



# 基于网络搜索数据的房地产 价格预测

国家统计局统计科学研究所

董倩

二零一四年十月二十九日



# 主要内容

研究背景和思路

理论分析框架

数据描述

房地产价格预测模型

基于网络搜索数据的房地产  
价格预测

结论与展望



# 一、研究背景和思路

---



研究背景



国内外研究现状



研究思路



# 研究背景

- 大数据时代的到来给政府统计带来了巨大机遇和挑战，中国政府统计已开始与企业合作，研究、试点大数据的应用工作。
- 政府统计部门对房价统计制度方法进行了一系列改革，改革后的房地产价格指数更加合理，但是由于该指数系列发布时间是每月中旬，其时效性仍无法满足大众的需求。
- 网络搜索数据通常代表经济主体的预期、预示着经济主体的下一步的经济行为，对其研究分析能够得出经济主体行为的趋势与规律。
- 利用网络搜索数据对房地产价格走势进行预测，是一种非常有效的工具和方法。



# 国内外研究现状

利用网络搜索数据进行预测在商业界和学术界均有诸多探索和研究。

- 百度公司与中科院：消费者信心指数。
- 百度预测：世界杯预测、高考预测、景点预测、城市预测和疾病预测。

应用网络搜索数据预测价格指数的研究还较少，而预测房地产价格的研究更少。

- 张崇等（2012）基于谷歌搜索数据的CPI预测具有一定的转折点预测能力，平均提前周期为4个月。
- Rajendra Kulkarni等人（2009）使用城市层面的谷歌搜索指数预测了20个城市经季节调整的case - shiller指数的变化情况。

