

住房价格统计中的样本代表性误差及其修正

吴璟 郑思齐 刘洪玉

(清华大学, 北京 100084)

摘要: 对住房价格变化的准确测量在当前背景下具有重要意义, 样本代表性误差则是其中必须解决的重要问题。基于住房和住房市场的特殊性, 分析住房价格统计过程中产生样本代表性误差的原因; 将样本代表性误差划分为系统性误差和随机性误差两个组成部分, 并比较其对我国住房价格统计的影响; 提出评估价格法、固定权重法、特征价格法 (包括截面建模形式和一次性建模形式) 等四种针对样本代表性误差的修正方法, 并通过北京市住房价格统计的数值模拟算例, 比较各种修正方法的有效性和可行性。研究结果表明: 样本代表性误差, 尤其是其中的随机性误差部分, 是当前制约我国住房价格统计结果质量的主要瓶颈之一, 而一次性建模形式的特征价格法则是一种有效的修正该误差的方法。研究成果有助于我国住房价格统计的改进和完善。

关键词: 住房市场; 住房价格统计; 样本代表性误差

中图分类号: F293.3 文献标识码: A

文章编号: 1000-131X (2009) 04-0140-05

Sample error in housing price statistics and its adjustments

Wu Jing Zheng Siqi Liu Hongyu

(Tsinghua University, Beijing 100084, China)

Abstract: Price is a key indicator in housing market, and sample error is a major problem in housing price statistics. This paper analyzes the causes and effects of sample error in housing price statistics in China, and establishes four possible adjustment methods, namely, the appraisal method, the fixed-weight method, the cross-sectional hedonic method and the pooled hedonic method. Then the validity and feasibility of these methods are compared empirically by taking a simulated housing price index in Beijing as an example. The results suggest that the sample error, especially its random error component, significantly affects the accuracy of housing price indicator in China at present, and the pooled hedonic method is the most applicable and efficient method for error adjustment.

Keywords: housing market; housing price statistics; sample error

E-mail: wujing00@mails.tsinghua.edu.cn

引言

住房价格是当前社会各界高度关注的热点问题, 而对住房价格水平及其变化的及时、真实和准确反映, 则构成了各种住房价格问题研究和讨论的基础性条件。因此, “加强市场监测, 完善市场信息披露制度” 近年来一直被视为政府针对住房市场的宏观调控政策的重要组成部分, 而住房价格统计又构成了其中的主体。

这样的背景无疑对住房价格统计的准确性提出了

更高的要求, 也就需要对住房价格统计过程中可能存在的各类误差进行更为深入的分析。样本代表性误差正是住房价格统计过程中面临的一个主要问题, 也是制约我国住房价格统计结果准确性的主要瓶颈之一, 但目前我国相关研究和统计实践中尚未对其给予充分重视。针对这一情况, 本文将对住房价格统计中样本代表性误差的根源、影响和修正方法进行系统分析, 以期有助于进一步提高我国住房价格统计的质量。

1 样本代表性误差的形成原因

住房价格统计中的样本代表性误差根源于住房商品的特殊性。首先, 住房是一种高度异质性商品。任何两套住房都不可能完全相同, 可能在区位条件、邻里环境、物理属性等方面存在巨大差异, 这些差异又

基金项目: 建设部研究开发项目 (439 06-K9-72)、国家自然科学基金 (70603017)

作者简介: 吴璟, 博士研究生

收稿日期: 2007-11-15 Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

将直接反映在其价格水平上。因此,任何一套住房的
市场价格,只有在其被交易时才可能被准确地观测和
记录下来。相反,对于日常生活中大量接触的普通同
质性商品而言,只需同一品种中的某一个体发生交
易,该品种内其他个体的价格就可由此推及。其次,
一套住房在其整个生命周期中通常只会被交易有限
的次数。换言之,在住房存量市场中,特定时间段内
只有一部分住房发生交易,特别是当报告期较短(例
如按月考察)时更是如此。以北京市为例,其 2006
年存量住房市场月均交易量约为 60 万 m^2 ,而同期
存量住房总量则超过 3 亿 m^2 ,即平均每月仅有不
到 0.2% 的存量住房发生交易。住房增量市场(即
新建商品住房市场)中也存在类似情况,特定时间
段内达成交易的住房单元仅是市场中可供交易单元
的一部分。仍以北京市为例,2006 年其新建商品
住房市场的平均吸纳周期约为 9 个月,即每个月
内仅有约 11% 的可销售单元发生交易。由此可见,
无论是存量住房市场还是增量住房市场,每一个报
告期内都只可能有一小部分住房单元的交易价格
被观测到。相反,在股票等交易极其频繁的证券
市场中,即使在有限的报告期内也可以认为市场
中的所有品种都会发生交易,并被观测到价格。

