

# 基于主成分分析法的商品住宅特征价格模型改进\*

黄古博, 李雨真

(武汉理工大学 经济学院, 湖北 武汉 430072)

**摘要** 特征价格模型作为一种对商品住宅价格进行有效分析和评估的工具而广泛应用于房地产领域。针对特征价格模型在实际应用中自变量数目较多、选择难度大且存在多重线性相关的问题, 提出采用主成分分析法对商品住宅特征模型进行改进, 排除变量间的线性相关对特征价格模型产生的误差。同时, 以武汉市轻轨沿线商品住宅楼盘作为研究对象, 构建了基于主成分分析的商品住宅特征价格模型, 并用该模型对商品住宅价格的预测结果与实际销售价格进行对比, 结果显示多数样本预测误差在 5% 以下, 验证了改进模型的优越性。

**关键词** 主成分分析法; 特征价格模型; 模型改进; 商品住宅; 实证分析

**中图分类号:** F287.8    **文献标识码:** A    **文章编号:** 1008-3456(2011)04-0093-05

商品住宅价格及其变动趋势的准确测度对于宏观经济运行和微观经济行为具有重要意义, 但由于非标准资产的异质性和个人资产交易的低频率性, 使得商品住宅价格的准确测度存在一定困难。在国外, 特征价格模型是一种广泛的应用于房地产领域, 能够对商品住宅价格进行有效的分析和评估工具。我国学者对该模型的研究仅有 10 多年的历史, 在很多方面如: 模型的函数选择、模型的应用领域等的研究还不够深入, 即使在房地产领域中的应用, 如对商品住宅价格及其变动趋势的估计, 也存在着许多诸如变量选择的问题<sup>[1]</sup>。主成分分析法, 强调指标变量间的差异性, 各指标变量权重系数反映了指标变量之间的内在结构关系, 具有客观性, 对自变量进行变换后可以形成相互独立的主成分, 能够减少变量间的信息交叉, 降低变量的选择难度, 因而经常被用于对多个变量的筛选。针对特征价格模型在商品住宅价格估计中的变量选择问题, 陈安明<sup>[2]</sup>在国内首次提出应用主成分分析法改进商品住宅的特征价格模型, 用来估计住宅的销售均价。遗憾的是作者没有进一步对改进模型的优越性给出有效的数据证明。

近几年, 有学者如周丽萍、宋永发等<sup>[3-4]</sup>实证研究了主成分分析法对住宅特征价格模型的改进效果。城市轨道交通对城市的区划作用明显, 对沿线住宅价值影响较大, 轨道交通也常被用作城市房产

市场细分的分界线<sup>[5-6]</sup>。本文从分析特征价格模型在商品住宅价格核算中的应用及存在的问题出发, 用主成分分析法对商品住宅特征价格模型进行改进, 并用改进的模型对武汉市沿轻轨线的 13 个楼盘进行销售均价估计, 通过模型估计值与实际销售均价的对比, 验证改进模型的优越性。

## 一、特征价格模型及其在商品住宅价格估算中的应用

### 1. 特征价格模型的建立

特征价格模型(hedonic price model, HPM)是一种经常被用来研究异质产品的差异特征与产品价格之间关系的模型。其理论基础是消费者理论及供求均衡模型, 其基本假定是商品异质性与特征市场隐含性<sup>[7]</sup>。特征价格模型认为消费者对异质商品的需求不是基于商品本身, 而是商品内含的特征或属性, 商品效用的高低由商品内含的特征或属性的数量决定。模型构建的基本思路是, 消费者愿意为某件商品支付的价格取决于他能够从商品的各种特征属性中获得的享受程度。某商品的特征价格模型为: 假设市场中所有家庭的收入水平和偏好类似, 那么这个市场中某商品的价格应当是该商品特征的函数, 即  $p = p(Z_1, Z_2, \dots, Z_n)$ , 式中  $p$  为该商品的价格;  $Z_1, Z_2, \dots, Z_n$  是该商品的特征。

收稿日期: 2011-03-06

\* 国家发展和改革委员会项目“中部地区城市发展与政策研究”(20093924); 中央高校基本科研业务费专项资金自主创新项目“高房价诱因的实证研究——以武汉市为例”。

