

Analysis of Profitability of Real Estate Listed Companies in Beijing, Shanghai and Shenzhen

—Based on Random Forest

Junqi Guan, Lin Li

School of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao Shandong
Email: imguanjq@163.com

Received: Feb. 26th, 2019; accepted: Mar. 14th, 2019; published: Mar. 21st, 2019

Abstract

According to the development of Beijing, Shanghai and Shenzhen real estate companies, choosing appropriate profitability indicators, the 57 listed real estate companies of Beijing, Shanghai and Shenzhen are classified, and the Random Forest model is established. The results show that the number of profit rankings of real estate listed companies in Beijing, Shanghai and Shenzhen is balanced, and the model by random forest has a high precision, which can effectively predict the profitability level of real estate listed companies in Beijing, Shanghai and Shenzhen. Based on the above conclusions, in response to the development of China's real estate industry, the corresponding policy recommendations for the future growth of the industry and how real estate companies improve their profitability are proposed.

Keywords

Real Estate, Listed Company, Profitability, Random Forest

京沪深房地产上市公司的盈利能力分析

——基于随机森林

官俊琪, 李 林

山东科技大学数学与系统科学学院, 山东 青岛

Email: imguanjq@163.com

收稿日期: 2019年2月26日; 录用日期: 2019年3月14日; 发布日期: 2019年3月21日

文章引用: 官俊琪, 李林. 京沪深房地产上市公司的盈利能力分析[J]. 金融, 2019, 9(2): 96-103.

DOI: [10.12677/fin.2019.92011](https://doi.org/10.12677/fin.2019.92011)

摘要

根据京沪深三地的实际情况, 选取合适的反映盈利能力的指标, 对三地57家房地产上市公司进行了盈利等级划分, 并建立了随机森林盈利能力分类模型。结果表明: 京沪深三地房地产上市公司盈利等级数量分布均衡, 且随机森林模型精度较高, 可以有效预测三地房地产上市公司的盈利能力等级。基于以上结论, 针对我国房地产行业的发展情况, 对于该行业的未来成长以及房地产公司如何提高盈利能力提出了相应政策建议。

关键词

房地产, 上市公司, 盈利能力, 随机森林

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

盈利能力能够体现企业利用现有资源获利能力, 其强弱能体现一个企业经营管理情况及未来发展前景, 是一个企业赖以生存的基础。通过对盈利能力分析能够了解目前企业的运行状况、发展状况以及暴露出来的一系列问题。所以, 盈利能力分析对于企业及其所在行业发展都有着重要的现实意义。近几年, 虽然房地产行业发展有所波动, 但仍然是关系国计民生的重要行业, 随着中国城镇化以及人口流动加剧, 对住宅的需求将逐步加大, 因此对房地产企业的发展现状有一个更好地了解, 才能对行业发展有更好地把握。京沪深三地在中国属于发达地区, 房地产企业众多, 发展迅速, 把握这三地的房地产企业发展可以对整个行业的发展有一个整体的认识。

关于企业盈利能力的研究, 国外开始的较早。Venkatraman 和 Ramanujam (1987) [1]认为企业盈利能力是一个包含三种绩效模式的多维度结构: 财务绩效; 商业绩效, 包括与增长和份额相关的部分, 以及与公司地位相关的部分; 组织效能, 包括与质量相关的部分, 以及与社会责任感相关的部分。Robert Kaplan 和 David Norton (1992) [2]提出的平衡记分卡提供了一个关注管理过程的框架, 其最大特点就在于不仅仅是各种财务指标和非财务指标的综合体, 而且通过因果关系将盈利能力评价的四个方面紧密地联系起来。Akhmedjonov 和 Balciizgi (2015) [3]研究了资本规模对土耳其银行盈利能力的影响, 在研究过程中用到了最小二乘等分析方法, 结果表明二者存在明显的影响关系。

国内学者侧重于利用各种方法建立盈利能力评价体系, 比如利用因子分析法、回归分析法提取指标的主要信息。种镇国、莫中杰(2010) [4]利用因子分析法对房地产上市公司 2009 年反映盈利能力的财务与非财务指标进行了实证分析并较为准确地分析出了房地产行业的盈利水平。王曼舒、刘晓芳(2013) [5]通过对 14 家上市银行在 2008~2010 年间的季度数据进行面板分析, 通过回归模型得出了净息差对企业的盈利能力影响较小的结论, 并指出银行要提高盈利能力除了做到降低成本外, 还要优化收益结构, 只有这样, 才能提高发展质量。也有利用层次分析法从多方面综合评价盈利能力的大小。王吉恒、王天舒(2013) [6]以万科三年的财务数据为例, 通过层次分析法建立企业盈利能力评价体系, 为盈利能力指标赋予权值, 进行了实证分析。

