



经济理论与经济管理

工作论文系列

Working Paper Series

城市间房价分化的影响因素研究

——基于机器学习方法

陈小亮 陈 衍 王兆瑞 肖争艳

ETBMWP2024013

- * 本刊编辑部推出工作论文项目，将“拟用稿”而尚未发表的稿件，以工作论文的方式在官网呈现，旨在及时传播学术成果，传递学术动态。
本刊所展示的工作论文，与正式刊发版可能会存在差异。如若工作论文被发现存在问题，则仍有被退稿的可能。各位读者如有任何问题，请及时联系本刊编辑部，期待与您共同努力、改进完善。
联系人：李老师；联系电话：010-62511022

城市间房价分化的影响因素研究^{*}

——基于机器学习方法

陈小亮 陈 衍 王兆瑞 肖争艳

[提 要] 2013年以来,中国一二三线城市房价增速呈现明显的分化走势,“因城施策”逐渐成为调控房价的主基调。要想“因城施策”,需要准确识别不同城市房价分化的主要影响因素。本文综合使用 Xgboost 等机器学习方法以及 SHAP 值解释性方法,基于 2009—2019 年全国 70 个大中城市的面板数据,测算并分析了一二三线城市多轮房价分化背后的主要驱动因素。研究结果表明:第一,预期因素在多轮房价分化期间都起到了重要的驱动作用;第二,货币政策本身并不是房价分化的主要驱动因素,但是可以通过影响预期进而导致房价分化;第三,需求因素和供给因素本身也不是房价分化的主要驱动因素,但是也可以通过影响预期进而导致房价分化。

有鉴于此,不同城市“因城施策”的关键在于,基于自身情况,从需求端或供给端出发,稳定公众对房价走势的预期。就一线城市而言,由于外来人口持续流入,而土地供给则相对偏少,因此处于住房供给相对紧缺的状态,这也使得一线城市居民预期房价呈现上涨趋势。要想稳定房价预期,应考虑在控制土地成本的前提下适当增加住房供给,以削弱住房供不应求所带来的房价上涨预期。此外,还要不断完善住房租赁市场,以减轻住房买卖市场的压力,进一步削弱一线城市的房价上涨预期。就二三线城市而言,由于这些城市往往面临人口流出的境况,居民预期房价会下跌。为了稳定房价预期,要进一步完善城市基础设施,提高教育、医疗和环境质量,从而改善居民购房需求。与此同时,住房供给相对充裕的二三线城市还要制定更加科学合理的人才引进计划,吸引人口流入,从而进一步稳定房价预期。

本文的边际贡献体现在两个方面。第一,虽然有一些文献尝试探寻不同城市房价分化的主要原因,但主要使用面板回归模型等传统方法,能够考察的影响因素个数和所识别的非线性关系相对有限。机器学习方法较好地克服了传统方法的不足之处,能够更准确地识别房价分化的主要影响因素。第二,虽然近年来中央一直呼吁“因城施策”,但是学术界对不同城市房价分化原因的研究尚不明确,已有文献主要研究某个因素对房价分化的影响是否显著,无法明确对比不同因素的影响究竟孰大孰小。本文使用 SHAP 值解释性方法可以直接对各个影响因素的相对重要性进行排序,从而识别房价分化的主要影响因素,为房地产调控“因城施策”给出更加明确的决策参考。

[关键词] 房价上涨;房价分化;房价预期;因城施策;机器学习方法

^{*} 陈小亮,中国社会科学院经济研究所,邮政编码:100836,电子信箱:chenxiaoliang2200@126.com;陈衍,中国人民大学统计学院;王兆瑞,中国人民大学经济学院;肖争艳(通讯作者),中国人民大学应用统计学研究中心、中国人民大学统计学院。本文得到国家自然科学基金面上项目(72073141)和国家自然科学基金专项项目(72141306)的资助。感谢匿名审稿专家的建设性意见,作者已经做了相应修改,本文文责自负。

一、引言

1998年住房市场改革以来，我国主要城市的房价在大部分时间都处于上涨状态，即便在2008年国际金融危机之后的十余年间，也经历了多轮房价上涨。不仅如此，2013年以来不同城市的房价走势呈现明显分化态势，并且带来了一些不利影响。一方面，面对一二三线城市房价频繁分化的情形，货币政策陷入两难境地。另一方面，加剧了部分城市房价大幅波动的风险。为了应对房价分化局面，近年来中央提出了“因城施策”的新思路。不过，目前尚缺乏“因城施策”的具体指导思路和决策参考。究其原因，已有研究尚未清晰地识别出不同城市房价上涨以及房价分化的主要影响因素，也就难以对症下药来“因城施策”调控房价。正因如此，探寻不同城市房价分化的影响因素具有重要意义，有助于更好地落实“因城施策”，从而促进不同城市房地产市场平稳发展。

面临不同城市房价分化的局面，已经有部分文献尝试识别背后的影响因素（详见文献综述部分）。不过，已有文献大都使用动态面板回归或者面板固定效应模型，而面板回归模型能够考察的变量个数有限，并且难以分析各个影响因素对房价的非线性影响，即便是王先柱和杨义武（2015）所使用的门槛回归模型，也只能较为有限地考察各个影响因素对房价的非线性影响。此外，不同时间段房价分化的程度和特征存在显著差异，但是传统实证方法难以详细对比分析不同时间段房价分化及其影响因素的差异。

与面板回归模型等传统实证方法相比，机器学习方法既能够更全面地考察影响房价的多个因素，又能深入挖掘各个影响因素对房价的非线性影响。进一步地，将机器学习方法结合SHAP值解释性方法（SHapley Additive exPlanations），可以计算出每一轮次房价分化期间各个因素对房价的影响，进而对不同因素的重要性进行排序。由此，即可识别出每一轮次房价分化的主要驱动因素。遵循这一思路，本文基于2009—2019年国家统计局重点关注的70个大中城市的面板数据，构建了包含需求因素、供给因素、宏观政策因素和预期因素四大类共计26个指标的基准指标体系，综合使用Xgboost等机器学习方法以及SHAP值解释性方法，测算并剖析了近年来中国一二三线城市房价分化的主要影响因素。

本文的边际贡献体现在两个方面。第一，本文较早地将机器学习方法应用到房价分化影响因素识别的研究当中，有助于弥补面板回归模型等传统方法的不足之处。虽然已经有一些文献尝试探寻不同城市房价分化的主要原因，但主要使用面板回归模型等传统方法，能够考察的影响因素个数和所识别的非线性关系相对有限。机器学习方法较好地克服了传统方法的不足之处，从而能够更准确地识别房价分化的主要影响因素。第二，本文直接对比了不同因素对房价分化的相对重要性，找到了房价分化的主要影响因素，从而为房地产调控“因城施策”提供了决策参考。虽然2016年以来中央一直呼吁“因城施策”，但是学术界对不同城市房价分化原因的研究尚不明确，已有文献主要研究某个因素对房价分化的影响是否显著，无法明确对比不同因素的影响究竟孰大孰小。本文使用SHAP值解释性方法可以直接对各个影响因素的相对重要性进行排序，从而识别房价分化的主要影响因素，为房地产调控“因城施策”给出更加清晰、明确的决策参考。

二、相关文献述评

房价分化的本质是不同城市房价上涨的趋势不同，因此分析房价分化的影响因素时，首先需

要梳理研究房价上涨影响因素的相关文献。国内外大量文献对影响房价上涨的因素进行了研究,总体而言主要关注了四类因素的影响。一是,需求因素。人口规模和人口密度的提升、居民收入水平的提升、基础设施的完善、教育和医疗水平的提高等因素都会增加居民的购房需求,进而推高城市房价(Takáts, 2012; 邹瑾等, 2015)。二是,供给因素。土地供给收紧和房屋建造成本上升会提高住房供给成本,进而促使房价上涨(Chow & Niu, 2015; Glaeser *et al.*, 2017)。三是,宏观政策因素。信贷放松和利率下降等宽松货币政策操作通常会推升房价(Jarociński & Smets, 2008)。地方政府面临的财政不平衡对于城镇房价的持续上涨也起到重要的推动作用,土地财政规模是连接两者的中间变量(王猛等, 2013; 宫汝凯, 2015)。四是,预期因素。居民对房价上涨预期的提升和乐观情绪的上涨,会对房价产生显著的促进作用(Abildgren *et al.*, 2019; 高波等, 2014)。

