

南宁市二手房价格影响因素分析及房价走势的预测*

Analysis and Prediction of the Factors That Influence the Price of the Second-hand House in Nanning

邝文竹, 刘琳

KUANG Wen-zhu, LIU Lin

(广西大学数学与信息科学学院, 广西南宁 530004)

(Department of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

摘要:选择南宁市江南区、西乡塘区、青秀区以及兴宁区的二手房销售数据,以地理因素为例,运用方差分析和 Fisher 最小显著差异法研究南宁市二手房售价的影响因素,并用回归分析和 BP 神经网络预测与其它城区有显著差异的城区的二手房售价随房龄变化的情况。结果发现,西乡塘区和江南区二手房平均售价的均值无显著差异 ($P > 0.05$),青秀区二手房平均售价的均值与江南区、西乡塘区、兴宁区的有显著差异 ($P < 0.05$),均值差分别为 0.23000、0.24310、0.08553,地理因素对二手房价格有显著影响,青秀区的二手房平均售价在 4 个城区中是最高的,房价随房龄的增加,呈不规则变化。

关键词:房价 回归分析 BP 神经网络

中图分类号:O212.1 **文献标识码:**A **文章编号:**1002-7378(2012)02-0110-03

Abstract: According to the study of second-hand house sale data with taking geography factors as sample, by variance analysis and Fisher least significant difference method, the factors that influence the price of second-hand house in Jiangnan, Xixiangtang, Qingxiu and Xingning districts are studied. The regression analysis and BP neural network are used to predict the relationship of second-hand house price changes with the house age. The method is valuable for reference and the research results are based on the facts.

Key words: house price, regression analysis, BP neural network

文献[1]利用经典的住房 Hedonic 价格模型,对影响住房价格的因素,如区位、环境、设计风格、建筑结构等,做具体的实证研究,并取得较好的结果。但是,目前南宁市二手房市场的销售案例中,记录的住房特征不全面,给分析和预测住房价格影响因素提出新的挑战。本文根据南宁市二手房销售数据,以城区地理因素为例,研究单位面积售价的影响因素,并预测二手房单位面积售价随房龄变化的情况,取得了较好的研究结果。

收稿日期:2011-05-24

修回日期:2011-07-09

作者简介:邝文竹(1984-),女,硕士,主要从事应用统计研究。

* 国家自然科学基金项目(11061002)和广西自然科学基金项目(2012GXNSFBA053010)资助。

1 数据来源及分析方法

1.1 二手房价格原始数据的收集及检验

分别从广西房产街和南宁市阳光居易房屋经纪有限责任公司,收集南宁市 4 个城区 2010 年全年的二手房销售案例作为样本。剔除异常值后,江南区有 73 个样本、西乡塘区有 126 个样本、青秀区有 135 个样本、兴宁区有 60 个样本。以城区地理因素为例,运用方差分析和 Fisher 最小显著差异法研究二手房价格的影响因素。通过 SPSS 软件将每个区的原始数据绘制图形,并进行皮尔逊 χ^2 检验。

1.2 二手房价格的回归分析

不同的二手房具有不同的特征,这些特征共同影响房价^[2]。结合这些住房特征,可以运用回归分析评估二手房的价格。以因素:房龄(x_1)、卧室个

数(x_2)、楼层(x_3)为例,用 y 与自变量做多元线性回归,回归模型为:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

$\beta_0, \beta_1, \beta_2, \beta_3$ 为回归系数。

1.3 二手房价格的 BP 神经网络预测

BP 神经网络在处理非线性问题上,显示出特殊的优越性,它具有良好的逼近能力、误差可以控制^[3],目前也有不少相关的应用。

下面以一个两层的 BP 神经网络为例,预测给定因素的二手房价格。为了提高神经网络的训练效率,需要对样本集数据作必要的预处理。城市住房价格遵循边际效用递减规律^[4,5]。因此,可以用指数形式将住房价格表示如下:

$$Y = \alpha X_1^{\beta_1} X_2^{\beta_2} \cdots X_k^{\beta_k}.$$

两边取对数后: $\ln Y = \ln \alpha + \sum_{i=1}^k \beta_i \ln X_i$ 。其中, Y 为住房价格, X 为住房特征, α 和 β_k 为系数。

在评价回归分析模型和 BP 神经网络模型的模型拟合效果时,选择两个评价指标。第一个是,均方

根误差: $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$; 第二个是,均

方百分比误差: $MSPE = \frac{1}{n} \sqrt{\sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i}\right)^2}$ 。

