

# 基于时间序列分析南京市二手房的定价模型

周洪伟

（南京晓庄学院信息工程学院 江苏·南京 211171）

**摘 要** 由于房地产普遍具有较强的异质性、不易流动性、不可分割性特征，致使对二手房资产的定价研究难度较大。时间序列分析是处理动态问题的一种重要的数学工具，而二手房的价格波动显然与时间相关，本人利用从搜房网收集的南京地区的2013年2月-2015年2月的二手房房价数据，利用时间序列分析作为工具建立了二手房价格预测的随机性模型，并把预测的价格和实际价格作了对比分析，研究发现两者的误差较小，对二手房交易的买卖双方能够提供一些合理的参考价值。

**关键词** 二手房 时间序列 模型 定价

中图分类号 F830

文献标识码 A

DOI:10.16400/j.cnki.kjdz.2016.10.067

## The Pricing Model of Second-hand Houses in Nanjing City Based on Time Series Analysis

ZHOU Hongwei

(College of Information Engineering, Nanjing Xiaozhuang University, Nanjing, Jiangsu 211171)

**Abstract** As a result of real estate generally has a strong heterogeneity, is not easy to mobility, non segmentation and other features, resulting in the price of second-hand housing assets is difficult to study. Time series analysis is to deal with dynamic problems an important mathematical tool, and fluctuations in the price of second-hand housing obviously and time related, I use from Soufangwang collected in Nanjing area in February 2013 2015 February second-hand housing prices data, using time series analysis as a tool to establish the second-hand housing prices to predict the random model; and at the analysis and comparison of the predicted and actual price, the study found that a smaller error of both, to second-hand housing transactions, the seller and the buyer can provide some reference value.

**Keywords** second-hand houses; time series; model; pricing

笔者从搜房网房天下(<http://nanjing.fang.com>)搜集了南京地区的二手房房价数据，得出南京二手房房价各个月份的平均价格，借此数据利用 Excel 和 Eviews 软件对南京二手房房价序列数据进行时间序列分析。

### 1 模型分析

任何事物发展都可能受偶然因素影响，为此要利用统计分析中加权平均法对历史数据进行处理。该方法简单易行，便于掌握，但准确性差，一般只适用于短期预测。时间序列预测一般反映三种实际变化规律：趋势变化、周期性变化、随机性变化。

#### 1.1 模型建立

复合型模型序列： $Y_t = C_t \times S_t \times T_t \times I_t$ ，其中  $S_t$  代表季节变动， $T_t$  代表长期趋势， $C_t$  代表循环变动， $I_t$  代表随机波动。南京二手房房价序列数据为  $Y_t$ ，运用加权平均的算法计算出南京二手房房价各个月份的平均价格，后续的二手房房价时间序列分析数据均取自于此。本文中计算  $y_t$  表示南京二手房价格（元/平方米）， $t$  表示月数，通过移动平均法尽可能消除随机波动项。按先易后难的步骤，先求出季节变动  $S_t$ 。之后用分离法即除法，将季节性从时间序列中剔除，接着求出趋势函数  $T_t$ 。继续使用分离法，然后使用移动平均消除随机波动  $I_t$ ，剩余部分就是循环波动  $C_t$ 。最后用剩余法求得随机波动  $I_t$ 。

##### 1.1.1 季节性分析

先观察 2013 年 2 月至 2015 年 2 月的二手房房价图，详细数据见附录表 1。

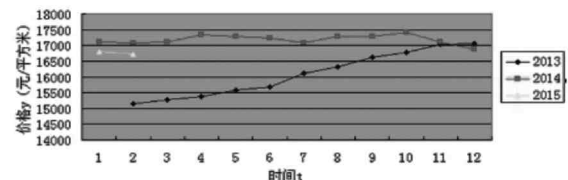


图 1 2013 年 2 月至 2015 年 2 月的二手房房价图

从图 1 观察，可能存在季节变动。下面对南京二手房房价进行季节分析，通过计算得到 1~12 个月的季节指数，如表 1。

表 1 1 月至 12 月季节指数表

月份	1	2	3	4	5	6
修正过的季节指数	0.987791	1.018414	0.994026	1.01906	0.999283	1.012686
月份	7	8	9	10	11	12
修正过的季节指数	0.995748	1.006701	0.993128	1.016265	0.989658	1.01347