这就意味着,进行住房价格统计时,只能使用
一部分单元的交易价格数据来代表或推断整个市
场的情况。这事实上可以视为一种抽样统计。根
据抽样调查理论,经过合理的试验设计后,完全有
可能通过只占目标总体一小部分的样本来推断目
标总体的性质。但是,其抽样过程必须满足一定
要求,例如样本容量应当具备一定规模、样本应
来源于目标总体中的随机抽取等^[1]。然而,住
房价格统计的“抽样”过程是在住房市场中自发
形成的,并不受测量主体的控制,因此通常难以
有效满足上述条件,导致多数情况下特定报告
期内交易单元的价格信息并无法合理代表整个
市场的情况,其中存在的误差就被称为样本代表
性误差^[2]。

2 样本代表性误差的类型和影响

国外文献中已经对住房价格统计中的样本代
表性误差问题进行了大量讨论,但这些研究主要
关注于其中的系统误差部分,即所谓的“样本选
择性偏误(sample selection bias)”。当前述“抽
样”过程存在非随机性,使得具有某一类特征
的住房在样本中被有规律的过度表达或表达不
足,并且此类住房的价格水平或变化趋势与市
场总体存在差异时,就将导致价格统计结果系
统性的偏离真实情况,作为一种系统误

差,样本选择性偏误将影响统计结果的有效性(效
度),因而受到国外学者的高度关注。但同时其研
究也表明,样本选择性偏误通常仅伴随特定方法
(例如重复交易法^[3-4])或特定数据源(例如抵
押物业估计数据^[5])的使用而出现,而这些方法
或数据源尽管已经在西方国家成为主流,但在
我国却尚未得到应用。因此,有理由认为样本
选择性偏误目前对我国住房价格统计的影响并
不明显。

相比之下,样本代表性误差中的随机误差部
分对我国住房价格统计的影响可能更为突出。这
种误差主要来源于样本容量不足,即报告期内
交易单元在总体市场中所占的比例过低,不足
以满足大数定律和中心极限定理的要求,使得
样本估计量受到样本分布情况的显著影响,无
法接近目标总体特征,但其偏离方向和程度不
具规律性,随各报告期内样本分布情况的不
同而改变。

特别是在当前我国新建商品住房市场的现
实背景下,这种误差的影响显得尤为重要。我国
新建商品住房市场目前以项目型集中供应的公
寓式住房单元为主,同一项目中的住房通常在
主要特征和交易价格上均十分相似,与此同时,
某一报告期内市场中交易活跃的项目往往相对
有限。因此,在每一个报告期内,可供住房价格
统计使用的样本(即交易单元)往往存在很高
的聚集度,表现为样本空间内的若干簇样本。
如图 1 所示,以成都、太原、杭州 3 个城市为例,
尽管这 3 个城市的市场规模各异,其月成交量
中约 80% 的单元都集中在约 20% 的项目中,
且市场规模越小,聚集度越高。而这些活跃项
目将对价格统计结果产生显著影响,特别是当
其中某个项目规模特别大,或者与普通项目的
差异性较大时,其影响尤为明显。许多城市的
住房价格统计中,都经常出现这样的情况:因
为某一个位于市中心的高档项目的推出而导
致该报告期的价格统计结果大幅上涨,或者因
为某一

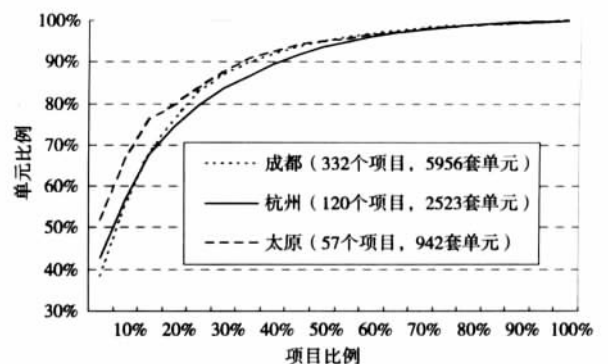


图 1 三城市住房交易单元构成情况的帕累托图 (2005.10)

Fig. 1 Concentration of housing units in major projects in three selected cities (October 2005)