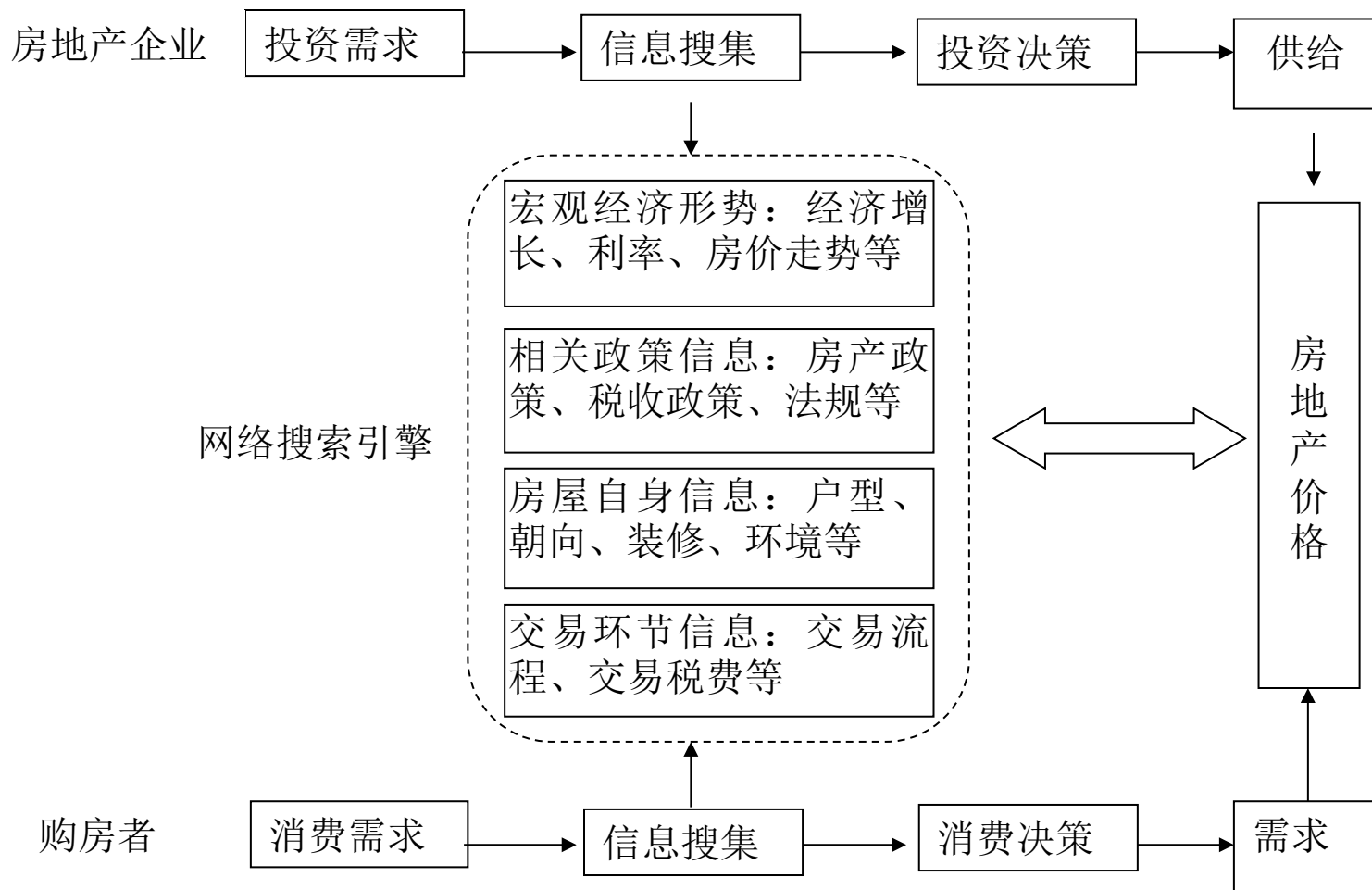


# 研究思路

- 为了解决房地产价格的时效性问题，本文尝试利用百度搜索数据预测我国主要大中城市的新建住宅销售价格指数和二手住宅销售价格指数。由于网络搜索数据可以实时获取，可以把影响价格变化的即时因素带入预测模型，这样在每月月初就可以得到上月的新房和二手房价格指数，比官方数据发布提前两周左右，弥补了传统统计数据信息发布滞后的问题，同时该预测数据也可以作为传统房地产价格统计数据的有益补充和参考。



## 二、理论分析框架





## 三、数据描述

---



研究对象



变量描述





# 研究对象

由于本文要利用百度搜索数据预测房地产价格，考虑到在一些规模较小或者经济不太发达的城市，人们对房地产信息的收集可能更多地通过广告、朋友介绍或房产中介等渠道，通过网络搜索房产信息相对较少，因此我们选取规模较大、经济比较发达、房地产交易相对活跃的6个城市作为研究对象。



一线城市

北京、上海、广州共3个。

二线城市

南京、沈阳、西安共3个。



# 变量描述

被解释变量



6个城市的二手住宅销售价格指数和新建商品住宅销售价格指数。采用2012年1月到2014年7月共31个月的月度同比数据，来源于国家统计局网站。



# 变量描述

## 解释变量：

- 首先根据房地产价格的影响因素，确定15个初始关键词；然后，利用百度搜索引擎的关键词自动推荐技术，得到关键词库；最后计算每个关键词与房地产价格指数的相关性系数，对关键词进行筛选。经过多次比较和筛选，得到以下关键词。
- **二手房：** 房价走势、房源、装修、房产网、公积金、房贷利率、房产税、房屋出租、房产中介、二手房、二手房交易流程、二手房交易税费。
- **新房：** 房价走势、房源、装修、房产网、公积金、房贷利率、新楼盘、保障房。



