

Research on the Price of Real Estate in China

—Based on ε -TSVR Model and VAR Model

Lingling Xie, Yu Zhang

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: xielingling1992@163.com

Received: Aug. 29th, 2015; accepted: Sep. 12th, 2015; published: Sep. 18th, 2015

Copyright © 2015 by authors 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

Abstract

China's price of real estate forecasts has been a hot livelihood issue, and scholars have paid much attention to it. In this paper, the monthly data of the national 2005-2013 years, in the study of China's real estate prices, are based on the use of vector auto regression VAR model and support vector regression (ε -TSVR) model, respectively to predict and compare the Chinese real estate prices. The results show that the average absolute error (MAE), the average absolute percentage error (MPE), the root mean square error (RMSE) value of the of ε -TSVR model are less than VAR model, which shows that the ε -TSVR model has better forecasting effects on the real estate prices in China.

Keywords

Real Estate Price, VAR Model, ε -TSVR Model

中国房地产价格的研究

—基于 ε -TSVR模型和VAR模型

谢玲玲, 张 雨

云南财经大学统计与数学学院, 云南 昆明
Email: xielingling1992@163.com

收稿日期: 2015年8月29日; 录用日期: 2015年9月12日; 发布日期: 2015年9月18日

摘要

中国的房地产价格的有效预测一直是民生的热点问题, 从而备受国内外学者的高度关注。本文选用全国2005~2013年的月度数据, 在研究中国房地产价格基础上, 采用向量自回归(VAR)模型和支持向量回归(ϵ -TSVR)模型, 分别对中国房地产价格进行预测, 并比较。研究结论表明: ϵ -TSVR模型的平均绝对误差MAE、平均绝对百分比误差MPE、均方根误差RMSE值都小于VAR模型, 说明 ϵ -TSVR模型对中国房地产价格的预测效果更佳, 在房价的预测中有较强的科学性和可行性。

关键词

模板房地产价格, VAR模型, ϵ -TSVR模型

1. 引言

房地产业俨然已成为中国国民经济的支柱性产业, 与经济发展, 社会生产及其他经济活动紧密相关。在今天的中国, 房地产行业带动了大量相关产业的迅猛发展。然而, 随着房地产业的迅速发展, 中国的房地产市场泡沫也越来越大, 从而中国的房价波动烈, 对中国经济的健康持续稳定发展有着重大不利的影响, 同时, 对人民群众的生活也而产生了较大大的影响作用。中国房价一直持续上升, 加重了居民购房的负担。因此, 对中国房地产价格进行预测不仅有利于房地产业与全面国民经济的持续、平稳、协调发展的实现, 具有重要的现实意义, 而且也已相当迫切。本文将利用历史数据构建合适的中国房地产价格预测模型, 综合考虑影响因素对中国房地产的价格进行预测分析, 以此判断未来中国房地产价格走势, 成为当下房地产价格研究中的重要任务。通过构造正确的房地产价格预测模型, 来判断未来的房地产业形势, 稳定房地产价格, 也就成为中国出台一系列调控房价的新政策的重要任务。

在房地产价格研究领域, 王彬以北京房地产价格为研究对象, 利用回归方程寻找影响房地产价格的主要因素, 研究表明人均收入和投资额是其主要影响因素[1]; 罗玉波运用分位数回归研究影响房价的因素[2]; 李勇、王有贵采用状态空间模型来分析影响中国房地产价格变动的主要因素[3]; 李芳、李秋娟运用马尔柯夫区制转换模型的自回归模型(MS-VAR)选取 2005~2012 年人民币汇率和房地产价格的月度数据研究人民币汇率与房地产价格的互动关系[4]。在自回归模型的运用研究领域, 李成、马国校利用月度银行间同业拆借利率数据建立自回归模型研究我国银行同业拆借市场问题[5]; 张卫平选取我国通货膨胀指数、货币量的季度数据, 建立自回归模型以研究其在预测我国通货膨胀的有效性[6]; 刘晓曙、郑振龙利用自回归模型验证商业银行的预测能力[7]。在支持向量回归的研究文献中, 彭显刚、王洪森将双子支持向量回归模型运用在电力系统的负荷预测中[8]; 曹慧运用支持向量回归机研究影响中国物价波动的主要因素[9]; 鲍漪澜运用支持向量机研究金融时间序列, 并对市场波动进行预测[10]。

