

基于 GIS 的房地产税税基批量评估

黄梦吟 郭化林(教授)

(浙江财经学院会计学院 杭州 310018)

【摘要】 本文首先对批量评估、自动评估模型、地理信息系统的相关概念和理论进行了阐述,基于此集中探讨了地理信息系统与税基评估的结合方式、步骤。其次,对影响房地产价格的特征和影响因素进行了分析,以此为基础,对影响因子进行量化分析时如何采用地理信息系统(GIS)的空间分析技术的原理优化进行了理论探索。最后,本文以宁波市二手商品住宅交易实例进行了实证研究,建立了结合 GIS 的批量评估模型。应用 GIS 中的 MapInfo 软件优化量化过程从而提高评估效率,并将量化结果录入存储,建立房产信息量化数据库,使房产的空间数据与属性数据建立对应关系,实现了批量自动评估又达到可视化效果。

【关键词】 房地产税税基评估 批量评估 自动评估模型 嗜好模型 地理信息系统

房地产交易活动越来越频繁,为了有效调控房地产市场,完善税制,2011年财政部决定在上海、重庆等地试点征收房地产税。其中,上海房地产税暂按应税住房市场交易价格的70%计算缴纳;重庆房地产税按以下计算公式:应税建筑面积×建筑面积交易单价×税率。从中可以看出两城市目前对新增房地产征税,并有逐渐把存量住房纳入征税范围的趋势,计税的基础均为交易价格,不能反映市场价值。笔者认为这是不合理、不公平的。采用当年的交易价格,不反映真实实现值,且如果逐年折旧这样形成的税收收入也是递减的,税基将逐年递减,无法保证公共支出的增长要求。

目前,我国的房地产估价还有以下亟待改进的地方:①人为因素介入过多,缺乏客观公正性;②效率低下,操作成本高;③缺乏系统性、统一性;④评估成果质量难以保证;⑤信息传播不及时、资源信息不能共享等。因而,探寻公平、高效的评估方法和技术成为亟须解决的问题。在批量评估中数据的采集、分析工作量大,过程中需要涉及大量的基础地理信息资料,最明显的地理信息表现在地产和房产的空间位置及与周围环境的地理关系。因此本文拟通过总结整理税基评估和批量评估理论,探索 GIS 在房地产税税基批量评估的实际应用,构建了批量评估的模型,并模拟应用于宁波市商品住宅的批量评估。

一、批量评估方法与自动评估模型

批量评估方法随着 20 世纪 70 年代以来计算机的发展、普及而出现并得到发展。由于市场信息充足,批量评估方法可以对评估客体的各因素进行分析,得到的分析值可运用回归分析方法,形成了较为准确的评估结果。批量评估是利用共同的基础数据和统一的标准方法,在给定时间对大批量房地产或资产进行估价的活动。它是将现代统计、数学技术等多种技术和系统综合并融入传统评估方法原理中,从而产生对一系列应税房地产的价值在较短时间内进行集中评估的新技术。

批量评估根据所评估群体资产的特征,一般选择适当的成本法、市场比较法或收益法作为模型设定层次,再根据所选择的模型和所能获得的数据,选择数理计量方法获得模型设定的系数,进而再推广应用到大批量房地产价值的评估。

在批量评估系统的应用中,自动评估模型(Automated Valuation Model,AVM)是关键技术。AVM 是一个很广义的概念,是指用于评估住宅类不动产价值的一系列计算机化计量经济模型,利用市场的历史数据,如近期交易价格、房地产特征信息等,形成大批房地产的个体价值。在批量评估方法的实现中,需要与计算机系统紧密结合,由此产生了“计算机辅助批量评估法”(Computer-Assisted Mass Appraisal,CAMA),其是指应用统一程序及统计学理论对大批量房地产进行估价的一种方法,由计算机利用回归技术进行市场分析,构造估价模型,在房地产相关的基本信息输入后,计算机就可以输出其评估值。

可供使用的自动评估模型有许多种,其中三种自动评估模型使用最为广泛,包括嗜好模型、重复销售模型和混合模型。其中嗜好模型(hedonic models)采用销售比较(市场)法,该模型在一定意义上以当地市场可比住宅的近期销售状况为基础,需要采集有关具体特征的信息以确定价值,包括居住面积和宗地规模、财产的年限和其他物理特征。当地市场上可比住宅的近期销售状况被用于估测评估标的价格。本文的研究正是基于此模型。