2 结果与分析

2.1 数据检验与分析

检验结果表明,江南区、西乡塘区和青秀区的样本总体均服从正态分布。比较特殊的是兴宁区的检验统计量: $\eta = 14.4786$, 临界值为 $\chi^2_{\alpha}(0.05) = 7.815$, 由于这一区的样本量相对较少,可能存在缺失数据,需要作进一步的方差齐性检验^[2]。

用 SPSS 软件计算 Levine 方差齐性检验,得 $F_{0.05} = 3,390 \approx 2.625, L = 8.04301 > 2.625$, 表明方差不齐。要解决该问题,可以把原始尺度下的测量值系统地转化为新尺度。对原始数据进行取自然对数的变换,将变换后的数据再次进行方差齐性的 Levine 检验,结果 $P = 0.06715 > 0.05$ 。可以认为经过对数变换后,4 个总体的方差都相等。用 SPSS 计算其峰度-偏度检验,结果变量的偏度为 0.155 (标准误差 0.309), 峰度为 0.129 (标准误差 0.608)。数据的偏度和峰度都接近于零,故可以认为该数据的总体服从正态分布。

以地理因素为例,对数据进行方差分析,检验统计量为 $F = 27.36157, P = 4.43924e - 16$, 临界值为

2.62779。按 $\alpha = 0.05$ 水平,拒绝 H_0 。故可以认为不同城区的地理因素对二手房价格有显著影响,这与南宁市各城区的实际情况相符。

对二手房销售价格的差异作进一步研究。按照 LSD 方法^[6],首先利用方差分析检验,令假设为 $H_0: \mu_1 = \mu_2 = \cdots = \mu_4$, 其检验的统计量为 $F = 27.36, \alpha = 0.05, df_1 = 3, df_2 = 390$, 由于 $27.36 > 2.63$, 拒绝 H_0 , 则至少有一个总体均值与其余的不同 ($P = 4.43924e - 16$)。对于给定的 $\alpha = 0.05, \mu_i$ 对 μ_j 的最小显著差异为 LSD_{ij} 。由于样本容量不同, ($n_i \neq n_j, i, j = 1, 2, 3, 4$), 所以每次计算的 LSD_{ij} 值也不同。直接用 SPSS 软件的多重比较功能计算,结果如表 1 所示。

表 1 “ln(单位面积售价)”的 LSD

城区 (i)	城区 (j)	均值差 (i - j)	标准误差	P
江南区	西乡塘区	0.01310	0.03534	0.71110
	青秀区*	-0.23000	0.03490	1.43e-10
	兴宁区*	-0.14448	0.04186	0.00062
青秀区	江南区*	0.23000	0.03490	1.43e-10
	西乡塘区*	0.24310	0.02976	4.37e-15
	兴宁区*	0.08553	0.03727	0.02230
西乡塘区	江南区	-0.01310	0.03534	0.71110
	青秀区*	-0.24310	0.02976	4.37e-15
	兴宁区*	-0.15757	0.03768	0.00004

* 表示在 $\alpha = 0.05$ 水平下有显著性差异。

根据表 1 中的 P 值,当 $P < 0.05$ 时,两个样本之间有显著性差异。在所有的成对比较中,只有西乡塘区和江南区的 $P = 0.71110 > 0.05$ 。即西乡塘区和江南区平均售价的均值无显著差异。由于西乡塘区和江南区既没有商业街市,又不是城市中心,所以这两个城区的住房格局和周围的居住环境相差不大。由表 1 的第二组值可见,青秀区和其他 3 个城区的 P 值均小于 0.05。青秀区平均售价的均值和其他 3 个城区有显著差异,而且青秀区与江南区、西乡塘区、兴宁区的均值差分别为 0.23000、0.24310、0.08553,均大于零。青秀区的二手房平均售价在 4 个城区中是最高的,这和青秀区的有利地理条件有很大关系。在种种条件的刺激下,青秀区的二手房价格也和该地的新楼盘价格一样出现上涨。

2.2 青秀区二手房价格分析与预测

2.2.1 回归分析

以广西房产街收集的二手房登记个案为依据,选取青秀区 2011 年以前建成的房屋的 175 个案例作为样本,用 SPSS 软件计算,观察回归分析输出表,在 0.05 的显著水平下 x_2 (卧室个数) 的系数显著性经验不通过 ($t = -0.9826, P = 0.3272 >$