即季节变动  $S_t$  可以用上述数据表示。

##### 1.1.2 趋势分析

将收集来的数据绘制一幅折线图，如图 2。

通过图像观察，显然这组数据并不平稳。下面  $t$  代表时间， $y_t$  代表二手房房价（元/平方米）。

求得  $y_t$  相关系数  $r = 0.745$ ，相关系数越接近 1 线性相关性越强，设线性函数  $y_t = \beta_1 t + \beta_0 + u_t$ ，其中  $u_t$  为独立的随机序列。

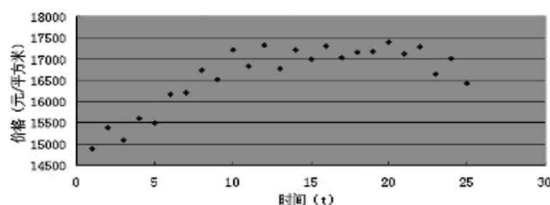


图2 2013年2月至2015年2月的二手房房价图  
利用 Eviews 工具算线性函数系数。

$$\hat{y} = 76.659t + 15600 \quad R^2 = 0.544$$

$$(5.35) \quad (73.23)$$

每个系数下面对应一个  $t$  检验值, 括号内为对应系数的  $t$  检验值, 比较式括号内检验值都取绝对值, 因为有两个变量自由度取  $n-2$ , 即 23。查表可得  $t_{0.05}(23) = 2.81$ 。在显著性水平 0.05 下, 检验值都大于临界值, 认定该系数值在统计上是显著的, 所以通过检验。称  $R^2$  为判定系数, 取值范围在 0 到 1 之间, 越接近于 1, 用模型来拟合数据的拟合程度越好。

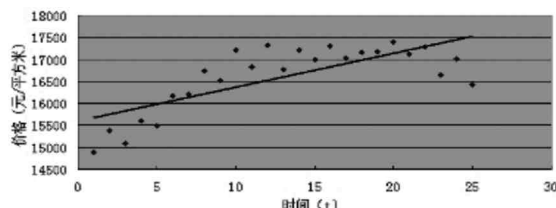


图3 2013年2月至2015年2月的二手房房价图

观察图3, 显然线性拟合效果不是太好。Y, lnt 的相关系数  $r = 0.879$ , lnt 与  $y$  线性正相关。接下来尝试一下对数函数拟合。

设对数函数  $\hat{y} = \beta_1 \ln(t) + \beta_0 + u_t$ , 其中  $u_t$  为独立的随机序列。  
用 Eviews 工具算函数系数, 计算结果如下  
对数函数  $\hat{y} = 798.66 \ln(t) + 14873.87 \quad R^2 = 0.773$ ;  
 $(8.856) \quad (66.458)$

对应系数的  $t$  检验值, 系数值在统计上是显著的, 所以通过检验  $R^2 = 0.773$  稍优于上一个线性函数。最后再尝试一下多项式函数拟合。

设多项式函数  $\hat{y} = \beta_2 t^2 + \beta_1 t + \beta_0 + u_t$ , 其中  $u_t$  为独立的随机序列。利用 Eviews 工具, 解得多项式函数:

