

李欣点 朱恩伟 刘洪玉 吴璟

# 城市同质化住房价格空间分布研究

## ——基于半参数特征价格模型分析

**内容提要：**本文基于特征价格理论，将住房的区位特征设定为地理坐标的函数，以非参数形式引入模型，同时将其他住房特征以参数形式引入模型，建立了半参数特征价格模型，并使用北京市 2011-2015 年间 119,745 条二手房成交数据进行了实证分析。研究表明：相较于传统方法，半参数特征价格模型在区位信息的捕捉能力、对样本数据的拟合效果和参数结果的可解释性上都存在明显的优势。进一步地，本文基于模型结果计算并绘制了同质化住房价格空间分布，展现了城市内部住房市场的微观结构，进而识别了住房价格的主要中心位于天安门西北方向的北海公园周边，次要中心位于中关村西侧、北四环西路附近。

**关键词：**半参数估计 特征价格模型 住房价格 空间分布

住房价格是区域经济发展情况的重要指标，同时也受到地方政府调控政策的影响。在实践中，一个城市内部不同区域间社会经济发展程度通常都存在显著差异，例如商业中心发展水平远高于城市郊区、旧城区改造和新城区建设会大幅促进局部经济发展等。同时，政府对不同区域的政策导向也可能有所不同，例如北京市不同区域分为首都功能核心区、城市功能拓展区、城市发展新区和生态涵养发展区，不同类型区域承担不同的城市功能。这就造成了城市内部不同区域间住房市场发展水平的差异。本文基于北京市 2011-2015 年间二手房交易数据，建立半参数特征价格模型估计了城市内部住房价格的空间分布，为观察住房市场微观结构提供了重要工具。其中，特征价格模型能够控制不同区域间住房的异质性，使得住房价格的空间分布同质可比。

### 一、相关文献评述

国内外学者对城市住房价格的空间分布规律进行了较为广泛的研究，普遍认为影响住房价格的最重要区位因素是到商业中心距离，其次是到最近交通枢纽的距离，此外也还可能包括医院、学校、公园等公共设施因素。Alonso(1964)最早提出“竞租理论”，基于单中心城市的假设对城市居民居住地选择行为进行分析，指出住房租金与离商业中心距离负相关。但随着现代城市发展，尤其是城市次中心的崛起和轨道交通的建设，使得单中心城市的假设不再符合实际情况。例如，Dubin 和 Sung(1987)使用美国城市的数据进行实证研究，发现城市次中心对住房价格的空间分布的影响不可忽视。Zheng 和 Kahn

(2008)使用 2004 年和 2005 年数据研究发现，北京市住房价格空间分布基本符合单中心结构，但地铁、公交、学校等公共设施对局部结构也存在非常重要的影响。于璐等(2008)使用相同的数据研究发现，住房价格随到商业中心的距离变化的规律在不同方向上存在异质性。这些研究都是先对住房价格分布结构进行假设，然后搜集数据进行验证，依赖大量的先验信息(次中心和公共设施的位置)，展示结果也不够直观。

另一类研究则直接对住房价格空间分布进行拟合，降低了对先验信息的依赖同时展示结果也更为直观。这类研究通常都使用了非参数或半参数方法来获得灵活的拟合效果。例如，Colwell(1998)使用非参数方法拟合了芝加哥商业中心地区的土地价格分布，并指出该分布的结构非常复杂，很难用简单的函数形式去表征。Clapp 等(2002)基于马塞诸塞州的房价数据，通过半参数方法估计特征价格模型，从而获得了同质化的住房价格空间曲面。Clapp(2004)则进一步将时间维度也放入模型的非参数部分。Geniaux 和 Napoleone(2008)和 Hill 和 Scholz(2013)分别利用法国和澳大利亚的数据建立了类似的半参数特征价格模型，并指出随着城市发展，住房价值的空间分布越来越复杂，单中心城市设定的局限性越来越明显。国内研究大部分都直接通过空间插值法直接拟合房价曲面，缺乏同质化的处理。例如，马敏蕾等(2008)使用北京市住房交易数据拟合了住房价格空间格局，李广娣和沈昊婧(2014)则拟合了沈阳市的住房价格空间分布。