本文主要选取 2005~2013 年月度住宅商品房平均销售价格数据, 综合考虑影响中国房地产价格的主要因素, 如利率、流动性过剩和国家财政预算收入。添加相关变量, 利用 ϵ -双子支持向量回归模型和向量自回归模型对中国的房地产价格进行预测, 并利用 2013 年 1 月~12 月的相关数据进行验证, 突出两个模型的预测效果。

2. 房地产价格影响因素

近几年, 随着房地产价格日渐上升, 房价涨幅迅速, 给人民生活带来了沉重的经济负担和压力, 影响了人民群众的生活水平, 据此分析影响房价的因素, 可以使得政府针对房地产行业采取切实有效的调

控措施, 以此实现房地产行业健康、持续、稳定的发展。

影响中国房地产价格的影响因素有很多, 本文主要考虑 3 个影响因素: 市场利率、国家财政收入和过剩流动性(EL)。

通过供给和需求两方面进行相互传导是利率变动对房地产价格的最主要影响方式。首先, 利率变化会对房地产市场的需求产生直接影响, 并且通过收入效应及替代效应两种转换方式完成。此外, 就投资性需求或者投机性需求, 利率上升时, 由于贷款融资产生的财务费用加上因为转手产生的交易费用, 房地产价格的上涨幅度必须存在一定程度的提高才能确保本金的收回。其次, 在供给方面, 利率直接反映了资金使用成本。若贷款利率不断增加时, 开发商的金融成本也随之增高, 但在物价没有上升的情况下, 资金的利息成本不能通过合理的方式转移到房价上, 房地产开发商的收益大幅减小, 开发商就会控制对房地产的投资规模, 导致房地产供给量下降。这样, 利率的变化通过供需双方的传导间接地影响了房地产的价格。

过剩流动性全称是指“流动性资产过剩”, 也就是流通中的的现金(M0)、狭义货币供应量(M1)及广义货币供应量(M2)等因素的过剩。简单的说, 政府采取扩张性货币政策时, 使货币供应量增加, 从而导致利率下降, 外币计价的存款收益高于国内本币存款的收益吸引力, 本币贬值, 国内房价与国外房价相比有所下降, 国外资流进购置国内房产, 增加了国内房地产需求, 导致国内房地产价格上升。

国家财政收入主要分两部分, 一部分是税收, 另一部分是以土地为租让金为核心的资源性收益。所以, 国家财政收入的高低直接影响房地产价格。

3. 模型介绍

3.1. ε -双子支持向量回归(ε -TSVR)

本文引入了一种新的支持向量回归方法, ε -双子支持向量回归(ε -TSVR)。 ε -TSVR 将先求得两个函数:

$$f_1(x) = w_1^T x + b_1 \quad (1)$$

$$f_2(x) = w_2^T x + b_2 \quad (2)$$

其中, x 为输入向量, $x_i \in R^n$, $w_1 \in R^n$, $w_2 \in R^n$, $b_1 \in R$ 且 $b_2 \in R$ 。由这两个函数的均值构成最终的回归函数, 设一共有 l 个训练样本, 在计算 $f_1(x)$ 和 $f_2(x)$ 时, ε -TSVR 分别极小化成如下的经验风险函数:

$$R_{emp}^{\varepsilon_1} [f_1] = \sum_{i=1}^l \max \left\{ 0, (y_i - f_1(x_i))^2 \right\} + c_1 \sum_{i=1}^l \max \left\{ 0, -(y_i - f_1(x_i) + \varepsilon_1) \right\} \quad (3)$$

