

南宁市房地产价格影响因素线性回归模型分析

■ 文 | 欧阳新宇 吕 津

摘要:房地产业是我国的支柱产业,文章选取了影响南宁市房地产价格的九个指标,通过构建多元线性回归模型对房地产价格的影响因素进行量化分析,结果表明,房地产价格主要受地区生产总值、年末总人口和房地产开发投资额的影响。

关键词:房地产价格 回归模型 南宁市

房地产业是我国经济的重要组成部分。作为西南地区的大型城市和交通枢纽,南宁市在房地产业的发展上对整个西南地区的房地产市场具有重要影响,因此研究南宁市的房价具有重要意义。文章定量分析了南宁市房价的影响因素,以期为南宁市的房价研究作出贡献。

1 模型构建

1.1 模型假设

为便于统计,避免数据误差,对模型进行基本假设:时间上,选取2002—2020年的数据;空间上,研究范围为南宁市全市;数据来源上,优先使用国家统计局数据;数据指标上,选取能量化的指标;方法上,使用多元回归模型和最小二乘法。

1.2 指标选择

将“商品房平均销售价”作为被解释变量,用 Y 表示。选取的变量为地区生产总值、在岗职工平均工资、年末总人口、居民消费价格指数、土地成交均价、房地产开发投资额、房地产开发企业购置土地面积、房地产开发企业竣工房屋土地面积、货币和准货币($M2$)供应量。

1.3 数据获取

通过查阅国家统计局、南宁市历年统计年鉴和前瞻数据

库获取数据,如表1所示。

2 实证分析

2.1 模型建立

根据所选取的指标,构建影响房价的模型,模型如下:

$$Y=C+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_7+\beta_8X_8+\beta_9X_9 \quad (1)$$

式中: C 为常数项; β 为系数; X 为解释变量。

2.2 多元线性初步回归

使用eviews7软件对数据进行分析,回归结果如图1所示。

由图1可得模型的回归方程如下:

$$Y=-7370.062+0.531656X_1+0.008542X_2+1.338955X_3+20.13194X_4+0.146013X_5-0.412075X_6-706174X_7-1.576309X_8+0.000366X_9 \quad (2)$$

图2揭示了该模型的拟合情况,结果表明整体拟合良好, \bar{R}^2 (调整后 R^2)达到了98.86%。同时, F 检验的统计量值为174.6650且显著性水平 P 值为0,说明解释变量对被解释变量的影响具有显著性。但大部分回归参数的 t 检验的统计量值不显著,这说明可能存在多重共线性问题。采用逐步回归法来解决。

表 1：南宁市 2002—2020 年商品房平均销售价格及影响因素数据

年份	商品房平均 销售价格 / (元·m ⁻²)	地区生产总 值 / 亿元	在岗职工平 均工资 / 元	年末总人 口 / 万人	居民消费价 格指数	土地成交均 价 / (元·m ⁻²)	房地产开发投 资额 / 亿元	房地产开发企 业购置土地面 积 / ×10 ⁴ m ²	房地产开发企 业竣工房屋面 积 / ×10 ⁴ m ²	货币和准货 币 (M2) 供应 量 / 亿元
	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
2002	2372	463.18	11363	634.68	433.5	1016.8300	24.71	80.6	118.8	185006.97
2003	2252	541.78	13173	641.67	438.7	1587.2300	39.48	103.6	215.03	221222.8
2004	2761.11	617.12	15447	648.85	455.8	774.3000	66.04	244.68	402.43	254107
2005	2605.03	727.9	17520	659.54	464	2293.1201	105.5	188.37	414.06	298755.7
2006	2873.42	880.11	20650	671.89	471	1969.1000	139.07	252.23	363.24	345577.9
2007	3404	1089.07	24789	683.51	493.6	2362.7800	187.46	351.61	419.77	403442.21
2008	3952	1320.43	29377	691.69	522.7	742.3600	202.46	287.5	456.13	475166.6
2009	4557	1524.71	32596	697.90	519	2493.8999	226.73	149.6	439.71	610224.5
2010	5135	1800.26	37042	707.37	536.1	4973.4102	322.82	190.06	519.25	725851.8
2011	5196.00	2211.44	40120	711.49	565	1974.2300	392.42	191.85	564.69	851590.9
2012	6002.89	2503.18	43847	713.50	579.7	5033.0200	362.73	101.36	664.53	974147.8
2013	6959.36	2845.60	48188	724.43	594.8	2902.8301	416.37	80.11	325.58	1106524.98
2014	6627.03	3148.32	54826	729.66	606.7	2493.2200	551.82	245.75	465.43	1228374.81
2015	6645.98	3147.91	63820	740.23	615.2	2898.8101	657.19	149.75	574.97	1392278.11
2016	6886.79	3405.99	68560	751.74	627.5	604.6000	854	192.76	471.61	1550066.67
2017	7776.40	3804.33	75481	756.87	637.5	3305.0000	958.09	165.95	578.22	1690235.31
2018	7782.17	4162.37	83452	779.82	650.9	3852.0000	1103.36	121.36	792.21	1826744.2
2019	8406.04	4506.56	90986	781.97	669.8	4915.0000	1461.08	289.82	710.77	1986488.82
2020	8604.82	4726.34	97079	791.38	686.5	4915.0000	1378.2	491.75	799.91	2186795.89

