

# 基于稀疏主成分的房地产市场发展分析

——以北京为例

刘超,吴丹丹,齐嘉悦

(北京航空航天大学数学与系统科学学院 数学、信息与行为教育部重点实验室,北京 10019)

**摘要:**采用稀疏主成分分析,对房地产市场发展的综合评价进行了探讨,并且以北京市房地产发展为例进行了分析。研究表明:(1)相比于主成分分析,稀疏主成分分析法更适合进行房地产市场综合评价。(2)基于该方法得到的结果基本符合市场的实际情况,更利于宏观调控,以保证房地产市场的健康有序发展。

**关键词:**房地产市场;综合评价;稀疏主成分

**中图分类号:** F293.3 **文献标志码:** A **文章编号:** 1673-291X(2012)30-0158-03

近年来,随着房地产市场的繁荣发展,房价持续走高,往往成为地区经济发展的重要支柱产业,这在北京尤其显著。越来越多的人开始关注房地产发展的影响因素,房地产业内部结构的合理性等等问题。而对房地产市场发展进行综合评价,不仅可以把握房地产市场发展状况,而且也可以为政府的宏观调控提供依据。然而,由于影响房地产市场发展的因素众多,而且都具有很强的相关性,因此在进行实证研究的时候可能会遇到两个方面的问题:一是所考虑的因素不全面,不能涵盖造成房地产市场长期波动的所有因素;二是如果考虑指标过多,这些指标之间同时存在相关性,也会增加对问题分析的复杂性。

本文采用稀疏主成分分析方法,对房地产市场发展综合评价进行了探讨,并且以北京市房地产发展1990—2010年的相关数据,分析了北京房地产市场的整体发展情况,并根据稀疏主成分表达式分析了影响房地产发展水平变动的因素。

研究表明,使用稀疏主成分分析法进行房地产市场综合评价是合适的。而且基于该方法对房地产市场发展状况进行分析的结果基本符合市场的实际情况。政府可以根据分析出的影响房地产市场发展因素进行有效的宏观调控,以保证房地产市场的健康有序发展。

## 一、模型与方法

由于影响房地产市场发展的因素众多,而且都具有很强

的相关性,而基于主成分分析(Principal Component Analysis,简称PCA)的综合评价能够有效解决以上问题,其优点在于:

(1)它能基于数据分析客观的确定权重,而且反映指标之间的内在结构关系;(2)它能有效地剔除不相关指标的影响;(3)它得出的综合指标(主成分)之间相互独立,不仅使指标数量大大减少,节省了指标收集的成本。

然而,主成分分析存在一些不足之处。其中典型的问题是,很多时候主成分难以解释,这使得主成分处理某些实际问题时,比如用该方法做综合评价,往往只能给出数据的统计解释,而很难给出对应的实际意义上的解释。为了使主成分易于解释,统计学家进行了深入研究。其中代表性的方法是使主成分的载荷产生稀疏解,从而增强可解释性。

典型的算法是Zou(2006)提出的稀疏主成分分析(Sparse Principal Component Analysis,简称SPCA),主要思想是:寻找一个已经做过PCA的回归优化框架。而且,回归框架应该通过应用Lasso(elastic net)惩罚而允许直接修正,这样就可以得到稀疏的得分。这样SPCA不仅达到了有效降维的目的,而且很大程度上简化了主成分的解释,使提取的主成分往往能对应于某些实际的含义,便于后续应用。

理论研究和仿真分析表明:SPCA是在PCA的基础上,牺牲主成分的部分方差贡献率,更进一步的对载荷进行稀疏,从而更清晰的揭示出数据的内在结构,显示出比PCA更

收稿日期:2012-06-18

基金项目:国家社会科学基金项目资助(11CTJ001);教育部人文社会科学基金项目基金资助(09YJC910001);北京航空航天大学基本科研业务费项目唯实人才培养基金资助(YWF-11-03-Q-001);北京航空航天大学大学生创新创业训练计划项目资助(111000689)

作者简介:刘超(1977-),男,湖北武汉人,讲师,博士,从事应用统计学研究;吴丹丹(1986-),女,河南驻马店人,硕士研究生,从事应用统计学研究;齐嘉悦(1992-),女,陕西西安人,本科生,从事数学研究。