和

$$R_{emp}^{\varepsilon_2} [f_2] = \sum_{i=1}^l \max \left\{ 0, (f_2(x_i) - y_i)^2 \right\} + c_2 \sum_{i=1}^l \max \left\{ 0, -(f_2(x_i) - y_i + \varepsilon_2) \right\} \quad (4)$$

其中, $c_1 > 0$ 和 $c_2 > 0$ 是参数, $\sum_{i=1}^l \max \left\{ 0, -(y_i - f_1(x_i) + \varepsilon_1) \right\}$ 和 $\sum_{i=1}^l \max \left\{ 0, -(f_2(x_i) - y_i + \varepsilon_2) \right\}$ 是损失函数。于是引入松弛变量 ξ , ξ^* , η 和 η^* , 通过引入正则化项 $\frac{1}{2}(w_1^T w_1 + b_1^2)$ 和 $\frac{1}{2}(w_2^T w_2 + b_2^2)$, ε -TSVR 的原始最优化问题可以表示为:

$$\begin{aligned} \min_{w_1, b_1, \xi, \xi^*} & \frac{1}{2} c_3 (w_1^T w_1 + b_1^2) + \frac{1}{2} \xi^{*T} \xi^* + c_1 e^T \xi, \\ \text{s.t.} & Y - (Aw_1 + eb_1) = \xi^*, \\ & Y - (Aw_1 + eb_1) \geq -\varepsilon_1 e - \xi, \quad \xi \geq 0, \end{aligned} \quad (5)$$

和

$$\begin{aligned} \min_{w_2, b_2, \eta, \eta^*} & \frac{1}{2}c_4(w_2^T w_2 + b_2^2) + \frac{1}{2}\eta^{*\top} \eta^* + c_2 e^T \eta, \\ \text{s.t.} & (Aw_2 + eb_2) - Y = \eta^*, \\ & (Aw_2 + eb_2) - Y \geq -\varepsilon_2 e - \eta, \quad \eta \geq 0, \end{aligned} \quad (6)$$

其中, $c_1, c_2, \varepsilon_1, \varepsilon_2$ 是正参数。

为解决式(5)和(6)的问题, 需要得到它们的对偶问题。

关于(5)式的拉格朗日问题由下式给出:

$$\begin{aligned} L(w_1, b_1, \xi, \alpha, \beta) &= \frac{1}{2}(Y - (Aw_1 + eb_1))^T (Y - (Aw_1 + eb_1)) + \frac{1}{2}c_3(\|w_1\|^2 + b_1^2) \\ &+ c_1 e^T \xi - \alpha^T (Y - (Aw_1 + eb_1) + \varepsilon_1 e + \xi) - \beta^T \xi \end{aligned} \quad (7)$$

其中, $\alpha = (\alpha_1, \dots, \alpha_l)$ 和 $\beta = (\beta_1, \dots, \beta_l)$ 是 Lagrange 乘法式子的向量, $w_1, b_1, \xi, \alpha, \beta$ 的条件由下式给出:

$$-A^T (Y - Aw_1 - eb_1) + c_3 w_1 + A^T \alpha = 0, \quad (8)$$

$$-e^T (Y - Aw_1 + eb_1) + c_3 b_1 + e^T \alpha = 0, \quad (9)$$

$$c_1 e - \beta - \alpha = 0, \quad (10)$$

$$Y - (Aw_1 + eb_1) \geq -\varepsilon_1 e - \xi, \quad \xi \geq 0, \quad (11)$$

$$\alpha^T (Y - (Aw_1 + eb_1) + \varepsilon_1 e + \xi) = 0, \quad \beta^T \xi = 0 \quad (12)$$

$$\alpha \geq 0, \beta \geq 0. \quad (13)$$

从式(9)中的 $\beta \geq 0$ 我们可以得出 $0 \leq \alpha \leq c_1 e$, 显然可以从式(7)和(8)中得出:

$$-\begin{bmatrix} A^T \\ e^T \end{bmatrix} Y + \left(\begin{bmatrix} A^T \\ e^T \end{bmatrix} \begin{bmatrix} A & e \end{bmatrix} + c_3 I \right) \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} A^T \\ e^T \end{bmatrix} \alpha = 0 \quad (14)$$

