

The Improvement of Hedonic Price Model Based on Factor Analysis

Xiaoyang Dong

School of Economics and Management, Hebei University of Technology, Tianjin
Email: dongxiaoyang99@sina.cn

Received: Apr. 19th, 2013; revised: May 10th, 2013; accepted: May 20th, 2013

Copyright © 2013 Xiaoyang Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Although hedonic price model is extensively used in real estate prices and the value assessment of non-market product, duplicate information along with numerous indexes which it concerned with not only makes its fitting accuracy declined but also prevents people giving a further interpretation to the influencing factors of the object's value. This paper makes factor analysis applied to Hedonic Price Model, replaces repeating index variable by irrelevant potential factors and explains the practical significance of potential factors by factor rotation so that the features of full information and simple variable are ensured. Hedonic price model can be improved with factor score applied to multiple regression model, which contributes to a higher prediction accuracy and a better interpretation of the influencing factors of value.

Keywords: Hedonic Price Model; Factor Analysis; Multiple Regression Model; Model Improvement

基于因子分析的特征价格模型改进

董晓阳

河北工业大学经济管理学院, 天津
Email: dongxiaoyang99@sina.cn

收稿日期: 2013年4月19日; 修回日期: 2013年5月10日; 录用日期: 2013年5月20日

摘要: 特征价格模型在房地产价格模型和对非市场产品的价值评估中得到广泛应用, 但是一般的特征价格模型不能有效地确定产品价值的影响因素, 涉及指标众多, 信息重复, 使拟合的准确性降低。本文将因子分析应用于特征价格模型, 用几个不相关的潜在因子替代了重复的指标变量, 并且用因子旋转解释潜在因子的现实意义, 从而保证了信息全面、变量简单。最终用潜在因子按照特征价格模型理论进行多元回归, 消除了多重共线性, 提高了回归准确性。

关键词: 特征价格模型; 因子分析; 多元回归模型; 模型改进

1. 引言

特征价格模型 (*hedonic price model*) 被广泛用于房地产的估价和对资源、环境等非市场产品价值的评估, 以得到异质产品差异特征与产品价格的函数关系^[1]。模型以产品为研究对象, 认为产品之所以能引起人们

享受程度的差异是因为产品本身所具有的特征, 尤其像房产、汽车、劳动力这样的异质产品是由一系列能影响效用的特征集合而成。不同的特征集合对应着不同的效用, 效用决定了商品的价格, 也可以理解为产品的每个特征都对应着一个隐含价格(市场上不能直

接观察到), 一系列的特征价格从而组成了产品的价格。尽管模型得到了广泛应用, 但模型也存在不足, 而国内关于模型改进的文章为数不多。本文针对其指标变量信息重复、价格特征难确定的问题, 将因子分析模型应用于特征价格模型, 以达到简化函数指标变量, 寻找价格特征的目的。

2. 特征价格模型

2.1. 特征价格模型的发展

“*Hedonic*”源于希腊语 *hedonikos*, 表示享乐, 这在经济学上指商品或服务的效用。*Hass*(1926) 首先进行特征价格模型的研究, 并提出了“*Hedonic*”这个概念。*Waugh*(1928) 分析了蔬菜质量对蔬菜的影响, *Andrew Court*(1939) 则利用特征价格模型研究了汽车各种特征的隐含价格^[2]。但是, 计算运算和存贮能力的制约, 数据收集技术的欠缺限制了特征价格模型在此阶段的发展。之后, *Lancaster*(1966) 进一步利用消费者理论阐述了特征价格模型, 他的模型假设商品的价格与特征成线性关系, 更适用于一般消费品。*Rosen*(1974) 就产品特征提出了市场均衡模型, 分析了消费者和生产者的选择、市场均衡的形成、特征价格回归的经验含义及指数结构, 为特征价格模型的建立和应用奠定了基础^[3]。