伴随着不同城市房价上涨态势的持续分化,越来越多的文献开始直接关注房价分化的成因,而且相关文献关注的影响因素基本上也可以归入上述四类因素。就需求因素而言,陆铭等(2014)研究发现,人口流动是导致城市房价分化的重要原因,人口净流入越多的城市房价增速也更快;陈淑云和唐将伟(2017)发现,公共服务的不均等化是导致城市房价分化的重要因素。就供给因素而言,韩立彬和陆铭(2018)通过对比不同城市土地供给政策研究发现,不同城市房价分化的根源在于土地供给在空间上与土地需求不匹配;刘诚和杨继东(2019)等发现,受禀赋条件的影响,不同城市会采取不同的土地供给策略,供地策略的差异会对不同城市房价产生异质性影响,从而导致房价分化。就宏观政策因素而言,涂红等(2018)基于2005—2016年间70个大中城市的数据研究发现,货币政策对房价的影响在不同省份的城市间存在显著差异,但是在同一省份内的城市间差异较小;倪鹏飞(2019)进一步指出,货币政策对房价分化的影响源于房地产市场的供需错配。就预期因素而言,王拉娣和安勇(2016)利用全国34个大中城市2000—2013年的面板数据研究表明,预期对城市房价的影响存在显著的地区差异。上述文献在分析房价上涨和不同城市房价分化的影响因素时,主要使用的是两类传统计量方法。一是,面板回归模型,主要包括面板固定效应模型、动态面板回归模型、门槛回归模型、面板Probit模型、面板协整模型等。二是,VAR类模型,主要包括传统VAR模型、SVAR模型、BVAR模型、TVP-VAR模型等。

虽然已有文献对房价上涨和不同城市房价分化的影响因素进行了大量研究,但是仍然存在一些不足和有待改进的地方。一是,已有文献在使用传统方法研究房价上涨和房价分化的影响因素时,主要通过回归系数显著性来判断某一因素对房价上涨是否有显著影响,而难以比较不同因素的影响究竟孰大孰小。这不仅难以有效辨别影响房价的主要因素,而且还会给宏观政策的制定带来一定困难。二是,导致房价分化的因素包括需求因素、供给因素、宏观政策因素和预期因素等多类因素,每一类因素下面又包含很多个具体指标,而且这些因素对房价的影响可能存在复杂的非线性关系,但是面板回归模型和VAR类模型等传统方法所能涵盖的变量个数有限而且对于变量间非线性关系的挖掘较为有限。

机器学习方法可以较好地弥补传统方法的上述不足。首先,机器学习方法通过在参数估计中加入正则化以及对自变量进行基函数变换,可以同时考虑多个自变量对被因变量的非线性影响。其次,机器学习方法不需要事先设定自变量与因变量间关系的函数形式,而是利用数据找到最优的函数形式,这避免了模型误设的问题。再次,由于机器学习模型的训练过程本质上就是最优化预测结果的过程,这使得机器学习方法具有较好的样本外预测性能。有鉴于此,本文将综合使用机器学习方法和SHAP值解释性方法,研究不同城市房价分化的主要原因,并在此基础上为房地产调控“因城施策”提供决策参考。

三、方法选取与模型构建

本文将综合使用 Xgboost 等多种机器学习方法和 SHAP 值解释性方法，分析城市房价分化的影响因素。本文将基于适用于面板数据的交叉验证方法和网格搜索手段，从多种机器学习方法中筛选出具有最优样本外预测能力的方法。此外，本文选择使用 SHAP 值解释性方法的原因是该方法不仅克服了变量替换法和 LIME 等解释性方法在统计性质上的一些不足，还能在每一个分样本上对变量重要性分别进行排序，同时计算出变量重要性大小和影响方向 (Aas *et al.*, 2021)。将 Xgboost 等机器学习方法和 SHAP 值解释性方法结合起来，能够更准确地识别不同轮次房价分化的主要影响因素。

(一) 机器学习方法和交叉验证方法

为了客观揭示不同城市房价分化的影响因素，尽可能地避免模型约束对结果造成的影响，本文将多种机器学习方法作为潜在分析方法，其中既包括线性的惩罚回归，也包括 SVR 和集成树模型等非线性方法，再根据样本外预测效果挑选最优的样本外预测方法作为基准的分析方法。

非线性机器学习方法方面，本文选择了以 Xgboost 和随机森林 (Random Forest) 为代表的集成树模型，以及支持向量回归 (SVR) 和 K 近邻回归 (KNN)，刻画各线城市房价增速与影响因素间的非线性关系。随机森林和 Xgboost 方法分别使用重抽样的 Bagging 集成法和二阶泰勒优化后的 Boosting 集成法，将多个回归树模型集成起来，既发挥了单棵回归树灵活训练的优势，又弥补了单棵回归树过拟合的不足。SVR 方法则融合了非线性核函数和正则化的手段，以此充分地刻画变量的非线性关系。而 KNN 方法则是基于样本距离的非线性方法，不涉及模型的训练，通过直接加权近邻样本中得到相应预测结果。

线性机器学习方法方面，本文选择了 SCAD (Smoothly Clipped Absolute Deviation) 和 Lasso 方法，二者都是在线性回归基础上加入惩罚项。Lasso 采用的是具有稀疏性的 l_1 正则化，SCAD 则采用具有 Oracle 一致性的平滑惩罚项。

为了筛选出适用于本文的模型，本文使用交叉验证的框架来寻找多种机器学习方法的最优超参数，并比较它们各自最优模型样本外的预测性能。由于本文使用的面板数据在时间和截面维度上存在相关性，无法直接使用 K 折交叉验证法 (Racine, 2000)。参考 Athey *et al.* (2019)、Babii *et al.* (2022) 的思路，本文采用垂直交叉验证法 (Vertical Cross-Validation) 对面板数据进行交叉验证，垂直交叉验证中训练集与测试集的样本在截面和时间上都不存在重复，从而能克服 K 折交叉验证在面板数据中的不足，具体做法是：①随机将 70 个城市划分为 K 组，每组有 s 座城市；②将第一组城市在最后 τ 个时期内的所有观测值，即在 $[T-\tau+1: T]$ 期间的观测值作为测试集，再将剩余组城市在 $[1: T-\tau]$ 期间的观测值作为训练集；③用训练集训练模型，评估模型在测试集上的预测效果；④重复第②步和第③步，遍历完所有 K 组城市。

关于截面维度参数 K ，参考文献常用的 5 折交叉验证做法 (Hastie *et al.*, 2009)，将 K 赋值为 5。关于交叉验证窗口长度 τ ，Athey *et al.* (2019) 对垂直交叉验证不同时间窗口长度的预测进行了评估，其结果显示交叉验证窗口长度 τ 在整体时间长度 T 的 9%~18% 这一范围内可以获得较为稳健的预测评估结果，本文所使用面板数据的整体时间长度为 44，这意味着交叉验证窗口长度在 4~8 这一范围都是合适的，本文选取这一范围内的中间值，即令 $\tau=6$ 。将参数赋值后，本文基于上述框架用网格搜索找出了 Xgboost 等六种方法的最优超参数，并将综合使用 R^2 、RMSE、MAE 三个指标评估这六种方法的预测效果，进而挑选出预测效果最优的方法。