目前, 利用盈利能力分类模型进行分析的较少, 同时在研究范围方面, 多以某一家上市公司的财务数据为例进行分析或者对于某地区的部分企业进行分析, 很少有学者将京沪深三地的房地产上市公司放在一起做实证研究。为此, 本文在已有研究的基础上, 采用随机森林对京沪深三地 57 家房地产上市公司的盈利能力进行实证分析。首先, 对 57 家房地产上市公司盈利能力进行一个等级划分; 然后, 用随机森林训练一个盈利能力分类模型; 最后, 由盈利能力分析结果, 为房地产行业发展提供政策建议。

2. 理论阐述

随机森林[7]指的是利用多棵决策树对样本训练集进行训练并预测的一种分类器, 在机器学习中, 随机森林是一个包含多个决策树的分类器, 并且其输出的类别是由个别树输出的类别的众数决定。在预测的时候, 输入一个样本, 每个决策树都会给出一个预测结果, 根据所有决策树的预测结果的众数来确定样本最终的类别值。在生成随机森林的过程中, 每一棵决策树都是通过一个独立抽取的样本建立的, 决策树之间的相关性会影响最终误差, 因此要尽量保证决策树之间不相关。

随机森林构建的具体步骤如下[8]:

- 1) 输入参数: 原始训练集样本的个数 N , 变量的数目 M , 节点分类变量 m ;
- 2) 用自助抽样法有放回地随机抽取 k 个自助样本集, 根据自助样本集, 构建 k 棵决策树, 每次未被抽到的样本组成了 k 个袋外数据, 即简称 OOB;
- 3) 在训练的过程中, 自助样本集逐渐生成一棵棵决策树, 在每个节点处, 随机选取 M 个特征值中的 m 个 (m 小于 M), 随后从 m 个特征中选取一个特征进行分支, 此过程要保证结点不纯度最小, 直至每个节点不纯度达到最小;
- 4) 生成多个决策树分类器, 输入需要预测的元组, 每个决策树都给出了一个分类结果, 根据所有决策树的分类结果得出最终的类别。

随机森林的优点如下[9]:

可以处理高维度的数据, 并且不用进行指标的选择; 模型训练完毕之后, 也可以得出数据每个指标的重要程度; 模型的训练速度很快, 并且可以进行并行化操作; 最小化了决策树之间的相关性, 所以模型分类精度较高, 与决策树相比, 可以产生高精度的分类器; 在训练过程中可以很好地估计缺失值, 因此缺失值对分类模型的准确率不会产生很大影响。

随机森林的缺点如下[10]:

对于噪声较大的数据进行分类时, 随机森林可能会产生过拟合现象从而影响分类的结果; 对于低维度数据, 可能不能产生很好的分类。

3. 京沪深房地产上市公司盈利能力实证分析

3.1. 指标选取及数据获取

根据科学性、全面性、可得性和同向性的指标选取原则[6], 选取的指标包括: 每股收益、毛利率、营业利润率、成本费用率、净资产收益率、总资产利润率、总资产净利率。

本文所用到的所有房地产公司盈利能力指标数据均来自中信证券股票交易软件中 2017 年的财务指标数据。

3.2. 盈利能力分级

在对盈利能力进行分级的时候, 以往的方法忽略了异常值的影响, 当存在特别大或者特别小的指标值时, 会使分级结果不合理, 故本文利用分位数进行分级, 这样分级使每个级别的公司数量比较均衡,

避免了偏大和偏小的异常值的影响。首先对于各个指标进行分级, 每个指标被分为三类, 指标分级之后, 再根据 7 个指标对房地产公司盈利能力进行分级。对于每个指标, 提取 4 分位数, 以每股收益为例:

每股收益: 1/4 分位数 0.14, 3/4 分位数 0.38; 将每股收益小于 0.14 的称为每股收益为弱, 在 0.14-0.38 的为中, 在 0.38 以上的为强。同时, 对于每个等级赋予一个分数, 强: 5 分, 中: 3 分, 弱: 1 分。依次类推, 其余 6 个指标都予以相同的处理, 分界点如表 1 所示。

Table 1. Quartile of the indicator

表 1. 指标的 4 分位数

变量	毛利率	营业利润率	成本费用率	净资产收益率	总资产利润率	总资产净利率	每股收益
1/4 分位数	22.16	8.69	11.5	3.06	1.07	0.65	0.14
3/4 分位数	46.54	32.49	67.64	7.56	3.65	2.51	0.38

