

基于协方差回归的烟台市二手房房价影响因素研究

郭 誉,于海生,柯佳武

(鲁东大学 数学与统计科学学院,山东 烟台 264039)

摘要:为深入研讨烟台市房地产市场的特点,本文利用网络爬虫技术在安居客网站上收集数据,通过有交互作用的多因素协方差模型确定影响烟台市二手房房价的12个主要因素,建立烟台市二手房房价的估值模型。利用Lasso回归模型进一步分析数值型变量,得出影响烟台市二手房房价的重要因素:行政区、物业费和所在楼层。

关键词:房地产;网络爬虫;协方差模型;Lasso回归

中图分类号: O213 **文献标志码:** A **文章编号:** 1673-8020(2022)03-0232-09

房地产业为我国经济发展的支柱性产业之一,该行业是否健康稳定发展关系到国家的国计民生^[1]。随着房地产市场的商业化加速,我国的房产价格开始直线上升,高昂的房价超出绝大部分居民的购置能力。同时,影响房价的诸多因素却不为大众所了解。经研究发现,房产价格波动是诸多不同却又相互关联的因素共同作用的结果,例如地区因素、政策因素、社会因素等,不同因素对房价的影响水平亦不同^[2]。因此,深入探讨房产市场的特征,并对主要影响因素进行定量分析,对于我国房地产市场稳步发展具有重要的战略意义^[3]。

面对城市新开售楼盘的昂贵售价,越来越多的购房居民将视线转移到二手房交易市场。然而,在交易过程中居民需要面对诸多问题,包括二手房房价影响因素,房源如何定价等。文献[4]使用层次分析法对影响因素进行权重分析,指出消费者应着重考虑中等物业费的住宅。文献[5]运用多元回归分析建模,以期科学指导住宅销售价格。文献[6]针对35个城市运用空间杜宾模型,研究不同权重下房价的空间溢出效应,结果表明这35个城市的房价存在明显空间正相关关系。

本文以山东省烟台市二手房为研究对象,运用方差分析等方法深入研究二手房的住宅特征对房产价格的影响,以期解决以下两个问题:

1) 价格问题。从烟台市实际案例出发,分析二手房房价影响机制,由此建立模型并运用案例数据进行估价,为居民购房或卖房价格做出合理预测^[7]。

2) 政府调控。将烟台市二手房房价信息进行可视化展现,以帮助政府部门直观了解房价分布情况以及影响房价分布的主要因素,从而准确切入房地产市场,为调控房产政策提供多角度思考^[8];同时根据研究结果,提升资源相对匮乏地区的基础建设及交通环境,缩小不同区域之间的基础设施差异。

1 问题背景

据国家统计局数据显示,2020年房地产业增加值为74 533亿元,房地产业对国民经济的贡献率为7.34%,相比于2011年增长了1.5%。纵观近十年,我国房地产业对经济的贡献率基本保持稳步增长态势^[9],该行业的良性发展对国民经济发展起到举足轻重的推动作用。

随着房地产行业的复苏,新发售楼盘价格亦开始回涨,居民购房压力骤增,因此性价比较高的二手

收稿日期:2021-08-03;修回日期:2022-03-28

基金项目:山东省研究生教育教学改革研究项目(SDYJG19073)

第一作者简介:郭誉(1998—),男,山东济宁人,硕士研究生,研究方向为经济与社会统计。E-mail: yoan_dec@outlook.com

通信作者简介:于海生(1978—),男,山东寿光人,副教授,硕士研究生导师,博士,研究方向为经济与社会统计。E-mail: qingzhu_yu@163.com

房市场得到更多人的青睐,这使得二手房交易数量紧追新发售房源,并形成相对完整的交易板块。二手房对抑制房价暴涨、增补房地产市场房源、避免新一轮的房产过热等现象有重要作用,其优点主要体现在以下方面:

1) 交通便利。随着城市的不断发展,新发售楼盘已进入“郊区化”,地理位置远离城市中心区域,购买新发售楼盘的居民在交通出行方面需要付出更多的时间成本。

2) 配套完善。二手房社区相应的配套设施较完善,民生所需供应充足,而新建商品房则需要时间发展周边环境。

3) 质量可靠无污染。二手房社区建设年代相对久远,房屋建造时用料充足、质量可靠,即便存在质量问题也易于观察;其次,装修产生的有害气体经过时间流逝挥发殆尽,对居住者身体健康无害处。