个郊区大盘的上市而导致该报告期的价格统计结果明显下降。这就使得住房价格统计结果的波动（特别是短期振荡）实际上在很多时候都源于这种“大盘效应”的影响，而并非市场景况真实变化的反映，但事实上后者才是住房价格统计的真正关注点。

综合上述分析，我国住房市场的特点决定了样本代表性误差中的随机误差部分是我国住房价格统计中面临的一项主要问题，而国外文献中大量讨论的系统误差部分（样本选择性偏误）的影响则相对有限。因此本文将以样本代表性误差中的随机误差部分作为主要分析对象，如无特别说明，以下提及的样本代表性误差也仅特指其中的随机误差部分。

3 样本代表性误差的修正方法

3.1 评估价格法

评估价格法的修正思路是，以住房评估价格取代交易价格，作为住房价格统计的依据。这种情况下，在各报告期内都有可能获取所有住房单元的价格信息，因而从根源上避免了样本代表性误差。美国房地产投资受托人理事会（NCREIF）编制物业收益率指数（NPI）、仲量联行在香港地区编制高档物业价格指数时，都采用了由专业估价人员定期对所有选定物业进行估价的方式。

但是，评估价格法的成本较高，难以大量使用，前述应用实例均以有限的物业组合作为考察对象，在我国规模庞大的住房市场中几乎不具备实施的可行性。其次，评估价格的准确性也经常受到质疑，可能存在“平滑偏误”等问题。因此，评估价格法目前在统计实践中的应用十分有限。

3.2 固定权重法

随机误差型样本选择性误差的特征是统计结果受到样本分布情况的显著影响，为此，另一种修正思路是在各报告期内固定样本的分布情况。通常的做法是，根据区位和物理特征对物业进行分组，在每个报告期内首先计算得到各细分组平均价格，再根据事先确定的权重（最常见的是固定沿用基准期权重，即拉氏公式法）求取各细分组价格的加权平均值^[6]。因其操作简便，固定权重法是目前国内应用最多的一种修正方法，香港差饷物业估价署发布的 RVD 指数，以及深圳市、天津市房地产主管部门编制的住房价格指数等，都采用了这一思路。

但固定权重法仍存在两点难以克服的障碍。首先是权重的确定，住房市场中通常无法准确取得各细分组的真实权重，而错误的权重将直接导致统计结果出

现系统误差（样本选择性偏误），这在拉氏公式法的应用过程中尤为突出；其次是细分组的划分，划分越细修正效果越好，但过多过细的分组不仅加大了统计成本，而且可能出现某些细分组在报告期内没有交易样本的情况。这些问题都影响了固定权重方法的效果。

3.3 特征价格法

特征价格法是基于特征价格模型进行住房价格统计的一种方法。如式(1)所示，首先通过各报告期内的交易样本，利用线性回归等方法分析各报告期内各种住房特征 (X_n) 对住房价格 (P) 的影响规律（即求解各种特征的特征价格 a_n ），再代入选定的住房特征组合（“标准住房”），求取其在各报告期内的价格，在此基础上构造价格统计量。由于考察对象（“标准住房”）在各报告期内保持一致，因此理论上不存在样本选择性误差。

$$P = c + \sum_{n=1}^N a_n X_n + \varepsilon \tag{1}$$

但实践中样本选择性误差的影响同样存在。特征价格法通常借助线性回归等方法求取各住房特征的特征价格，在样本量不足——特别是在出现样本聚集时，回归过程中很有可能出现不准确的估计结果^[7]。以区位条件的影响为例（图 2），由于要素替代规律的作用城市中住房价格梯度通常呈内凹曲线（图中实线所示）。但仅借助于某一报告期内的样本时，这种规律通常难以得到准确反映。图 2 就提供了一种较为极端的情况，当某一报告期样本大量聚集时，个别高杠杆点将对回归结果产生显著影响^[7]，并使得回归得到的住房价格梯度线（图中虚线所示）明显偏离于真实情况，从而导致住房价格统计结果出现误差。