$$\hat{y} = -9.1526t^2 + 313.13t + 14619 \quad R^2 = 0.9695$$

$$(-16.298) \quad (20.817) \quad (172.256)$$

对应系数的  $t$  检验值, 当  $t$  值为负时先取绝对值再比较, 系数值在统计上是显著的, 所以通过检验。

通过  $R^2$  的数值和图4展示, 易得出这组数据拟合出的三个函数中多项式函数拟合效果最好。即趋势 T 选择

$$\hat{y} = -9.1526t^2 + 313.13t + 14619.$$

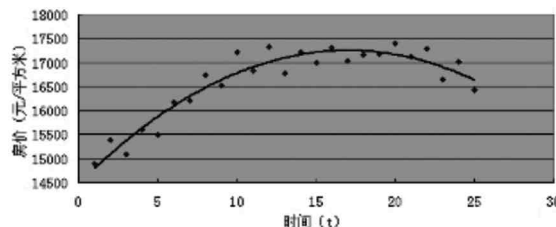


图4 2013年2月至2015年2月的二手房房价图

### 1.1.3 周期性分析与随机波动分析

季节变动(S)和趋势变动(T)都已研究好, 现在需要考虑是否具备周期性, 通过剩余法的思想和原理, 从时间序列中分别剔除掉季节变动(S)和趋势变动(T), 剩下循环波动(C)和不规则波动(I), 然后进行平滑, 将不规则波动(I)尽量消去。所剩下的结果即为循环波动值(C)。

观察图5 循环波动值在上下波动, 可以看出在 2013-2015 可能存在一个大周期和几个小周期。不过时间序列较短, 周期性的分析还有待于更长的时间序列数据来验证。

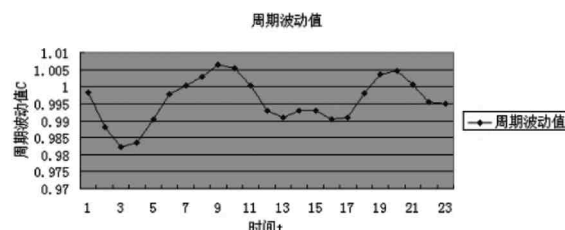


图5 周期波动值

### 1.2 模型评估

模型定为  $Y_t = C_t \times S_t \times T_t \times I_t$ , 通过均方误差来评判模型预测的好坏, 均方误差越小, 模型预测越准确。均方误差用 MSE 表示, 求得均方误差为 34306.512, 为保持单位一致将均方误差开根号求得 185.22 元/平方米。误差还是比较小的。拟合效果较好。

按照上述模型可以得到 2015 年 3 月、4 月以及 5 月的预测值分别为 16276.52 元/平方米, 16419.71 元/平方米, 15932.09 元/平方米。从搜房网搜集了 2015 年 3 月、4 月真实房价分别为 16695 元/平方米、16624 元/平方米。3 月预测值与真实值相差 418.48 元/平方米, 4 月预测值与真实值相差 204.29 元/平方米, 两个月误差都在 500 元/平方米以内, 可以接受。

### 2 小结

笔者利用从搜房网上搜集的南京地区的 2013 年 2 月-2015 年 2 月的二手房房价数据, 考虑影响二手房的几个重要的因素: 经济环境因素、区位因素和建筑因素的基础上, 利用多元统计分析中时间序列分析的相关理论, 通过建立随机模型, 给出二手房价格的预测公式, 并作了统计的显著性检验, 希望该模型的分析方法能给二手房交易的买卖双方能够提供一些合理的参考价值。

附录表 1 南京某地区的二手房房价数据表

年份 月份	2013	2014	2015
1		17103	16806
2	15150	17084	16719
3	15280	17108	
4	15375	17325	
5	15585	17294	
6	15690	17242	
7	16102	17081	
8	16306	17292	
9	16612	17281	
10	16782	17408	
11	17026	17105	
12	17057	16872	

### 注释

贾俊平. 统计学(第二版)[M]. 北京: 清华大学出版社, 2006.7.

张万宏. 非平稳时间序列的预测方法研究[J]. 控制理论与控制工程, 2007.5.8.

Jonathan D. Cryer, Kung-Sik Chan. 时间序列分析及应用 R 语言[M]. 潘红宇等译. 北京: 机械工业出版社, 2011.1.

李敏, 陈胜可. Eviews 统计分析及其应用[M]. 北京: 电子工业出版社, 2011.5.