作者简介: 黄古博(1990-), 男; 研究方向: 房地产金融。E-mail: 531413400@qq.com

## 2. 用于商品住宅估价的特征价格模型

商品住宅非常明显异质性及构成使用价值的产品特征存在明显的差异,正好符合特征价格模型的基本假定。因此,自 1967 年 Ridker 把特征价格理论应用到住房市场分析以来,国外很多学者开展了大量的相关研究。如 Robert<sup>[8]</sup>就曾利用该模型专题研究了城市轨道交通对周边房地产价值的影响。商品住宅的价格变动包括随商品住宅特征改变而引起的价格变动及随时间变化而引起的价格变动。商品住宅特征主要包括住宅的区位特征、住宅本身的建筑特征和住宅的环境特征。住宅的区位特征包括住宅与购物中心的距离、住宅附近是否有公共交通、住宅与上班地点的距离等;住宅本身的建筑特征包括住宅面积和各房间的大小、住宅所在楼层、住宅朝向、住宅浴室的个数等;住宅的环境特征包括住宅周边的学校情况、基础设施配套情况、绿化情况、物业管理情况等。商品住宅特征价格模型的一般形式为:

$$P_i = P(S_i, N_i, Q_i) \quad (1)$$

式(1)中, $P_i$ 为商品住宅的销售价格; $S_i$ 为商品住宅的建筑特征向量; $N_i$ 为商品住宅的环境特征向量; $Q_i$ 为商品住宅的区位特征向量。

特征价格模型函数表达形式没有一个特有理论定式,研究者考虑到自变量和因变量的相互关系,可以采用包括指数函数、对数函数、线性函数、二次函数等多种函数形式。其中,线性函数以其计算简便、结果精确而被多数国内学者采用,由线形函数形式表示的商品住宅特征价格模型为:

$$P = a_0 + \sum a_i Z_i + \epsilon \quad (2)$$

式(2)中, $a_0$ 为常数项; $a_i$ 为商品住宅特征的影响系数; $Z_i$ 为商品住宅特征值; $\epsilon$ 为随机误差。

对方程(2)的各个特征变量求偏导数,就得到商品住宅各特征对应的特征价格,商品住宅的总价格等于各个特征价格的代数和。

## 3. 商品住宅特征价格模型估计中存在的问题

(1)市场细分导致有偏差的估计。统一的城市住宅市场是应用特征价格模型的隐含前提。理论上特征价格的研究并不需要细分住宅市场,但现实中的住宅市场往往是由几个不同类型的子市场共同组成的。由于消费者需求结构不同、地域偏好存在差异及相关信息的不对称使得整个城市范围内的住宅市场不可避免地出现细分市场。不同的细分市场对应着不同的特征价格函数,不同的特征价格函数对商品住宅的价格估算肯定不同。如果以统一的城市

住宅市场对应的特征函数来分析细分市场,那么特征价格模型对商品住宅的价格估算肯定会有偏差。要比较准确地估算商品住宅的价格,必须对整个城市范围内的住宅市场进行合理细分。因此,怎样合理地细分市场,成了特征价格模型应用的关键。细分市场方法很多,有的研究者采用区位和政治上的分界特征来细分住宅市场,有的采用社会学或人口统计学上的特征对住宅市场进行细分。问题是目前还无法给出标准来细分子市场和评价细分的差异性。

(2)解释变量间存在多重共线性问题。当解释变量之间存在完全线性相关,那么可以说解释变量之间存在完全多重共线。这个时候,就不能根据样本做任何的假设检验,因为从存在完全多重共线关系的解释变量,根本就没有办法计算出所有参数的唯一估计值。解释变量完全多重共线性的情形不多,但在采用实际调查数据进行分析时,自变量之间共线性程度往往很高。多重共线性问题在实际中总是或多或少的存在,只是存在的程度不同而已。通常用方差膨胀因子法来检验解释变量的多重共线性,解释变量对应的方差膨胀因子大于 10,就可以断定解释变量之间存在着严重的共线性问题。解决多重共线性的问题的关键是从模型中删掉那些不重要的解释变量。