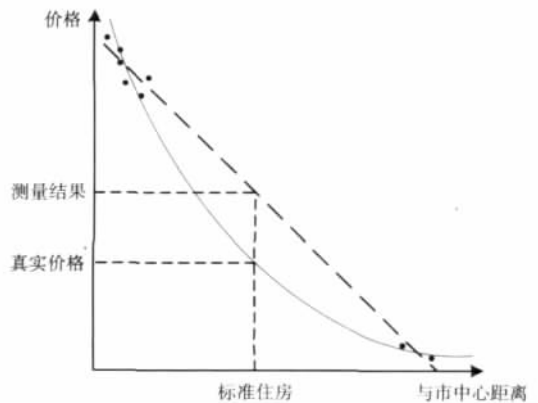


图 2 样本代表性误差对特征价格法的影响
Fig. 2 Effect of sample error on the hedonic method

这一问题可以通过选择适宜的特征价格法形式来克服。使用截面建模形式的特征价格法，即在各报告期内分别建立特征价格模型时，回归结果受单个报告期内样本分布情况的影响较大，上述误差可能较为显

著。而使用一次性建模形式的特征价格法，即利用所有报告期数据建立统一的特征价格模型时，则可以通过多个报告期样本在样本空间上的叠加，弱化样本分布的影响，使得回归结果更接近于真实情况，从而在更大的程度上避免样本代表性误差的影响。

4 实证分析

下面对上述各种修正方法的效果进行实证检验。现实市场中，由于无法获取住房价格的“真实”走势，因此难以准确判明样本代表性误差的影响程度，也难以直观比较各种修正方法的效果。为此，本文引入一个基于北京市住房市场的数值模拟算例进行说明。

首先，基于 2002~2005 年间北京市各住房项目的属性和平均销售价格数据，对住房价格的决定机制进行研究。研究中发现，与城市中心区距离 (Distance) 是影响住房项目价格 (Price) 的最主要因素，且定量研究后发现存在如下梯度关系：

$$\ln(\text{Price}) = 9 - 0.35 \ln(\text{Distance}) + \mu$$

$$\mu \sim N(0, 0.25) \quad (2)$$

在此基础上，假设式(2)反映出的价格-区位关系是影响住房价格的惟一确定因素，同时，假设市场中住房价格保持每报告期 5% 的环比增幅，则其“真实”价格轨迹如图 3 中粗实线所示。进一步的，假设该市场的总体规模为 200 个项目，按与市中心距离均匀分布 (城市半径限定为 40 km)。各报告期内均约有 15% 的项目发生交易，且各项目的交易概率相同。利用 SPSS 中的样本随机抽取功能和 Excel 中的正态分布随机数发生器，模拟上述过程，得到连续 12 个报告期内的交易样本。某一次模拟中，得到交易样本的分布情况如表 1 所示。

表 1 各报告期交易样本与城市中心距离的平均值
Table 1 Average distances to CBD of the samples in each period (km)

报告期	平均距离	报告期	平均距离
T1	15.18	T7	16.38
T2	16.16	T8	20.08
T3	17.96	T9	24.00
T4	15.24	T10	16.86
T5	19.68	T11	14.82
T6	24.56	T12	22.10

下面针对这些样本进行价格统计，各种方法得到的价格统计结果与事先设定的真实价格轨迹越接近，该方法的准确性程度越高，反之则较低。基于可行性角度考虑，首先排除评估价格法，因此最终选用拉氏公式法、截面建模特征价格法和一次性建模特征价格

法三种修正方法分别进行价格统计，同时引入未进行任何修正的加权平均法 (各样本权重相等，即等价于简单算术平均法) 进行比照。各种方法的计算结果以价格指数的形式表现，如图 3 所示。表 2 进一步给出了各种方法得到的住房价格环比涨幅与真实值 (5%) 之间的差异。

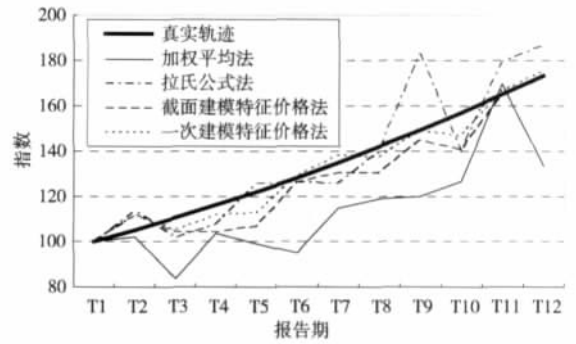


图 3 住房价格统计结果与真实趋势的比较 (T1=100)
Fig. 3 Comparison between the statistical results and the real price path (T1=100)

表 2 各种方法统计结果与真实趋势的误差
Table 2 Results of different adjustment methods

方法	环比涨幅范围	环比涨幅方差
加权平均法	[-21.41% , 33.71%]	259.38%
拉氏公式法	[-23.45% , 31.47%]	233.19%
特征价格法 (截面建模形式)	[-6.87% , 18.14%]	59.52%
特征价格法 (一次建模形式)	[-6.00% , 14.62%]	40.10%

