

基于 X-12-ARIMA 和 AR-GARCH 模型的房价波动研究

聂淑媛

(洛阳师范学院 数学科学学院,河南 洛阳 471934)

摘要:以 SAS 软件为工具,以 2007 年 1 月至 2015 年 6 月北京市新建住宅价格指数序列为样本,构建时间序列模型进行实证研究.结果表明,基于 X-12-ARIMA 模型和 AR(2)-GARCH(1,1)模型的复合模型是拟合房价的最优模型.房价序列存在明显的季节特征和典型的波动聚集性,X-12 季节调整方法和异方差模型显著有效,拟合相对误差不超过 0.4%.对房价的短期预测表明,近期内房价仍保持 3%~5% 的增长态势,且外部因素对房价的影响程度远远大于房价自身的波动冲击力.

关键词:房价;X-12-ARIMA 模型;AR-GARCH 模型;季节调整

中图分类号:O212

文献标志码:A

自货币化和市场化住房制度实施后,全国各地的房价持续走高且波动剧烈,从 1998 年房改开始到 2010 年的 10 多年间,我国的房价指数年平均涨幅维持在 1% 左右,随后的短短几年内则迅速上涨到 5% 左右,北京、上海、深圳等一线城市更是房价暴涨,年平均增长率甚至高达 20%.房价的快速增长不仅催生了房地产投资增长过快、过热、居民生活压力过大等诸多社会问题,而且直接影响到国家经济运行的稳定性.

房价波动始终是学界研究的热点问题,李仲飞等^[1]从成本推动和需求拉动的角度探讨了推动房价上涨的根本原因;王立平等^[2]基于时变参数模型理论解析房价变动的缘由;张宇青等^[3]基于三元 VAR-ARCH-BEKK 模型研究了经济预警指数、国房景气指数与 CPI 指数的波动溢出效应;顾莹等^[4]利用加权马尔可夫链预测房价;赵华平等^[5]基于 35 个大中城市静态和动态空间面板模型实证研究了城市宜居性特征对商品住宅价格的影响;葛龙^[6]使用 GARCH 和 COPULA 模型预测分析了天津房地产市场的发展态势;文献^[7]利用 AR(3)-ARCH(1)模型研究了郑州市新建住宅价格指数序列等.但目前鲜有文献构建 X-12-ARIMA 模型探究房价波动的季节调整效应,而国家统计局公布的“国房景气指数”,是反映我国房地产业基本运行状况和发展变化趋势的重要统计指标之一,在计算指标数值时也率先使用 X-11 程序进行季节因素调整,故利用 X-12 方法分析房价的季节波动特征是可行的技术路线.基于此,本文以已有研究为基础,综合运用 X-12-ARIMA 模型和 GARCH 族模型,实证研究并适度预测房地产价格的总体发展态势,具有重要的现实意义.

1 理论模型简介

1.1 X-12-ARIMA 模型

作为目前比较成熟和应用最广泛的季节调整方法,X-12-ARIMA 模型由美国普查局季节调整首席研究员 Findley 等^[8]于 1998 年在 X-11 过程基础上编制,大致分为两个阶段,第一阶段主要是建立预调整模块 regARIMA,利用带回归元的 ARIMA 模型对原始数据做丰富的预处理,包括检测、修正各类异常值和离群值,估计、消除历法效应,然后提供向前预测和向后预测,以保证数据的完整性;第二阶段主要是把第一阶段所产生的回归误差导入 X-11 模块进行季节调整,生成季节指数,并对季节调整效果进行严格的诊断检验.模型的

收稿日期:2016-03-14;修回日期:2016-06-20.

基金项目:河南省软科学研究计划项目(152400410152);河南省科技厅计划项目(162400410122);河南省高等学校重点科研项目(16B110008).

作者简介(通信作者):聂淑媛(1974-),女,河南台前人,洛阳师范学院副教授,博士,研究方向为时间序列分析的应用, E-mail:nsynsy2008@163.com.