二、地理信息系统介绍

地理信息系统(GIS)是指在计算机软、硬件系统的支持下,对地球表层空间中的有关地理数据的位置、形态、分布等各种信息进行采集、存储、分析、管理、显示和描述的技术系统和决策支持系统,该系统应具备数据获取、空间数据分析、空间信息推导、结果表示等基本功能。

空间分析是 GIS 的关键技术,是 GIS 区别于一般数字制

图系统的主要标志之一。它是指用于分析地理事物的一系列技术,其分析结果依赖于地理信息的空间分布,并将结果直接面向最终用户。本文研究的房地产税评估具有时间和空间的属性特征,特别是空间分析在评估过程中起到了重要的作用,因此可以视为基于 GIS 的房地产税批量评估研究。

MapInfo 是 GIS 常用的软件工具,能将分析对象的地理信息资料、属性数据、空间数据与地图紧密地联结起来,实现文字、图形、图像信息完美结合。MapInfo 的精华是其分析、查询功能,并且能将分析结果以统计图、专题图、表格、图形或图像的方式形象地表达。不仅如此,MapInfo 还可以直接或间接地与一些常用的数据库系统(如 Microsoft Excel)进行动态数据连接、交换,实现在地图的基础上对数据库进行操作,完成数据库与地理图形的有机结合。

三、GIS 与税基评估的结合方式

地理信息系统是将数据库技术、计算机图形图像处理、空间分析结合,功能强大、直观。而我国房地产税税基评估存在数据缺乏系统性、客观性、公正性等缺点。将房地产税税基评估与 GIS 结合(如图 1 所示),充分利用 GIS 的数据管理能力和空间分析功能进行各项税基评估工作,两者结合相得益彰。因此,GIS 与房地产税基评估相结合,是充分利用二者各自的优越性、实现优势互补的有效措施。

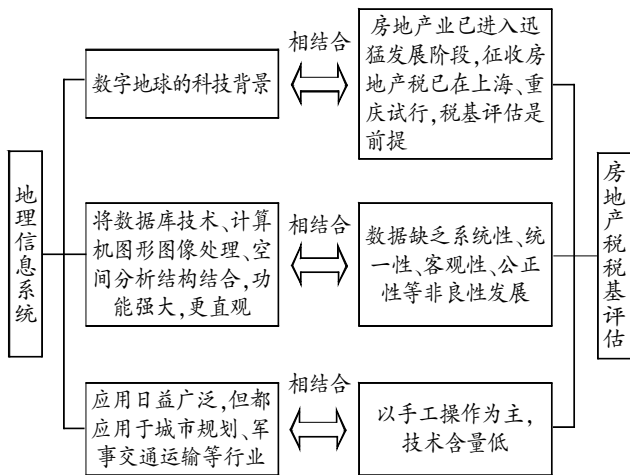


图 1 GIS 与房地产税基评估的结合

其应用流程如图 2 所示:

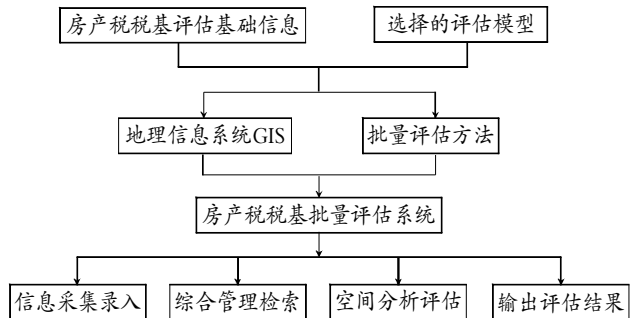


图 2 基于 GIS 的房地产税基批量评估信息系统

图 2 说明:①将处理后的卫星图像进行数字化,确定所评估区域内房地产的位置特点。②将整理后的税基评估的基础信息数据(包括空间数据和属性数据)输入到 GIS 中,即每个评估对象为一个目标,相关信息一一对应到评估对象以实现空间数据和属性数据的一体化管理。③评估过程中,选取有效的评估模型,应用 SPSS 或者 EVIEWS 等统计软件完成,可再将得出的评估结果存入相应房产的属性表中。④最后利用编程技术,集合成基于 GIS 的房地产税基批量评估系统,实现评估的信息化,必要时更新数据,便可高效地完成评估工作。⑤利用 GIS 相关软件,对评估成果进行直观呈现。