正如理论分析中所预测的，由于样本容量规模有限，统计结果受到样本分布情况 (表 1) 的显著影响，因此各种方法得到的统计结果均与事先设定的市场真实轨迹存在一定差异。其中，加权平均法未对样本代表性误差做任何修正，故其受影响程度最大，某一报告期内的统计结果可能高估 (例如 T10 至 T11) 或者低估 (如 T7 至 T8) 市场价格的真实变化幅度，甚至可能得出与真实趋势完全相反的结果 (如 T11 至 T12)。

三种修正方法中，固定权重的拉氏公式法的统计结果略好于加权平均法，但其改善程度有限，特别是在某些报告期内 (如 T9)，如果样本聚集程度很高，且其分布与基准期差异较大，统计结果将出现显著误差。两种形式的特征价格法都取得了较高的效果，其中一次性建模法的结果最接近真实轨迹，截面建模法次之，与理论分析结果相符。但需要指出的是，该算例中假设样本均匀分布且出现概率相同，在很大程度上避免了图 2 所示的模型估计误差。实际应用过程中，随着样本聚集性程度的加大，截面建模法的效果可能受到更大的影响，从而使得一次性建模法的优势更为显著。

5 结论和建议

若干住房经济学领域的研究成果都证明，住房价格的变化并不是一个随机游走过程，相反存在着巨大的惯性，而价格统计结果中反映出的短时间内的剧烈波动，多数实际上是统计方法的局限性所引入的噪音^[8-9]。样本代表性误差正是这种噪音的主要来源之一。住房商品的异质性和交易不频繁性等特征，决定了基于交易数据进行住房价格统计时，样本代表性误差几乎是无法根除的。

对于住房价格统计结果的使用者而言，在我国目前多数住房价格统计均未针对样本代表性误差进行有效修正的情况下，有必要对统计结果的局限性和可能存在的误差有充分认识。特别是对于政府管理部门而言，不应应对短期内出现的住房价格波动过于敏感，特别是应避免根据住房价格的短期变化而过快作出调控决策。否则，如果像图 3 所反映的那样，出现了统计结果和真实情况的背离，那么这种基于错误认识作出的不当决策，很有可能将会带来巨大的负面影响。

另一方面，对于住房价格统计的实施者而言，则应更加重视样本代表性误差，努力通过恰当的修正方法弱化其影响。本文的理论分析和数值模拟结果表明，评估价格法、固定权重法等均不能取得理想的效果，而特征价格法——尤其是其中的一次性建模法，则更为有效。

吴 璟 (1981-)，男，博士研究生。主要从事房地产经济学研究。

郑思齐 (1977-)，女，副教授。主要从事房地产经济学、城市经济学研究。

刘洪玉 (1962-)，男，教授。主要从事房地产经济学研究。

参 考 文 献

[1] 杜子芳. 抽样技术及其应用 [M]. 北京: 清华大学出版社, 2005(Du Zifang. Sampling techniques and practices[M]. Beijing: Tsinghua University Press, 2006 (in Chinese))

[2] Haurin R, Hendershott H. House price indexes: issues and results[J]. Real Estate Economics, 1991, 19(3): 259-269

[3] Clapp M, Giaccotto C. Which real estate price index is right for mortgage-backed securities [J]. Real Estate Finance, 1995, 12(2): 41-47

[4] Genesove D, Mayer C. Loss aversion and seller behavior: evidence from the housing market[J]. The Quarterly Journal of Economics, 2001, 116(4): 1233-1260

[5] Case K, Shiller R. Is there a bubble in the housing market [J]. Brookings Papers on Economic Activity, 2003, (2): 299-362

[6] 徐国祥. 统计指数理论及应用[M]. 北京: 中国统计出版社, 2004 (Xu Guoxiang. Statistical index theory and its applications[M]. Beijing: China Statistics Press, 2004 (in Chinese))

[7] 吴喜之, 田茂再. 现代回归模型诊断[M]. 北京: 中国统计出版社, 2003(Wu Xizhi, Tian Maozai. Modern diagnose techniques of regression models[M]. Beijing: China Statistics Press, 2003(in Chinese))

[8] Case K, Shiller R. The Efficiency of the market for single-family homes [J]. American Economic Review, 1989, 79 (1): 125-137

[9] Shiller R. Macro markets: creating institutions for managing society's largest economic risks [M]. New York: Oxford University Press, 1998