基本形式为:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^Dx_t = \theta(B)\Theta(B^s)\epsilon_t,$$

其中, B 为滞后算子, s 为季节周期, d, D 分别是非季节差分次数和季节差分次数, $\phi(B), \Phi(B^s)$ 表示非季节和季节性自回归系数多项式, $\theta(B), \Theta(B^s)$ 则表示非季节和季节性移动平均系数多项式, $\{\epsilon_t\}$ 是零均值白噪声序列.

随着 X-12-ARIMA 软件在控制季节调整效果、处理异常值和识别季节模式等方面展现的优势, 学者们纷纷根据本土化传统节日因素, 研发和使用凸显本国特色的 X-12 程序, 李剑等^[9] 基于 X-12-ARIMA 模型和 ARCH 类模型研究了中国粮食价格的波动特征, 吴岚等^[10] 则探讨了 X-12 季节调整技术在物价波动方面的应用.

1.2 GARCH 族模型

为修正 1982 年 Engle 所创建 ARCH 模型的拟合精度问题, 1986 年, 美国经济学家 Bollerslev 提出了广义自回归条件异方差模型, 简称 GARCH 模型. 模型构造 $x_t = f(t, x_{t-1}, x_{t-2}, \dots)$ 用于拟合确定性信息, 且

$$\epsilon_t = \sqrt{h_t}e_t, h_t = \omega + \sum_{i=1}^p \eta_i h_{t-i} + \sum_{j=1}^q \lambda_j \epsilon_{t-j}^2 \quad [11].$$

该模型发展迅速, 且逐步衍生出 AR-GARCH、GARCH-M 和 TGARCH 等 GARCH 族模型, 构成当前国际上研究金融市场资产定价的前沿理论. 近年来, 陆续有学者开始探讨上述模型在居民消费价格指数、房价测度等领域的应用, 如张宇青、葛龙和笔者等分别构建不同的 GARCH 模型研究了全国、天津市与河南省的房价波动效应.

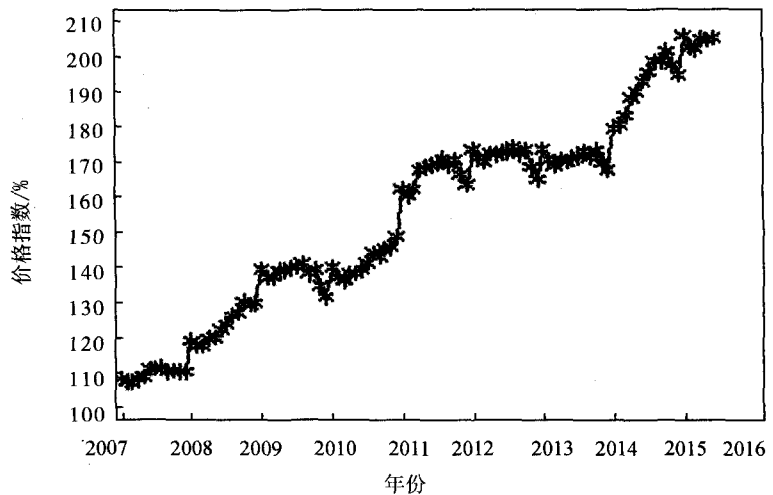


图1 北京市新建住宅价格定基比指数时序图

2 变量选取与数据来源

本文以北京市新建住宅价格指数为指标, 为确保数据统计口径的一致性, 全部数据均来源于国家统计局网站, 由于时间序列季节调整通常采取定基比指数为样本, 而我国自 2011 年才开始公布以 2010 年为基数的新建住宅价格定基比指数, 数据量不足以满足季节调整需要, 故本文选取以上年价格为 100 的新建住宅价格同比数据, 兼顾到数据的时效性和易得性, 选择样本区间为 2007 年 1 月至 2015 年 6 月, 但该序列本身已剔除了部分季节因素, 故首先对原序列进行预处理, 样本数据转化为以 2007 年各月为基数的定比指数序列.