四、房地产价格影响因素分析

考虑到房地产供求关系,构建房地产价格影响因素体系,如表 1 所示。由于各个因素本身具有动态性,所以这些因素对房地产价格的影响也不是固定的,随着时间、地区、房地产用途的不同,它们的影响作用也各不相同。比如,本来是影响较小的因素,可能会成为主导因素;相反,主导因素也会成为次要因素,而且它们对房地产价格的影响程度有的可以量化,有的则难以量化,只能凭借评估师的经验加以判断。

表 1 房地产价格影响因素体系

一般因素	经济因素	经济发展因素、财政金融因素、产业结构因素
	社会因素	人口因素、社会福利水平、家庭结构、社会治安状况
	行政因素	土地适用的制度及政策、城市规划及发展战略、税收、行政隶属、交通管制
	心理因素	购买或出售的心态、讲究风水、价值观
区域因素	商业服务繁华程度	服务业繁华状况、商业中心的位置关系
	道路通达程度	道路系统的通畅程度、道路的级别
	交通便捷程度	公共交通的完善程度和便利程度
	城市设施状况	基础设施、生活设施、文化娱乐设施
个别因素	环境状况因素	绿化、噪声、水质量、空气质量
	土地	区位、面积、形状、地质、地形、容积率、用途、使用期限
	建筑物	结构、材料、设计、设备、施工质量、法律限制

目前在批量评估中应用最为广泛的是直接市场法,它是采用了统计校准技术的自动评估模型中的嗜好模型(hedonic models),本文以该模型为基础,它是利用房产特征变量与房产市场价值间存在的因果关系而建立的模型。

假设在公平、公正、公开的环境中,房地产价格具有客观性,从某种程度上说影响房地产价格的因素及影响程度也应是客观的。因此,可以说存在一个房地产价格的数学模型,可以将影响房地产价格的因素结合到该模型中:

$$y_i = \alpha_i(x_1, x_2, x_3 \dots x_i, \dots x_n) (i=1, 2, 3, \dots n) \quad (1)$$

式中: y_i 表示房地产价格; x_i 表示影响房地产价格的各种因素; α_i 表示影响房地产价格变化的权重系数。

线性函数、对数函数、半对数函数和对线性函数批量评估中经常采用的四种简单的函数形式,称之为基本函数形式。

1. 线性形式。自变量和因变量均以线性形式进入模型:

$$y_i = \alpha_0 + \sum \alpha_i x_i + \varepsilon \quad (2)$$

其中, ε 为随机误差项。

2. 对数形式。自变量和因变量均以对数形式进入模型:

$$\text{Ln}y_i = \alpha_0 + \sum \alpha_i \text{Ln}x_i + \varepsilon \quad (3)$$

3. 对数线性形式。自变量采用线性形式, 因变量采用对数形式, 即:

$$\text{Ln}y_i = \alpha_0 + \sum \alpha_i x_i + \varepsilon \quad (4)$$

4. 半对数形式。自变量采用对数形式, 因变量采用线性形式, 即:

$$y_i = \alpha_0 + \sum \alpha_i \text{Ln}x_i + \varepsilon \quad (5)$$

五、基于 GIS 的房地产价格影响因素分析

房地产价格受众多的影响因素的综合作用, 不过可以提取一部分的影响因素来解释价格, 从而通过批量评估技术来弥补单宗评估中的成本法、市场法、收益法的不足。在对各个影响因素进行量化的过程中, 应用传统的评估方法, 耗费时间长、结果不精确。譬如, 评估房产所在位置是否受道路、水流、景观等因素在一定范围的影响, 应用传统的方法很难估计, 或是应用主观评价, 这是不精确的。应用 GIS 的空间分析功能, 可达到事半功倍的效果。

GIS 的空间分析功能如缓冲区分析、叠加分析、路径分析等分析模型均可用于特征价格评估, 其中缓冲区分析在其中起到了重要作用。譬如, 为了对评估对象所在区域的交通设施进行量化, 可以利用缓冲区分析功能查询出该交易案例一定范围内所有的道路、公交站点是否在一定区域内, 并能快速定位查询这些道路的相关信息 and 公交站点的最短距离等情况, 为量化评分提供充足的数据支持。