本文将 Clapp 等(2002)提出的半参数特征价格

模型应用到北京市住房价格的研究中,并直观展示了城市内住房价格的空间分布。尤其值得注意的是,该模型能够控制不同位置住房质量的差异,因而获得的价格空间分布是同质可比的,弥补了国内研究的不足。

## 二、同质化住房价格空间分布理论分析

住房是一种典型的异质性商品。除了区位以外,不同的住房在面积、朝向、装修、楼层、建筑结构和形式、建成时间和社区配套等方面都可能存在显著差异,这些住房特征的差异会导致住房质量的不同,进而对住房价格产生显著的影响。因此,不同位置住房的价格并不能直接比较。例如,城市中心附近的住房价格通常都高于城市郊区的住房,但靠近城市中心的住房通常建成时间较早,因而折旧较多,所以实际上城市中心和郊区的住房价格差距比实际观察到的更大。在时间维度住房价格指数的编制中,通常都需要进行同质化处理,也就是消除不同时期交易的住房特征和质量的差异,使指数能够反映住房市场本身的变化。因此,在研究住房价格空间分布时,也需要排除不同位置住房特征和质量的差异,进行类似的同质化处理。

本文以地理坐标为自变量的非线性函数直接表征住房的区位信息,构建半参数特征价格模型。具体地,模型设定为

$$\ln(P_{it}) = g(z_{i1}, z_{i2}) + \sum_{k=1}^M \beta_k x_{ik} + \sum_{t=1}^T \delta_t D_t + \varepsilon_{it} \quad (2)$$

其中,  $P_{it}$  表示第  $i$  套住房在时间  $t$  观察到的交易价格,  $x_{ik}$  表示第  $i$  套住房第  $k$  个特征的取值;  $D_t$  是月度哑元,当住房交易发生在  $t$  期时取值为 1,其他情况取值为 0;  $\beta_j$  和  $\delta_t$  均为待估计系数;  $\varepsilon_i$  为随机扰动项。  $z_{i1}$  和  $z_{i2}$  分别是第  $i$  套住房的经纬度坐标,  $g(\cdot)$  是待估计的非线性函数。该模型可以使用 Robinson-Stock 两阶段法进行估计。许多研究指出,这种半参数形式能够非常灵活地拟合复杂的住房价格空间分布,并且避免潜在的遗漏变量和函数设定误差等问题。

可以发现,通过以参数形式引入住房特征  $x_{ij}$ ,该模型可以控制不同位置处住房特征差异对住房价格的影响,因此非参数部分的估计结果  $\hat{g}(z_1, z_2)$  可以代表坐标  $(z_1, z_2)$  处的同质化住房价格。具体地,假设存在完全相同的两套住房分别位于  $(z_{i11}, z_{i12})$  和  $(z_{i21}, z_{i22})$  处,则其估价结果分别为:

$$\begin{aligned} \ln(\hat{P}_{i_1}) &= \hat{g}(z_{i_11}, z_{i_12}) + \sum_{j=1}^M \hat{\beta}_j x_{i_1j} \\ \ln(\hat{P}_{i_2}) &= \hat{g}(z_{i_21}, z_{i_22}) + \sum_{j=1}^M \hat{\beta}_j x_{i_2j} \end{aligned} \quad (3)$$

由于  $x_{i_1j} = x_{i_2j}, j=1, 2, \dots, M$ , 可得:

$$\frac{\hat{P}_{i_1}}{\hat{P}_{i_2}} = \exp(\hat{g}(z_{i_11}, z_{i_12}) - \hat{g}(z_{i_21}, z_{i_22})) \quad (4)$$

因此,任意两组坐标间  $\hat{g}(\cdot)$  函数值之差的自然指数就是两地同质化住房价格的比值。

## 三、城市同质化住房价格空间分布实证数据和结果

### (一) 数据描述

本文的数据包含北京市 2011-2015 年间共 119,745 条二手住房交易记录,每一条记录不仅包含住房的成交价格、成交时间和地理位置,还详细记录了房源层面和小区层面的各项特征,可用于构建特征价格模型。