3 实证研究

3.1 房价的季节性波动特征分析

首先运行 SAS 软件的 gplot 程序, 创建北京市新建住宅价格定基比指数 (记为 x_t) 的时序图, 直观考察图 1 可知, 序列不存在潜在的异常值、离群值和交易日效应, 呈现出强劲的增长趋势, 并带有一定的季节波动性, 每年年末达到全年最低值, 次年初反弹.

为精确剖析房价的季节特征, 继续运行 proc X12 过程识别 regARIMA 模型. 根据 transform 语句的运行结果和最小信息量 AIC, SBC 准则, 确定对序列进行对数变换并采用乘法季节调整模型. 利用 identify 语句, 综合比较 X-12-ARIMA 模型中非季节差分次数 $d=1, 2$ 和季节差分次数 $D=0, 1, 2$ 所有组合的自相关函数 ACF 和偏自相关函数 PACF, 结合残差序列自相关函数的 χ^2 检验, 最终确定 $d=2, D=1$. 再次运行

proc X12 程序,利用 arima 语句 model 命令的形式变换,对 p, q, P, Q 分别取值于 0、1、2 时的所有组合模型,进行参数估计与拟合优度检验^[12],确定最佳模型为 $ARIMA(0, 2, 2) \times (0, 1, 1)_{12}$:

$$(1-B)^2(1-B^{12})\ln x_t = (1-0.20215B-0.23662B^2)(1-0.64741B^{12})\epsilon_t,$$

(2.02) (2.37) (8.07)

其中,括号内是对应参数的 t 值,上述模型中的参数均通过了显著性检验。

X12 程序的输出结果非常丰富,限于文章篇幅,本文只列出经过 x-12-ARIMA 过程季节调整后的房价序列(记为 y_t ,见表 1)和最终的季节因子时序图(图 2),图 2 证实房价序列确实存在季节波动性,且波峰波谷相对稳定,波谷为年末,12 月降至全年最低点,11 月次之,波峰为每年的 1 月,这与根据图 1 初步判断的房价季节波动特征相吻合。进一步观察图 2 可知,从 2007 年至今,波峰和波谷的季节指数明显下降,分别由 2007 年的 3.866、-5.304 降至 2014 年的 2.941、-7.366,但波动幅度持续加大,年度极差从 2.38% 增加到 3.54%,这符合北京市愈来愈强的实际房价波动状况。

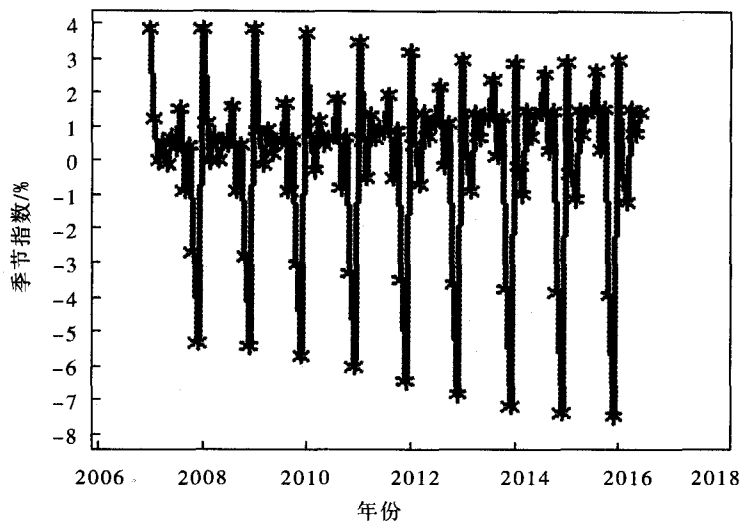


图2 北京市新建住宅价格季节指数时序图

3.2 对季节调整后的房价序列拟合 AR-GARCH 模型

对表 1 中的 y_t 序列进行 ADF 和 PP 平稳性检验,结果显示序列在无漂移项自回归、有漂移项自回归和带趋势项回归三种类型下都显著非平稳,且各种处理均无法实现残差平稳。故对 y_t 序列进行差分运算(记为 $d_{if} y_t$),ADF 和 PP 检验结果表明 $d_{if} y_t$ 序列在 0.05 的水平下显著平稳,进一步利用 SAS 软件的 archtest 命令检验其异方差性,PQ 检验和 LM 检验结果显示 $d_{if} y_t$ 序列具有较高阶的 ARCH 效应,适合采用 ARCH 类模型进行分析。