将 7 个指标进行等级划分之后, 计算得到每个公司的盈利能力得分, 如表 2 所示。

Table 2. Profitability score

表 2. 盈利能力得分

公司名称	盈利能力得分	公司名称	盈利能力得分	公司名称	盈利能力得分	公司名称	盈利能力得分
泛海控股	23	南山控股	15	中航地产	9	深深房 A	23
金融街	23	大龙地产	7	皇庭国际	27	中粮地产	23
绵石投资	33	香江控股	27	招商蛇口	23	华联控股	35
荣丰控股	9	华业资本	35	世联行	27	东旭蓝天	13
财信发展	23	万通地产	19	天地源	17	中洲控股	17
三湘印象	11	北京城建	15	中华企业	19	中房股份	7
中关村	11	首开股份	17	ST 匹凸	17	光大嘉宝	31
中国国贸	31	金地集团	23	光明地产	19	新黄浦	31
世纪星源	9	空港股份	15	ST 新梅	29	浦东金桥	33
万科 A	25	华丽家族	17	华远地产	21	万业企业	35
深振业 A	23	市北高新	13	上实发展	21	外高桥	15
深物业 A	29	绿地控股	17	城投控股	31	张江高科	27
沙河股份	7	华鑫股份	19	信达地产	15	电子城	27
陆家嘴	31	京能置业	23	世茂股份	23	上海临港	23
北辰实业	21						

根据每个公司的得分, 将盈利等级分为三类: 强、中、弱。最高得分 $\max = 35$, 最低得分 $\min = 7$, 计算盈利能力等级分界点[11]:

$$A_1 = \min + \frac{\max - \min}{3} * 1 = 16.33, \quad A_2 = \min + \frac{\max - \min}{3} * 2 = 25.67$$

将小于等于 16 的盈利等级定为弱, 在 17~25 之间的定为中, 大于 25 的定为强, 各等级的公司数量如表 3 所示。

Table 3. Number of companies with various profit levels

表 3. 各个盈利等级的公司数量

盈利等级	强	中	弱
数量	16	25	16

据此, 我们分地区统计不同盈利等级的公司数量, 见表 4。

Table 4. Number of companies with different profit levels in each region

表 4. 各地区不同盈利等级的公司数量

地区	盈利能力强	盈利能力中	盈利能力弱	总计
北京	3	8	7	18
上海	8	10	4	22
深圳	5	7	5	17

从等级划分结果可以看出: 在北京的 18 个房地产上市公司中, 盈利能力被判定为强的只有 3 个, 仅占北京地区房地产上市公司总数的 16.7%, 而盈利能力为弱的上市公司为 7 个, 占总数的 38.9%。可见与深圳上海相比, 北京的房地产上市公司盈利状况最差。三地中房地产上市公司盈利状况最好的地区为上海, 盈利能力被判定为强的占到了 36%, 而被判定为弱的仅占到了 18%。而深圳地区房地产公司的盈利情况介于北京和上海之间, 盈利能力被判定为强的与弱的都各占到了 29%。

近年来, 为了满足北京地区的人口增长, 北京地区的房地产过度开发, 价格持续攀升, 而随之带来的是投资减少、发展空间小、增长缓慢等一系列问题, 导致北京地区的房地产上市公司盈利状况不佳, 但是也有部分表现良好的公司。

上海地区的房地产就有很大的投资前景, 上海地区作为全球性的大都市经济、政治、文化协同发展, 各个方面的发展都比较全面。2016 年, 上海 GDP 总量达 2.75 万亿, 全国城市排名第一。此外, 第三产业的增加值占比 70% 以上, 产业结构合理。因此, 上海地区不仅经济经济发展迅速, 发展质量也有所保证。稳定并持续发展的经济环境更加有利于房地产行业的快速稳定发展, 因此上海地区的房地产上市公司的盈利能力也尤为突出。

深圳地区房地产上市公司盈利能力表现良好, 近几年深圳经济发展迅速, 房地产行业也在发展中稳步前行, 2017 年较 2016 年相比深圳地区商品房投资以及销售额有所下降, 但仍然在行业中处于领先地位。

3.3. 随机森林盈利能力分类模型

在确定了公司盈利等级以后, 利用样本数据建立随机森林盈利能力分类模型, 用于判断公司盈利等级。我们将原始数据分为训练集与测试集两部分, 在京沪深三地 57 个房地产上市公司中以其中 40 个作为训练集, 训练随机森林盈利能力分类模型, 其余 17 个为测试集, 用于测试随机森林盈利能力分类模型的精度。