根据 CREIS 中指数据库二手房版数据统计,2021年2月,烟台市二手房累计成交962套,同比上升411.7%,其中成交面积为8.6万 m^2 ,同比增长408.9%。2021年1—2月,烟台市二手房成交面积为23.9万 m^2 ,同比增长95.6%。可见,烟台市二手房市场发展良好,居民需求高涨。

2 描述性统计

2.1 分层等距抽样

本文研究数据来源于山东省烟台市2021年1—5月的二手房源,共包含数据44 236条。考虑到数据代表性、研究可行性、成本合理性,采用分层等距抽样方法在安居客网站上进行数据爬虫^[10]。其中,样本量 n 的计算公式为

$$n = \frac{N u_{\alpha/2}^2 S^2}{N d^2 + u_{\alpha/2}^2 S^2}, \quad (1)$$

这里, N 为总体规模, α 为置信水平, $u_{\alpha/2}$ 是标准正态分布上侧面积为 $\alpha/2$ 时对应值, d 为绝对误差限度, S^2 为总体方差。由于 S^2 未知,利用文献[11]中总体比例估计问题的结论,取 $S^2 = 0.25$,计算可得 $n \approx 1485$ 。

利用分层等距抽样确定烟台市9个行政区域的样本量,再对每个行政区域的房源信息进行等距抽样。在安居客烟台市二手房版面的房源信息显示中,规定每间隔6个房源抽取一个样本,最终得到1485个样本,部分行政区的样本量分布如图1所示。

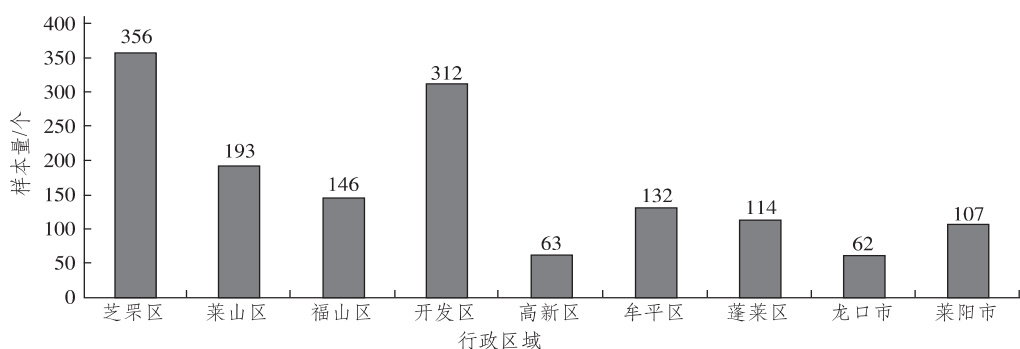


图1 烟台市部分行政区样本量

Fig.1 Sample size of some administrative districts in Yantai

2.2 指标选取

影响二手房房价的因素有很多,本研究结合文献[12]、实际调查和网络爬虫,选取13个指标作为解释变量,具体指标见表1。

表 1 二手房房价影响因素
Tab.1 Influencing factors of second-hand housing prices

影响因素	变量	说明
房价	Y	烟台市二手房价格
行政区	X_1	芝罘区、莱山区、福山区等 9 个行政区
街道位置	X_2	每个行政区分为 1~26 个不等的街道
房龄	X_3	从房屋交付使用至 2021 年 1 月 1 日间隔年限
朝向	X_4	南、北、西北、南北等朝向
户型	X_5	一室一厅、两室一厅等户型
面积	X_6	房屋建筑面积
楼层总数	X_7	房屋所在楼的楼层总数
所在楼层	X_8	低、中、高
电梯	X_9	是、否
装修程度	X_{10}	毛坯、简装、精装、豪华装
容积率	X_{11}	小区的地上总建筑面积占用地面积的比率
绿化率	X_{12}	小区绿化面积占小区总面积的比率
物业费用	X_{13}	房屋每月每平方米的物业费

2.3 均价分析

对烟台市二手房进行价格分析后发现,房产价格大致呈正偏态分布,如图 2 所示。由图 2 可以得到:约 60%的二手房单价处于 $6000 \sim 14\ 000$ 元 \cdot m^{-2} ;结合整体房产价格的集中趋势,二手房均价为 $10\ 199$ 元 \cdot m^{-2} ,众数为 $10\ 000$ 元 \cdot m^{-2} ,中位数为 9630 元 \cdot m^{-2} 。由此可以推测,烟台市二手房单价大致在 $10\ 000$ 元 \cdot m^{-2} 上下浮动。