2.2. 特征价格模型的理论基础

该模型建立在消费者理论和供需均衡模型之上。一般情况, 对于消费者, 特征商品的效用函数为严格凹函数, 即随着某一特征增加, 该特征的边际效用越低, 效用决定价格, 所以边际需求价格也越低, 消费者愿意支付的价格和特征属性值之间的关系用消费者标价曲线表示; 对于生产者, 在最优生产组合和一定利润条件下, 其愿意提供商品的价格为凸函数, 即价格随着某一特征特征值的增加而增加且边际产量也增加, 此关系用供应曲线表示。如图 1 所示, 当特征价格曲线与供给曲线和需求曲线相切时, 产品特征的供求达到均衡, 得到均衡价格, 从而得到特征价格随特征属性值而变化的变化情况。

2.3. 国内模型改进现状

特征价格模型目前存在的问题主要是模型形式

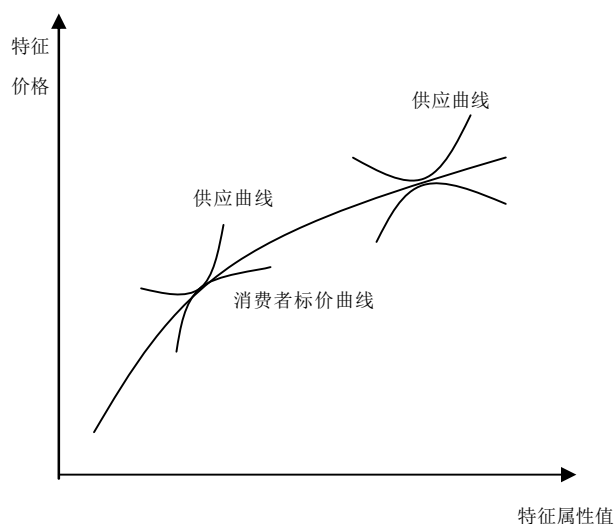


Figure 1. A price curve of a feature attribute
图 1. 某一特征属性的特征价格曲线

的选择、参数估计方法的改进和变量的选择与处理。国内对于特征价格模型的改进主要是形式选择、估计方法和建模方法三方面的改进。周丽萍(2008)将神经网络引入特征价格模型, 以实现函数形式的自动选择^[4]; 王一兵(2005)、冷凯君(2005)、郑其敏(2006)尝试性采用半参数法和非参数法解决模型估计的问题^[5]; 陈安明(2006)、宋永发(2008)^[6]、黄古博(2011)^[7]将主成分分析引入特征价格模型重组特征变量, 改善变量的相关性。

2.4. 基于因子分析模型改进的意义与思路

本文将因子分析应用到特征价格模型中, 是对特征价格模型变量的选择与处理的改进, 以期达到将重复、繁杂的指标变量提炼、简化, 消除模型多重共线性的目的。本文对模型的改进主要针对特征价格模型存在的以下问题:

- 1) 影响价格的特征变量众多, 增加了计算的难度, 也给模型参数的估计和模型的预测带来不利影响。
- 2) 各变量的相关程度较高, 很可能导致函数的近似多重共线性或多重共线性, 引起模型建立的误差。比如土地价值评估中用到的建筑面积与房间数量、离市中心距离与人口密度、绿化率和空气质量, 这些自变量相关程度很高, 影响模型建立。
- 3) 特征价格模型不能有效寻找产品价值的主要影响因素, 也不能进一步对产品价值的影响因素进行

有效解释。对于比较成熟的研究领域，特征因素仍有探讨的必要，例如，虽然房地产价值的关键因素经大量研究已经确定为区位特征、结构特征和邻里特征，但是在研究新型住宅项目时，我们仍有必要探究该项目是否受其它因素影响，并总结影响因素的含义。对于比较陌生的评估领域，特征因素更不应由主观确定的，而是应该通过对多种变量的研究得到，而特征价格模型缺乏确定主要影响因素和解释变量的功能。