(二) SHAP 值解释性方法

SHAP 值解释性方法的核心思想是将预测值分解为各个影响因素的贡献大小，即变量的 SHAP 值。SHAP 值同时反映了各个因素对被解释变量影响的大小和方向，在本文中指的是各个因素对房价的影响大小和方向。根据 SHAP 值解释性方法，机器学习模型 $f(\cdot)$ 在训练样本 x_{ct} 上的预测值 $\Delta h p_{ct} = \hat{f}(x_{ct})$ 可以分解为：

$$\Delta h p_{ct} = \varphi_{ct}^0 + \sum_{p=1}^P \varphi_{ct}^p \tag{1}$$

其中， φ_{ct}^0 为模型预测的参考值， φ_{ct}^p 为第 p 个房价影响因素的 SHAP 值。

关于参考值 φ_{ct}^0 ，本文将 0 作为模型预测房价增速的参考值。一是从 SHAP 值的意义来看，模型预测参考值 φ_{ct}^0 衡量了在不考虑所有因素对房价影响的条件下，第 t 时期第 c 座城市实际房价增速的大小 (Aas *et al.*, 2021)。二是本文需要同时研究一二三线城市房价的影响因素，但一二三线城市房价上涨的情况又各有不同，除了 0 以外难以找到统一的基准。就单个因素的 SHAP 值而言， $\varphi_{ct}^p > 0$ 说明第 p 个因素在训练样本 x_{ct} 上对房价有着拉动作用， $\varphi_{ct}^p < 0$ 则相反。 φ_{ct}^p 的绝对值 $|\varphi_{ct}^p|$ 越大，说明该变量对房价增速的影响越重要。

本文参考 Štrumbelj & Kononenko (2014) 的做法对每个因素的 SHAP 值进行测算。假定 Ω 为由 P 个因素组成的变量全集， U 为不包含第 p 个因素的变量集，由此根据训练样本 x_{ct} 和 U 、 \bar{U} 两个集合构造出用于模拟计算的样本 x_{ct}^U 。其中样本 x_{ct}^U 中的观测点按照变量是否在 U 中被分成两部分，变量集 U 中影响因素的观测值保持不变，而 \bar{U} 中因素的样本观测值则是从样本 x_{ct} 外的训练样本中随机抽取而得。所以 $f(x_{ct}^{U \setminus \{p\}}) - f(x_{ct}^U)$ 刻画了样本 x_{ct} 中第 p 个因素的观测值被随机替换后，样本 x_{ct} 的预测结果的变化。通过多次重复以上计算过程，可以近似地模拟出剔除第 p 个因素后样本预测值的变化，估计出第 p 个因素对房价增速预测值 $\Delta h p_{ct}$ 的贡献。由于全集 Ω 中抽取 $|U|$ 个不包含 $\{p\}$ 的因素，一共可以产生 $\frac{|\Omega|!}{|U|! (|\Omega| - |U| - 1)!}$ 种组合，所以考虑所有组合加权下第 p 个因素的 SHAP 值计算为：

$$\varphi_{ct}^p = \sum_{U \subseteq \Omega \setminus \{p\}} \frac{|U|! (|\Omega| - |U| - 1)!}{|\Omega|!} \cdot \frac{1}{N} \sum_{n=1}^N (f(x_{ct,n}^{U \setminus \{p\}}) - f(x_{ct,n}^U)) \tag{2}$$

然而以上的抽样过程需假设影响因素之间相互独立，这极大地限制了方法的应用范围。对此，Aas *et al.* (2021) 提出在进行模拟抽样时，将高斯核函数作为抽样样本与样本 x_{ct} 之间相似度的度量，并以此为权重加权各个模拟样本的结果，进而充分利用所有样本信息将不同变量相关性考虑在内。所以本文将采用 Aas *et al.* (2021) 改进后的 SHAP 值解释性方法作为机器学习模型的解释性方法。

(三) 研究思路与测算过程

本文结合 Xgboost 等机器学习方法和改进后的 SHAP 值解释性方法，计算不同时期一二三线城市房价上涨的主要影响因素，进而探寻每一轮房价分化背后的主要驱动因素。详细研究思路分为四步：

1. 假设一二三线城市集合分别为 C_1 、 C_2 、 C_3 ；根据一二三线城市平均房价增速是否大于 0，划分出房价上涨的区间集合 T_{C_1} 、 T_{C_2} 、 T_{C_3} 。基于垂直交叉验证筛选出样本外预测最优机器学习方法，再使用所有城市样本重新训练模型，确保模型充分利用所有观测信息。

2. 使用 Aas *et al.* (2021) 改进的 SHAP 值解释性方法计算第 p 个房价影响因素的 SHAP

值： $(\varphi_{c_1}^p, \varphi_{c_2}^p, \dots, \varphi_{c_r}^p)$ 。

3. 将样本时间段划分为 2009—2012 年、2013—2014 年、2015—2017 年和 2018—2019 年（划分依据详见下文），分别记为 T_{D1} 、 T_{D2} 、 T_{D3} 、 T_{D4} ，根据每座城市所属的城市类别、各类城市房价上涨的区间，借鉴 VAR 模型中方差分解的思想，加总影响因素在各个分析时间段内各类城市房价上涨区间的 SHAP 值，从而计算出第 p 个影响因素在不同分化时期对一二三线城市房价上涨的贡献率大小：

$$CR_{C_j, T_{D_u}}^p = \frac{\sum_{c \in C_j} \sum_{t \in T_{C_j} \cap T_{D_u}} \varphi_{c,t}^p}{\sum_{p=1}^P \sum_{c \in C_j} \sum_{t \in T_{C_j} \cap T_{D_u}} \varphi_{c,t}^p}, j=0,1,2,3; u=1,2,3,4 \quad (3)$$

进一步地，按照每个因素所属的大类因素，加总得到四个大类因素在不同房价分化时期对一二三线城市房价增速的贡献率。

4. 根据每轮分化时期的分化情况，找出该轮分化中房价增速波动最大的城市类，将该类城市房价的主要影响因素作为引起该轮分化的主要因素，并进一步通过对比分析探寻每一轮房价分化背后的主要驱动因素。

四、指标体系和数据说明

（一）房价分化时期界定

关于房价分化，主要是指同一时期不同城市间的房价增速存在明显差异的现象（倪鹏飞，2019；刘诚和杨继东，2019）。虽然各界普遍认为近年来一二三线城市房价增速出现分化，但是尚无研究对各线城市房价分化的时期进行明确界定。本文基于近年来一二三线城市房价增速走势以及各界普遍认可的房价分化时期，提出房价分化的两点界定标准。一是，在一定时期内，一线城市房价平均增速与二三线城市的房价平均增速相差达到 5 个百分点及以上。二是，一线与二三线城市房价平均增速差值达到 5 个百分点及以上所持续的时间达到 1 年及以上。考虑到 2008 年国际金融危机前后，中国经济走势和房地产市场走势呈现出不同态势，本文重点考察危机之后的房价分化状况。

结合图 1 可以看出，2009—2012 年间一二三城市的房价增速走势较为同步，但是从 2013 年开始呈现明显的分化趋势。而且，样本期内的大多数时间里二线城市和三线城市的房价增速较为接近，没有呈现出明显的分化趋势。有鉴于此，本文认为城市间房价分化主要表现为一线城市和二三线城市之间的分化，可以将二三线城市合并在一起加以分析。总体来看，截至 2021 年一线与二三线城市之间的房价增速已经出现了四轮明显分化，第一轮分化出现在 2013—2014 年间，第二轮分化出现在 2015—2017 年间，第三轮分化出现在 2018—2019 年，第四轮分化出现在 2020—2021 年间。

