

# 用逐步回归法解释北京房价

李昕垚

2016 年 8 月 17 日

## 目录

|                 |   |
|-----------------|---|
| 1 前言            | 1 |
| 2 逐步回归          | 1 |
| 2.1 逐步回归结果      | 1 |
| 2.2 $t$ 检验进一步降维 | 2 |
| 2.3 真实单价和拟合单价对比 | 2 |
| 2.4 残差对比        | 3 |

## 1 前言

现阶段大家对于房价的判断基本是 **定性的**，如通过 **M2** 和 **政策** 来判断房价的涨跌。本文目的是通过 **逐步回归法** 定量的解释 **房价上涨的因素**。

## 2 逐步回归

回归分析是统计学习最重要的组成部分，虽然前提假设多，但凭借结果稳定、可解释性强等优良性质经久不衰。逐步回归法是回归方法的一种，主要用于自变量降维。如这次试验，开始时选择了 53 个自变量，通过逐步回归法可以降到 25 个，再通过参数的  $p$  值可以降到 3 个，方便原因的解釋。

### 2.1 逐步回归结果

| 统计量       | 值      | p 值                     |
|-----------|--------|-------------------------|
| F         | 88.68  | $3.003 \times 10^{-12}$ |
| 调整后 $R^2$ | 0.9821 | —                       |

2.2 t 检验进一步降维

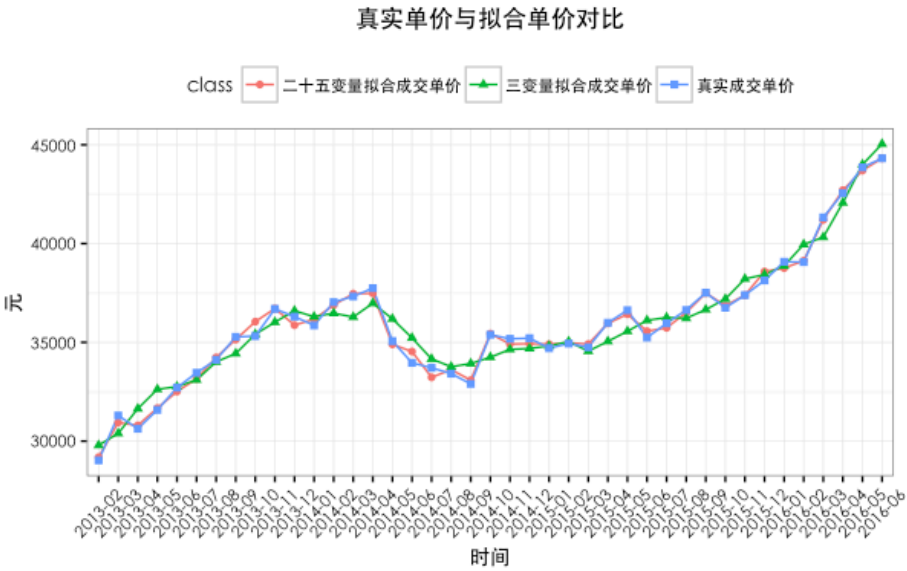
考虑到现有模型自变量数依旧较多，通过  $t$  值可以进一步筛选自变量，代价是损失了少量的  $R^2$ 。回归结果还是较为满意的，虽然损失了  $R^2$ ，但是回归方程整体的显著性有所提高，参数的显著性也有所提高。

| 统计量       | 值      | p 值                   |
|-----------|--------|-----------------------|
| F         | 234.3  | $2.2 \times 10^{-16}$ |
| 调整后 $R^2$ | 0.9459 | —                     |

设本月成交单价为  $y$  元，上月新增客源单价为  $x_1$  元，上月 M2 为  $x_2$  亿元，上月国内信贷为  $x_3$  亿元，则回归方程为：

$$y = 1.258 \times 10^4 + 8.446 \times 10^{-1} \times x_1 - 2.785 \times 10^2 \times x_2 + 2.431 \times 10^{-2} \times x_3$$

2.3 真实单价和拟合单价对比



从上图可以看出，选用 25 个变量的拟合效果非常好，选用 3 个变量的拟合效果稍差。

2.4 残差对比