定义 $G = [A \ e]$, $v_1 = [w_1^T, b_1]^T$, 式(13)可以表示为:

$$-G^T Y + (G^T G + c_3 I) v_1 + G^T \alpha = 0 \quad (15)$$

或者

$$v_1 = (G^T G + c_3 I)^{-1} G^T (Y - \alpha) \quad (16)$$

将式(16)运用 K.K.T.的条件和拉格朗日公式, 可以得到式(5)的对偶问题:

$$\begin{aligned} \max_{\alpha} & -\frac{1}{2} \alpha^T G (G^T G + c_3 I)^{-1} G^T \alpha^T + Y^T G (G^T G + c_3 I)^{-1} G^T \alpha - (e^T \varepsilon_1 + Y^T) \alpha \\ \text{s.t.} & 0 \leq \alpha \leq c_1 e, \end{aligned} \quad (17)$$

当式(17)中的 c_3 用 σ 代替时, σ 只是一个固定的正数, 所以结构风险最小化原则不能在用 c_3 时完全反映出来。

用同样的方法, 将式(17)代入 Lagrange 函数, 便得到了式(7)的 Wolfe 对偶问题:

$$\begin{aligned} \max_{\gamma} & -\frac{1}{2} \gamma^T G (G^T G + c_4 I)^{-1} G^T \gamma^T + Y^T G (G^T G + c_4 I)^{-1} G^T \gamma - (e^T \varepsilon_2 + Y^T) \gamma \\ \text{s.t.} & 0 \leq \gamma \leq c_2 e, \end{aligned} \quad (18)$$

其中 c 是拉格朗日乘数, 广义增量 $v_2 = [w_2^T \ b_2]^T$ 由下式可以得出:

$$v_2 = (G^T G + c_4 I)^{-1} G^T (Y + \gamma) \quad (19)$$

一旦从式(18)中求出式(5)和(6)的 (w_1, b_1) 和 (w_2, b_2) , 那么函数 $f_1(x)$ 和 $f_2(x)$ 就可以求得解。那么估计的回归量构造如下:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}(w_1 + w_2)^T x + \frac{1}{2}(b_1 + b_2) \quad (20)$$

本文研究的中国房地产价格的数据是非线性的, 由以下两个相近的函数:

$$\begin{aligned} f_1(x) &= K(x^T, A^T) \mu_1 + b_1 \\ f_2(x) &= K(x^T, A^T) \mu_2 + b_2 \end{aligned} \quad (21)$$

其中 K 是核参数, 构建原问题如下:

$$\begin{aligned} \min_{\mu_1, b_1, \xi} & \frac{1}{2} c_3 (\mu_1^T \mu_1 + b_1^2) + \frac{1}{2} \xi^T \xi^* + c_1 e^T \xi, \\ \text{s.t.} & Y - (K(A, A^T) \mu_1 + e b_1) \geq -\varepsilon_1 e - \xi, \quad \xi \geq 0, \\ & Y - (K(A, A^T) \mu_1 + e b_1) = \xi^*, \end{aligned} \quad (22)$$

$$\begin{aligned} \min_{\mu_2, b_2, \eta} & \frac{1}{2} c_4 (\mu_2^T \mu_2 + b_2^2) + \frac{1}{2} \eta^T \eta^* + c_2 e^T \eta, \\ \text{s.t.} & (K(A, A^T) \mu_2 + e b_2) - Y \geq -\varepsilon_2 e - \eta, \quad \eta \geq 0, \\ & (K(A, A^T) \mu_2 + e b_2) - Y = \eta^*. \end{aligned} \quad (23)$$

