.638X_8 + 0.827X_9 \quad (5)$$

(3) 对数模型多元回归方程:

$$\ln P = -0.724 + 0.004X_1 + 0.055X_2 - 0.007X_3 + 0.066X_4 + 0.002X_5 + 0.088X_6 + 0.220X_7 +$$

$$0.041X_8 + 0.964\ln X_9 \quad (6)$$

式(5)与式(6)为多元线性回归分析最终确定的回归方程,其中 P 代表二手房价格, X_1, X_2, \dots, X_9 分别代表总层数、装修程度、楼龄、教育设施、公交站、有无 CBD、有无地铁、生活、娱乐设施和面积。对各个特征变量对房价的预期影响进行回归参数公式符号的验证分析,验证结果符合预期值。经过对比,线性模型中各项标准误差较大,可能是因为定量变量面积的数值较大,从而引起了数值较小的定性变量的回归参数的畸变。对数模

型各个回归参数的标准误差远远小于线性模型,在一定程度上反映出对数模型有更好的解释能力。

3.4 残差分析与模型应用

(1) 残差分析

在模型假定基础上建立判定系数的计算和显著性检验,这就考虑到模型的假定基础的真实性问题。如果模型的假定不真实,那么模型的适用性就遭受怀疑,残差分析是证实模型假定真实性的基本方法,两个模

型的绝大多数标准化残差随机落于-2 和+2 之间,满足方差齐次性的假设,可认为两个模型的假定都合理。

(2) 参数回代差价分析

目前共有 1 494 个武汉市二手房样本案例,通过所得出的两个模型回归公式做参数回代,得出计算出的理论价格,并且与相应的实际二手房价格做差价作为横坐标,纵坐标为差价数值出现的频率,形成频率分布直方图,如图 2 所示:

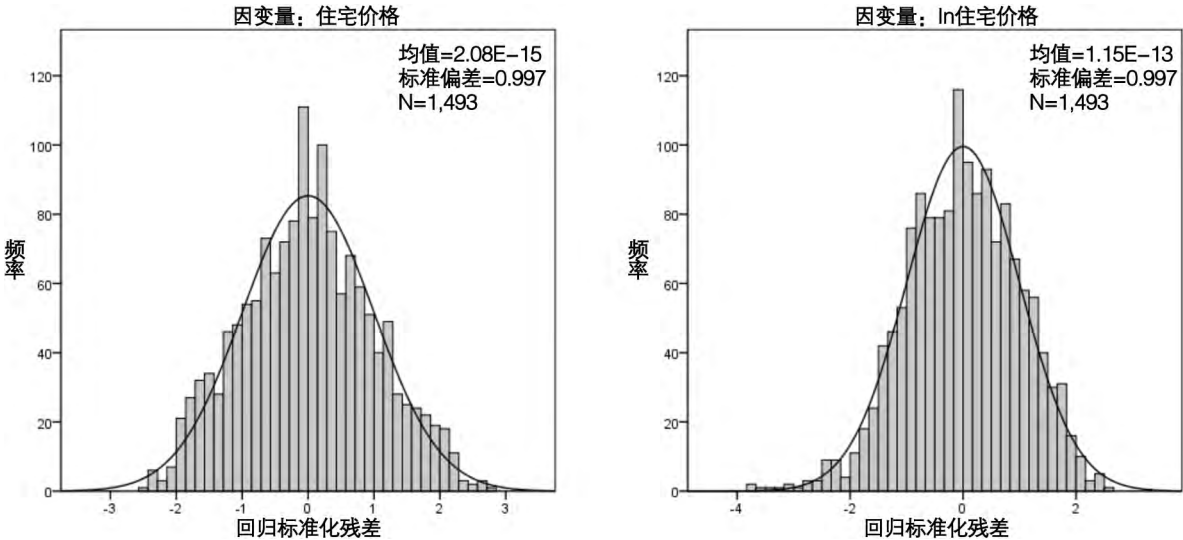


图 2 武汉市线性模型与对数模型差价频率直方图

线性模型和对数模型分别对应差价频率直方图都满足正态分布,每平方米价格差价大多数都集中在 1 500元以内,对比可见对数模型的差价频率曲线更为陡峭,且超出 2 000元范围的差价相比于线性模型要少得多,模型拟合效果更好。

(3) 特征变量影响程度的区域分析

各个二手房特征价格的量纲不同,故采用标准化回归系数(去除量纲)来衡量自变量的相对重要性,可将自变量对因变量的相对影响程度大小进行排序,如表 7 所示。

(4) 实例验证分析

随机选取武汉市某一二手房案例来进行参数回代反算二手房价格,线性模型回归方程案例计算出来的二手房总价差价为 2 万元,二手房每平方米价格差价 172 元。对数模型回归方程案例计算出来的二手房总价差价为 1 万元,二手房每平方米价格差价 85 元,可见这个案例拟合效果不错。从整体看,线性模型每平方米价格差价均值在 870 元,对数模型的每平方米价格差价均值在 820 元,由差价频率分布直方图可看出每平方米价格差价大多数分布在 1 500元以内。