## 四、房地产价格预测模型

---



背景模型



预测模型构建



# 背景模型

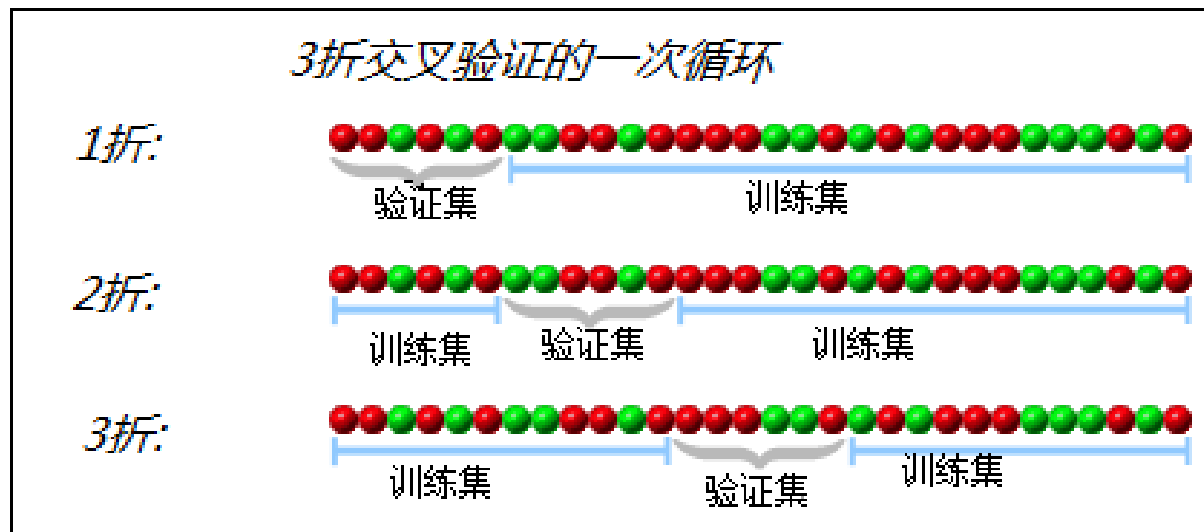
## 交叉验证技术

线性回归	回归树	Bagging	神经网络
混合性线性回归	随机森林	m-Boosting	支持向量机



# 预测模型构建： 交叉验证技术下的八种模型

- 针对每个城市，我们用3折交叉验证技术，采用8种模型进行预测，并对各种模型的预测结果进行比较。3折交叉验证的一次循环如图所示。

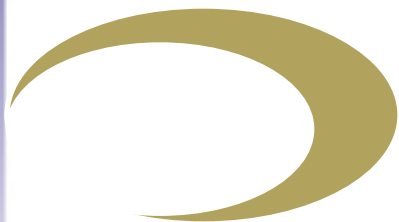





## 五、基于网络搜索数据的房地产价格预测



二手住宅销售价格指数预测



新建商品住宅销售价格指数预测



# 影响6个大中城市二手房价格的主要关键词搜索指数

城市	搜索关键词
北京	房价走势、房源、装修、公积金、二手房交易流程、房屋出租
上海	房价走势、房源、装修、房贷利率、二手房交易流程、二手房交易税费、房产中介、房屋出租
广州	装修、房产网、公积金、二手房交易流程、房屋出租
南京	装修、房产网、公积金、房贷利率、二手房、房产税、房屋出租
沈阳	房价走势、装修、公积金、房贷利率、二手房交易税费、二手房、房产税
西安	房价走势、装修、房产网、公积金、二手房交易流程、房产税、房屋出租