(3)解释变量的错误设定。解释变量的错误设定,指的是相关的解释变量被忽略或者不相关的解释变量被引入。在构建商品住宅特征价格模型中,研究者往往围绕商品住宅的特征选择对应的解释变量,而某个商品住宅特征很多,往往难以穷尽,在进行解释变量选择时经常出现漏选或误选的情况,对选择解释变量进行数据采集时有时只能近似地进行量化。既然所有特征价格模型的估计在某种程度上都存在错误设定,那么在模型中把那些能够产生较多效用的属性考虑到回归方程中就可以满足要求。

## 二、主成分分析法改进的商品住宅特征价格模型

### 1. 主成分分析法

主成分分析法,最早是由美国心理学家 Charles Spearman 于 1904 年提出的,是因子分析法的一种特殊形式。它从解释变量方差的角度来研究变量的解释力,假设所有解释变量的方差能基本上被主成分所解释,那么在分析与评价指标变量时,就可以用

少数几个潜在的相互独立的主成分指标变量的线性组合来反映多个实测指标变量的主要信息。主成分分析法有两大优点:一是降低了变量指标的选择难度。评价指标变量的选择直接关乎评价结果,在实际中评价指标变量之间往往存在相互影响的关系,很难选择合理的解释变量。主成分分析法提取的主成分不影响解释结果,降低了变量维度,从而降低了变量选择难度。二是更为准确对影响因素进行估算与评价。强调差异性原理的主成分分析法,指标变量的权重系数反映了指标变量之间的内在结构关系,系数确定不受主观因素的影响,具有客观性。自变量在经过变换后消除了评价指标之间的相关影响,减少了信息交叉。采用主成分分析法得到的分析评价结果具有客观性和准确性<sup>[9]</sup>。在特征价格模型的构建中引入主成分分析法,建立基于主成分分析的特征价格函数,可以进一步改进特征价格模型的评估效果<sup>[10]</sup>。

## 2. 基于主成分分析法的特征价格模型

设  $n$  个住宅样本的  $m$  个特征变量所构成的矩阵如下:

$$X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{pmatrix}$$

对上述矩阵进行标准化处理得:

$$Z_j = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (3)$$

式(3)中:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

$$i = 1, 2, \dots, n, j = 1, 2, \dots, m$$

根据公式  $r_{jk} = \frac{1}{n-1} \sum_{i=1}^n Z_{ij} Z_{ik}$  得出住宅数据的相关矩阵  $R = (r_{jk})_{m \times m}$ , 由此计算相关矩阵的特征根为:  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ , 相应的特征向量  $L_j = (l_{j1}, l_{j2}, \dots, l_{jm})^T, j = 1, 2, \dots, m$ 。于是标准化后的特征变量转换成主成分:

$$F_j = ZL_j = l_{j1}Z_1 + l_{j2}Z_2 + \dots + l_{jm}Z_m$$

$$j = 1, 2, \dots, m \quad (4)$$

若主成分累计贡献率:

$$\sum_{j=1}^k (\lambda_j / \sum_{j=1}^m \lambda_j) \geq 85\% \quad (k < m)$$

则选择前  $k$  个主成分为主成分变量。

以住宅销售价格  $P$  作为因变量,以主成分变量作为自变量,采用线性函数建立商品住宅特征价格

模型为:

$$P = \alpha_0 + \alpha_1 F_1 + \alpha_2 F_2 + \dots + \alpha_k F_k + \varepsilon \quad (5)$$