2.4 房龄对房价的影响

烟台市二手房房龄分布从 2022 年交房的 -1 年到 1985 年所建房屋的 36 年,核密度分布情况见图 3。由图 3 可以看到,二手房房龄呈现双峰分布,峰值主要集中在 1999 年和 2019 年。这是因为 1998 年我国取消福利分房政策后,居民住房正式开启商品化道路,房地产行业增加值占国内生产总值的比例从 1998 年的 4% 增长至 2020 年的 7.3%;此外,近几年烟台市房产行业发展良好,政府批准多块土地用于兴建住房^[13]。

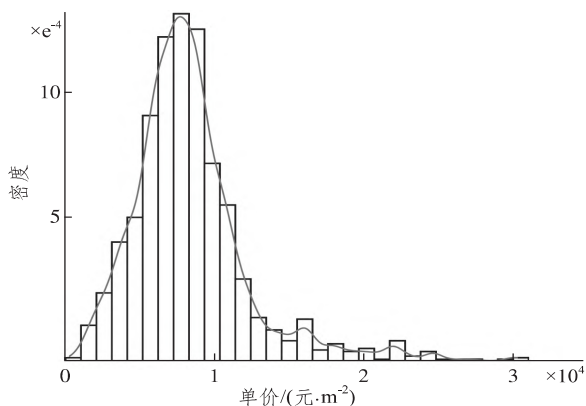


图 2 烟台市二手房单价分布

Fig.2 Second-hand housing unit price distribution in Yantai

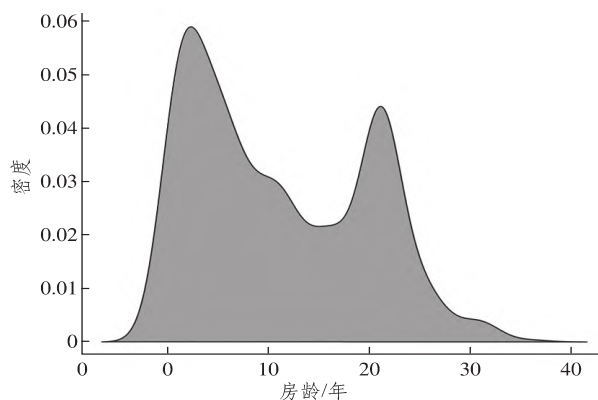


图 3 烟台市二手房房龄的核密度分布

Fig.3 The kernel density of the age of second-hand housing in Yantai

2.5 电梯和装修程度对房价的影响

分析电梯和装修程度两个离散变量对房价的影响,得到图 4。通过图 4 可以看出,有电梯的房源价格略高于无电梯房源;豪华装修的二手房房价最高,毛坯房二手房房价最低,简装房的价格略高于精装