# 二手住宅销售价格指数预测

## 6个大中城市二手房价格最优预测模型

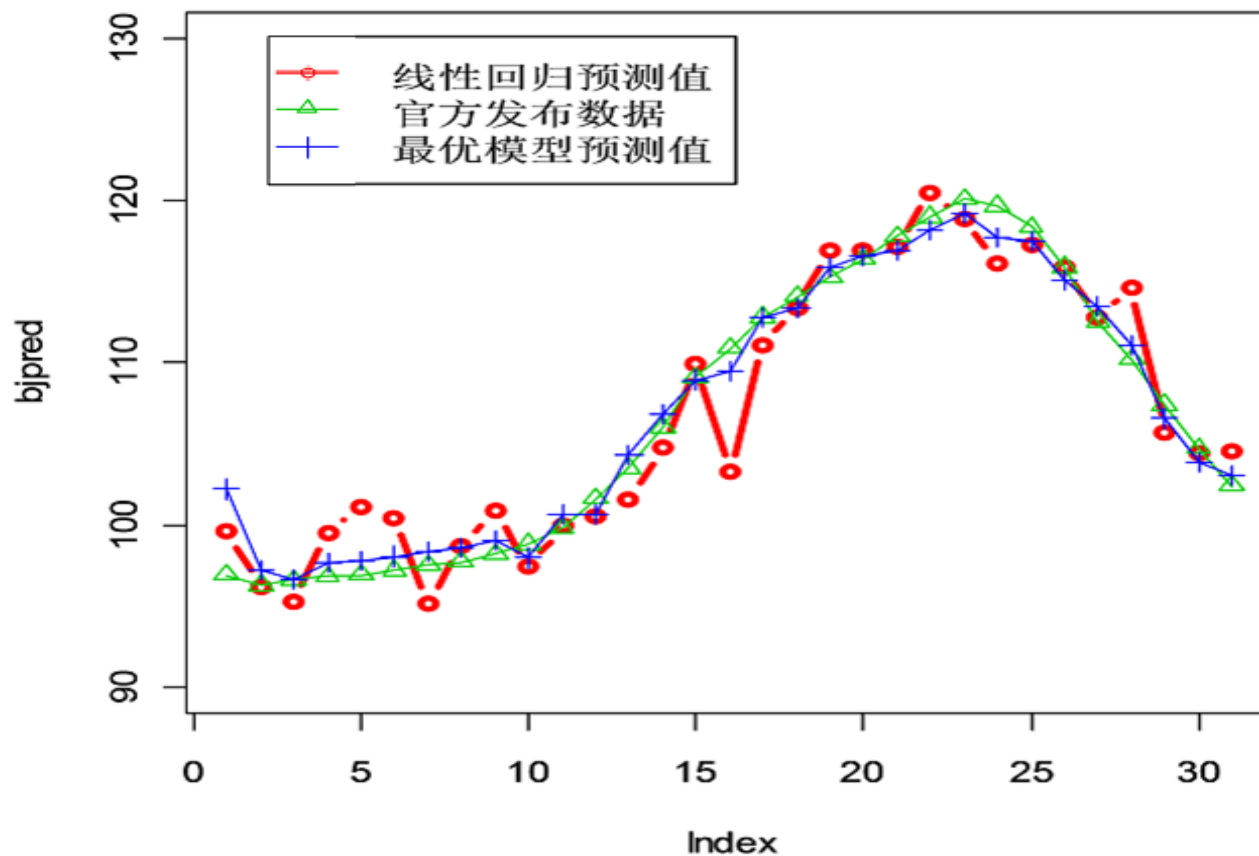
序号	城市	拟合度最优模型	稳定性最优模型
1	北京	随机森林	随机森林
2	上海	SVM	SVM
3	广州	SVM	SVM
4	南京	SVM	SVM
5	沈阳	SVM	SVM
6	西安	SVM	SVM