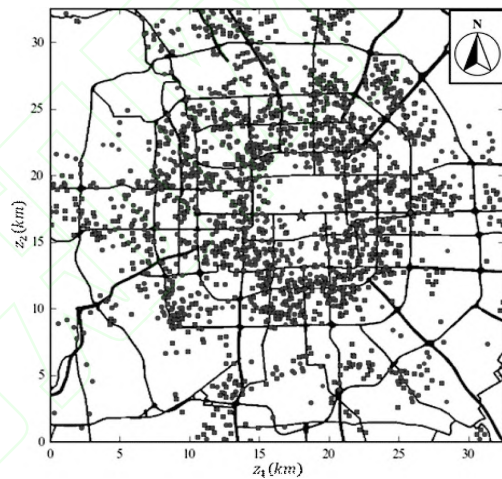


图1 样本空间分布散点图

图1展示了样本空间分布的散点图。其中,住房地理坐标的原始数据格式是经度和纬度,本文以东经 116.187 度、北纬 39.756 度处为坐标原点,以千米为单位重新定义纵横坐标,并记为  $z_1$  和  $z_2$ 。

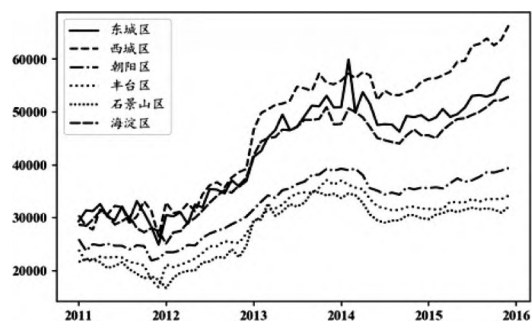


图2 北京“城六区”二手住房交易均价

图2展示了使用样本数据计算的北京市“城六区”各月二手住房交易均价。可以发现,各区住房价格整体上表现出较为相似的变化趋势,但在价格水平和变化幅度上仍然有明显的区别。从 2011 年初至 2015 年末,东城、西城和海淀区的住房价格上涨幅度明显高于朝阳、丰台和石景山区,均价差异从期初

的约 5600 元 / 平方米增长到期末的约 23400 元 / 平方米。2015 年末,西城区二手住房交易均价最高,达 66400 元 / 平方米;石景山区最低,仅为 32100 元 / 平方米。各区域住房子市场存在明显的异质性,表明了城市内部住房市场微观结构分析的必要性。

### (二)模型结果比较与分析

表 1 展示了在特征价格模型中,使用不同方法引入区位信息的模型估计结果。其中,(1)列没有引入区位信息,作为参照组;(2)列引入了到市中心和到最近地铁站的距离;(3)列引入了行政区哑元;(4)列同时使用了这两种区位变量;(5)列将住房的区位信息设定为地理坐标的非线性函数,建立了半参数特征价格模型。