基于因子分析的特征价格模型改进，涵盖了主成分分析简化变量的优点，在此基础上增加了解释和分析变量的功能，为研究非经济因素等隐性变量对价格的影响提供便利。

特征价格模型基于因子分析改进后的建模过程如图 2。

3. 基于因子分析的特征价格模型改进

3.1. 模型的适用前提

改进后的特征价格模型的使用前提主要有市场条件、研究对象和变量要求三个方面，其中市场条件和研究对象是特征价格模型本身的适用要求，变量要求是为了保证模型改进后的优化效果明显。

1) 市场条件

a) 统一市场

由于不同市场中的消费者的偏好、需求结构不同，不同的市场信息传递障碍，影响因素的重要程度有差异，所以不同的市场应建立不同的特征价格模型，使模型更有针对性。

b) 完全竞争

市场是完全竞争的，从而使市场同质，市场信息

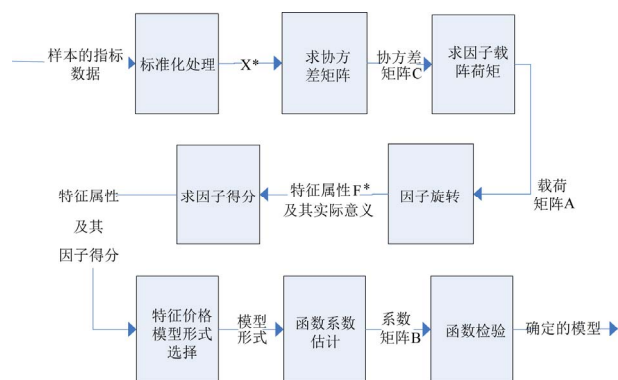


Figure 2. The process of the improved model
图 2. 改进后的建模过程

充分共享，市场中需求者和供给者充足且能自由出入市场。例如，如果在房产的供应量远远超过人们的需求量的情况下，将特征价格模型应用于房产价格估计，那么房产的价格不会对区位特征等自变量敏感，函数将不再适用。特征价格模型应该建立在消费者理性、充分了解产品特征并且产品的供应和需求对产品的价格敏感的基础上。

2) 研究对象

产品是异质的，而且所研究产品的特征因素应是连续变化的，以方便函数性质的应用和解释。

3) 变量要求

应选取较多的样本和指标变量，以便更准确地研究影响产品价值的因素。在探索新产品价值的影响因素，处理变量相关程度较高问题上，该模型更有优势。

3.2. 模型的建立

3.2.1. 建立因子分析模型

选定待研究的产品，并选取有可能影响产品价值的指标变量，这些指标变量是否能影响产品价值及其确切的影响程度未知。在因子分析模型中，每一个指标变量都可以表示成公共因子的线性函数与特殊因子的和，公共因子和特殊因子都是用来反映显在变量的“抽象”变量，它们是不可直接测量但又客观存在的影响因素。模型的矩阵表示是：

$$X = AF + \varepsilon$$

其中：

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pm} \end{pmatrix}$$

$$F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

模型中 X 是指标变量， A 是因子载荷矩阵， a_{ij} 是因子“载荷”，表示第 i 个变量在第 j 个因子上的负荷。 F 是不可直接测量但又客观存在的公共因子，公共因子经过变换可以得到产品的特征变量。 ε 是影响每个变量的特殊因子。

该模型的建立需满足以下条件：

1) $m \leq p$ ，即公共因子的个数不超过指标变量的个数；

2) $\text{cov}(F, \varepsilon) = 0$ ，公共因子和特殊因子不相关；

3) $\sum_F I_m$ ，各公共因子不相关且方差为 1；

4) $\sum_{\varepsilon} = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix}$ ，即各个特殊因子不相关，方差可以不等。

3.2.2. 确定特征因素及其实际意义

1) 原始数据的标准化处理

选取 n 个样本产品，代入对应的 p 个指标变量值，从而得到原始数据的矩阵为

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p-1} & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p-1} & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np-1} & x_{np} \end{bmatrix}$$

标准化处理公式：

$$X_j^* = \frac{X_j - E(X_j)}{\sqrt{D(X_j)}} \quad i = 1, \dots, p$$

其中

$$E(X_j) = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad D(X_j) = \sum_{i=1}^n (x_{ij} - \bar{x}_{ij})^2,$$