## 三、改进模型对商品住宅价格估算效果的实证检验

### 1. 变量选择与数据来源

变量分为自变量和因变量。因变量,一般采用楼盘商品住宅的实际成交价格均值。自变量,从对商品住宅价格有影响的3个主要因素——建筑因素、环境因素和区位因素中选取。①建筑特征变量分析。商品住宅价格与住宅建筑本身的特征相关,住宅建筑包括住宅所在楼盘的整体建筑状况和住宅本身的建筑状况。楼盘整体规划用地面积、楼盘总建筑面积、楼盘类型及住宅建筑面积、住宅房间数、住宅装修程度等是建筑特征中重要的变量,建筑年龄、建筑朝向、装修程度、住宅所在楼层等也影响商品住宅的价格。②环境特征变量分析。环境特征主要包括住宅小区及周边的生态环境和地方政府提供的公共服务。一般的,住宅小区内或周边有良好的景观效果,绿化好、有公园、有水景等,地方政府提供的公共服务设施齐全,有学校、医院等,消费者愿意支付更高的费用,住宅销售价格会上升。此外,物业管理方面对房地产价格也有影响。③区位特征变量分析。房地产最重要的3点,即“区位,区位,还是区位”,区位因素对商品住宅项目价格有重要的影响。商品住宅的区位决定了商品住宅到城市中心、交通站点、购物中心等的距离,直接影响到居民的出行和可达性。一般的,公共交通良好和可达性较好的商品住宅,其销售价格较高。

本文以武汉市轻轨沿线距离车站2 km 以内的13个住宅项目为分析样本,获取相关变量数据。变量、变量涵义及数据获取方式见表1。

### 2. 模型的构建

(1)主成分分析。将武汉市轻轨沿线13个楼盘对应的16个特征属性样本数据,运用SARS软件进行主成分分析,其相关系数矩阵特征值与主分量系数见表2。

从表2可知,前6个特征值的累积贡献率为90.4%,超过85%,可以选择前6个为主分量。根据因子载荷及显著性,可以看出决定  $F_1$  大小的变量主要是  $X_1, X_2, X_3, X_4, X_5$ ; 决定  $F_2$  大小的变量主要是  $X_8, X_{11}, X_{12}, X_{13}$ ; 决定  $F_3$  大小的变量主要是  $X_6, X_7, X_{12}, X_{14}, X_{15}$ ; 决定  $F_4$  大小的变量主要是  $X_{14}$ ,

表 1 变量涵义与数据获取方式

变量	变量涵义	数据获取方式
<b>建筑物特征</b>		
X <sub>1</sub>	楼盘整体规划用地面积/万 m <sup>2</sup>	采用实际数据
X <sub>2</sub>	楼盘总建筑面积/万 m <sup>2</sup>	采用实际数据
X <sub>3</sub>	楼盘类型	根据各楼盘定位不同,高档取 5,中档取 3,低档取 1,混合型则取相应平均值
X <sub>4</sub>	住宅装修程度	根据实际情况,精装修取 5,初装修取 3,毛坯房取 1
X <sub>5</sub>	容积率	采用实际数据
X <sub>6</sub>	是否有现房	已入住 2 年以上取 5,入住 2 年以内取 3,正在建设或拟建设取 1
X <sub>7</sub>	停车位状况	停车位充足取 5,有停车位但不充足取 3,无停车位取 1
<b>环境特征</b>		
X <sub>8</sub>	绿化率	采用实际数据
X <sub>9</sub>	是否有公园	小区内有公园取 5,附近有公园取 3,周边没有公园取 1
X <sub>10</sub>	是否有水景	近江流取 5,小区内部有水景的取 3,周边无水景取 1
X <sub>11</sub>	社区附近是否有中小学	有重点中小学取 5,有一般中小学取 3,没有取 1
X <sub>12</sub>	社区附近是否有医院	有市级重点及以上医院取 5,有市级普通医院取 3,有小区内门诊或社区医院取 1
<b>区位特征</b>		
X <sub>13</sub>	社区附近是否有购物中心	临近商业中心取 5,附近有大型购物中心取 3,附近无大型购物中心取 3 取 1
X <sub>14</sub>	社区附近公交线路状况	有 3 条以上的公交线路取 1,若 3 条以下取 0
X <sub>15</sub>	到最近轻轨车站步行距离/m	采用实际数据
X <sub>16</sub>	乘车到大型客运站的运行时间/s	采用实际数据