表 1 住房区位信息的半参与参数建模方法比较

|                 | (1)                  | (2)                  | (3)                  | (4)                  | (5)                  |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| log(住房价格)       | OLS                  | OLS                  | OLS                  | OLS                  | Robinson-Stock       |
| log(住房面积)       | -0.175***<br>(0.005) | -0.096***<br>(0.004) | -0.110***<br>(0.004) | -0.085***<br>(0.004) | -0.095***<br>(0.004) |
| 楼层比             | 0.194***<br>(0.028)  | 0.152***<br>(0.024)  | 0.167***<br>(0.023)  | 0.163***<br>(0.022)  | 0.179***<br>(0.020)  |
| 楼层比二次项          | -0.081***<br>(0.023) | -0.089***<br>(0.020) | -0.102***<br>(0.018) | -0.108***<br>(0.018) | -0.124***<br>(0.016) |
| 特殊楼层:地下室        | -0.341***<br>(0.030) | -0.410***<br>(0.027) | -0.458***<br>(0.026) | -0.461***<br>(0.026) | -0.487***<br>(0.025) |
| 特殊楼层:顶层         | -0.080***<br>(0.005) | -0.052***<br>(0.004) | -0.053***<br>(0.004) | -0.044***<br>(0.004) | -0.039***<br>(0.003) |
| 楼层类型:高层         | -0.045***<br>(0.005) | -0.027***<br>(0.004) | -0.023***<br>(0.004) | -0.020***<br>(0.004) | -0.016***<br>(0.003) |
| 楼层类型:超高层        | -0.088***<br>(0.007) | -0.052***<br>(0.006) | -0.046***<br>(0.006) | -0.039***<br>(0.005) | -0.033***<br>(0.005) |
| 二居室             | 0.038***<br>(0.003)  | 0.017***<br>(0.003)  | 0.028***<br>(0.002)  | 0.016***<br>(0.002)  | 0.021***<br>(0.002)  |
| 三居室             | 0.104***<br>(0.005)  | 0.044***<br>(0.004)  | 0.052***<br>(0.004)  | 0.026***<br>(0.004)  | 0.023***<br>(0.003)  |
| 四居室及以上          | 0.176***<br>(0.010)  | 0.073***<br>(0.009)  | 0.076***<br>(0.008)  | 0.036***<br>(0.008)  | 0.026***<br>(0.007)  |
| 结构类型:塔楼或板塔结合    | 0.096***<br>(0.002)  | 0.016***<br>(0.002)  | 0.025***<br>(0.002)  | 0.004***<br>(0.001)  | 0.003***<br>(0.001)  |
| 建成年代:2001-2005年 | -0.036***<br>(0.002) | -0.037***<br>(0.002) | -0.014***<br>(0.002) | -0.004***<br>(0.002) | 0.042***<br>(0.002)  |
| 建成年代:2006-2010年 | -0.023***<br>(0.003) | 0.021***<br>(0.002)  | 0.004***<br>(0.002)  | 0.046***<br>(0.002)  | 0.092***<br>(0.002)  |
| 建成年代:2011年以后    | 0.011***<br>(0.005)  | 0.069***<br>(0.005)  | 0.114***<br>(0.004)  | 0.134***<br>(0.004)  | 0.162***<br>(0.004)  |
| log(小区面积)       | -0.025***<br>(0.001) | -0.003***<br>(0.001) | -0.002***<br>(0.001) | 0.007***<br>(0.001)  | 0.010***<br>(0.001)  |
| log(到市中心距离)     |                      | -0.365***<br>(0.002) |                      | -0.262***<br>(0.002) |                      |
| log(到最近地铁站距离)   |                      | -0.048***<br>(0.001) |                      | -0.052***<br>(0.001) |                      |
| 朝向:南北           | -0.058***<br>(0.005) | 0.024***<br>(0.004)  | 0.033***<br>(0.004)  | 0.059***<br>(0.004)  | 0.057***<br>(0.003)  |
| 其他控制变量          | 是                    | 是                    | 是                    | 是                    | 是                    |
| 时间固定效应          | 是                    | 是                    | 是                    | 是                    | 是                    |
| 行政区固定效应         | 否                    | 否                    | 是                    | 是                    | 否                    |
| 样本量             | 119,745              | 119,745              | 119,745              | 119,745              | 119,745              |
| R <sup>2</sup>  | 0.270                | 0.471                | 0.597                | 0.654                | 0.735                |

注:(1)表中括号内为稳健标准误,\*\*\* \*\* \*分别表示在1%、5%和10%水平下显著;(2)楼层类型以“多层”为对照组,居室数量以“一居室”为对照组,结构类型以“板楼”为对照组,建成年代以“2000年以前”为对照组,朝向以“北”为对照组;(3)为控制篇幅,表中省略容积率、绿化率和除“南北”外其他朝向哑元的系数估计结果。

从模型拟合度来看,(1)至(5)列的R<sup>2</sup>逐步提高,显示了模型对区位信息捕捉能力不断增强。(1)列的R<sup>2</sup>为27.0%,说明区位信息以外的变量最多仅能解释房价的四分之一左右。(2)、(3)和(4)列通过传统方式逐步引入区位信息后,R<sup>2</sup>分别提升了