# 二手住宅销售价格指数预测

## 北京二手房价格预测模型效果图

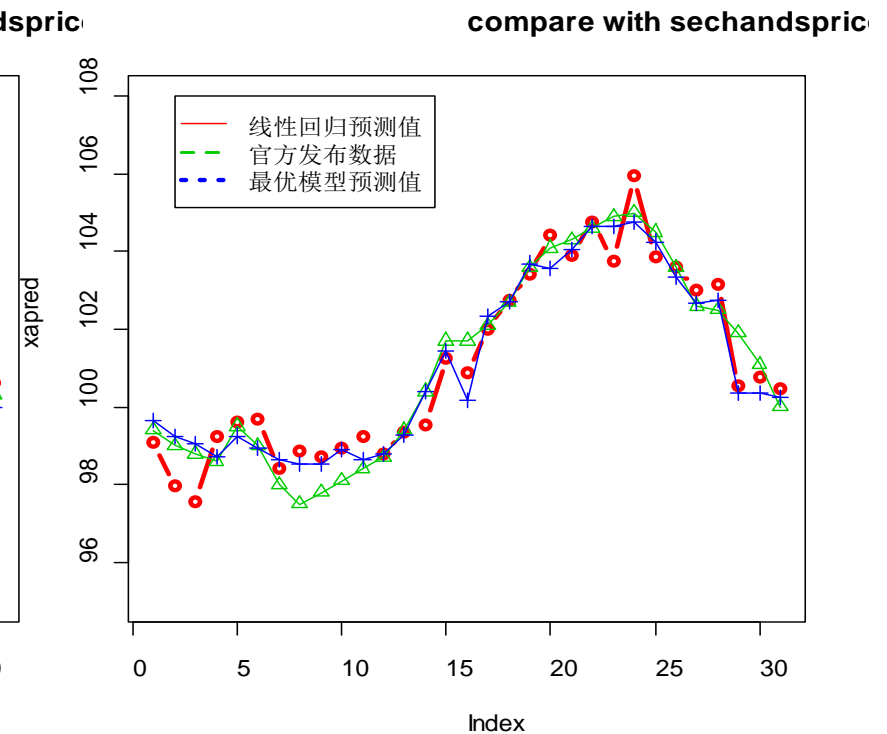
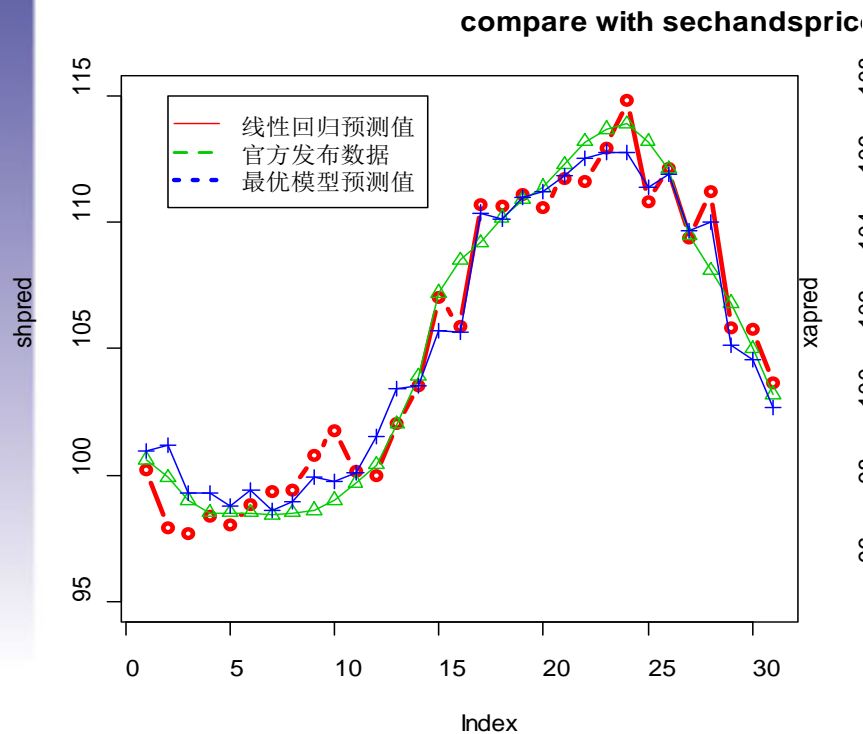
compare with sechandspricetb






# 二手住宅销售价格价格指数预测

## 上海和西安二手房价格预测模型效果图





# 影响6个大中城市新房价格的主要关键词搜索指数

城市	搜索关键词
北京	房价走势、房源、装修
上海	房源、装修、保障房
广州	装修、公积金、房贷利率、保障房
南京	房价走势、房产网、公积金、房贷利率
沈阳	房价走势、装修、公积金
西安	装修、房产网、公积金、房贷利率