强大的优势。而且,与PCA方法相比较,SPCA中稀疏的载荷可以有效避免指标重叠信息,从而更好的降维,使得评价结果更加真实可靠。

本文的研究思路是通过SPCA这种改进的主成分方法,对各种影响因素进行综合整体考虑,以反映房地产市场的整体发展情况,并根据主成分表达式分析其变动原因。

## 二、变量与数据

### (一) 变量

一个地区的房地产市场发展往往受到多方面因素的影响。主要包括:(1)地区规模与经济发展水平。通常,地区规模越大,对房地产的需求也越大,因为更多的购房者会带动房地产业。另外,房地产市场的发展与地区经济发展水平具有很强的相关关系。(2)房地产市场自身的活跃程度。这主要体现在房地产企业的数量,房地产建筑面积的变动情况。(3)对房地产的需求。主要体现为购房者的收入水平,即对房地产的有效支付能力。(4)金融业等其他相关条件。金融业提供的资金支持对房地产市场发展有着重要作用。其中,中长期贷款与房地产开发有密切关系,可以反映金融业对房地产市场发展的影响程度。

基于以上的讨论,本文选择以下指标:

表2 稀疏主成分和主成分分析的方差贡献率

	SPCA		PCA	
	SF1	SF2	F1	F2
方差的贡献率 (%)	87.7	3.8	89.5	7.7
累积方差贡献率 (%)	87.7	91.5	89.5	97.3
主成分的权重 (%)	95.8	4.2	92.0	8.0

表3 稀疏主成分和主成分分析的载荷表

变量名		SPCA		PCA	
		SF1	SF2	F1	F2
GDP (亿元)	x1	0.280	0.244	0.280	0.288
常住人口 (万人)	x2	0.285	0	0.284	0.188
房地产开发投资 (亿元)	x3	0.288	0	0.288	0.182
住宅开发投资 (亿元)	x4	0.287	0	0.287	0.114
房地产开发企业个数 (个)	x5	0.287	0	0.287	0.005
商品房施工面积 (万平方米)	x6	0.286	0	0.287	-0.182
商品房竣工面积 (万平方米)	x7	0.273	-0.043	0.273	-0.343
住宅竣工面积 (万平方米)	x8	0.267	-0.339	0.268	-0.377
当年商品房新开工面积 (万平方米)	x9	0.255	-0.305	0.256	-0.401
商品房销售面积 (万平方米)	x10	0.267	-0.356	0.267	-0.357
商品房销售额 (亿元)	x11	0.278	0	0.278	0.221
人均可支配收入 (元)	x12	0.286	0	0.285	0.191
中长期贷款余额 (万元)	x13	0.265	0.777	0.264	0.411

可以看出,SPCA提取具有较强的综合作用。前两个成分的累积贡献率达到91.5%,基本可以认为代表了所选指标的绝大部分信息。因此对这两个稀疏主成分进行分析,可以较好的反映北京市房地产市场的整体发展情况以及历史变化趋势。而且,影响房地产市场发展的因素作用较平均。其中,SPCA的第一稀疏

表1 指标体系

指标	符号
GDP (亿元)	x1
常住人口 (万人)	x2
房地产开发投资 (亿元)	x3
住宅开发投资 (亿元)	x4
房地产开发企业个数 (个)	x5
商品房施工面积 (万平方米)	x6
商品房竣工面积 (万平方米)	x7
住宅竣工面积 (万平方米)	x8
当年商品房新开工面积 (万平方米)	x9
商品房销售面积 (万平方米)	x10
商品房销售额 (亿元)	x11
人均可支配收入 (元)	x12
中长期贷款余额 (亿元)	x13

### (二) 样本和数据来源

根据影响房地产市场发展的主要因素,考虑到数据的可得性和完整性,选择数据时间范围为1990—2010年,均来源于《北京统计年鉴》(限于篇幅,这里没有给出原始数据)。

## 三、实证分析

根据样本的相关系数矩阵可知,指标彼此之间相关性非常强。稀疏主成分分析的结果(见表2和表3)。为了对比,同时给出了主成分分析的结果。

主成分的载荷系数非常接近,一种可能的解释是北京市房地产市场在不同的时期,受到某一因素的影响较多,而从长期来看这些因素的影响比较平均。从第二稀疏主成分可以看出,中长期贷款余额对房地产市场发展的影响更重要一些,这与目前中国房地产市场发展与金融行业紧密联系的现状是吻合的。