20.1%、32.7%和38.4%。这一结果说明了传统方法的有效性,少量区位变量的引入就能够显著提升对被解释变量的解释能力。(5)列的R<sup>2</sup>为73.5%,相比(1)列提升了46.5%,相比传统方法中拟合效果最好的(4)列也提升了8.1%。这说明使用地理坐标和非参数设定能够更好地拟合住房价格的空间分布,相比传统方法捕捉到了更多有效的区位信息——其中有些信息可能是传统方法难以观测的。

表1中各列模型大部分的参数估计结果与预期是一致的,但在某些变量上表现出一定差异。以取值为“南北”的朝向哑元为例,其对照组取值为“北”,显然南北通透的住房相比北向住房应当有明显的溢价,因而系数估计结果预期为正。在实际估计结果中,(1)列估计结果显著为负,而(2)至(5)列估计结果均显著为正。再以建成年代取值为“2001-2005年”的哑元变量为例,其对照组取值为“2000年以前”,显然建成年代更新的住房应当价格更高,因而系数预期为正。但是,(1)至(4)列估计结果均显著为负,即认为2001-2005年间建成的住房整体上相比2000年前建成的住房存在明显的折价,这一结果是难以解释的。只有(5)列对该变量的系数给出了正向且显著的估计结果。

以上结果显示,相比传统方法,使用地理坐标和非参数设定能够给出更为合理的变量估计结果,说明该方法能够更为有效地捕捉区位信息。

### (三)同质化住房价格空间分布

由式(4)可知,半参数特征价格模型的非参数部分 $\hat{g}(\cdot)$ 就是同质化住房价格的空间分布指标,能够反映完全相同的住房在不同空间位置上的价格差异。记函数 $\hat{g}(z_1, z_2)$ 的最小值 $\hat{g}_{\min}$ ,则任意点 $(z_1, z_2)$ 处的同质化价格指数为:

$$PI(z_1, z_2) = \exp(\hat{g}(z_1, z_2) - \hat{g}_{\min}) \times 100 \quad (5)$$

图3展示了北京市同质化住房价格指数的等高线图,图4是该曲面的三维展示图。整体上来看,北京市同质化住房价格的分布呈现一个主要中心加一个次要中心的格局。其中,主要中心并不完全与天安门重合,而是位于其西北方向的北海公园周边,距离

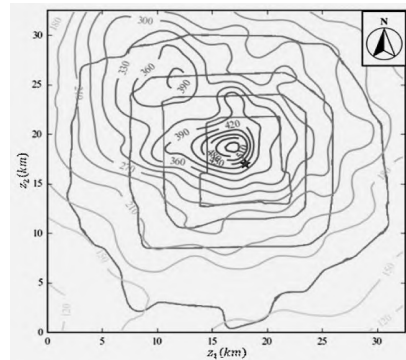


图3 北京市同质化住房价格空间分布等高线图

天安门约2千米;次要中心位于中关村西侧、北四环西路附近。除此以外,住房价格的主要中心向北二环和北三环之间、西三环玉渊潭公园附近有延展的趋势,这两处是未来城市中心的候选。

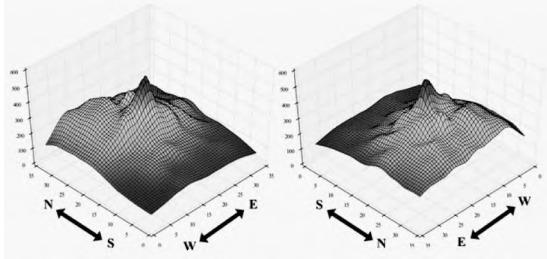


图4 北京市同质化住房价格空间分布三维图

以上结果说明,北京市的房价分布不适用单中心城市结构的假设,同时,基于先验信息选定的城市中心往往距离真实房价分布中心有一定的偏差。此外,通过建立半参数特征价格模型,也可以更新研究者对城市空间结构的认知,有助于构建更为合理的参数化模型。

#### 四、结论与建议

本文将住房的区位特征设定为地理坐标的非线性函数,以非参数形式引入模型,同时将其他住房特征以参数形式引入模型,建立了半参数特征价格模型。实证结果显示,该模型在区位信息的捕捉能力、对样本数据的拟合效果和参数结果的可解释性上都存在明显的优势。基于模型结果绘制的同质化住房价格空间分布显示,北京市同质化住房价格分布的主要中心位于天安门西北的北海公园附近,次中心位于中关村西侧、北四环西路附近。