其中, c_1, c_2, c_3, c_4 是正参数, 用相同的方法可以得到其对偶问题, 如下:

$$\begin{aligned} \max_{\alpha} & -\frac{1}{2} \alpha^T H (H^T H + c_3 I)^{-1} H^T \alpha^T - (Y^T + e^T \varepsilon_1) \alpha + Y^T H (H^T H + c_3 I)^{-1} H^T \alpha \\ \text{s.t.} & 0 \leq \alpha \leq c_1 e, \end{aligned} \quad (24)$$

$$\begin{aligned} \max_{\gamma} & -\frac{1}{2} \gamma^T H (H^T H + c_4 I)^{-1} H^T \gamma^T + (Y^T - e^T \varepsilon_2) \gamma - Y^T H (H^T H + c_4 I)^{-1} H^T \gamma \\ \text{s.t.} & 0 \leq \gamma \leq c_2 e, \end{aligned} \quad (25)$$

其中 $H = [K(A, A^T) \ e]$, 那么广义向量 $v_1 = [\mu_1^T \ b_1]^T$ 和 $v_2 = [\mu_2^T \ b_2]^T$ 可以由下式得:

$$v_1 = (H^T H + c_3 I)^{-1} H^T (Y - \alpha), \quad (26)$$

$$v_2 = (H^T H + c_4 I)^{-1} H^T (Y + \gamma), \quad (27)$$

若计算出式(22)和式(23)中的 (u_1, b_1) 和 (u_2, b_2) , 函数 $f_1(x)$ 和 $f_2(x)$ 便可以获得, 则构造如下估计的回归量:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}(u_1^T + u_2^T) K(A, x) + \frac{1}{2}(b_1 + b_2) \quad (28)$$

3.2. VAR 模型

在预测中国房地产价格方面, 本文构建了 VAR 模型(向量自回归)来做预测研究。向量自回归模型

(Vector Auto Regressive)是 Sims 在十九世纪八十年代提出的。VAR 模型是研究多个相关的经济变量互动关系的常用模型,因而 VAR 模型用来预测相时间序列系统和随机干扰项对系统变量动态冲击的分析,然后理解不同经济冲击对经济变量产生的影响机制。

一般地,考虑一组时间序列变量: $y_{1t}, y_{2t}, \dots, y_{mt}$, 我们将其定义成 $n \times 1$ 维向量 Y_t :

$$Y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{bmatrix}, t=1, 2, \dots, T, \quad T \text{ 为总期数。}$$

则滞后 p 阶的 VAR 模型形式为:

$$Y_t = C + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t \quad (29)$$

C 为 $n \times 1$ 维常数向量, Φ_i 为 $n \times n$ 维自回归系数矩阵, ϕ_{ij} 为其矩阵中的元素。 ε_t 为 $n \times 1$ 维白噪声序列, 其中每个元素均是非自相关的。

一个两变量且滞后期为 1 自回归模型形式为:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (30)$$

其中, $\varepsilon_{1t}, \varepsilon_{2t} \sim N(0, \sigma^2)$, $Cov(\varepsilon_{1t}, \varepsilon_{2t}) = 0$ 。

$$Y_t = C + \Phi_1 Y_{t-1} + \varepsilon_t \quad (31)$$

则滞后 k 期的模型为:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_k Y_{t-k} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Omega) \quad (32)$$

VAR 模型虽然看着比一般模型较为复杂,但是它的求解却不繁琐。常见的, VAR 模型估计求解方法有最大似然估计(MLE)和最小二乘估计(OLS),最大似然估计和最小二乘估计求解得到的各参数是一致的。

4. 实证分析

4.1. 相关数据说明

本文选取的是 2005~2013 年的月度住宅商品房价格为因变量,样本点数量共为 108 个。数据来自 2005 年~2013 年的中国统计年鉴。其中将 2005 年~2012 年的月度数据,共 96 个样本点,用来进行模型的拟合。将剩下的 12 个样本点,2013 年 1 月~12 月的月度住宅商品房价格与用于模型预测的验证和比对。

自变量选取的是 2005~2013 年的月度过剩流动性和全国银行间同业拆借市场利率,国家财政收入。其中全国银行间的同业拆借市场利率可以从中国人民银行官网统计数据栏中获取。过剩流动性通常用广义货币供应量(M2)来表示。广义货币供应量(M2)和国家财政收入的全国月度数据均可从中国国家统计局统计年鉴中获得。