在盈利能力分类模型的训练过程中, 有两个参数要首先确定好: 节点分类指标个数、随机森林中决策树的个数。以上两个参数是构建随机森林盈利能力分类模型中的两个重要参数, 也是决定随机森林盈利能力分类模型预测能力的两个重要参数。

首先, 确定利用训练集训练随机森林盈利能力分类模型时决策树节点分类指标个数。分类指标一共有 7 个, 从 1 个指标开始, 逐个测试, 找出误差最小的指标个数。利用 R 软件进行操作, 决策树用系统默认值, 得到预选指标个数与误差值关系, 如表 5 所示。从表 5 可以看出, 节点指标数为 2 时, 误差最小, 因此取节点分类指标个数为 2。

Table 5. The relationship between the number of classification indicators and the error

表 5. 分类指标个数与误差的关系

预选指标个数	1	2	3	4	5	6	7
误差	0.2982	0.2632	0.3158	0.2807	0.3158	0.2807	0.2807

然后, 需要确定第二个重要参数: 随机森林中决策树个数。使用 R 软件的绘图函数绘制决策树个数与分类误差的关系图, 如图 1 所示。图 1 中列出了三个等级“强”(good)、“中”(mid)、“弱”(bad)各自在预测的时候, 决策树数量与误差的关系, 同时有总的“预测结果”(total)。可以看出, 决策树在 300 棵以前, 误差非常不稳定, 在 300 棵以后所有预测误差均趋于稳定, 而且误差基本降到最低。

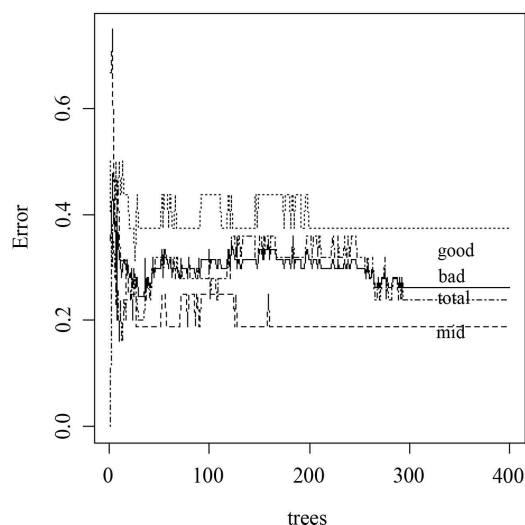


Figure 1. The relationship between the number of decision trees and the error
图 1. 决策树数量与误差的关系

基于以上分析, 取节点分类指标数为 2、决策树数为 300 棵构建随机森林盈利能力分类模型。然后, 利用测试集的 17 个京沪深房地产上市公司盈利能力指标数据进行分类准确率测试, 最终测试结果如表 6 所示。

Table 6. Test results
表 6. 测试结果

	强	中	弱	误差
强	5	0	0	0
中	0	8	0	0
弱	0	1	3	0.25

从测试结果可以看出, 盈利能力为强与中的公司被很好预测出来, 没有误差, 盈利能力为弱的四个中有一个被错误地预测为弱, 17 个测试样本预测准确率 94%, 说明所构建模型准确率较高, 能可靠预测京沪深房地产上市公司的盈利能力。

3.4. 重要性分析

由于分类指标有 7 个, 目前并不知道那个指标对于盈利能力等级划分重要。所以, 利用 R 软件中的随机森林包确定影响盈利能力的重要指标, 重要性度量见表 7。

表 7 中, 精度平均减少值与节点不纯度平均减少值都是测量分类指标重要性的重要指标, 两个值越大, 代表分类指标越重要, 即对于京沪深房地产上市公司盈利能力等级划分的影响也就越大。根据各个指标的精度平均减少值与节点不纯度平均减少值, 对盈利能力影响大小从高到低依次为: 净资产收益率、营业利润率、成本费用率、总资产净利率、总资产利润率、毛利率、每股收益。

Table 7. Indicator importance measure
表 7. 指标重要性度量

变量	精度平均减少值	节点不纯度平均减少值
每股收益	1.932725	2.176060
毛利率	2.014887	2.303949
营业利润率	8.687835	4.539417
成本费用率	8.485846	4.559157
净资产收益率	10.47783	5.392063
总净资产利润率	2.360471	2.949022
总资产净利率	6.940584	3.710166