基于研究结果,本文提出以下几点建议:

1.将同质化处理引入住房价格空间结构的研究中。同质化处理能够剔除住房质量的差异对价格的

影响,使得不同位置的住房价格比较得到更为严谨、有意义的结果。

2.在特征价格模型中,使用非参数或半参数方法捕捉和控制住房价格的区位信息。住房价格空间结构通常都较为复杂,而非参数或半参数方法能够灵活地拟合这种结构,在提升拟合效果的同时提高其他变量系数估计的准确性。

3.根据城市内不同区域住房市场的发展水平,施行差异化政策。根据本文的同质化住房价格空间分布,北京市不同区域住房市场发展不够均衡,东侧和南侧住房市场明显落后于西侧和北侧。因此,可以建议政府加强东侧和南侧基础设施建设,创造更多就业机会,引导职住平衡;同时重点加强西侧和北侧住房市场调控,抑制住房价格过快增长,从而实现各区域住房市场均衡发展。

参考文献:

- [1]Alonso, W. Location and Land Use[J].1964.
- [2]Dubin, R. A., Sung, C. Spatial Variation in the Price of Housing: Rent Gradients in Non-Monocentric Cities[J]. Urban Studies, 1987, 24(3).
- [3]Colwell, P. F. A Primer On Piecewise Parabolic Multiple Regression Analysis Via Estimations of Chicago Cbd Land Prices [J]. The Journal of Real Estate Finance and Economics,1998,17(1).
- [4]Clapp, J. M., Kim, H. J., Gelfand, A. E. Predicting Spatial Patterns of House Prices Using Lpr and Bayesian Smoothing[J]. Real Estate Economics, 2002,30(4).
- [5]Clapp, J. M. A Semiparametric Method for Estimating Local House Price Indices[J]. Real Estate Economics, 2004, 32(1).
- [6]Geniaux, G., Napoleone, C. Semi-Parametric Tools for Spatial Hedonic Models: An Introduction to Mixed Geographically Weighted Regression and Geoadditive Models[J]. Hedonic Methods in Housing Markets,2008.
- [7]Zheng, S., Kahn, M. E. Land and Residential Property Markets in a Booming Economy: New Evidence From Beijing[J]. Journal of Urban Economics,2008,63(2).
- [8]于璐,郑思齐,刘洪玉.住房价格梯度的空间异质性及影响因素[J].经济地理, 2008, 28(3).
- [9]马敬雷,吕斌,冯长春.基于Gis基础上的北京住房价格空间格局研究[J].中国国土资源经济, 2008,21(12).
- [10]Hill, R. J., Scholz, M. Incorporating Geospatial Data Into House Price Indexes: A Hedonic Imputation Approach with Splines [J]. Ottawa-Group, Statistics Denmark, Copenhagen,2013.
- [11]李广娜,沈昊婧.城市住房价格的空间分布格局研究——以沈阳市为例[J].现代城市研究,2014(2).

(作者单位:清华大学建设管理系)

## Distribution of Constant-quality Housing Price

—Based on a Semi-parametric Hedonic Pricing Model

**Abstract:**Based on hedonic pricing theory,a semi-parametric hedonic model is built in this paper, in which the location characteristics are specified as a nonlinear function of the geographic coordinates in a nonparametric form, while other characteristics are specified in a parametric form. The empirical models are estimated by 119,745 transaction data from 2011 to 2015 from the second housing market in Beijing, and the results show that the semi-parametric hedonic model has significant advantages in location information capture ability, fitting goodness and parameters interpretability, when compared with the traditional models. According to the model results, the spatial distribution of the constant-quality housing price index is computed and visualized, which demonstrated the within-city microstructure of housing market. It could be identified that Beihai Park, which is located in the northwest of Tiananmen Square, is the main center of housing price. The secondary center is located in the proximity of Zhongguancun.

**Keywords:**Semi-parametric estimation; Hedonic Pricing Model; Housing prices; Spatial distribution