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7370.062	7255.804	-1.015747	0.3363
X1	0.531657	0.896199	0.593235	0.5676
X2	0.008542	0.065732	0.129947	0.8995
X3	1.338955	13.53806	0.098903	0.9234
X4	20.13194	10.66095	1.888381	0.0916
X5	0.146013	0.060136	2.428027	0.0381
X6	-0.412075	1.151250	-0.357937	0.7286
X7	-0.706174	0.709449	-0.995384	0.3456
X8	-1.576309	0.758454	-2.078318	0.0674
X9	-0.000366	0.002470	-0.148248	0.8854
R-squared	0.994307	Mean dependent var	5305.160	
Adjusted R-squared	0.988615	S.D. dependent var	2181.591	
S.E. of regression	232.7800	Akaike info criterion	14.04348	
Sum squared resid	487678.8	Schwarz criterion	14.54056	
Log likelihood	-123.4131	Hannan-Quinn criter.	14.12761	
F-statistic	174.6650	Durbin-Watson stat	2.498218	
Prob(F-statistic)	0.000000			

图 1：最小二乘法（OLS）回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.032696	1.246624	2.432727	0.0280
X1	0.000309	7.48E-05	4.135222	0.0009
X6	-0.000785	0.000130	-6.033798	0.0000
X3	0.007216	0.001957	3.687023	0.0022
R-squared	0.980030	Mean dependent var	8.483276	
Adjusted R-squared	0.976036	S.D. dependent var	0.461296	
S.E. of regression	0.071410	Akaike info criterion	-2.256089	
Sum squared resid	0.076491	Schwarz criterion	-2.057260	
Log likelihood	25.43285	Hannan-Quinn criter.	-2.222439	
F-statistic	245.3746	Durbin-Watson stat	2.510707	
Prob(F-statistic)	0.000000			

图 2：保留的解释变量最小二乘法（OLS）回归结果

2.3 逐步回归

使用 eviews7 软件进行分析，结果如表 2 所示。

通过观察逐步回归的结果，可以看出 X_7 的 $\text{Pro}(t)$ 值大于 0.05，表示 X_7 的 t 统计量没有通过显著性检验，通过观察 R^2 ，可以看出 X_5 对被解释变量拟合优度的值最小，因此舍弃 X_5 和 X_7 。

按照解释变量贡献大小逐一引入的方式进行检验。在引入解释变量时，如果其回归参数没有通过 t 检验，则将其从模型中排除；反之亦然。这样可以有效地选择出对被解释变量产生重要影响的解释变量。