净资产收益率、营业利润率与成本费用率这三个指标对于京沪深三地房地产上市公司处于什么样的盈利等级有着重要影响，对于房地产上市公司盈利能力的划分占据主导地位。同时，提高利润总额、净利润，降低营业成本对于提高房地产上市公司的盈利能力至关重要。因此，在实际操作中就要求企业完善管理运营机制，提高生产效率，采用科学合理的营销方案，使每一单位成本所带来的收益最大化，从而提高企业盈利能力。

4. 政策建议

基于本文的结论，并结合目前房地产上市公司的盈利能力现状，对房地产行业发展提出如下建议：

1) 坚持宏观调控方向

目前，在国家宏观政策调控的大背景下，房地产行业发展下行，房地产投资放缓，但是房地产作为刚性需求，政策因素稍有变动，房价仍然面临高速增长的趋势。所以，国家政府有关部门对于房地产行业的调控政策必须坚持实行，同时还要保证经济的持续平稳发展，从而有效抑制房地产投资过热，控制房价，杜绝炒房。

2) 加快实施房地产税收制度改革

房地产的税收制度改革对于房地产市场的调节非常重要。首先，要以降低政府干预为原则，以市场化为导向，在此基础上建立起完善的税收制度。推进房产税的实施，从而使房地产市场达到一种供需平衡的状态，通过税收的调节控制房价，使房地产市场能够健康发展。

3) 上市公司进驻三四线城市，开发有潜力的地产

目前，京沪深三地房地产上市公司开发房地产以一线城市以及发达的二线城市为主，这些发达地区的房产已经被过度开发，发展空间小，要注重三四线城市的房地产建设，这些地区多数处于高速发展阶段，发展基数小，要将这些地区的房地产潜力开发出来，不但能在很大程度上提高公司的盈利能力，促进公司发展，而且最大化房产公司的盈利潜力，从而很大程度上促进整个行业的发展。

致 谢

本论文的完成离不开各位老师、同学和朋友的关心与帮助。在此要感谢各位老师针对论文所提出的宝贵意见，还要感谢同门的师兄师妹们，在科研过程中给我以许多鼓励和帮助。回想整个论文的写作过程，虽有不易，却让我除却浮躁，经历了思考和启示，也更加深切地体会了所学专业的精髓和意义，因此倍感珍惜。

基金项目

山东科技大学研究生导师指导能力提升计划立项项目(KDYC17018)，山东科技大学专业学位研究生

教学案例库建设项目(SKDYAL17010)。

参考文献

- [1] Ramanujam, V. and Venkatraman, N. (1987) Planning System Characteristics and Planning Effectiveness. *Strategic Management Journal*, **8**, 453-468. <https://doi.org/10.1002/smj.4250080505>
- [2] Kaplan, R.S. and Norton, D.P. (1992) The Balanced Scorecard—Measures that Drive Performance. 187-204.
- [3] Akhmedjonov, A. and Izgi, B.B. (2015) If Bank Capital Matters, Then How? The Effect of Bank Capital on Profitability of Turkish Banks during the Recent Financial Crisis. *Macroeconomics & Finance in Emerging Market Economies*, **8**, 160-166. <https://doi.org/10.1080/17520843.2014.940988>
- [4] 种镇国, 莫中杰. 基于因子分析的中国房地产上市公司绩效评价[J]. 经济研究导刊, 2010(36): 74-75.
- [5] 王曼舒, 刘晓芳. 商业银行收入结构对盈利能力的影响研究——基于中国 14 家上市银行面板数据的分析[J]. 南开管理评论, 2013, 16(2): 143-149.
- [6] 王吉恒, 王天舒. 基于企业层次分析法的盈利能力分析——以万科企业股份有限公司为例[J]. 哈尔滨商业大学学报(社会科学版), 2013(4): 3-10.
- [7] Han Jiawei. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2012: 249-250.
- [8] 黄文, 王正林. 数据挖掘: R 语言实战[M]. 北京: 中国工信出版集团, 2014.
- [9] Liaw, A. and Wiener, M. (2002) Classification and Regression by Random Forest. *R News*, **23**, 18-22. <https://doi.org/10.1057/9780230509993>
- [10] Pavey, T.G. and Gilson, N.D. (2016) Field Evaluation of a Random Forest Activity Classifier for Wrist-Worn Accelerometer Data. *Journal of Science & Medicine in Sport*, **20**, 75-80. <https://doi.org/10.1016/j.jsams.2016.06.003>
- [11] 刘慧. 基于数据挖掘的电力上市公司盈利能力分析[D]: [硕士学位论文]. 北京: 华北电力大学, 2008: 35-36.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-0967, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: fin@hanspub.org