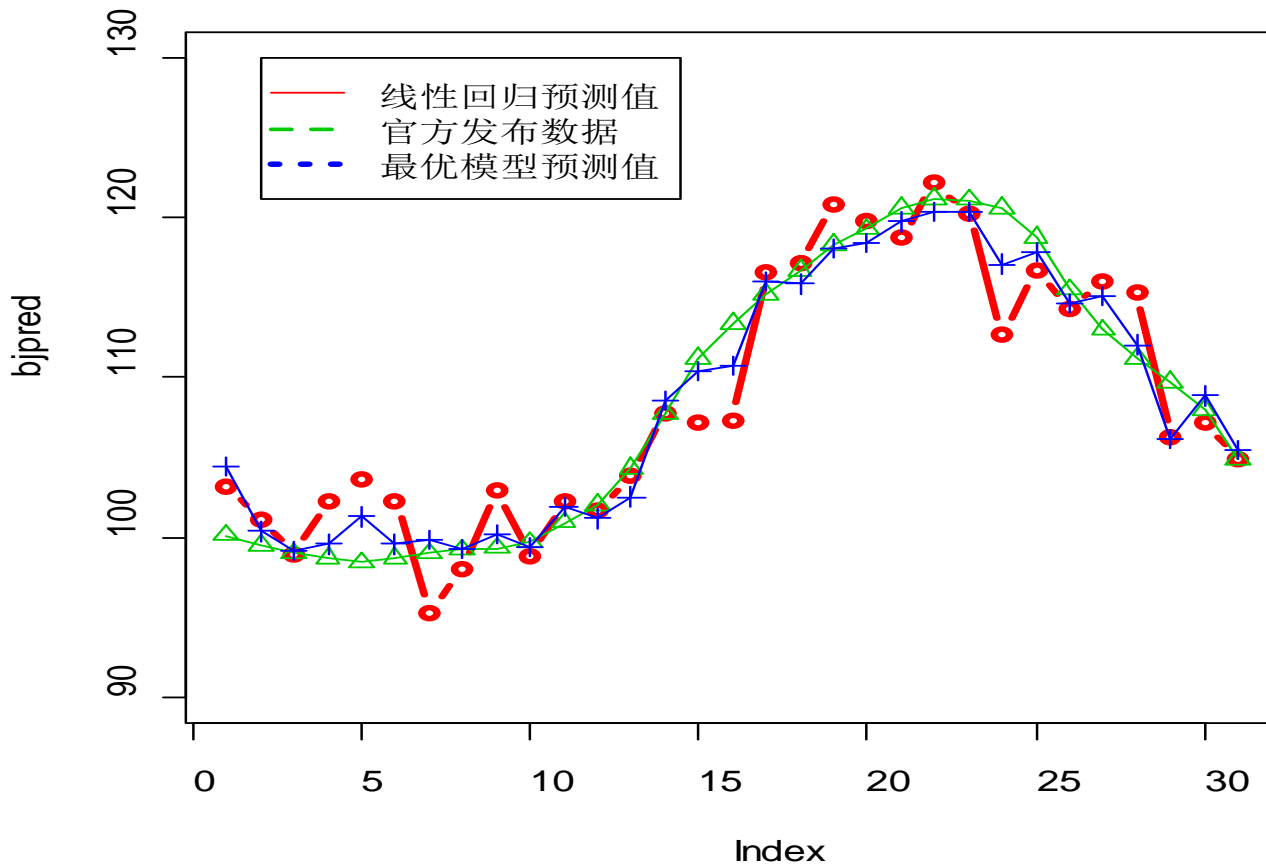
# 6个大中城市新房价格最优预测模型

序号	城市	拟合度最优模型	稳定性最优模型
1	北京	随机森林	随机森林
2	上海	SVM	SVM
3	广州	随机森林	随机森林
4	南京	SVM	SVM
5	沈阳	SVM	SVM
6	西安	随机森林	随机森林