使用上述方式，确定了 X_1 、 X_6 和 X_3 对被解释变量有显著影响，然后应用最小二乘法进行回归分析，结果如图 2 所示。

由图 2 可得回归方程如下：

$$Y=3.032696+0.000309X_1-0.000785X_6+0.007216X_3 \quad (3)$$

2.4 模型检验

2.4.1 经济意义检验

X_1 的值表明当地区生产总值上升 1%，商品房平均销售价格上升 0.031%，这与理论上 GDP 和商品房平均销售价

表 2：单因素一元线性回归一览表

回归结果	R^2	Adjusted- R^2	D-W 值	F 值	Prob (F)	t 值	Prob (t)
X_1	0.975484	0.974042	1.049555	676.4287	0.000000	26.00824	0.0000
X_2	0.935363	0.931561	0.539421	246.0089	0.000000	15.68467	0.0000
X_3	0.953123	0.950365	0.997500	345.6502	0.000000	18.59167	0.0000
X_4	0.982756	0.981742	0.905089	968.8434	0.000000	31.12625	0.0000
X_5	0.390652	0.354808	0.801512	10.89866	0.004218	3.301312	0.0042
X_6	0.836211	0.826575	0.402864	86.79203	0.000000	9.316224	0.0000
X_7	0.034030	-0.022792	0.086008	0.598886	0.449632	0.773877	0.4496
X_8	0.609687	0.586727	1.014402	26.55476	0.000080	5.153535	0.0001
X_9	0.946863	0.943737	0.579390	301.9285	0.000000	17.40484	0.0000

格成正比的关系是一致的。 β_6 的值表明当房地产开发投资额下降 1%，商品房平均销售价格上升 0.079%，在这种情况下，地产投资的减少导致供应减少，最终影响商品房价，所以房地产开发投资额与房价成负相关关系，这符合经济现实意义。 β_3 的值表明当年年末总人口上升 1%，那么商品房平均销售价格会上升 0.72%，这证明了人口与房价之间的正相关性，人口流入带来需求增加，推动房价上涨。

2.4.2 统计推断检验

(1) 拟合优度检验。由图 2 中看出模型对样本的拟合效果好（可决系数 $R^2=0.980030$ ， $\bar{R}^2=0.976036$ ）。

(2) 方程显著性检验（F 检验）。已知显著性水平 $\alpha=5\%$ ，在 F 分布表中的临界值为 $F_{0.05}(3,15)=3.29$ ，由于 $F > F_{0.05}(3,15)$ 且 $\text{Prob}(F\text{-statistic}) < 0.05$ ，所以回归方程是显著的，解释变量对被解释变量有显著影响，并且通过了检验。

(3) 变量显著性检验（t 检验）。已知显著性水平 $\alpha=5\%$ ，在 t 分布表中的临界值为 $t_{0.025}(15)=2.131$ ，由于 $t(\beta_1)=4.135222 > t_{0.025}(15)$ ， $|t(\beta_6)|=6.033798 > t_{0.025}(15)$ ， $t(\beta_3)=3.687023 > t_{0.025}(15)$ 。因此，在 95% 的置信水平下，解释变量对被解释变量的影响是显著的，并且通过了检验。

2.5 结论

通过逐步回归后，得到了房价主要的影响因素是地区生产总值、年末总人口和房地产开发投资额。当地区经济增长时，总收入增加，购房需求增加，房价上涨。地区人口增加也会导致购房需求增加，推动房价上涨。房地产投资额的增

加会导致供给增加，进而导致房价下跌，反之亦然。

3 结束语

房价的影响因素很多，有些因素可量化，有些因素不可量化（如政策因素），很难分析出这个因素对房价的影响有多大。不同人对房价影响因素的认识可能有所不同，不同学者使用不同的研究方法得出结论也可能有所不同。同样，即使使用相同的研究方法分析不同地区的情况，结果也可能不同。文章通过回归模型分析影响南宁市房价的因素，得出影响其房价的主要因素是地区生产总值、年末总人口和房地产开发投资额。

参考文献

- [1] 中华人民共和国统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2021.
- [2] 南宁市统计局. 南宁统计年鉴 2021[M]. 北京: 中国统计出版社, 2021.
- [3] 邱婷婷. 福州市房地产价格影响因素研究[J]. 福建技术师范学院学报, 2021, 39(6): 597-603.
- [4] 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.

作者 / 欧阳新宇

本科 研究方向为产业经济学
单位 闽南理工学院

通讯作者 / 吕津

博士 教授
研究方向为产业经济学
单位 闽南理工学院