对特征变量的准确量化是模型能否成功的关键, 要充分利用 GIS 的数据管理、查询和分析功能为评估人员提供辅助支持。在 GIS 的支持下, 可以采用当地的基础地理数据库, 然后应用 GIS 的各种查询和分析功能来对该交易案例周围的环境、设施和物业进行分析和评价赋分。具体来说, 对于某评估对象的自然环境、生活配套和教育配套设施的评价赋分, 可以利用 GIS 的缓冲查询功能查询出一定范围内对应的特征变量指标数据, 得到一定范围内是否有医院、学校、超市、景观等信息及具体对应指标的属性信息, 也测算出其距离并进行赋值。

六、案例研究: 在宁波市住宅价格中的应用

1. 确定最优模型。单身公寓、联体排屋及别墅只能满足小部分人的需求, 因此本文剔除了这三种类型的住宅, 通过筛选剔除不合理的样本数据后, 最后选取了宁波市 2011 年 5 月份海曙区、江东区 and 江北区三个行政区域共 150 份的住宅价格数据, 每个行政区域各 50 份。

首先对宁波市评估区域进行数字化, 即对小区、道路、水流、公交站点等各种基础设施进行数字化并按其特征分层存放, 为之后的评估工作做好铺垫工作。

根据宁波市的特点, 选取以下指标作为该市房地产税税基评估的量化因素, 如表 2 所示。

表 2 住宅价格批量评估影响因素量化方法

影响因素	量化方案
行政区域	海曙区、江东区、江北区, 分别赋值为 1, 2, 3
商业服务中心	小区离商业服务中心的实际距离, 以公里为单位, 保留到小数点后一位
公交	在 500 米内住宅小区是否有公交站点或轻轨站点, 如果有则为 1, 否为 0
轻轨	在 1000 米内住宅小区是否有轻轨站点, 如果有则为 1, 否为 0
超市、菜场	在 1000 米内住宅小区是否有大型超市、菜场, 如果有则为 1, 否为 0
医院	在 1000 米内住宅小区是否有医院, 如果有则为 1, 否为 0
文体中心	在 1000 米内住宅小区是否有文体中心, 如果有则为 1, 否为 0
教育	在 1000 米内住宅小区是否有好口碑的中小学, 如果有为 1, 否为 0
环境景观	在 1000 米内住宅小区是否有旅游景点、自然景观, 如果有为 1, 否为 0
建筑面积	交易信息中的建筑面积直接赋值
建筑结构	钢结构、砖混结构、钢筋混凝土结构分别赋值为 1, 2, 3
卧室个数	卧室个数
卫生间数	卫生间个数
建筑类型	多层、小高层和高层分别赋值为 1, 2, 3
装修	毛坯房、简装、精装修和豪华装修分别赋值为 1, 2, 3, 4
房龄	2011 年与住宅小区建成交付年代的差值
内部环境设施	分等级赋值: 差、一般、较好、好, 分别赋值为 1, 2, 3, 4
楼层	该样本房屋所在的楼层
总楼层	该样本房屋所在的楼一共有几层

期间进行量化分析时, 便可借助 GIS 专业软件进行优化分析, 如 MapInfo。以月湖为例, 以 1000 米为缓冲区的单位长度, 用 MapInfo 软件进行缓冲区分析, 如图 3 所示。

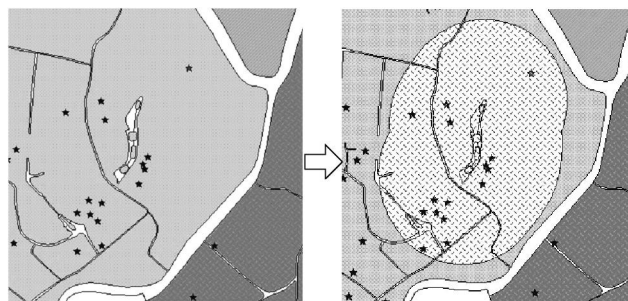


图 3 月湖缓冲区分析示意图

对月湖进行缓冲区分析后, 在其为中心的 1000 米内的住宅小区高亮显示, 并可对其属性“景观”标示为 1。其他景观如日湖、天一阁等景观, 以及后文中的交通、外部的配套设施等相关属性均可进行类似的操作, 以达到提高评估效率的目的。