4.2. ε -TSVR 模型拟合

ε -TSVR 模型的预测本文选取的核函数为高斯核函数:

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

在运用 ε -TSVR 模型回归时本文运用十折交叉验证的方法测试算法准确性。每次所做试验都会得到相应正确率。

预测时, 本文利用 2005~2012 年月度的中国房地产价格、银行间同业拆借利率、M2 货币供应量、国家财政收入数据进行分析, 得到模型的拟合值和实际值, 如图 1。

对 ε -TSVR 模型进行拟合准确度的定量分析, 其中常用的判断拟合准确度的性能计算公式如下:

$$\text{SSE (误差平方和): } \text{SSE} = \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$\text{SST (总平方和): } \text{SST} = \sum_{i=1}^m (y_i - \bar{y}_i)^2$$

$$\text{SSR (回归平方和): } \text{SSR} = \sum_{i=1}^m (\hat{y}_i - \bar{y}_i)^2$$

$$\text{NMSE (正则均方误差): } \text{NMSE} = \text{SSE}/\text{SST} = \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2}$$

正则均方误差用来测量估计真实值与预测值之间的偏差。当 NMSE 值愈小时, 预测的时间序列值与实际值愈接近, 即预测效果愈好; NMSE 值越大, 则预测效果越差。

$$R^2 \text{ (决定系数): } R^2 = \text{SSR}/\text{SST} = \frac{\sum_{i=1}^m (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2}$$

进一步对 ε -TSVR 模型进行拟合准确度的定量分析, 判断拟合准确度的性能计算公式的计算结果如表 1。

$R^2 = 0.9918$ 表明在房地产价格的总变差中, 被估计的拟合方程所解释的比例为 99.18%, 即因变量(房地产价格)的 99.18% 可由模型确定。而且, 如图 1 所示, ε -TSVR 模型下实际值和拟合值的曲线完全重叠, 说明模型的拟合度非常高。

4.3. ε -TSVR 模型预测及结果分析

运用支持回归向量机构建模型, 根据 2013 年 1 月~12 月 12 个月度数据得到新的房地产价格的预测值, 预测效果见图 2。可见根据 ε -TSVR 模型得到的拟合值与实际值非常接近, 几乎在一条线上, 预测效果非常好。

进一步对 ε -TSVR 模型进行预测准确度的定量分析, 判断预测准确度的性能计算公式的计算结果如表 2。

$R^2 = 0.999$ 表明在房地产价格的总变差中, 被估计的拟合方程所解释的比例为 99.99%, 即因变量(房地产价格)的 99.99% 可由模型确定。说明模型拟合度较高, 预测效果非常好。

不管是拟合还是对未来房价的预测, ε -TSVR 模型得到的效果都非常好。

4.4. VAR 模型拟合

运用 VAR 构建的模型, 根据 2005~2012 年月度数据得到新的房地产价格的拟合值, 拟合效果见图 3。可见拟合值与实际值总体来说比较接近, 拟合效果较好。

$R^2 = 0.929$ 表明在房地产价格的总变差中, 拟合方程所解释的比例为 92.9%, 即房地产价格的 92.9% 可由模型所确定, 可以看出方程的拟合度较高, 见表 3。