表 2 相关系数矩阵特征值与主成分系数

λ	特征值	累计贡献率	变量	主成分					
				F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>
λ <sub>1</sub>	4.968	0.311	Z <sub>1</sub>	0.383	-0.101	0.149	0.179	-0.293	0.047
λ <sub>2</sub>	3.602	0.536	Z <sub>2</sub>	0.341	-0.183	0.140	0.122	-0.306	0.026
λ <sub>3</sub>	2.211	0.674	Z <sub>3</sub>	-0.352	-0.152	0.117	-0.181	-0.147	0.188
λ <sub>4</sub>	1.535	0.769	Z <sub>4</sub>	-0.317	0.029	0.310	0.133	-0.309	0.270
λ <sub>5</sub>	1.911	0.844	Z <sub>5</sub>	-0.370	-0.078	0.245	0.038	-0.021	0.357
λ <sub>6</sub>	0.953	0.904	Z <sub>6</sub>	-0.108	0.194	-0.432	-0.258	-0.337	0.363
λ <sub>7</sub>	0.566	0.939	Z <sub>7</sub>	0.268	0.064	0.347	0.242	0.202	0.322
λ <sub>8</sub>	0.415	0.965	Z <sub>8</sub>	0.206	0.376	-0.050	-0.257	-0.206	0.048
λ <sub>9</sub>	0.302	0.984	Z <sub>9</sub>	0.209	-0.296	-0.016	-0.212	0.452	0.235
λ <sub>10</sub>	0.205	0.997	Z <sub>10</sub>	0.248	-0.231	-0.097	-0.069	0.020	0.609
λ <sub>11</sub>	0.033	0.999	Z <sub>11</sub>	0.172	0.442	-0.015	0.018	0.298	0.127
λ <sub>12</sub>	0.018	1.000	Z <sub>12</sub>	0.017	0.389	0.346	0.299	-0.074	0.039
λ <sub>13</sub>	0.000	1.000	Z <sub>13</sub>	-0.205	0.361	0.139	-0.046	0.297	0.068
λ <sub>14</sub>	0.000	1.000	Z <sub>14</sub>	-0.006	0.011	0.368	-0.563	0.178	0.000
λ <sub>15</sub>	0.000	1.000	Z <sub>15</sub>	-0.064	0.197	-0.432	0.348	0.154	0.271
λ <sub>16</sub>	0.000	1.000	Z <sub>16</sub>	-0.262	-0.294	-0.085	0.360	0.253	-0.033

X<sub>15</sub>, X<sub>16</sub>; 决定 F<sub>5</sub> 大小的变量主要是 X<sub>2</sub>, X<sub>4</sub>, X<sub>6</sub>, X<sub>9</sub>; 决定 F<sub>6</sub> 大小的变量主要是 X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>, X<sub>10</sub>。

(2)武汉市轻轨沿线商品住宅特征价格模型。建立房价 P 与各主成分的回归函数  $P = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \beta_4 F_4 + \beta_5 F_5$ , 将主成分值(F<sub>i</sub> 值)代入回归函数, 利用逐步回归法进行参数估计, 将没有通过 5% 显著性检验的参数去掉, 再进行第二次回归,

表 3 通过 5% 显著性检验的参数估计结果

参数	估计值	标准误	F 值	P 值
β <sub>0</sub>	6 415.38	71.78	7 986.20	0.000 1
β <sub>1</sub>	-90.00	33.46	7.23	0.027 5
β <sub>2</sub>	423.21	37.09	130.18	0.000 1
β <sub>4</sub>	491.26	54.32	81.80	0.000 1
β <sub>5</sub>	198.83	65.67	9.17	0.016 4

得到 5% 显著性检验的参数估计, 结果见表 3。

由此, 得到主成分分析法改进的武汉市轻轨沿线商品住宅特征价格模型为:

$$P = 6415.38 - 90.00F_1 + 423.21F_2 + 491.26F_4 + 198.83F_5 \quad (6)$$

### 3. 模型预测

利用改进模型式(6)对武汉市轻轨沿线的 13 个商品住宅进行价格预测, 预测结果见表 4。

由表 4 可知, 除少数样本外, 多数样本预测误差在 5% 以下, 可见该模型对武汉轻轨沿线商品住宅价格预测比较精准。同时说明, 用主成分分析法能够有效改进商品住宅特征价格模型。