房。从实际情况考虑,购买二手房的多数居民在购房后会对房屋进行改造,相对于精装来说,简装更便于重新翻修。

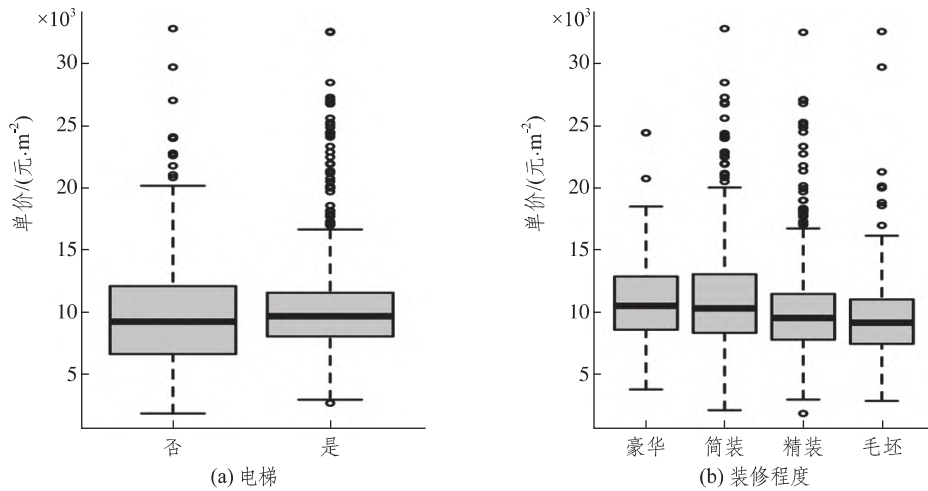


图4 电梯和装修程度对房屋单价影响

Fig.4 The effect of elevators and the degree of decoration on housing unit price

2.6 小区配置对房价的影响

将绿化率、物业费合并为小区配置进行综合分析,分析结果见图5、6。由图5可以看到,随着绿化率增加,二手房对应的平均房龄反而降低,推测高绿化率的小区属于近几年新建楼盘出售的二手房。此外,对于相对繁荣且房价较高的商圈来说,其绿化率普遍较低,从而使开发成本降低;对于偏远区域,虽然绿化率较高,但由于楼盘购买力较低、人口流动少、商圈覆盖少等因素,高绿化率区域的房价反而会较低。

图6为物业费用对房屋单价的影响情况,可以看出,物业管理费越高,房价越高。这与实际情况相符,因地理位置、内部环境、配套管理等影响因素,高级住宅小区房价相对要高。

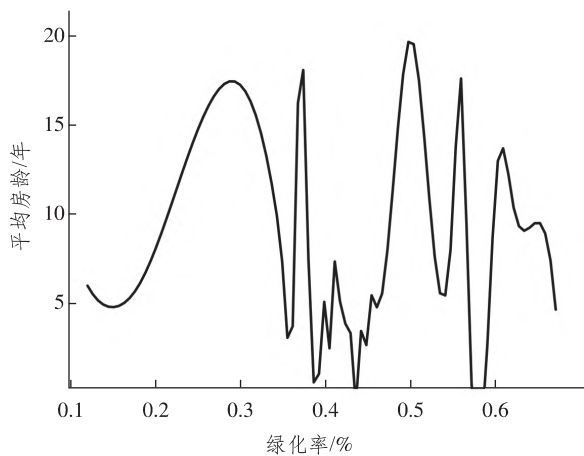


图5 不同绿化率对应的平均房龄
Fig.5 Average house age corresponding to different greening rates

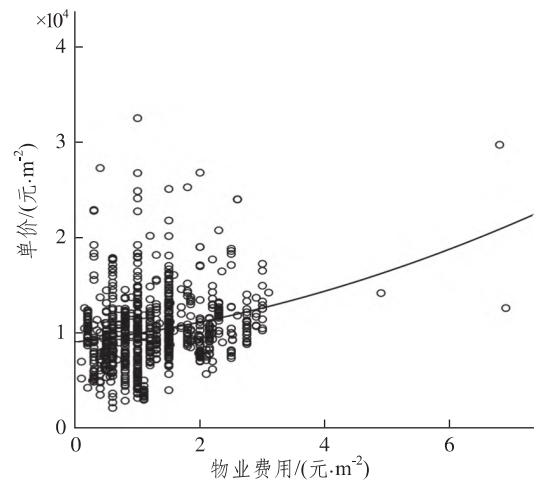


图6 物业费用对房屋单价的影响
Fig.6 The influence of property cost on housing unit price

3 协方差模型

3.1 不考虑交互作用的多因素协方差模型

考虑表1中13个影响因素,构建模型如下:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \beta_4 \times X_4 + \beta_5 \times X_5 + \beta_6 \times X_6 + \beta_7 \times X_7 + \beta_8 \times X_8 + \beta_9 \times X_9 + \beta_{10} \times X_{10} + \beta_{11} \times X_{11} + \beta_{12} \times X_{12} + \beta_{13} \times X_{13} + \varepsilon, \quad (2)$$

其中, β_0 为截距项, $\beta_i (i = 1, 2, \dots, 13)$ 为回归系数, ε 为误差项。

通过 R 语言对模型(2)进行协方差分析,结果见表2。

表2 协方差模型分析结果
Tab.2 Covariance model analysis results

影响因素	自由度	均方和	F 值	Pr 值	显著性
行政区	8	1 211 886 593	334.675 1	< 2.2E-16	***
街道位置	40	130 659 274	36.082 9	< 2.2E-16	***
房龄	1	11 240 489	4.865 1	0.027 7	**
户型	31	21 928 762	6.055 9	< 2.2E-16	***
面积	1	258 374 450	71.352 8	< 2.2E-16	***
朝向	5	2 197 063	0.950 9	0.466 3	
所在楼层	2	21 643 677	5.977 1	0.002 6	***
楼层总数	1	4 281 917	1.182 5	0.277 0	
电梯	1	85 264 673	23.546 7	1.36E-06	***
装修程度	3	40 389 169	11.153 9	3.12E-07	***
容积率	1	1 175	0.000 3	0.985 6	
绿化率	1	1 039 917	0.287 2	0.592 1	
物业费用	1	42 129 252	11.634 4	0.000 7	***