则 $(X_1^*, X_2^*, \dots, X_p^*)$ 是标准化之后的数据。

2) 求指标值的协方差矩阵

协方差矩阵是任意两指标变量协方差构成的矩阵，表示指标之间的相关关系，用 c_{ij} 表示 X_i^* 和 X_j^* 之间的协方差，则

$$c_{ij} = \text{cov}(X_i^* - X_j^*) = \frac{\sum_{i=1}^n (X_i^* - \bar{X}_i^*)(X_j^* - \bar{X}_j^*)}{n-1}$$

由此得到协方差矩阵为

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1p} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & c_{p3} & \cdots & c_{pp} \end{bmatrix}$$

3) 因子载荷矩阵的求解^[8]

对因子载荷矩阵的估计方法很多，这里主要介绍

主成分法估计。首先解特征方程 $|\lambda I - C| = 0$ ，用雅可比法求出特征值 $\lambda_j (j = 1, 2, 3, \dots, p)$ ，并使其按大小顺序排列，即 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ，然后分别求出其对应的正交化特征向量 t_1, t_2, \dots, t_p 。由于影响产品价值的特征属性个数未知，可根据需要自定产品特征属性的个数，在此定为 m 个。为了简化，略去后 $p - m$ 项的贡献，只取前 m 项特征向量。在 SPSS 中，可以从特征值大于 1 和固定因子数两种抽取方式中选择一种方式，完成特征向量的抽取。

可以求得因子载荷矩阵

$$A = (\sqrt{\lambda_1} t_1, \sqrt{\lambda_2} t_2, \dots, \sqrt{\lambda_m} t_m)$$

由因子载荷矩阵可得公共因子的“共同度”，用“共同度”来描述公共因子表达指标变量信息的完整度，即用因子分析法得到的特征属性表达指标变量信息的完整度。“共同度”表示为 h_i^2 ， h_i^2 接近于 1，说明公共因子几乎说明了变量的全部信息。

$$h_i^2 = \sum_{j=1}^m a_{ij}^2 (j = 1, 2, \dots, p)$$

公共因子的相对重要性可以用其对所有变量的方差贡献率之和衡量，表示为 S_j 。

$$S_j = \sum_{i=1}^p a_{ij}^2 (j = 1, 2, \dots, p)$$

对于主成分法估计的因子载荷阵，

$S_j = \lambda_j (j = 1, 2, \dots, m)$ ，表明了每个公共因子对结果的影响程度，影响产品价格的特征属性从公共因子中产生。

4) 因子旋转

为了更好的解释对象价值的影响因素，公共因子应该被赋予实际含义，公共因子的实际含义往往能反映出产品的特征属性。如果根据特征根和特征向量求得的因子载荷矩阵难以看出公共因子的含义，就要将公共因子做因子旋转，使公共因子对某些指标变量关联性大，某些指标变量关联性小，从而确定公共因子的含义。例如，在利用因子分析探究牙膏价格的特征因素时发现，某公共因子和指标变量中的“清新口气分数”以及“洁白牙齿分数”等功能性指标成强正相关，由此可以推断该公共因子的意义为“功能特征”。

对于任一正交矩阵 B ，满足 $B^T B = I$ ，所以：

$$X = AF + \varepsilon = A(B^T B)F + \varepsilon$$

可以得到新的载荷矩阵 $A_* = AB^T A_*$ ，其公共因子为 $F_* = BF$ 。在 SPSS 软件中，可以自己选择因子旋转的方法，一般选择最大方差法进行因子旋转，从而得到有现实含义的公共因子。