（二）房价影响因素指标体系的构建

本文基于经济学理论以及已有文献，构建了房价影响因素的指标体系。总体而言，房价的影响因素包括需求因素、供给因素、宏观政策因素和预期因素四大类。

1. 需求因素。需求因素将会影响住房需求，进而影响城市房价走势，本文主要考虑三个方面的需求因素。一是人口因素。主要体现在人口密度、人口结构和人口流动上。本文使用城市常住人口/城市建成区面积衡量城市人口密度对房价的影响，用人口自然增长率衡量人口年龄结构变化对房价的影响，用城镇化率衡量城乡人口流动对房价的影响。此外，还参考韩立彬和陆铭

(2018)的方法,使用(常住人口-户籍人口)/常住人口测度移民比率,进而计算出移民比率的增长率,从另一个角度反映人口流动对房价的影响。二是居民收入水平。参考已有文献做法,使用人均GDP增速、职工平均工资增速和人均可支配收入增速衡量居民收入对住房需求的影响。此外,还将城市GDP增速纳入指标体系。三是教育医疗等公共服务水平。参考倪鹏飞(2019)、刘诚和杨继东(2019)等文献的做法,分别使用人均床位数、人均道路面积、城市绿化覆盖率和AQI空气质量指数衡量城市医疗服务水平、教育水平、交通水平和环境因素对住房需求的影响。

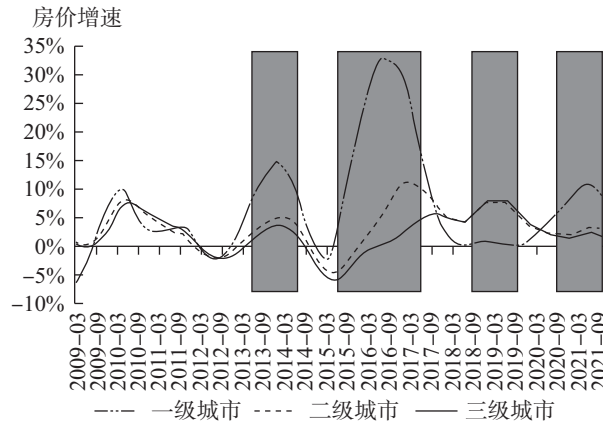


图1 一二三线城市房价增速与房价分化状况

注:相关数据源自国家统计局公布的70个大中城市二手住宅销售价格同比指数。阴影部分表示的是四轮房价分化时期。

2. 供给因素。市场上商品住宅的供给量和供给成本由开发商的供给行为和政府住宅用地的供给行为共同决定。借鉴韩立彬和陆铭(2018)的做法,使用单位土地购置价格同比增速、住宅房屋造价同比增速和商品房竣工面积三个指标反映开发商供给行为,使用住宅用地供应面积反映政府的住宅用地供给行为。由于部分城市的住宅用地供应面积和房屋竣工面积波动较大,导致面积增速无法准确地表示城市住宅用地和房屋供应量的变化,所以本文选择使用面积指标本身来刻画不同城市住宅用地供应和房屋供应的变化。

3. 宏观政策因素。就货币政策而言,参考涂红等(2018)等文献的做法,使用全国M2同比增速、新增个人住房贷款占新增贷款总额的比重以及个人住房贷款同比增速反映央行的数量型货币政策对房价增速的影响,使用个人住房贷款平均利率反映央行的价格型货币政策对房价增速的影响。就财政政策而言,参考已有文献的做法,使用土地出让金与一般公共预算收入的比值衡量各个城市对土地财政的依赖度,由此反映地方政府财政政策对房价增速的影响。此外,本文还进一步引入各城市的赤字率水平和债务率涨幅(陈彦斌等,2015)反映财政政策操作对房价增速的潜在影响。

4. 预期因素。一方面,参考高波等(2014)等的做法,使用适应性预期加以测度。将适应性预期的计算公式设定为: $\dot{P}_{t+1}^e = 0.5 \cdot \dot{P}_t^e + 0.5 \cdot \dot{P}_t$, 其中 \dot{P}_t^e 为第 t 期对房价增速的预期, \dot{P}_t 为第 t 期的房价增速。另一方面,央行在每个季度的《城镇储户问卷调查报告》中发布“房价预期指数”,以衡量调查人群中预期下季度房价上涨的比例,本报告使用该调查指标进一步测度预期因素对房价增速的影响。

上述四类因素,共计26个指标,构成了本文研究不同城市房价分化影响因素的指标体系。虽然与许多使用机器学习的文献所考量的上百个甚至更多指标相比有所偏少,但是本文的指标体系已经较为全面地涵盖了倪鹏飞(2019)、韩立彬和陆铭(2018)、刘诚和杨继东(2019)等文献

陈小亮等：城市间房价分化的影响因素研究

所考察的房价分化影响因素。之所以本文没有纳入更多变量，是因为纳入过多变量不仅会影响模型计算结果的准确性（Hastie *et al.*，2009；Aas *et al.*，2021），而且还会降低结果的可解释性。为了兼顾模型指标体系的丰富程度和测算结果的可解释性，本文构建了中等规模的指标体系。

（三）数据来源和预处理

本文所使用的房价增速以及影响房价增速的指标主要来自国家统计局、中国人民银行、生态环境部、Wind、国信房地产信息网和《中国城市统计年鉴》。基于本文所研究的问题，并综合考虑数据可得性和数据质量，本文将样本时间跨度界定为2009年一季度至2019年四季度，这一时间段较好地涵盖了2013—2019年间发生的三轮房价分化，因此能够对近年来城市房价分化的主要因素进行较为深入的分析。

在实证分析之前，还需要对相关数据进行预处理。第一，部分城市的人均GDP等个别指标存在缺失值，本文使用MissForest方法对缺失值进行填补。第二，将月度 and 年度指标转换为季度指标。本文使用的指标中，人均GDP增速等多个指标是年度数据，参考张春华等（2017）的做法，使用三次样条法进行插值处理。此外，将住宅用地出让面积和土地出让金总额在每季度内三个月的数值进行加总进而得到季度数据，而房价增速则需要将每个季度内三个月的平均增速作为季度值。第三，对各个变量的观测值进行Z-score标准化处理，使得模型训练不受量纲影响。

五、主要实证结论

本文的主要研究目标是，探寻一二三线城市之间多轮房价分化的主要影响因素。考虑到不同轮次房价分化的表现不同，需要对每一轮房价分化背后的影响因素进行详细剖析。进一步地，对于某一轮次的房价分化，需要先分析该轮次房价分化的具体表现和特征，进而找到其背后的驱动因素。本文将首先挑选出适用的机器学习方法，然后遵循上述思路探寻2013—2014年、2015—2017年和2018—2019年三轮房价分化各自的驱动因素，在此基础上提炼多轮房价分化背后的共同影响因素。

（一）各种机器学习方法的预测性能对比与方法选取

表1对比了Xgboost等六种机器学习方法的预测性能，从中可以得到两点主要结论。第一，对本文而言，非线性方法的预测性能明显优于线性方法。以 R^2 为例，Xgboost等四种非线性方法的 R^2 均超过了50%，而SCAD和Lasso这两种线性方法的 R^2 均低于40%。第二，在四种非线性方法中，Xgboost和随机森林的表现明显占优，而且Xgboost方法表现最佳。不管是从 R^2 的数值来看，还是从MSE和MAE的数值来看，Xgboost方法表现都是最佳的，随机森林次之。有鉴于此，本文在基准回归中选择使用Xgboost方法开展研究。为了保证实证结果的可靠性，本文还将在稳健性检验中进一步使用随机森林方法开展研究。

表1 不同机器学习方法在本文的预测性能对比

机器学习方法		R^2	MSE	MAE
非线性方法	Xgboost	78.79%	0.20	0.29
	随机森林	73.65%	0.21	0.32
	SVR	70.28%	0.27	0.35
	KNN	55.29%	0.42	0.42
线性方法	SCAD	38.96%	0.52	0.56
	Lasso	35.61%	0.54	0.57