0.05)。剔除 x_2 , 再用 SPSS 计算, 通过观察回归分析的方差分析, $F=16.9421, P=1.922e-07$, 说明回归方程高度显著, x_1, x_3 整体上对 y 有高度显著的影响。回归方程为

$$y = 7565.511 - 48.587x_1 + 35.729x_3. \quad (1)$$

由方程(1)看出, 房龄对单位面积售价产生负的影响。但是二手房的价格预测涉及的不确定因素众多, 预测的过程往往是非线性的, 所以回归分析只能初步估计二手房的价格走势。

2.2.2 BP神经网络预测

BP神经网络的拟合效果评价的各项指标值比线性回归拟合的各项指标值小, 显示出较强的拟合能力。

从图1可以看出, 青秀区的房价随着房龄的增加, 呈现出不规则的变化。报告期为2010年, 可以把房龄为2~50年的预测结果分为3个时间段分析。

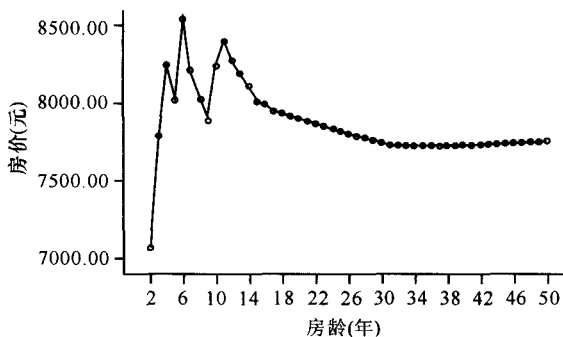


图1 单位面积房价预测

在房龄为2~10年时, 房价的折旧费并不是有规律的递增或递减, 这与房价由于折旧而逐年递减不相符合。原因是由于最近几年国家出台相关的房地产政策, 影响二手房的销售价格。例如2007建成的房屋(房龄为3年)价格到了2010年会因折旧而降低, 但是其下降的速度没有房价上涨的速度快, 所以出现了上面的反常情况; 在房龄为10~40年时, 二手房的单位面积销售价格基本持平, 略微有规律的折旧变化。价格随着房龄的增大而递减, 而且“递减值逐渐”变小。可以认为, 这个时间段的二手房价格受国家政策调控或者其它外界因素的影响较少, 其价格变化是由房屋的自然老化、破损而引起的折旧; 在房龄为40~50年时, 单位面积销售价格上涨,

与实际情况不符。这可能与BP神经网络固有的缺陷: 过拟合和推广能力差有关, 它们造成对40年房龄的房价预测不精确。

Vapnik提出的支持向量机^[7]则可以较好地解决上述缺陷问题。此外, 处理非线性问题时还可以运用非线性回归分析、时间序列模型、灰色系统模型等, 根据实际问题采用不同的模型以获得更高的预测精度^[8]。

3 结束语

本文以2010年全年的南宁市二手房市场数据为例, 对南宁市二手房价格影响因素进行分析及预测。发现南宁市不同城区的地理因素对二手房价格有显著影响。青秀区的房价随着房龄的增加, 呈现出不规则的变化。

参考文献:

- [1] 邹高禄. 二手房价格对于住房特征和区位变化敏感性分析[J]. 西南师范大学学报, 2005, 30(3): 552-555.
- [2] Dipasquale D, Wheaton W C. Urban economics and real estate markets [M]. Englewood Cliffs. US. New Jersey; Prentice-Hall, 1996, 35-36, 65-72, 189-192.
- [3] 许东, 吴铮. 基于MATLAB 6. X的系统分析与设计——神经网络[M]. 第2版. 西安: 西安电子科技大学出版社, 2002.
- [4] Harrison D M, Smersh G T, Schwartz A L, et al. Environmental determinants of housing prices: the impact of flood zone status[J]. Journal of Real Estate Research, 2001, 21(1/2): 3-20.
- [5] 古扎拉蒂. 计量经济学[M]. 第3版. 林少宫, 译. 北京: 中国人民大学出版社, 2000.
- [6] 张忠占. 统计学方法与数据分析引论[M]. 北京: 科学出版社, 2003.
- [7] 阮连法. 一种新的二手房价格评估方法[J]. 统计与决策, 2004(1): 120-121.
- [8] 苏一丹, 李桂. 流体神经网络模型在规划物流陪送运输方案中的应用[J]. 广西大学学报, 2002, 27(3): 204-206.

(责任编辑: 陈小玲)