表 1 季节调整后的北京市新建住宅价格指数 y_t 序列

年份	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
2007	104.51	106.03	106.96	107.63	109.07	110.48	110.69	109.85	111.15	110.24	112.95	115.70
2008	115.19	116.57	117.59	119.09	120.17	121.67	123.52	124.78	128.07	129.89	132.28	135.10
2009	135.69	136.19	137.57	138.16	138.85	139.19	139.49	139.48	139.06	138.79	137.90	137.17
2010	136.44	136.53	136.63	137.00	137.80	138.60	140.45	142.36	143.96	144.67	148.96	154.89
2011	159.07	159.87	162.76	166.55	168.01	168.22	168.70	169.01	169.78	169.89	170.06	170.03
2012	170.49	170.92	170.87	171.18	171.46	171.70	171.72	172.11	172.43	172.57	172.29	172.10
2013	170.81	170.52	169.75	169.46	169.47	169.84	170.30	170.78	171.32	172.07	173.62	175.07
2014	176.69	180.93	184.36	187.02	189.45	191.89	194.35	196.40	198.57	200.31	201.42	202.16
2015	203.05	203.16	203.39	203.78	204.09	204.16	-	-	-	-	-	-

由于 SAS 软件没有提供识别相对最优 ARCH 模型的命令,本文在综合考虑残差序列自相关性、AIC 和 SBC 信息量、模型可决系数 R^2 和参数能否通过显著性检验的基础上,经反复尝试,最终对 $d_{if} y_t$ 序列拟合了关于延迟因变量回归且无回归常数项的 $AR(2)$ -GARCH(1,1)模型,输出结果见图 3,模型口径为:

$$\begin{cases} y_t - y_{t-1} = 0.8967(y_{t-1} - y_{t-2}) + v_t, \\ v_t = -0.4857v_{t-1} - 0.2169v_{t-2} + u_t, \\ u_t = \sqrt{h_t}\varepsilon_t, \varepsilon_t \stackrel{iid}{\sim} N(0, 0.96734), \\ h_t = 0.1093 + 0.8768\varepsilon_{t-1}^2 + 0.1838h_{t-1} + 0.1838h_{t-1}. \end{cases}$$

The AUTOREG Procedure

GARCH Estimates

SSE	96.7336936	Observations	100
MSE	0.96734	Uncond Var	.
Log Likelihood	-123.41164	Total R-Square	0.90316
SBC	274.454306	AIC	258.823285
Normality Test	15.99435	Pr > ChiSq	0.2960

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
lagdifyt	1	0.8967	0.0310	28.92	<.0001
AR1	1	0.4857	0.1680	2.89	0.0038
AR2	1	0.2169	0.1151	1.88	0.0595
ARCH0	1	0.1093	0.0771	1.42	0.1565
ARCH1	1	0.8768	0.3387	2.59	0.0096
GARCH1	1	0.1838	0.1777	1.84	0.0464