(二) 每一轮城市房价分化的影响因素分析

表 2 使用 Xgboost 方法分别测算了每一轮房价分化时期各类因素对一二三线城市房价的贡献率。此外,为了更好地进行对比分析,表 2 还测算了 2009—2012 年这段时期(未分化时期)各类因素对一二三线城市房价的贡献率。

表 2 每一轮房价分化时期各类因素对一二三线城市房价贡献率 (%)

Panel A: 2009—2012 年间和 2013—2014 年间各类因素的贡献率							
影响因素		一线城市		二线城市		三线城市	
		2009—2012	2013—2014	2009—2012	2013—2014	2009—2012	2013—2014
需求因素	经济因素	22.64	14.60	14.38	17.15	14.94	24.39
	人口因素	19.21	10.97	12.81	14.67	9.83	11.74
	公共服务因素	11.10	21.16	16.93	20.42	18.05	26.13
供给因素	供给成本因素	4.59	11.94	3.43	8.08	2.25	-0.45
	供给面积因素	10.48	10.57	6.97	8.81	2.91	4.64
政策因素	货币政策因素	24.43	-8.06	11.82	-4.08	18.73	6.59
	财政政策因素	13.76	4.15	14.54	7.53	12.86	10.54
预期因素	预期因素	-6.21	34.68	19.12	27.43	20.43	16.44

Panel B: 2015—2017 年间和 2018—2019 年间各类因素的贡献率							
影响因素		一线城市		二线城市		三线城市	
		2015—2017	2018—2019	2015—2017	2018—2019	2015—2017	2018—2019
需求因素	经济因素	2.99	9.65	7.74	9.85	20.56	10.20
	人口因素	3.18	6.69	9.83	8.74	7.72	5.68
	公共服务因素	17.83	9.54	6.67	10.74	20.68	9.99
供给因素	供给成本因素	2.49	-2.58	3.00	-0.50	7.25	5.84
	供给面积因素	6.88	6.83	4.24	5.04	-3.15	0.99
政策因素	货币政策因素	28.99	10.27	44.41	8.51	87.51	15.14
	财政政策因素	6.15	13.19	6.18	10.92	2.00	8.10
预期因素	预期因素	31.50	46.42	17.93	46.71	-42.57	44.06

注:经济因素包括人均 GDP 增速、职工平均工资增速、人均可支配收入增速和城市 GDP 增速;人口因素包括人口自然增长率、城镇化率和移民占比涨幅;公共服务因素包括人均床位数、人均财政教育支出同比增速、人均道路面积、绿化覆盖率和城市 AQI 空气质量指数;供给成本因素包括单位土地价格同比增速和商品住宅房屋造价同比增速;供给面积因素包括住宅用地供应面积和商品房竣工面积;货币政策因素包括全国 M2 增速、新增个人住房贷款同比增速和住房贷款利率;财政政策因素包括各城市土地财政依赖度、赤字率和债务率涨幅;预期因素包括各城市的房价上涨预期和全国房价上涨预期指数。表 3 同。

第一,2013—2014 年城市房价分化的影响因素分析。2013—2014 年间这一轮房价分化的主要表现是,一线城市房价增速明显高于二三线城市,二线城市房价增速略高于三线城市,而且 2013—2014 年间一线城市的房价增速超过了 2009—2012 年,二三线城市的房价增速则低于 2009—2012 年。因此,要想找到本轮房价分化的主要原因,关键是找到一线城市房价上涨的主

陈小亮等：城市间房价分化的影响因素研究

要驱动因素。

基于表 2 中 Panel A 的测算结果，通过将 2009—2012 年房价未分化时期与 2013—2014 年房价分化时期各个因素对房价增速的贡献率进行对比知，2013—2014 年间一线城市房价上涨主要源自三方面原因。其一，预期因素对一线城市房价增速的贡献率从 2009—2012 年间的 -6.21% 升至 2013—2014 年间的 34.68%，涨幅超过 40 个百分点。这是因为，2013—2014 年间一线城市房价上涨预期有所提高。其二，需求因素中的公共服务对一线城市房价增速的贡献率从 11.10% 提高到了 21.16%，涨幅也超过 10 个百分点，主要是因为这一时期一线城市环境质量有了大幅提高，使得其对房价上涨的贡献率提高。其三，供给因素中的供给成本对一线城市房价上涨的贡献率从 4.59% 升至 11.94%，上涨了 7.35 个百分点，因为这一时期北上广的土地成本明显上涨。

2013—2014 年，二线城市房价上涨的主要影响因素与一线城市较为一致，但是贡献率相对偏低。就预期因素而言，从 2013 年初到 2014 年上半年，二线城市房价预期经历了持续的上涨过程，预期的上涨也是推动这一轮二线城市房价上涨的主要因素，其贡献率从 2009—2012 年的 19.12% 上升到 2013—2014 年的 27.43%，涨幅为 8.31 个百分点。就需求因素而言，经济、人口和公共服务因素的贡献率都有所上升，其中公共服务的贡献率上升幅度较大，达到了 3.49 个百分点，这主要是由人均公园绿地面积贡献率的提高所致。就供给因素而言，由于 2013—2014 年间二线城市单位土地价格出现了明显上涨，从而使得单位土地价格贡献率较 2009—2012 年上涨了 4.28 个百分点。与之不同，2013—2014 年间三线城市房价增速的主要影响因素是需求因素，需求因素对三线城市房价增速的贡献率达到了 62.25%，比 2009—2012 年间提高了近 20 个百分点，也显著高于一线和二线城市。预期因素对三线城市 2013—2014 年房价增速的贡献率则下降了 4 个百分点。

综上可知，2013—2014 年间一二三线城市房价分化主要是源自三方面原因。第一，预期因素是本轮房价分化的最主要影响因素。第二，需求因素中的公共服务对一二三线城市房价增速的贡献率都出现了一定上涨，并且在一线城市的贡献率涨幅要高于二三线城市，这表明公共服务因素在本轮房价分化中也发挥了一定作用。第三，供给因素中的土地成本对一二线城市房价上涨也起到了重要的推动作用，并且在一线城市中的贡献率涨幅要高于二线城市，这表明单位土地价格在一定程度上导致了本轮房价分化。此外，货币政策和财政政策对一二三线城市房价增速的贡献率均明显下降，而且一线城市降幅高于二三线城市，所以政策因素不是导致本轮房价分化的关键因素。

第二，2015—2017 年城市房价分化的影响因素分析。2015—2017 年间这一轮房价分化的主要表现是，一线城市房价增速大幅高于二三线城市，二线城市增速也高于三线城市，并且一二三线城市在此期间的房价增速整体而言都超过了 2009—2012 年和 2013—2014 年。因此，要想找到一二三线城市房价分化的主要原因，最关键的仍然是要找到一线城市房价增速大幅上涨的主要驱动因素。

基于表 2 中 Panel B 的测算结果，通过将 2015—2017 年与 2013—2014 年各个因素对房价增速的贡献率进行对比分析可以发现，2015—2017 年间一线城市房价上涨主要源自预期因素和货币政策两个方面因素的驱动。其一，预期因素对一线城市房价增速的贡献率在 2015—2017 年为 31.5%，虽然相比 2013—2014 年略有下降，但是与二三线城市预期因素的贡献率相比明显偏高。其二，货币政策对一线城市房价增速的贡献率出现了大幅上涨，从 2013—2014 年间的 -8.06% 升至 2015—2017 年间的 28.99%。将预期因素与货币政策因素加总之后可知，二者对一线城市 2015—2017 年房价增速的贡献率达到了 60.49%，是一线城市房价上涨的主要驱动因素。相比之下，需求因素和供给因素对一线城市 2015—2017 年间房价增速的贡献率明显下降。