注: **表示在1%显著性水平下呈显著, ***表示在0.1%显著性水平下呈显著。

由表2可以看到,在显著性水平为5%的条件下,朝向、楼层总数、容积率、绿化率4个因素没有通过显著性检验,需从模型(2)中剔除,从而得到修正模型:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \beta_5 \times X_5 + \beta_6 \times X_6 + \beta_8 \times X_8 + \beta_9 \times X_9 + \beta_{10} \times X_{10} + \beta_{13} \times X_{13} + \varepsilon. \quad (3)$$

通过 R 语言对模型(3)进行协方差分析,得到具体分析结果,见表3。

表3 修正的协方差模型分析结果
Tab.3 Modified covariance model analysis results

影响因素	自由度	均方和	F 值	Pr 值	显著性
行政区	8	1 211 886 593	333.674 9	< 2.2E-16	***
街道位置	64	130 659 274	35.975 1	< 2.2E-16	***
房龄	1	11 240 489	4.880 2	0.027 5	**
户型	31	21 928 762	6.037 8	< 2.2E-16	***
面积	1	258 374 450	71.139 5	< 2.2E-16	***
所在楼层	2	21 964 761	6.047 7	0.002 4	***
电梯	1	83 546 062	23.003 2	1.80E-06	***
装修程度	3	37 395 803	10.296 4	1.05E-06	***
物业费用	1	32 067 428	8.829 3	0.003	***

注: **表示在1%显著性水平下呈显著, ***表示在0.1%显著性水平下呈显著。

通过表3可以发现,在显著性水平为5%的条件下,模型(3)中的变量均通过显著性检验。其中,调整前可决系数 R^2 为0.923 6,调整后 R^2 为0.859 7,说明模型拟合优度较好。因此,烟台市二手房的价格

方程为:

$$Y = 9\,169.145 + \beta_1 \times X_1 + \beta_2 \times X_2 + 7.676 \times X_3 + \beta_5 \times X_5 + 25.75 \times X_6 + \beta_8 \times X_8 + 629.584 \times X_9 + \beta_{10} \times X_{10} + 205.638 \times X_{13} + \varepsilon \quad (4)$$

由方程(4)可以得到以下结论:

- 1) 房龄与二手房单价呈正相关。在其他因素保持不变的情况下,房龄越大,价格越高。但相关程度不大,房龄每增加1年,单价会增加7.676元·m⁻²。
- 2) 面积与二手房房价呈正相关。面积每增加1m²,房价会增加25.75元·m⁻²,即面积越大,房价越高。
- 3) 是否配备电梯与二手房房价呈正相关。在其他因素保持不变的情况下,配备电梯的房源会比没有电梯的房源单价高629.584元·m⁻²。
- 4) 物业费用与二手房房价呈正相关。物业费用每增加1元,二手房单价会上涨205.638元·m⁻²。这与实际情况相符,因为物业费用越高,小区环境和品质越好,房价就越高。

3.2 考虑交互作用的多因素协方差模型

在不考虑交互作用的多因素协方差模型(4)中,剔除的因素并非对房价毫无影响,更大可能是该因素与其他因素产生交互作用,进而影响房价。因此,直接剔除这些因素,所得模型也可能并非最优结果。为此需考虑自变量间的交互作用对目标变量的影响,提出变量间可能存在的3种交互作用。“房龄”与“楼层总数”可能存在交互作用,当楼层总数小于6层时,房龄可能会较高,当楼层总数大于7层时,房龄可能会较低。“房龄”与“容积率”可能存在交互作用,房龄较小的房源容积率可能更大,房龄较大的房源容积率可能更小。“电梯”与“所在楼层”可能存在交互作用,当二手房有电梯时,其所在楼层可能较高,否则,所在楼层可能较低。基于各因素间的交互作用,建立如下模型:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \beta_4 \times X_4 + \beta_5 \times X_5 + \beta_6 \times X_6 + \beta_7 \times X_7 + \beta_8 \times X_8 + \beta_9 \times X_9 + \beta_{10} \times X_{10} + \beta_{11} \times X_{11} + \beta_{12} \times X_{12} + \beta_{13} \times X_{13} + \beta_{14} \times X_3 \times X_7 + \beta_{15} \times X_3 \times X_{11} + \beta_{16} \times X_9 \times X_8 + \varepsilon \quad (5)$$