将样本房产的属性数据与其对应的空间属性相结合, 即将上述的信息输入 GIS 中, 实现房产信息的空间分布与本身的属性特征一一对应, 如图 4 所示。

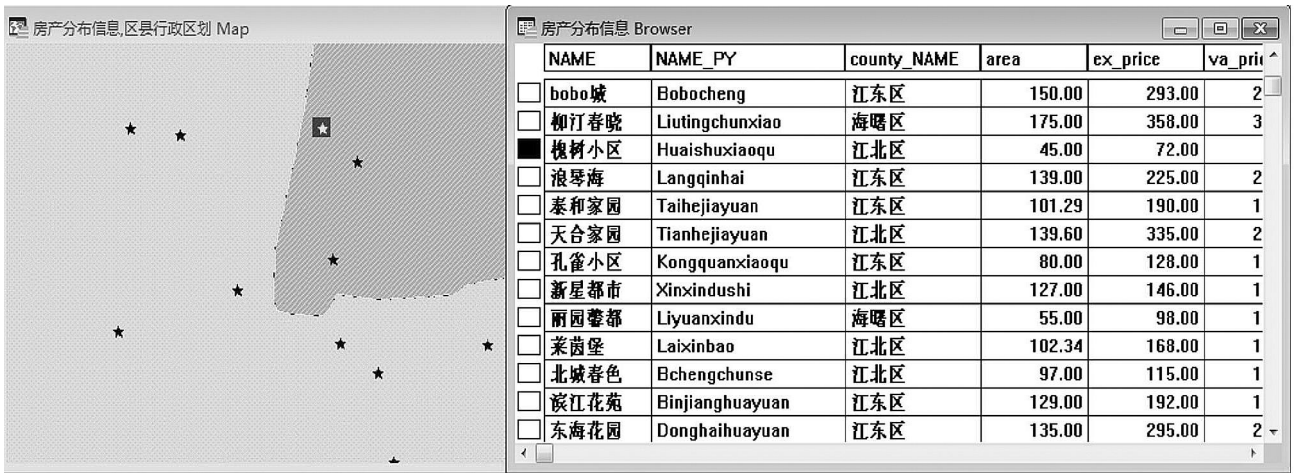


图4 空间信息与属性信息对应示意图

当点击图形中的房产样本点时,相关的属性数据会高亮显示,即可查看该房产的相关信息。当要修改目标房产的信息时,无论是从地图上还是在属性表中查找,都可实现更新,并自动关联,这样既可从地图上查询到该房产的信息,又可以根据属性信息查看房产的地理位置。

本文以150份的房价数据作为因变量,各个量化后的影响因素作为自变量,根据公式2至公式5,采用SPSS18.0软件进行模型的拟合分析。

表3 模型结果分析汇总表

	R	R ²	调整R ²	标准估计的误差
线性	0.943	0.889	0.884	28.905 663
对数	0.968	0.937	0.932	0.132 680
对数线性	0.960	0.921	0.915	0.148 025
半对数	0.929	0.863	0.853	32.530 841

对以上四种模型进行回归分析比较后,根据拟合优度调整后的R²排列优劣依次为:对数形式、对数线性形式、线性形式、半对数形式。因此,选择的最优模型为对数形式的嗜好模型。

2. 最优模型分析。对数形式的嗜好模型进行显著性和方差分析。F值较大,说明自变量造成的因变量的变动远远大于随机变量对因变量造成的影响。在SPSS中可用F值的相伴概率值Sig.检验。

从表4中可知,回归方程方差分析的显著性检验值小于0.001,说明方程是高度显著的,拒绝全部系数均为0的原假设,通过了F检验。因变量和自变量确实存在线性关系,可以使用线性模型,模型对样本数据的拟合在统计上是有意义的,回归方程有效。

表4 方差分析

模型	平方和	自由度	均方和	F	Sig.
回归	35.649	9	3.961	319.489	0.000
残差	1.686	136	0.012		
总计	37.336	145			

进一步检验回归系数的显著性,分别验证各回归系数是否与零无显著差异,采用T检验,如表5所示。各回归系数(除常数外)的Sig值(p值)均小于显著性水平0.05下,因此认为自变量能够很好地解释因变量。