图3 $d_{it}y_t$ 序列拟合GARCH模型输出结果

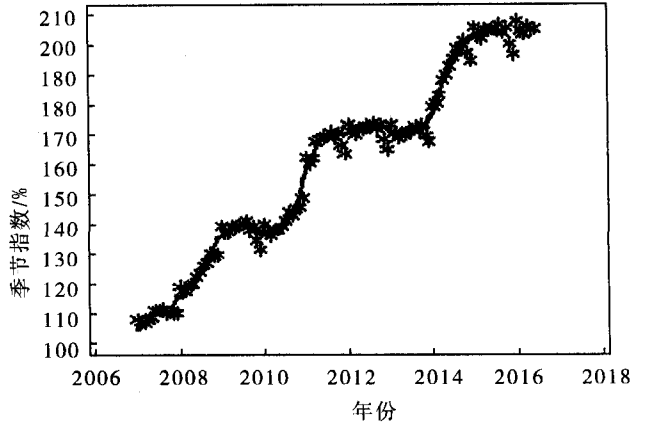


图4 北京市新建住宅价格拟合值与原始值对比图

检验结果显示除 AR(2)-GARCH(1,1)模型中的常数项不显著外,其它变量均高度显著,且正态性检验的概率 P 值 0.2960 远远大于显著性水平 0.05,与假定 GARCH 模型残差项服从正态分布相吻合,模型拟合有效.

3.3 对模型拟合效果的评价研究

为检验模型的拟合效果,本文将样本区间修订为 2007 年 1 月至 2014 年 6 月,综合使用上述模型进行外推预测,具体方案是首先利用 AR(2)-GARCH(1,1)模型对季节调整后的房价 y_t 序列进行拟合,然后把拟合结果回代到 X-12-ARIMA 过程中,结合 X-12-ARIMA 过程分解的季节因子、趋势成分和不规则因子,最终给出 2014 年 7 月至 2015 年 6 月北京市新建住宅价格 x_t 序列的拟合值,结果见表 2.对拟合值与预留房价数据进行对比,并计算不受量纲影响的相对误差指标数值,结果显示,相对误差值均控制在 0.35% 以内,其中,2015 年 6 月的房价真实值为 205.6,拟合值为 205.69,绝对误差只有 0.09,相对误差不超过 0.05%.图 4 进一步输出了北京市新建住宅价格拟合值和原始序列值的对比图,两曲线几乎重合,拟合效果比较理想.

表 2 基于 X-12-ARIMA 和 AR(2)-GARCH(1,1)复合模型和残差自回归模型二的房价拟合效果比较表

时间	基于复合模型的房价拟合数据			基于残差自回归模型二
	真实值	拟合值	相对误差/%	的相对误差/%
2014-07	195.8	196.291	0.251	3.39
2014-08	199.0	199.396	0.199	1.36
2014-09	198.9	199.005	0.053	2.77
2014-10	201.8	201.909	0.054	0.63
2014-11	197.6	198.110	0.258	3.79
2014-12	194.8	195.390	0.303	0.52
2015-01	206.0	206.111	0.054	1.89
2015-02	202.8	203.504	0.347	1.65
2015-03	202.3	202.547	0.122	3.02
2015-04	205.3	205.419	0.058	3.21
2015-05	204.9	205.029	0.063	1.77
2015-06	205.6	205.690	0.044	4.79

3.4 对不同建模方法的比较分析与房价预测

鉴于拟合时间序列数据的有效模型是非唯一的,且在文献[7]中,笔者曾分别创建了以时间变量 t 和以延迟因变量为因子的残差自回归模型、ARIMA 模型、以及 AR(3)-ARCH(1)四类模型拟合 2009 年 3 月至

2013 年 3 月的郑州市新建住宅价格指数序列,并通过综合对比分析,最终确定了最优拟合模型.本文结合北京市新建住宅价格指数 x_t 序列的基本特征——显著的长期递增趋势和一定的季节波动规律,尝试拟合各类时间序列模型^[7,11-12],结果见表 3.根据检验统计量易知,以延迟因变量为因子的残差自回归模型二是相对最优模型.由于 SAS 系统没有直接输出上述复合模型的 AIC, SBC 和 R^2 等统计量数值,故对模型二,也采取前文 3.3 中的拟合效果检验方法——直接计算 x_t 序列拟合值与预留数据的相对误差,并与表 2 中基于复合模型得出的相对误差进行对比,结果发现,基于残差自回归模型二的相对误差值均超过了 0.5%,最大值为 4.79%,平均高达 2.879%,显著大于由复合模型得出的相对误差.