其中, $\beta_i (i = 1, 2, \dots, 16)$ 为回归系数, $X_3 \times X_7$ 、 $X_3 \times X_{11}$ 、 $X_9 \times X_8$ 为交互作用项。

通过R语言对模型(5)进行协方差分析,结果见表4。

表4 存在交互作用的协方差模型分析结果

Tab.4 Covariance model analysis results with interaction

影响因素	自由度	均方和	F值	Pr值	显著性
行政区	8	1 211 886 593	339.483 4	< 2.2E-16	***
街道位置	64	130 659 274	36.601 3	< 2.2E-16	***
房龄	1	2 404 891	4.280 2	0.045 6	**
户型	31	21 928 762	6.142 9	< 2.2E-16	***
面积	1	258 374 450	72.377 9	< 2.2E-16	***
朝向	9	3 646 404	1.021 5	0.420 3	
所在楼层	2	21 643 677	6.063	0.002 4	***
楼层总数	1	4 281 917	1.199 5	0.273 6	
电梯	1	85 264 673	23.885	1.15E-06	***
装修程度	3	40 389 169	11.314 1	2.48E-07	***
容积率	1	1 175	0.000 3	0.985 5	
绿化率	1	1 039 917	0.291 3	0.589 5	
物业费用	1	42 129 252	11.801 6	0.000 6	***
房龄: 楼层总数	1	57 226 873	16.030 9	6.57E-05	***
房龄: 容积率	1	134 561	0.037 7	0.846 1	
所在楼层: 电梯	1	22 688 511	6.355 7	0.011 8	**

注: **表示在1%显著性水平下呈显著, ***表示在0.1%显著性水平下呈显著。

由表4可以得到,考虑交互作用后,朝向、容积率、绿化率以及房龄与容积率的交互作用仍没有通过显著性检验,但楼层总数与房龄的交互作用对房价有显著影响。因此,修正模型(5)得到:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \beta_5 \times X_5 + \beta_6 \times X_6 + \beta_7 \times X_7 + \beta_8 \times X_8 + \beta_9 \times X_9 + \beta_{10} \times X_{10} + \beta_{13} \times X_{13} + \beta_{14} \times X_3 \times X_7 + \beta_{16} \times X_9 \times X_8 + \varepsilon \quad (6)$$

通过 R 语言对模型(6)进行协方差分析,得到具体结果,见表5。表5表明,各变量对房价均有显著影响。因此,模型(6)为二手房房价的最终估值模型。

表5 存在交互作用修正后的协方差模型分析结果

Tab.5 Covariance model analysis results after the presence of interaction correction

影响因素	自由度	均方和	F 值	Pr 值	显著性
行政区	8	1 211 886 593	339.676 1	< 2.2E-16	***
街道位置	64	130 659 274	36.622 1	< 2.2E-16	***
房龄	1	2 404 891	4.280 2	0.045 6	**
户型	31	21 928 762	6.146 3	< 2.2E-16	***
面积	1	258 374 450	72.419	< 2.2E-16	***
所在楼层	2	21 964 761	6.156 4	0.002 2	***
楼层总数	1	4 449 180	1.247	0.264 3	
电梯	1	82 613 174	23.155 4	1.66E-06	***
装修程度	3	39 726 465	11.134 8	3.20E-07	***
物业费用	1	39 075 497	10.952 4	0.001	***
房龄: 楼层总数	1	59 100 175	16.565	4.97E-05	***
所在楼层: 电梯	1	21 611 141	6.057 3	0.014	**

注: ** 表示在 1% 显著性水平下呈显著, *** 表示在 0.1% 显著性水平下呈显著。

4 变量重要性分析

由于线性回归模型预测精度较差,且无法直观判断变量重要性大小,因此本文采用 Lasso 回归对数据作进一步分析。Lasso 回归在解决数据特征维度过高问题的同时优化模型,使模型相对简单^[12]。由于 Lasso 回归只能对数值型变量进行分析,所以需要对类别行变量进行转换,结果如表6所示。

表6 变量转换

Tab.6 Variable conversion

影响因素	观测值	对应编码
行政区	芝罘区、莱山区、福山区、开发区、高新区、牟平区、蓬莱区、莱阳市、龙口市	9、8、7、6、5、4、3、2、1
所在楼层	低、中、高	3、2、1
电梯	是、否	1、0
装修程度	豪华、精装、简装、毛坯	4、3、2、1