表4 改进模型对武汉市轻轨沿线的商品住宅价格预测结果

楼盘序号	观察价格/(元/m <sup>2</sup> )	预测价格/(元/m <sup>2</sup> )	标准误	预测值价格 95%置信区间/(元/m <sup>2</sup> )	残值	预测误差/%
1	8 600	8 423	159	8 058~8 791	177	2.1
2	8 700	8 775	182	8 354~9 195	-75	-0.8
3	7 000	7 150	112	6 893~7 408	-150	-2.1
4	5 600	5 695	170	5 303~6 087	-95	-1.7
5	6 000	5 832	131	5 530~6 134	168	2.8
6	5 800	5 681	205	5 209~6 153	119	2.1
7	5 600	5 940	87	5 739~6 141	-340	-6.1
8	5 800	6 202	136	5 888~6 515	-402	-6.9
9	5 700	5 437	142	5 111~7 564	263	4.6
10	5 600	5 617	133	5 311~5 923	-17	-0.3
11	5 800	5 605	167	5 221~5 988	195	3.4
12	5 700	5 741	194	5 294~6 188	-41	-0.7
13	7 500	7 303	216	6 803~7 802	197	2.6

## 四、结 语

引入主成分分析法建立商品住宅特征价格模型,可以尽量减少商品住宅特征属性间存在的相关性,实证检验改进模型对商品住宅价格的预测比较精准,多数样本预测误差在5%以下。采用主成分分析法将选中的特征属性仍然保留在模型中,便于根据具体商品住宅特征属性进行定价。

## 参 考 文 献

- [1] 张立新,姜吉坤,温海珍.基于特征价格模型的商品住宅价格评估方法应用[J].统计与决策,2010(21):157-159.
- [2] 陈安明.基于主成分分析的住宅项目特征定价模型[J].重庆大学学报:自然科学版,2006,29(6):145-147.
- [3] 周丽萍,李慧民,樊胜军.基于主成分分析的住宅特征价格模型的实证应用[J].西安电子科技大学学报:社会科学版,2008(2):68-73.
- [4] 宋永发,王颖,王建东,等.基于主成分分析法的特征价格模型[J].价值工程,2008(4):134-138.
- [5] 梁青槐,孔令洋,邓文斌.城市轨道交通对沿线住宅价值影响定量计算实例研究[J].土木工程学报,2007(4):98-102.
- [6] 史玉芳,李慧民.西安市城市轨道交通对沿线住宅价格的影响研究[J].西安建筑科技大学学报:自然科学版,2010(2):231-235.
- [7] 贾生华,温海珍.房地产特征价格模型的理论发展及其应用[J].外国经济与管理2004(5):42-44.
- [8] ROBERT C. Rail transit and joint development: land market impacts in Washington, DC and Atlanta[J]. Journal of the American Planning Association,1994,60(1):83-94.
- [9] 云虹.主要成分分析法在承诺评价模型中的应用[J].管理学报,2006,3(5):538-542.
- [10] 刘先通,袁长迎,段宝福,等.SPSS10.0 统计分析软件与应用[M].北京:国防工业出版社,2002.

# Improvement of Commercial Housing Price Model Based on Principal Component Analysis

HUANG Gu-bo, LI Yu-zhen

(School of Economics, Wuhan University of Technology, Wuhan, Hubei, 430072)

**Abstract** Hedonic price model, a tool to analyze and assess commercial housing price, is widely used in the field of real estate. Based on the fact that there are more multiple choice linear and more independent variables in practical application, this paper proposes that principal component analysis should be adopted to improve characteristic model of commercial housing and exclude the error generated by the linear correlation between the variables of the hedonic price model. Meanwhile, this paper chooses commercial residential real estate along Wuhan rail as the research object, constructs the commercial housing hedonic price model based on principal component analysis and compares the predicting results of commercial housing price and actual sales prices. The result shows that the majority of the sample forecast errors is under 5%, which verified the advantages of the improved model.

**Key words** principal component analysis; hedonic price model; model improvement; commercial residential buildings; analysis

(责任编辑:金会平)