5) 因子得分的计算

由于因子分析模型在实际应用时要求公共因子数小于指标变量数，所以只能对因子得分进行估计。SPSS 软件中有很多估计方法可供选择，常用的是回归估计。设 R 是 X 的相关系数矩阵，在汤姆森回归法中，因子得分可以由下式得到：

$$\hat{F}_* = A_*^T R^{-1} X$$

至此，由因子分析模型得到了 m 个意义明确、不相关的公共因子，并且得到其因子得分。在特殊因子可忽略的情况下，公共因子可以作为影响产品价值的特征因素，因子得分为特征属性的取值，从而根据已知的研究对象的价值和因子得分，建立特征价格模型。

3.2.3. 建立特征价格模型

1) 模型函数形式的选择

用 P 表示产品价格， β_0 表示特征变量以外影响价格的控制变量， β_i 表示特征价格， ε 表示误差项，则特征价格函数模型的三种形式表现如下^[9]：

a) 线性函数

$$P = \beta_0 + \sum \beta_i C_i + \varepsilon$$

b) 半对数函数

$$\ln P = \beta_0 + \sum \beta_i C_i + \varepsilon$$

c) 对数函数

$$\ln P = \beta_0 + \sum \beta_i \ln C_i + \varepsilon$$

三种形式中，线性形式是早期学者研究特征价格模型时应用的函数形式，G. C. Hass(1922) 曾成功应用该形式评价农地的价格，但是无法反映特征商品效用函数边际递减规律。半对数形式应用比较广泛，回归系数对应着特征价格与产品总价格和比值。对数函数服从从边际效用递减的规律，回归系数对应特征变量的价格弹性。从三种形式中选择其中的一种形式，将样本各价格特征对应的因子得分作为 C_i 与样本对应的价格一同带入选定的函数形式，以求得函数系数，选

择形式注意自变量和因变量的数量级不应相差太大。

2) 函数系数的估计

确定函数形式后，需要带入数据，以一定准则对系数进行估计。最小二乘法是常用的估计回归参数的方法，最小二乘法以残差平方和最小为原则，现以线性函数为例来说明函数参数的估计。设 $e_i = P - \hat{P}$ ， \hat{P} 是价格用函数的估计值。残差平方和 SSE 的表达式为

$$SSE = \sum_{i=1}^n e_i^2$$

$$e_i = P - C\hat{B}$$

其中 $B = [\beta_0 \beta_1 \beta_2 \dots \beta_m]^T$ ，根据 $(AB)' = B'A'$ 这一性质， $(P' C \hat{B}) = \hat{B}' C' P$ ，欲求 SSE 最小，则

$$\frac{\partial SSE}{\partial \hat{B}} = -2C'P + 2C' C \hat{B} = 0$$

得到 $\hat{B} = (C' C)^{-1} C' P$ ，从而可以得到特征价格模型的参数。回归参数是产品样本价格值 P 的线性组合，具有无偏估计性、最小方差性。

在实际应用中，首先用普通最小二乘法拟合，如果出现问题，从模型的选用、自变量的选取和估计方法上改进，比如当模型存在异方差性时采用加权最小二乘法、当模型存在自相关性时可以采用广义最小二乘法。

3.3. 函数检验

1) 经济检验，检验参数估计值的符号和大小是否符合经济理论，自变量之间、自变量与因变量的关系与预期的是否一致。比如，土地价值和环境质量成正比，如果环境质量前的特征价格为负，则需查找原因及修正。

2) 拟合优度检验，检验样本回归直线与观测值之间的拟合程度。拟合优度由样本决定系数(或称为复相关系数)得到，越接近，拟合程度越高。

$$R^2 = \frac{SSR}{SST}$$

其中 SSR 为回归平方和，

$$SSR = \sum_{i=1}^n (\hat{P}_i - \bar{P})^2$$

SST 为总离差平方和，

$$SST = \sum_{i=1}^n (P_i - \bar{P})^2$$

P_i 是各实际样本价格的值, \hat{P}_i 是由模型得到的价格值, \bar{P} 是实际价格的平均值。当自变量多时, 复相关系数也会变大, 为了确保复相关系数能体现拟合优度, 引入用自由度修正了的复相关系数 R_α^2 , n 是样本数, m 是函数中自变量的个数^[10]。又因 $SST = SSR + SSE$,