根据 R 语言的 glmnet 函数,可以利用广义交叉验证法确定最优调和参数 λ 值, lambda.min 表示模型误差最小时对应的 $\log \lambda$ 取值。为保留尽可能多的数据信息,本文选择该值进行分析,得到降维后系数不为 0 的解释变量。其中,电梯、容积率、绿化率降维后系数为 0,与上文研究结果相同。将引入模型的数值型变量按重要性排序,分别得到系数路径解(图7)和特征系数值(图8)。综合图7、8,将影响二手房单价的变量按重要性排序,依次为行政区、物业费、所在楼层、房龄、装修程度、面积、楼层总数。

5 结语

通过对烟台市二手房数据进行建模分析发现,影响烟台市二手房房价的重要因素分别为:行政区、物业费和所在楼层。具体来看,市中心区域的房价明显高于市周边区域;物业费每提高 1 元 $\cdot m^{-2}$,房价就会上涨 341.61 元 $\cdot m^{-2}$;所在楼层每提升一个单位,房价上涨 161.85 元 $\cdot m^{-2}$ 。

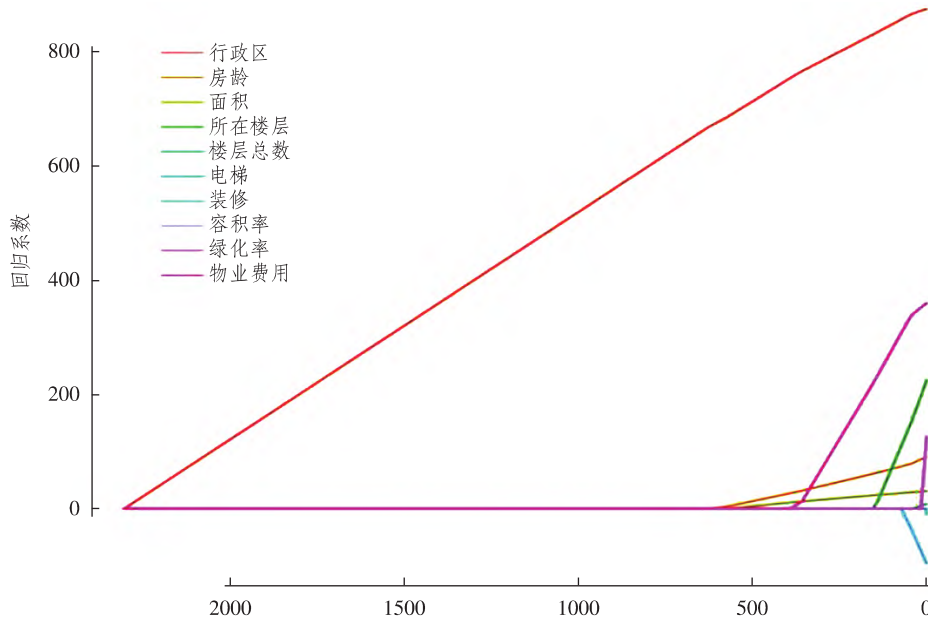


图7 Lasso 回归系数路径解

Fig.7 Lasso regression coefficient path solution

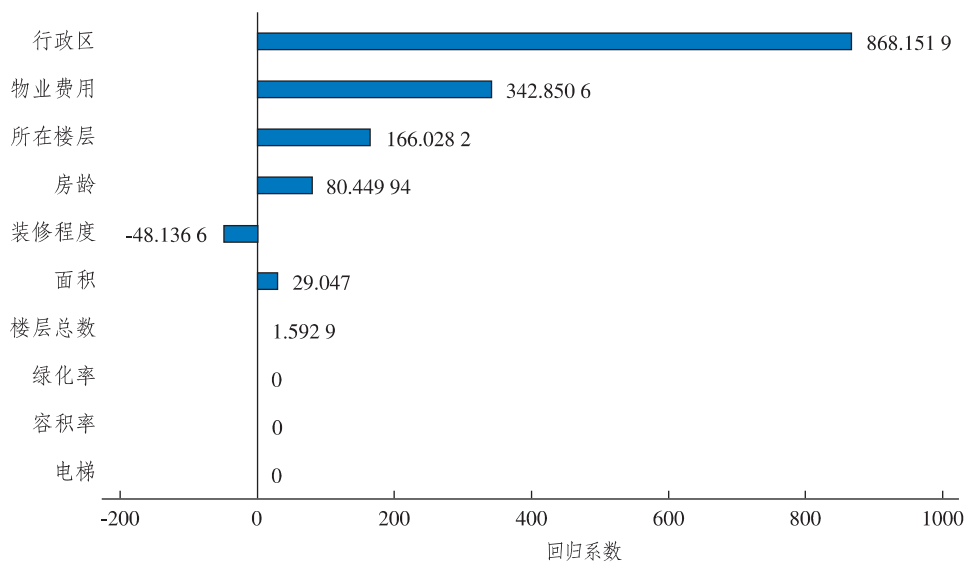


图8 不同影响因素的特征系数值

Fig.8 Characteristic coefficient values of different influencing factors

房地产行业作为我国经济发展的支柱性产业,其健康发展对我国宏观经济至关重要。首先,政府部门需建立更为严格的监督管理机制,在不引起经济衰退等问题的情况下,适当降低住宅房地产投资的数量和强度,或出台针对低收入群体的购房优惠政策,可以在一定程度上缓解房价增长^[14]。其次,烟台市二手房产权问题相对较多,影响市场的快速发展,政府应积极引导相关利益方,依法解决产权纠纷。最后,中介机构作为买卖双方的沟通桥梁,是房产信息的主要来源,在二手房市场中发挥着重要作用,所以应规范中介机构信息收集。总之,政府部门应严格规范行业标准,加强对相关机构的监督管理,引导住宅市场和中介机构健康发展。

参考文献:

[1] 黎永红.我国房地产市场特征和影响房价的因素分析[J].商讯,2021(8):153-154.

- [2] 丁力人.我国房产价格影响因素的分析[J].市场周刊(理论研究),2015(5):32-35.
- [3] 周佳琪,金百锁.基于空间网络自回归变点模型的合肥市房地产价格影响因素分析[J].中国科学院大学学报,2020,37(3):398-404.
- [4] YEE G F,SUFAHANI S F,WAHAB M H A, et al.Analytical Hierarchical Process (AHP): statistical index and factors that influence the housing price in Kuala Lumpur [J].Journal of Physics: Conference Series,2021,1793(1):12-28.
- [5] YU B.Research on influencing factors of average sales price of residential commercial housing in major cities of China [J].IOP Conference Series: Earth and Environmental Science,2020,608(1):12-20.