综上所述,拟合北京市新建住宅价格指数序列的最优模型是由 X-12-ARIMA 模型和 AR(2)-GARCH (1,1)模型构建的复合模型,表 4 列出了利用该模型预测的 2015 年 7 月至 2016 年 6 月 x_t 序列定基比指数,总体来看,短期内房价仍将处于上升态势,通过把定基比指数还原为同比指数计算可得,房价上升幅度保持在 3%~5%左右,比上一年的增幅略小,但依然维持波峰波谷分别为年初和年末的季节性特征不变.

表 3 对北京市新建住宅价格指数 x_t 序列拟合的各类时间序列模型

模型	理论模型结构	拟合模型表达式	检验统计量		
			AIC	SBC	R^2
残差自回归模型一:以时间变量为因子	$\begin{cases} x_t = T_t + S_t + \epsilon_t, \\ T_t = \beta_0 + \beta_1 t + \dots + \beta_k t^k, \\ \epsilon_t = \phi_1 \epsilon_{t-1} + \dots + \phi_p \epsilon_{t-p} + a_t. \end{cases}$	$\begin{cases} x_t - S_t = 0.1575t + \epsilon_t, \\ \epsilon_t = 1.1585\epsilon_{t-1} - 0.2342\epsilon_{t-5} + a_t. \end{cases}$	324.59	332.09	0.9923
残差自回归模型二:以延迟因变量为因子	$\begin{cases} x_t = T_t + S_t + \epsilon_t, \\ T_t = \beta_0 + \beta_1 x_{t-1} + \dots + \beta_k x_{t-k}, \\ \epsilon_t = \phi_1 \epsilon_{t-1} + \dots + \phi_p \epsilon_{t-p} + a_t. \end{cases}$	$\begin{cases} x_t - S_t = 0.9121(x_{t-1} - S_{t-1}) + \epsilon_t, \\ \epsilon_t = 0.8343\epsilon_{t-1} + a_t. \end{cases}$	313.34	318.32	0.9925
模型三:加法季节模型	$(1-B)^d(1-B^s)^D x_t = \frac{\theta(B)}{\phi(B)} \epsilon_t.$	$(1-B)(1-B^{12})x_t = (1-0.32907B^2 - 0.4477B^{12})\epsilon_t.$	348.06	353.02	-
模型四:异方差 AR-GARCH 模型	$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + v_t, \\ v_t = \sum_{k=1}^m \beta_k v_{t-k} + u_t, \\ u_t = \sqrt{h_t} \epsilon_t, \\ h_t = \omega + \sum_{i=1}^p \eta_i h_{t-i} + \sum_{j=1}^q \lambda_j \epsilon_{t-j}^2. \end{cases}$	$\begin{cases} x_t = 106.78 + 0.9696t + v_t, \\ v_t = 0.8719v_{t-1} + u_t, \\ u_t = \sqrt{h_t} \epsilon_t, \epsilon_t \stackrel{iid}{\sim} N(0, 11.056), \\ h_t = 8.423 + 0.2903\epsilon_{t-1}^2. \end{cases}$	541.51	554.64	0.9864

注:模型一和模型二中的季节因子 S_t 根据文献[11]中的方法计算得到,本文没有具体列出 S_t 序列.另外, SAS 系统不直接输出模型三的可决系数 R^2 变量.

表 4 北京市新建住宅价格定基指数预测值(2015-07-2016-06)

2015 年	7 月	8 月	9 月	10 月	11 月	12 月
预测值	206.82	208.897	207.99	210.647	206.185	203.139
2016 年	1 月	2 月	3 月	4 月	5 月	6 月
预测值	211.738	208.194	210.242	209.848	209.917	211.408

4 结 语

本文以 SAS 软件为工具,对北京市新建住宅价格指数序列拟合了不同的时序模型,实证研究和比较分析表明,基于 X-12-ARIMA 模型和 AR(2)-GARCH(1,1)模型的复合模型是最优拟合模型,房价序列存在明显的季节波动性,波峰波谷分别稳定在每年的 1 月和 12 月, X-12 季节调整技术操作方便、效果显著.房价序列还存在典型的波动聚集性, AR(2)-GARCH(1,1)模型中两个系数 λ_1 和 η_1 均在 0.05 的水平下显著,且二者之和为 1.0606,非常接近于 1,说明房地产市场的外部因素和房价序列的滞后项对房价的作用效应较