表5 对数模型系数表

模型	非标准化系数		标准系数	t	Sig.
	B	标准误差	Beta		
(常量)	0.165	0.148		1.112	0.268
区域	-0.054	0.012	-0.087	-4.375	0.000
lnCBD距离	-0.078	0.012	-0.141	-6.396	0.000
公交站	0.167	0.032	0.106	5.149	0.000
学校	0.081	0.025	0.070	3.184	0.002
景观	0.074	0.022	0.072	3.345	0.001
ln面积	1.054	0.031	0.804	33.614	0.000
ln装修	0.073	0.021	0.075	3.566	0.001
ln房龄	-0.079	0.021	-0.116	-3.730	0.000
ln内部环境设施	0.112	0.028	0.121	3.962	0.000

由于影响房价的各个影响因素的单位不同,它们的影响程度无法直接比较。但是,标准化后的回归系数(Beta)是在所有变量标准化后得到的,它具有可比性,使用绝对值进行各因素影响程度的排序,依次为:lnCBD距离、ln内部环境设施、ln房龄、公交站、区域、ln面积、ln装修、景观、学校。

在α=0.05的显著性水平下,上述变量进入模型,则有:

$$\ln P = 0.165 - 0.054 \times \text{区域} - 0.078 \ln \text{CBD距离} + 0.167 \times \text{公交站} + 0.081 \times \text{学校} + 0.074 \times \text{景观} + 1.054 \ln \text{面积} + 0.073 \ln \text{装修} - 0.079 \ln \text{房龄} + 0.112 \ln \text{内部环境设施}$$

整理后,宁波市二手商品住房批量评估后所确定的房地产税税基与各个影响因素之间的关系为:

$$P = \text{CBD距离}^{-0.078} \times \text{面积}^{1.054} \times \text{装修}^{0.073} \times \text{房龄}^{-0.079} \times \text{内部环境设施}^{0.112} \times e^{0.165 - 0.054 \times \text{区域} + 0.167 \times \text{公交站} + 0.081 \times \text{学校} + 0.074 \times \text{景观}}$$