- [6] YALAN S.Research on spatial spillover effects and influencing factors of urban housing price in China based on gravity weights [J].Academic Journal of Business & Management,2020,2(6):147-155.
- [7] 高华.武汉市三环线内二手房价空间分异及影响因素研究[D].武汉:武汉大学,2020.
- [8] 潘添翼,贾德铮.上海二手房房价影响因素[J].中国市场,2019(5):29-32.
- [9] 胡建华.我国房地产价格影响因素分析研究[D].北京:对外经济贸易大学,2019.
- [10] 蒲东齐,邵丁玲,吴婷,等.基于网络搜索行为对商品房价格的短期预测[J].信息通信,2018,4(1):17-19.
- [11] 杜子芳.抽样技术及其应用[M].北京:清华大学出版社,2005.
- [12] 王海泉.武汉市二手房价格评估研究[D].武汉:华中师范大学,2018.
- [13] 张仁勇,秦伟山.烟台市中心城区房地产类项目供地量的演变规律、影响因素及调控路径[J].鲁东大学学报(自然科学版),2019,35(4):360-364.
- [14] MENG M S,ZHI F L.Econometric analysis of macro influencing factors of housing prices in China: comparison of two period data based on EVIEWS [J].Global Journal of Economics and Business Administration,2019,3(21):22-28.

Influencing Factors of Second-hand Housing Prices in Yantai City Based on Covariance Regression

GUO Yu, YU Haisheng, KE Jiawu

(School of Mathematics and Statistics Science, Ludong University, Yantai 264039, China)

Abstract: In order to deeply study the characteristics of real estate market in Yantai City, this paper used web crawler technology to collect data on Anjuke website, and determined 12 main factors that affect the housing prices of second-hand houses through an interactive multi-factor covariance model. Moreover, the valuation model of second-hand housing price in Yantai City was established. Finally, Lasso regression model was introduced to further analyze the numerical variables. It is concluded that the important factors affecting the housing prices of second-hand houses in Yantai City are administrative area, property fees and floor.

Keywords: real estate; web crawler; covariance model; Lasso regression

(责任编辑 顾建忠)