# 新建商品住宅销售价格指数预测

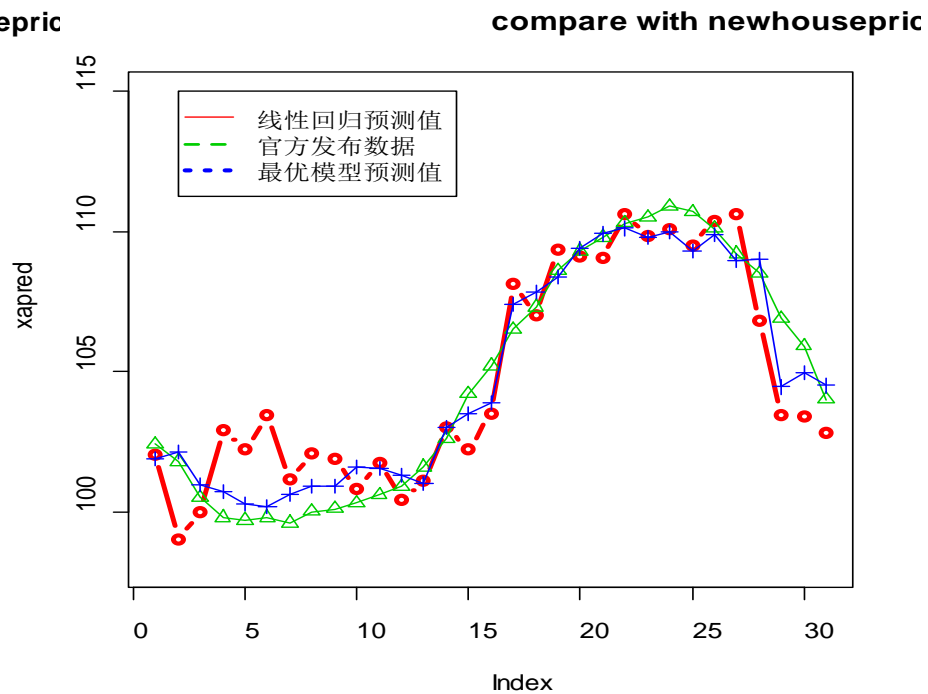
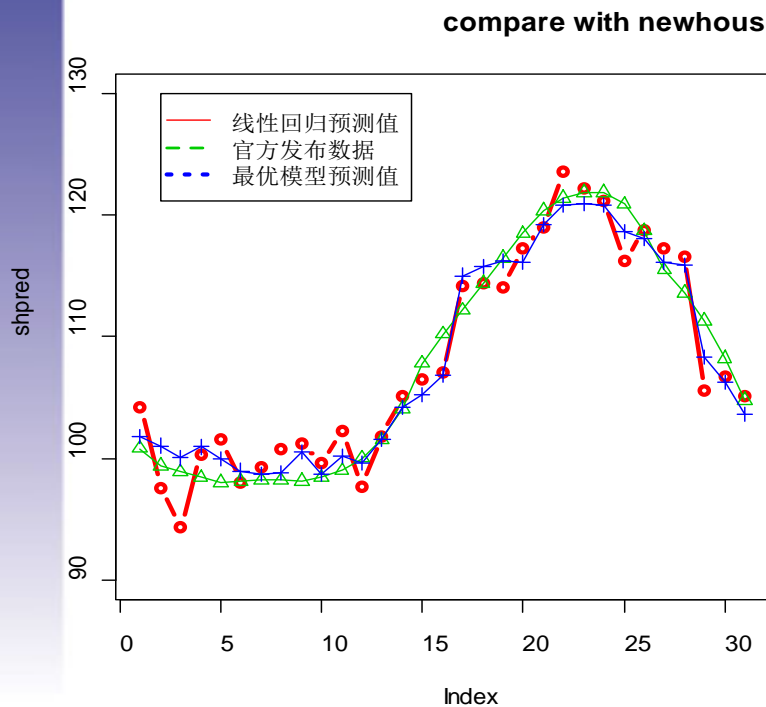
## 北京市新房价格指数实际值与预测值对比图

compare with newhousepric



# 新建商品住宅销售价格指数预测

## 上海和西安新房价格指数实际值与预测值对比图





## 六、结论与展望



研究结论



创新点



研究展望





# 研究结论

本文基于百度搜索指数，采用交叉验证技术和8种模型成功地拟合和预测了6个大中城市的新房和二手房价格指数，预测标准均方误差和均方误差最低达到0.0232，不但具有良好的预测效果，而且与传统预测方法和官方调查数据相比具有很强的时效性。由于百度搜索指数每日实时更新，因此基于本文的预测模型在每月1日即可得到上月的二手住宅和新建住宅销售价格指数的预测数，比官方统计数据提前了两周。



# 创新点

第一，采用网络搜索数据对房地产价格进行预测，在国内的相关研究还很少。采用网络搜索数据进行预测不但具有较好的预测效果，而且与传统预测方法和官方调查数据相比具有很强的时效性。



# 创新点

第二，通过运用交叉验证技术在8种模型中选择的最优模型成功地预测了6个城市的二手住宅和新建住宅销售价格指数。总体来看，线性回归模型和最优模型的预测结果与实际值的走势都基本一致，但是最优模型的预测值与实际值更接近。



# 创新点

第三，由于可获取的百度搜索指数是从2011年1月开始，截止到2014年7月，每个城市每个关键词可获得的数据量非常小，做同比处理后仅31个。为了弥补数据量较少可能带来的偏差，本文采用了3折交叉验证技术，保证了最终预测结果的精确性和可靠性。



# 研究展望

本文的研究思路和方法还可以进一步拓展到其他官方统计的月度公布数据，如CPI、居民收入、居民消费支出等，都可以通过选取适当的相关度较高的关键词，采用本文的研究方法进行预测，结果不但具有很强的时效性，对官方调查数据也是有益的补充。此外，根据月度劳动力就业情况的调查数据，还可以对月度失业率进行预测，弥补城镇登记失业率的不足。

而且，可以预见，随着网络搜索指数数据量的积累，模型预测精度将会越来越高。



谢谢! = ]