3. GIS在模型中的应用。将最终确定的批量评估公式输入到MapInfo中,利用其强大的属性数据处理功能,也可得到最后的评估结果,与SPSS计算出来的结果一致,如图5所示。

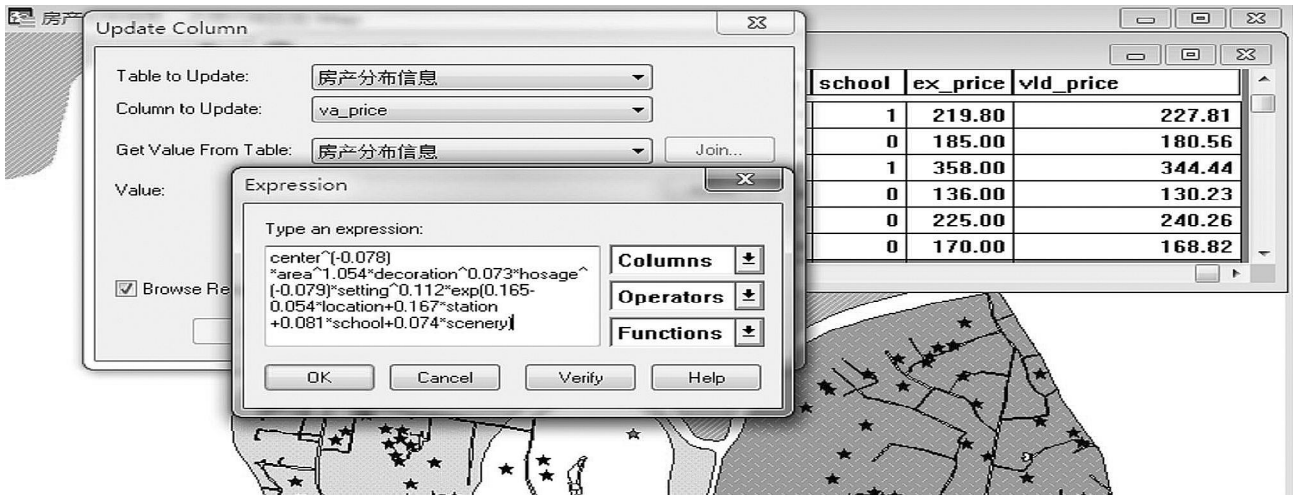


图5 MapInfo 评估结果示意图

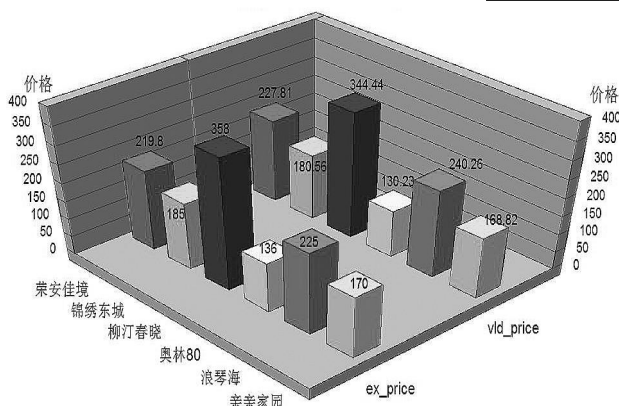


图6 交易价格和评估价格比较示意图

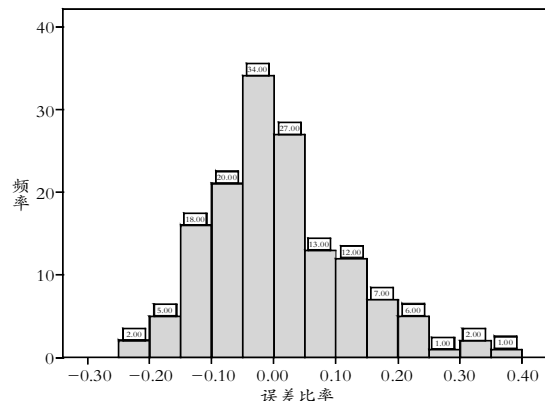


图7 误差比率频率直方图

可利用 MapInfo 的统计功能,绘制各种统计图,如图 6 为交易价格和评估价格的比较图,效果直观(由于样本数据繁多,只挑选了几个样本数据进行演示)。

此后,如果将新增的房产数据更新至该数据库中,只要在图上画出位置就能自动计算 CBD 距离,面积,离学校距离、公交站等大多数空间信息,只需填写房龄、装修、楼层等个别因素这样少数的几个字段,房龄每年数据库会自动更新,这样根据确定的自动评估模型,就能批量评估出应税房产的税基。从而实现评估工作数据存档和管理,实现批量评估方法和 GIS 技术相结合,提高评估效率。

4. 模型评价。对回归方程的预测值与实际值(挂牌价格)进行误差比率分析,即预测值减去实际值再除以实际值。最终的评估结果误差比率的绝对值在 5% 以内的样本数量为 61 个,占总样本的 41.78%;误差比率的绝对值在 10% 以内的样本数量为 97 个,占总样本的 66.44%;误差比率的绝对值在 15% 以内的样本数量为 122 个,占总样本的 83.57%,如图 7 所示。因此,方程的回归结果可以接受。

5. 实例分析小结。方程显示,CBD 距离、房龄、所在区域与评估值呈反方向变化:即住宅距离 CBD 越近,房价越高,这也符合现实中的实际情况,越靠近 CBD,购物、休闲、出行都十分便利,自然房价越高;楼龄越大,在其他同等条件下住

宅价格越低,随着时间变化,房屋的装修设计过时、结构老化等问题会导致住宅价格下降。所在区域,由于数值 1、2、3 分别代表了宁波市的三个老城区,即海曙区、江东区、江北区,海曙区是宁波市的中心城区,三个区的繁华发展程度依次递减,因此平均的房价也依次递减。

其他影响因素——面积、装修、内部环境设施、公交站、学校、景观等都与住宅的评估值同方向变化,这也符合实际情况。面积越大、小区内部的环境设施越好、附近有公交站点、学校和美丽的景观设施,都会对住宅价格产生正向的影响。

其他未纳入模型的因素,并不能说它们对住宅价格没有影响,只是由于样本数量的限制,未在本模型中显著体现。

综上,该回归方程作为应税评估模型,在理论和实际上都通过了解释,符合建立模型的要求,能够基本评估房产的应税价值,从而简化了确定税基的过程

【注】本文系 2011 年浙江省大学生科技创新活动计划(新苗人才计划)项目(编号:2011R413032)的阶段性研究成果。

主要参考文献

1. 玄永生,王建忠,王余丁.我国房产税税基评估问题研究.环渤海经济瞭望,2011;4
2. 卢莉,宋伟.应用 GIS 技术进行房地产估价探讨.经济体制改革,2002;5