在二线城市，2015—2017年货币政策对房价增速贡献率是最高的，达到了44.41%，这意味着仅货币政策因素就能够解释这一期间二线城市房价增速上涨的接近一半。与一线城市类似，由于在此期间全国范围内的货币政策环境相对宽松，住房贷款利率下降、个人住房贷款同比增速上升，从而驱动二线城市房价明显上涨。除了货币政策，其他因素对2015—2017年间二线城市房价增速的贡献率均低于2013—2014年间的贡献率。其中，公共服务对二线城市房价增速贡献率的降幅最大，达到了15个百分点左右，预期因素和经济因素贡献率的降幅也接近10个百分点。在三线城市，货币政策同样是本轮房价上涨最主要的影响因素，而且其贡献率高达87.51%。预期因素对三线城市房价增速的贡献率同样大幅下跌至-42.57%。究其原因，2015年三线城市房价预期明显下降，但房价增速走势是上升的，预期下降没有带来房价增速的下降，从而导致预期因素的贡献率为负。

将上述分析综合起来可知，2015—2017年间一二三线城市房价增速分化的主要驱动因素是预期因素。在此期间，预期因素对一线城市房价增速的贡献率维持在高位，对二线城市房价增速的贡献率下降了近10个百分点，对三线城市房价增速的贡献率更是大幅下降了近60个百分点。不难发现，预期因素贡献率的变化与一二三线城市房价分化的趋势保持一致，这说明预期因素在本轮房价分化中发挥着重要作用。需要特别注意的是，虽然货币政策因素的贡献率在一二三线城市都显著上升，并且货币政策是二三线城市2015—2017年房价上涨的最主要影响因素，但是它并没有使得二三线城市的房价增速超过一线城市，这表明货币政策因素并不是本轮房价分化的最主要因素。此外，需求因素和供给因素对一二三线城市房价增速的贡献率均出现了一定程度的下降，而且在一线城市的降幅最大，可见它们并不是导致房价分化的重要因素。

第三，2018—2019年城市房价分化的影响因素分析。在最近10年间出现的四轮房价分化中，2013—2014年、2015—2017年以及2020—2021年三轮房价分化都表现为一线城市房价增速高于二三线城市，只有2018—2019年这一轮房价分化表现为二三线城市房价增速高于一线城市。要想找到本轮房价分化的原因，既要分析二三线城市房价增速上涨的驱动因素，又要分析一线城市房价增速明显放缓的原因。

从表2中Panel B的测算结果可知，对一线城市而言，2018—2019年间对其房价增速贡献率最高的是预期因素，贡献率达到了46.42%，明显超过了2013—2014年和2015—2017年房价分化时期的贡献率。事实上，2018—2019年间一线城市房价预期一直表现出下降趋势，房价预期指数从2018年一季度末的8.46持续下滑至2019年四季度末的0.34。这根源于中央以及各个城市对房价的高强度管控。在高强度管控之下，一线城市的居民普遍预期房价难以保持2015—2017年间的涨势，进而使得一线城市房价增速明显放缓，2018—2019年间一线城市房价增速平均为0.57%，而2015—2017年房价增速均值达到了17.31%，峰值更是达到了33.03%。与预期因素形成鲜明对比的是，货币政策对一线城市房价增速的贡献率从2015—2017年的28.99%下降到2018—2019年的10.27%，这主要是因为中央为了防范房价泡沫化风险而采取的货币政策紧缩操作。2018—2019年住房贷款利率平均为5.61%，较2015—2017年上升了0.69个百分点。

在本轮房价分化期间，二线和三线城市房价增速主要影响因素及其走势与一线城市高度相似。一方面，预期因素对房价增速的贡献率大幅升高。预期因素对二线和三线城市2018—2019年间房价增速的贡献率分别高达46.71%和44.06%，都是贡献率最高的影响因素，比2015—2017年间高出了数十个百分点。从二三线城市房价预期指数来看，2018—2019年二三线城市房价预期指数平均为6.49和6.1，较2015—2017年分别提高了3.39个和6.65个百分点。究其原因，由于一线城市出台了一系列调控房地产市场过快上涨的政策，使得人们普遍预期一线城市房价将会下降，而二三线城市则因为人才引进计划带来的人口流入和棚改货币化带来的住房需求而

陈小亮等：城市间房价分化的影响因素研究

出现了预期房价上涨的局面。另一方面，货币政策对房价增速的贡献率大幅下降。在二线城市，货币政策对房价增速的贡献率从 44.41% 下降到 8.51%。在三线城市，货币政策对房价增速的贡献率更是从 87.51% 骤降至 15.14%，这同样源自货币政策的紧缩操作。

综上可知，2018—2019 年间的这一轮房价分化主要是由预期因素引起的。在一二三线城市，预期因素对房价增速的贡献率都大幅升高，但是在二三线城市预期对房价增速的贡献率涨幅更高，这与二三线城市房价增速高于一线城市的分化现象是一致的。一线城市预期因素贡献率上升并没有带来一线城市房价显著上涨，是因为居民预期房价增速下降而非上涨，这与一线城市房价的走势也是一致的。值得注意的是，在这一轮房价分化过程中，货币政策对一二三线城市房价增速的贡献率均出现了大幅下降，而且在二三线城市的降幅比在一线城市更大，这表明货币政策并不是本轮房价分化的主要驱动因素。

（三）三轮房价分化时期主要影响因素的对比分析

在依次分析了 2013—2014 年、2015—2017 年和 2018—2019 年的三轮房价分化的重要影响因素之后，可以进一步总结提炼出三轮房价分化的共同驱动因素以及各类因素作用的相对大小，主要包括三点结论。

第一，预期因素在三轮房价分化期间都起到了非常重要的驱动作用。具体来讲，对于 2013—2014 年间的房价分化而言，一线城市的公共服务改善程度以及净流入人口数量高于二三线城市，从而使得一线城市房价上涨预期高于二三线城市。对于 2015—2017 年间的房价分化而言，全国范围内的货币政策环境较为宽松，在一线城市宽松的货币政策显著推升了房价上涨预期，大量投资性和投机性资金涌入，使得一线城市房价飙升。不过，货币政策并没有带动二三线城市尤其是三线城市的房价上涨预期，2015—2017 年一线城市房价上涨预期指数比二三线城市分别高出了 13.25 和 16.9。对于 2018—2019 年间的房价分化而言，二三线城市由于人才引进计划带来的人口流入和棚改货币化带来的住房需求而出现了预期房价上涨的局面，最终导致二三线房价涨势超过一线城市这一罕见局面。

第二，货币政策本身并不是房价分化的主要驱动因素，但是可以通过影响预期进而导致房价分化。从 2013—2014 年间的房价分化来看，货币政策在一二三线城市中的贡献率均比 2009—2012 年有所下降，尤其在一线城市中的降幅还超过了二三线城市，这与一线城市房价增速明显高于二三线城市所导致的房价分化现象并不一致。从 2015—2017 年间的房价分化来看，货币政策在一二三线城市中的贡献率均有所上升，在三线城市的贡献率甚至达到了 87.51%，高于一线城市 58.52 个百分点，但这并没有带来三线城市房价的大幅上涨，反而一线城市房价增速大幅高于二三线城市。从 2018—2019 年间的房价分化来看，货币政策对二三线城市房价的贡献率比 2015—2017 年大幅下降，降幅远大于一线城市，然而这并没有使得二三线城市房价增速低于一线城市，相反一线城市房价持续低迷。综上可以看出，每一轮房价分化中货币政策因素在一二三线城市贡献率的变化都与房价分化现象不一致，因此货币政策本身并不是房价分化的主要驱动因素。不过，货币政策的变化可以带来预期的变化，进而引发房价分化。例如，2015—2017 年间，全国层面货币政策普遍宽松，这导致大量投资性和投机性资金涌入一线城市，一线城市房价上涨预期得到强化，而二三线城市并没有足够的资金支持驱动房价预期上涨，由此导致了本轮房价分化的发生。