$$R_\alpha^2 = 1 - \frac{SSE/(n-m-1)}{SST/(n-1)}$$

3) 显著性检验, 包括回归方程的显著性检验和回归系数的显著性。回归方程的显著性检验的检验内容为产品价值和所得的特征属性之间的线性关系, 是总体显著性检验, 检验全部特征变量对价格的共同影响是否显著, 用 F 检验法, 在置信水平 α 条件下, $F > F_\alpha(m, n-m-1)$ 时方程因变量和自变量之间的线性关系显著。回归系数的显著性的检验内容是产品价值和每个特征属性之间的线性关系, 用 t 检验法, 在置信水平 α 条件下, $|t_{\hat{\beta}_i}| > t_{\alpha/2}(n-m-1)$ 时表明 $\hat{\beta}_i$ 和产品价值有显著的线性关系, 其中 n 为样本数量, m 为自变量的数量。

4) 其他检验, 包括序列相关检验、异方差检验和特征向量的多重共线性检验。特征价格模型最后转化为多元回归问题, 这些检验都是针对多元回归建立的假设进行的检验, 序列相关性(自相关性)使模型参数估计不具有最优性, 使变量的显著性检验没有意义, 使模型预测失效, 可以用 $D-W$ 法检验; 异方差性使回归系数估计值不具有最小性, 可以用 $White$ 等检验方法; 多重共线性会造成无法得到参数估计量或影响普通最小二乘法的估计效果, 可以用相关系数检验法检验^[11]。

4. 模型的问题及推广

在特征价格模型基础上作基于因子分析法的改

进, 其本质是用多元回归技术和因子分析法完善特征价格模型。首先利用因子分析得到影响各指标的公共因子, 然后通过因子旋转使公共因子具有明确意义, 从而得到影响产品价格的特征因素, 最后对各样本公共因子的因子得分和样本价格进行拟合, 得到特征价格模型。该模型整合了重复的信息, 简化并提炼了变量, 便于拟合, 并可以对价格的影响因素进一步解释。改进后的模型既可以同特征价格模型一样, 应用于交通对土地价格的影响、淡水资源的价值等非市场产品的评价, 应用于研究非市场因素对产品价格的影响, 也适用于多指标产品预测等需要多元回归的案例。但是, 针对特征价格模型而言, 初始指标的选择和信息的收集仍然是值得探讨的问题。此外, 基于因子分析法改进之后, 该模型函数参数所具有的现实含义和因子得分的实际意义, 也是值得探讨的问题。

参考文献 (References)

- [1] R. Cervero, C. D. Kang. Bus rapid transit impacts on land uses and land values in Seoul, Korea. UC Berkeley Center for Future Urban Transport, 2009.
- [2] 郑捷奋. 城市轨道交通与周边房地产价值关系研究[D]. 北京: 清华大学, 2004.
- [3] 陈亮. 住宅项目特征定价模型研究[D]. 成都: 西南交通大学, 2007.
- [4] 周丽萍, 李惠民, 肖静. 神经网络在特征价格模型中的应用[J]. 西北工业大学学报, 2008, (4): 189-192.
- [5] 罗晓娟. 特征价格模型的发展应用研究[J]. 技术经济与管理研究, 2012, (6): 11-15.
- [6] 陈永发. 基于主成分分析法的特征价格模型[J]. 价值工程, 2008, (4): 134-137.
- [7] 黄古博, 李雨珍. 基于主成分分析法的商品价格[J]. 华中农业大学学报(社会科学版), 2011, (4): 93-97.
- [8] 曾五一. 统计学(中国版)[M]. 北京: 北京大学出版社, 2006.
- [9] 王旭育. 城市住宅特征价格模型的理论分析[J]. 上海管理科学, 2006, (4): 68-69.
- [10] 朱建平. 应用多元统计分析[M]. 北京: 科学出版社, 2006.
- [11] 贾生华, 温海珍. 房地产特征价格模型的理论发展及其多元统计分析[J]. 外国经济与管理, 2004, (5): 42-44.