对数模型下武汉各区标准化回归系数及特征影响程度分析表

表 7

特征变量	武汉市		特征变量	江汉区		武昌区		江夏区	
	标准化回归系数	影响程度排序		标准化回归系数	影响程度排序	标准化回归系数	影响程度排序	标准化回归系数	影响程度排序
总层数	0.004	6	总层数	0.000 094	9	0.005	3	0.001	9
ln 面积	0.964	1	ln 面积	0.950	1	0.955	1	0.909	1
装修程度	0.055	5	装修程度	0.024	8	0.050	5	0.042	6
楼龄	-0.007	8	楼龄	-0.017	5	-0.008	4	-0.023	4
教育设施	0.066	3	教育设施	0.116	3	0.027	8	0.070	5
生活娱乐设施	0.041	4	生活娱乐设施	0.037	6	0.027	6	0.025	8
公交站	0.002	9	公交站	-0.002	7	0.003	7	0.015	3
有无 CBD	0.088	7	有无 CBD	0.390	4	0.020	9	0.073	7
有无地铁	0.220	2	有无地铁	-	2	-	2	0.269	2

4 结 论

本文基于特征价格模型,结合武汉二手房数据,采用多元回归方法,评估了二手房价格的影响因素,主要工作与贡献为:

(1) 利用挂牌数据、看房价网数据、百度 API 数据查询获取了武汉市及其江汉区、江夏区、武昌区的二手房各个特征变量的数据并且进行筛选、量化和整合。并对影响因素进行了排序。

(2) 分别构建了可供实际应用的线性模型、对数模型对二手房价格评估进行分析比较,经过模型拟合优度、标准误差、残差分布与差价频率图的对比分析,对数模型拟合效果更好,其回归方程解释能力更强。

在数据分析时,有些问题需要进一步研究。如有些数据仍不容易获取,如居民的收入水平、小区环境质量、小区居民素质等,可能会导致特征变量因素不全面进而影响模型的拟合。以后的研究应重点学习如何量化人文因素与环境质量这些未有特定量化标准的特征变量。

参考文献

- [1] 张鑫. 基于特征价格的二手房价格评估方法研究[D]. 杭州: 浙江大学, 2007.
- [2] Court A T. Hedonic Price Indexes with Automotive Examples, In The Dynamics of Automobile Demand, New York: General Motors, 1939.
- [3] 李志辉. 基于 Hedonic 模型的武汉住宅特征价格研究[D]. 武汉: 华中农业大学, 2008.
- [4] Paliwal R, Geervarghese G A, Babu P R, et al. Valuation of Landmass Degradation Using Fuzzy Hedonic Method: A Case Study of National Capital Region [J]. Environmental & Resource Economics, 1999, 14(4): 519~543.
- [5] Jonathan P, Russell J A, Andrew G, et al. The neurophysiological bases of emotion: An fMRI study of the affective circumplex using emotion-denoting words [J]. Human Brain Mapping, 2009, 30(3): 883~895.
- [6] Baltagi B H, Bresson G. Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris [J]. Journal of Urban Economics, 2011, 69(1): 24~42.
- [7] 蒋一军, 袁江辉. 房地产价格指数与 Hedonic 模型 [J]. 中国资产评估, 1996(3): 30~32.
- [8] 温海珍, 贾生华. 二手房的特征与特征的价格——基于特征价格模型的分析 [J]. 浙江大学学报·工学版, 2004(10): 1138~1149.
- [9] 黄古博, 李雨真. 基于主成分分析法的商品住宅特征价格模型改进 [J]. 华中农业大学学报: 社会科学版, 2011(4): 93~97.
- [10] 李恒凯, 王秀丽, 刘小生. 基于 GIS 和 PCA 的住宅房产特征价格模型 [J]. 测绘科学, 2012, 37(2): 119~122.
- [11] Butler R V. The specification of hedonic indexes for urban housing [J]. Land Economics, 1982, 58, 94~108.
- [12] Lee J, Kwak S J, List J A. Average Derivative Estimation of Hedonic Price Models [J]. Environmental & Resource Economics, 2000, 16(1): 81~91.
- [13] 马思新, 李昂. 基于 Hedonic 模型的北京住宅价格影响因素分析 [J]. 土木工程学报, 2003(9): 59~64.
- [14] 周丽萍. 商品住宅特征价格模型与指数的应用研究 [D]. 西安: 西安建筑科技大学, 2008.

The Influence Factors research of Second-hand House Price in Wuhan City Based on the Multiple Regression Method

Miao Ge¹, Li Yingbing^{1,2}, Yuan Fei¹

(1. School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; 2. Spatial-temporal Big Data Research Center, School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China)

Abstract: Second hand housing prices are subject to regional factors, housing factors, market factors and other factors. A lot of attentions are paid to the real estate property of the differences and amendments for the traditional evaluation methods. There is a strong subjectivity. Based on hedonic price model, the multiple linear regression model and multiple regression model are used to establish the function relationship between the price of second-hand house and location factors, housing factors and other influencing factors. By using Wuhan second-hand housing listing data, the data from the website of kanfangjia (www.kanfangjia.cn) and the data from the website of Baidu, making the screening, quantification and integration the various characteristics of second-hand housing variables of Wuhan, and to compare and analyze the influence degree of each characteristic variable on the price of second-hand house, and the results of the experiment fit well with the actual situation.

Key words: second hand housing; hedonic price; price evaluation; influencing factor; multiple regression method