第三，需求因素和供给因素本身也不是房价分化的主要驱动因素，但是也可以通过影响预期进而导致房价分化。从实际情况来看，一方面，不管是一线、二线还是三线城市，需求因素和供给因素的贡献率之和整体而言呈现下降趋势，这在一线城市体现得尤为明显。另一方面，一线城市房价增速的大涨大跌是几轮房价分化的核心原因。由此可见，需求因素和供给因素本身并不是

房价分化的主要原因。事实上，需求因素和供给因素的贡献率整体下降，并不是因为需求因素和供给因素有所恶化，而是因为预期因素和货币政策因素的贡献率之和提高，导致需求因素和供给因素的贡献率被动下降。不过，需求因素和供给因素可能通过影响预期进而影响房价走势。例如，2018—2019年间，由于二三线城市出台大量的人才引进计划，吸引了人口流入，进而推升了二三线城市的房价上涨预期，而这正是本轮房价分化的重要原因之一。

需要补充说明的是，韩立彬和陆铭（2018）、刘诚和杨继东（2019）等研究认为，供给因素是导致房价分化的重要因素，这与本文结论有所不同。其主要原因在于，已有文献研究的时间区间大多集中在2000—2010年，本文主要分析的是2013年以来的三轮房价分化情况，而在此期间供给因素并没有发生显著变动。此外，已有文献在研究供给因素的影响时，并没有考虑预期因素，本文发现相对于供给因素，预期因素是导致房价分化更重要的因素。而且，供给因素和需求因素可能通过影响预期进而导致房价分化。

（四）稳健性检验

为了确保实证结果的稳健性，本文在基准回归的基础上，进行了稳健性检验。^① 一是对基准指标体系进行扩充。在基准回归中，本文选取的货币政策指标是全国层面的指标，在稳健性检验中，本文进一步将各个城市金融机构人民币贷款余额增速的数据纳入指标体系，用以衡量不同城市对货币政策的异质性反应。二是考虑宏观政策因素的时滞性，引入政策因素的滞后期。无论是货币政策还是财政政策，其对房价的传导具有一定的时滞。有鉴于此，本文在稳健性检验中将货币政策和财政政策指标的当期数据替换为滞后半年的数据，进而消除政策时滞可能产生的影响。稳健性检验的结果详见表3，将表3与表2的基准结果对比可以发现，在增加城市层面信贷指标并考虑政策时滞之后，基准回归结果仍然成立，可见本文的主要结论较为稳健。囿于篇幅限制，不再详述。

六、结论与政策建议

本文基于2009—2019年国家统计局重点关注的70个大中城市的面板数据，构建了包含26个指标的基准指标体系，综合使用Xgboost等机器学习方法和SHAP值解释性方法，研究了不同城市房价分化的主要原因。研究发现，第一，预期因素在多轮房价分化期间都起到了重要的驱动作用。第二，货币政策本身并不是房价分化的主要驱动因素，但是可以通过影响预期进而导致房价分化。第三，需求因素和供给因素本身也不是房价分化的主要驱动因素，但是也可以通过影响预期进而导致房价分化。

基于本文的实证结论可知，要想“因城施策”促进房价平稳增长，各个城市应该基于自身情况，从需求端或供给端出发稳定居民对房价的预期。就一线城市而言，由于外来人口持续流入，而土地供给则相对偏少，因此长期处于住房供给相对紧缺的状态，这也使得一线城市居民预期房价呈现上涨趋势。要想稳定房价预期，应考虑在控制土地成本的前提下适当增加住房供给，以削

^① 除此之外，本文还进行了另外三方面稳健性检验。第一，将适应性预期的计算公式中滞后一期房价赋值权重从0.5调整为0.3和0.7。第二，将机器学习方法由Xgboost方法替换为随机森林方法。第三，在基准回归中使用的数据截至2019年，在稳健性检验中将数据更新到2020年。之所以将更新数据后的计算结果作为稳健性检验加以展示，而非作为基准结果，主要是因为，2020年的数据因新冠疫情冲击而产生了大幅波动，可能导致模型计算结果出现偏误。将相关稳健性检验结果与表2对比可知，本文的核心结论仍然较为稳健。囿于篇幅限制，不再列示相关实证结果，留存备索。

陈小亮等：城市间房价分化的影响因素研究

弱住房供不应求所带来的房价上涨预期。此外，还要不断完善住房租赁市场，以减轻住房买卖市场的压力，进一步削弱一线城市的房价上涨预期。就二三线城市而言，由于这些城市往往面临人口流出的境况，居民预期房价会下跌。为了稳定房价预期，要进一步完善城市基础设施，提高教育、医疗和环境质量（罗俊等，2022），从而改善居民购房需求。与此同时，住房供给相对充裕的二三线城市还要制定更加科学合理的人才引进计划，吸引人口流入，从而进一步稳定房价预期。

表 3 稳健性检验的结果 (%)

Panel A: 2009—2012 年间和 2013—2014 年间各类因素的贡献率						
影响因素	一线城市		二线城市		三线城市	
	2009—2012	2013—2014	2009—2012	2013—2014	2009—2012	2013—2014
需求因素	49.73	43.50	39.46	52.40	37.73	63.37
经济因素	24.83	14.16	15.33	17.64	15.92	25.73
人口因素	17.25	10.71	11.71	15.22	8.34	12.41
公共服务因素	7.64	18.64	12.42	19.54	13.46	25.23
供给因素	19.29	23.17	13.57	17.80	6.93	6.19
供给成本因素	7.73	12.27	5.92	8.51	4.12	1.59
供给面积因素	11.56	10.90	7.64	9.29	2.80	4.60
政策因素	43.33	-7.70	28.89	-4.97	37.44	4.38
货币政策因素	29.97	-11.35	15.25	-12.68	25.43	-6.16
财政政策因素	13.36	3.65	13.64	7.71	12.01	10.54
预期因素	-12.34	41.03	18.08	34.77	17.91	26.07

Panel B: 2015—2017 年间和 2018—2019 年间各类因素的贡献率						
影响因素	一线城市		二线城市		三线城市	
	2015—2017	2018—2019	2015—2017	2018—2019	2015—2017	2018—2019
需求因素	22.57	15.17	22.39	24.86	43.65	21.23
经济因素	3.40	8.46	7.62	9.26	20.05	9.84
人口因素	3.03	3.31	10.21	8.20	7.09	4.34
公共服务因素	16.13	3.40	4.56	7.40	16.51	7.05
供给因素	10.02	6.52	7.80	5.84	7.72	8.06
供给成本因素	3.15	-0.23	3.73	0.84	-2.25	7.43
供给面积因素	6.87	7.89	4.07	5.00	9.98	0.64
政策因素	37.19	37.00	54.58	25.91	88.27	29.24
货币政策因素	32.09	24.64	49.37	16.20	87.18	21.59
财政政策因素	5.10	12.36	5.21	9.71	1.09	7.65
预期因素	30.23	41.30	15.23	43.39	-39.64	41.47

就货币政策而言，则需要保持“连续性、一致性、稳定性”，以进一步帮助各个城市稳定房

价预期。正如本文实证结果所示，一方面，房价增速对货币政策的变动较为敏感，另一方面货币政策本身并不是房价分化的主要驱动因素，但是可以通过影响预期进而导致房价分化。“保持房地产金融政策的连续性、一致性、稳定性”能够更好地稳定房价预期。而且，根据宏观政策“三策合一”新理论框架，长期来看，货币政策不应该将控房价作为主要目标，而应该聚焦稳增长（陈彦斌，2022）。至于控房价，应该更多地使用宏观审慎政策加以应对（马勇和黄辉煌，2021），不断完善宏观审慎政策能够更好地防范投资性和投机性资金导致的房价剧烈波动，从而稳定房价预期。