强,且具有一定的持续性,影响的消失速度缓慢.模型中 λ_1 远大于 η_1 ,表明与房价相关的外在变量,如居民消费水平CPI、地区生产总值GDP、土地价格和房地产开发投资额等,对房价的影响程度显著强于房价自身的波动冲击,体现了房地产市场的不确定性和各经济指标之间的关联性.

基于此,建议政府有关部门应及时、准确地把握国内外经济的发展动向,设计以房价长期波动为切入点的调控政策,减少或避免急刹车式的短期措施,加大相关经济指标的调控力度,以促进房地产市场的健康、稳定和可持续发展.

参 考 文 献

- [1] 李仲飞,张 浩.成本推动、需求拉动——什么推动了中国房价上涨?[J].中国管理科学,2015,23(5):143-149.
- [2] 王立平,申建文.房价变动的一个理论解析——基于时变参数模型的实证[J].华东经济管理,2014,28(10):87-91.
- [3] 张宇青,周应恒,易中懿.经济预警指数、国房景气指数与CPI指数波动溢出实证分析——基于三元VAR-ARCH-BEKK模型[J].统计与信息论坛,2014,29(3):36-41.
- [4] 顾 莹,夏乐天.加权马尔可夫链在房价预测中的应用[J].重庆理工大学学报(自然科学版),2013,27(8):125-130.
- [5] 赵华平,张所地.城市宜居性特征对商品住宅价格的影响分析——基于中国35个大中城市静态和动态空间面板模型的实证研究[J].数理统计与管理,2013,32(4):706-717.
- [6] 葛 龙.基于GARCH和COPULA模型的天津房地产市场预测[D].天津:天津大学,2008.
- [7] 聂淑媛,武新乾.基于时间序列模型的河南省房地产价格研究[J].郑州大学学报(理学版),2014,46(1):58-62.
- [8] Findley D F, Monsell B C, Bell W R, et al. New capabilities and methods of the X-12-ARIMA seasonal adjustment program[J]. Journal of Business and Economic Statistics, 1998, 16:127-177.
- [9] 李 剑,宋长鸣,项朝阳.中国粮食价格波动特征研究——基于X-12-ARIMA模型和ARCH类模型[J].统计与信息论坛,2013,28(6):16-20.
- [10] 吴 岚,朱 莉,龚小彪.基于季节调整技术的我国物价波动实证研究[J].统计研究,2012,29(9):61-65.
- [11] 王 燕.应用时间序列分析[M].北京:中国人民大学出版社,2012.
- [12] 谢龙汉,尚 涛.SAS统计分析与数据挖掘[M].北京:电子工业出版社,2012.

Analysis on the Housing Price Fluctuation Based on the X-12-ARIMA Model and AR-GARCH Model

NIE Shuyuan

(College of Mathematical and Science, Luoyang Normal University, Luoyang 471934, China)

Abstract: Based on SAS software, the paper empirically studies the newly built housing price indices of Beijing from Jan. 2007 to Jun. 2015 by establishing all kinds of time series models. The study results show that the optimal fitting model is the compound model of X-12-ARIMA model and AR(2)-GARCH(1,1) model. There is the clear seasonal fluctuation and significant clustered volatility in housing price series. The whole effect of carrying on seasonal adjustment by X-12 method is obvious and the heteroskedastic model is particularly valid. The relative error of forecasting is less than 0.4%. The results of the short-term forecasting indicate that the housing price will rise in the range of 3%-5% in the near future. Meanwhile, the influence degree of the external factor on the housing price is distinctly stronger than that of the internal impact of the housing price series itself.

Keywords: housing price; X-12-ARIMA model; AR-GARCH model; seasonal adjustment