值得注意的是,PCA 的前两个主成分的累积贡献率达到 97.3%,高于 SPCA 的 91.5%,但是,SPCA 的结果更易于解释,更准确地反映了数据的内在结构,因为 PCA 的第二主成分难以明确看出这种现象中长期贷款余额对房地产市场发展趋势的突出影响。

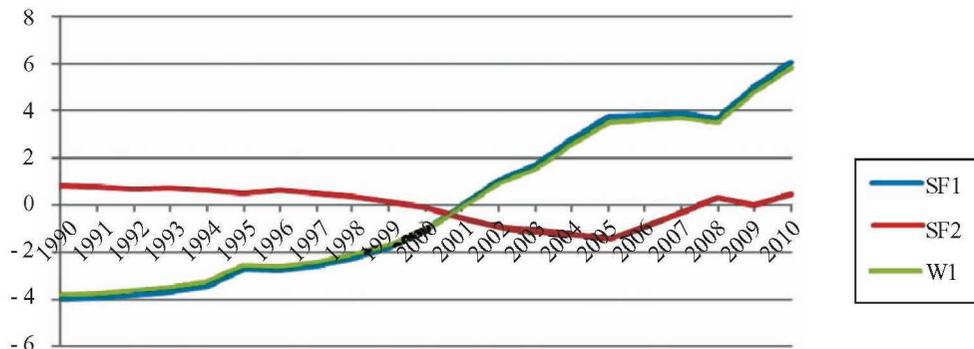


图 1 房地产市场发展趋势图

从图 1 可以看出,北京市房地产市场基本上保持了稳定发展,而且呈现出快速增长趋势。同时,综合得分和第一稀疏主成分得分的趋势基本一致,这表明各种影响因素对房地产市场的影响从长期来看比较平均。但是,综合得分和第一稀疏主成分得分的趋势是不完全一致,这是由于第二稀疏主成分发挥了作用。这可以从上页表 3 中看出,去掉载荷为 0 的部分,可以发现,GDP 和中长期贷款余额对综合得分的影响是正向的,其中,中长期贷款余额的影响更大。而商品房竣工面积、住宅竣工面积、当年商品房新开工面积、商品房销售面积对综合得分的影响是负向的。这个结论可以从综合得分与 13 个(经过标准化处理后的)指标的对比中得到验证。这表明得到的综合得分确实反映了房地产市场的发展状况,而且,它的变化趋势也可以揭示出其他各种经济指标的发展趋势,这主要是通过稀疏载荷得以实现的。

参考文献:

[1] Zou,H.,Hastie,T.,and Tibshirani,R.(2006),Sparse principal component analysis[J].Journal of Computational and Graphical Statistics, 15,265-286.  
 [2] Y.Wang and Qiang Wu.,Sparse PCA by iterative elimination algorithm[J].Advances in Computational Mathematics,2012:137-151.  
 [3] 钱争鸣,陈伟彦.中国工业经济效益指标评价与主成分分析的实证研究[J].统计研究,1999,(7):49-52.  
 [4] 刘超,吴丹丹,杨考.一种新的高维数据降维方法[J].统计与咨询,2012,(4):16-17.  
 [5] 北京市统计局.北京统计年鉴:1991—2011[K].北京:中国统计出版社,1991—2011.

The Real Estate Market Comprehensive Evaluation based on Sparse Principal Component Analysis  
 ——An Example of Beijing

LIU Chao, WU Dan-dan, QI Jia-yue

(School of Mathematics and Systems Science and LMIB of the Ministry of Education, Beihang University, Beijing 100191, China)

**Abstract:** Based on the sparse principal component analysis, this paper has discussed the real estate market comprehensive evaluation, and the Beijing real estate development as an example is analyzed. The research results show that (1) compared to the principal component analysis, sparse principal component analysis is more suitable for the real estate market comprehensive evaluation. (2) the results based on the method are basically accordant with the practical situation, this is more conducive to the macroeconomic regulation and control to ensure the healthy and orderly development of the real estate market.

**Key words:** real estate market; comprehensive evaluation; sparse principal component analysis

[责任编辑 陈凤雪]