参考文献

- 陈淑云、唐将伟，2017：《公共服务供给不均等加剧了国内房价分化吗？——基于我国286个地级及以上城市面板数据的实证》，《经济体制改革》第4期。
- 陈彦斌，2022：《宏观政策“三策合一”新理论框架》，《经济研究》第11期。
- 陈彦斌、郭豫媚、陈伟泽，2015：《2008年金融危机后中国货币数量论失效研究》，《经济研究》第4期。
- 高波、王辉龙、李伟军，2014：《预期、投机与中国城市房价泡沫》，《金融研究》第2期。
- 宫汝凯，2015：《财政不平衡和房价上涨：中国的证据》，《金融研究》第4期。
- 韩立彬、陆铭，2018：《供需错配：解开中国房价分化之谜》，《世界经济》第10期。
- 刘诚、杨继东，2019：《土地策略性供给与房价分化》，《财经研究》第4期。
- 陆铭、欧海军、陈斌开，2014：《理性还是泡沫：对城市化、移民和房价的经验研究》，《世界经济》第1期。
- 罗俊、黄佳琦、刘靖姗、石荣浩，2022：《保障性住房、社会身份与公平感——来自社区实地实验的证据》，《经济理论与经济管理》第10期。
- 马勇、黄辉煌，2021：《双支柱调控的金融稳定效应研究》，《经济理论与经济管理》第9期。
- 倪鹏飞，2019：《货币政策宽松、供需空间错配与房价持续分化》，《经济研究》第8期。
- 涂红、徐春发、余子良，2018：《货币政策对房价影响的区域差异：来自多层混合效应模型的新证据》，《南开经济研究》第5期。
- 王拉娣、安勇，2016：《居民收入差距、预期对城市房价的影响》，《经济问题探索》第12期。
- 王猛、李勇刚、王有鑫，2013：《土地财政、房价波动与城乡消费差距——基于面板数据联立方程的研究》，《产业经济研究》第5期。
- 王先柱、杨义武，2015：《差异化预期、政策调控与房价波动——基于中国35个大中城市的实证研究》，《财经研究》第12期。
- 张春华、高铁梅、陈飞，2017：《经济时间序列频率转换方法的研究与应用》，《统计研究》第2期。
- 邹瑾、于焯华、王大波，2015：《人口老龄化与房价的区域差异研究——基于面板协整模型的实证分析》，《金融研究》第11期。
- Aas, K., M. Jullum, and A. Løland, 2021, “Explaining Individual Predictions when Features are Dependent: More Accurate Approximations to Shapley Values”, *Artificial Intelligence*, 298 (9): 1–24.
- Abildgren, K., N. L. Hansen, and A. Kuchler, 2018, “Overoptimism and House Price Bubbles”, *Journal of Macroeconomics*, 56 (6): 1–14.
- Athey, S., M. Bayati, G. Imbens, et al., 2019, “Ensemble Methods for Causal Effects in Panel Data Settings”, *AEA Papers and Proceedings*, 109 (5): 65–70.
- Babii, A., E. Ghysels, and J. Striaukas, 2022, “Machine Learning Time Series Regressions with an Application to Nowcasting”, *Journal of Business & Economic Statistics*, 40 (3): 1094–1106.
- Chow, G. C., and L. Niu, 2015, “Housing Prices in Urban China as Determined by Demand and Supply”, *Pacific Economic Review*, 20 (1): 1–16.
- Glaeser, E., W. Huang, Y. Ma, and A. Shleifer, 2017, “A Real Estate Boom with Chinese Characteristics”, *Journal of Economic Perspectives*, 31 (1): 93–116.

陈小亮等：城市间房价分化的影响因素研究

Hastie, T., R. Tibshirani and J. Friedman, 2009, *The Elements of Statistical Learning*, New York: Springer-Verlag.

Jarociński, Marek, and Frank R. Smets, 2008. “House Prices and the Stance of Monetary Policy”, *Federal Reserve Bank of St. Louis Review*, 90 (4): 339 – 365

Racine, J., 2000, “Consistent Cross-validators Model-selection for Dependent Data: Hv-block Cross-validation”, *Journal of Econometrics*, 99 (1): 39 – 61.

Štrumbelj, E., and I. Kononenko, 2014, “Explaining Prediction Models and Individual Predictions with Feature Contributions”, *Knowledge and Information Systems*, 41 (3): 647 – 665.

Takáts, E., 2012, “Aging and House Prices”, *Journal of Housing Economics*, 21 (2): 131 – 141.

(责任编辑：李振新)

IDENTIFICATION OF INFLUENCING FACTORS OF HOUSING PRICE DIVERGENCE IN DIFFERENT CITIES: BASED ON MACHINE LEARNING METHODS

CHEN Xiaoliang¹ CHEN Kan² WANG Zhaorui³ XIAO Zhengyan^{2,4}

- (1. Institute of Economics, CASS;
2. School of Statistics, Renmin University of China;
3. School of Economics, Remin University of China;
4. Center for Applied Statistics of Renmin University of China)

Summary: Since 2013, the growth rate of housing prices in China’s Tier 1, Tier 2, and Tier 3 cities has exhibited a clear trend of differentiation, and “city-specific policies” have gradually become the mainstay of housing price regulation. To implement “city-specific policies,” it is necessary to accurately identify the primary factors that affect the differentiation of housing prices in different cities. This paper comprehensively employs machine learning methods such as Xgboost and SHAP value interpretability methods, based on panel data of 70 large and medium-sized cities nationwide from 2009 to 2019, to calculate and analyze the primary driving factors behind the multi-round differentiation of housing prices in Tier 1, Tier 2, and Tier 3 cities. The research results indicate that: First, expected factors played a crucial driving role during the multi-round differentiation of housing prices; Second, monetary policy itself is not the primary driving factor of housing price differentiation, but it can cause housing price differentiation by affecting expectations; Third, demand factors and supply factors themselves are not the primary driving factors of housing price differentiation, but they can also cause housing price differentiation by affecting expectations.

In view of this, the key to “city-specific policies” in different cities is to stabilize the public’s expectations of housing price trends based on their own situation, starting from the demand side or the supply side. For Tier 1 cities, due to the continuous influx of population outside the city and relatively scarce land supply, they are in a state of relatively tight housing supply, which also

makes residents' expected housing prices show an upward trend. To stabilize housing price expectations, it is necessary to consider appropriately increasing housing supply under the premise of controlling land costs, so as to weaken the expected housing price increase caused by the imbalance between housing supply and demand. In addition, government should continue to improve the housing rental market to alleviate the pressure on the housing buying and selling market, and further weaken the expected housing price increase in Tier 1 cities. For Tier 2 and Tier 3 cities, since these cities often face the situation of population outflow, residents' expected housing prices will decline. In order to stabilize housing price expectations, it is necessary to further upgrade urban infrastructure, improve education, medical and environmental quality, and improve residents' housing purchase demand. At the same time, Tier 2 and Tier 3 cities with relatively abundant housing supply should also formulate more scientific and reasonable talent introduction plans to attract population inflows, so as to further stabilize housing price expectations.

The marginal contribution of this paper is reflected in two aspects. First, although some literature attempts to explore the primary reasons for the differentiation of housing prices in different cities, traditional methods such as panel regression models are mainly used. And the number of factors that can be examined and the identified nonlinear relationships are relatively limited. Machine learning methods can better overcome the shortcomings of traditional methods and more accurately identify the primary factors affecting housing price differentiation. Second, although the central government has been calling for "city-specific policies" in recent years, the academic community's research on the reasons for the differentiation of housing prices in different cities is not yet clear. Existing literature mainly studies whether a certain factor has a significant impact on housing price differentiation, and cannot clearly compare the relative importance of different factors. This paper uses SHAP value interpretability methods to directly rank the relative importance of various influencing factors, to identify the primary factors affecting housing price differentiation and provide more clear decision-making references for real estate regulation and control.

Key words: housing price rising; housing price divergence; housing price expectation; city-specific policies; machine learning method