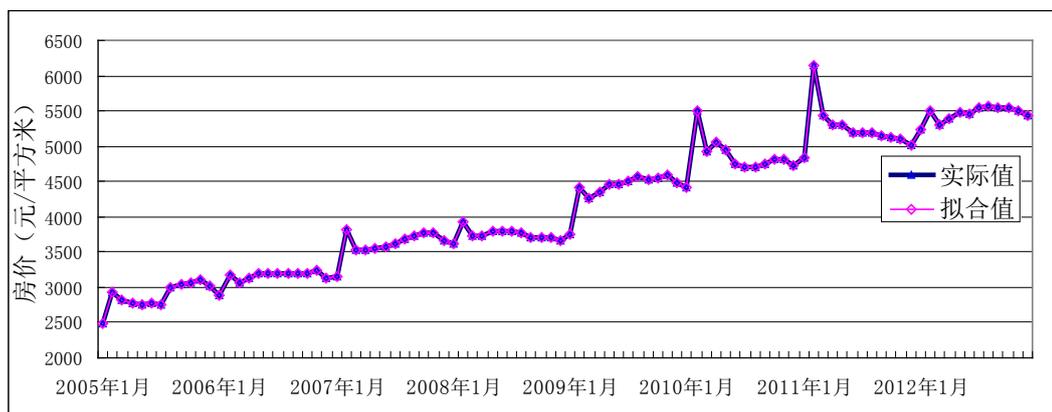


Figure 1. Price data of actual value and the fitting value comparison of ε -TSVR model

图 1. ε -TSVR 模型下房价数据实际值与拟合值比较

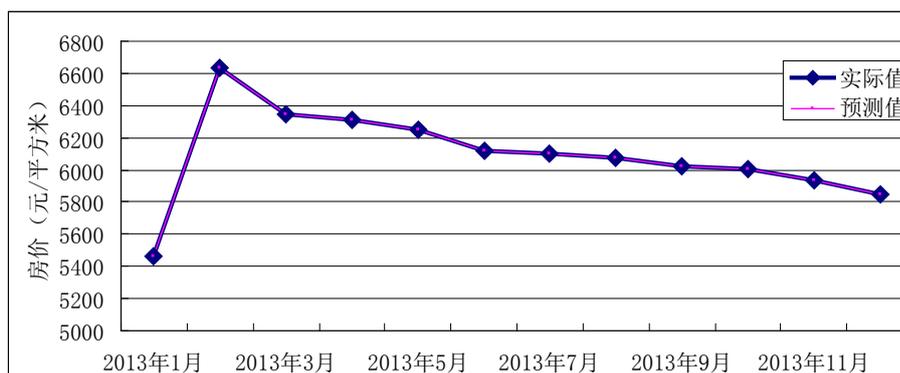


Figure 2. Price data of actual value and the predictive value comparison of ε -TSVR model

图 2. ε -TSVR 模型下的房价预测值与实际值比较

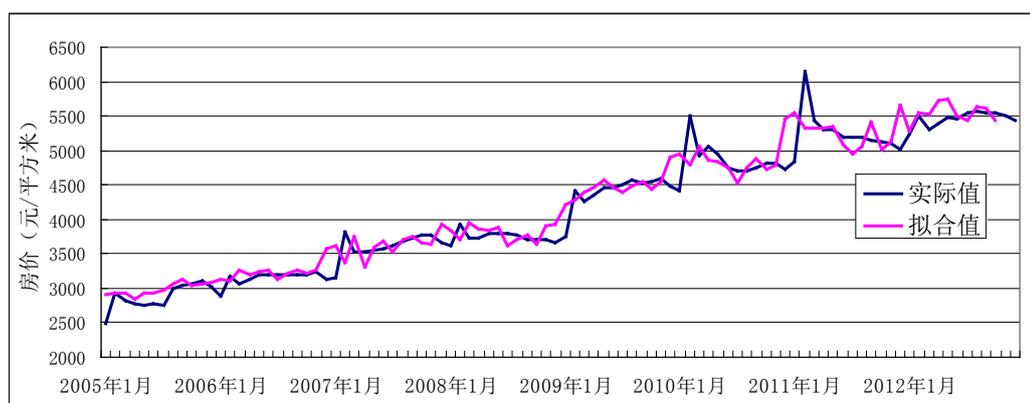


Figure 3. Price data of actual value and the fitting value comparison of VAR model

图 3. VAR 模型房地产价格实际值与拟合值比较

Table 1. Calculating value of the performance formula of fitting accuracy

表 1. 判断拟合准确度的性能公式计算值

SSE	SST	SSR	NMSE	R^2
2058.2	12,236	12,136	1.6821e-005	0.992

Table 2. Calculating value of the performance formula of the forecast accuracy
表 2. 判断预测准确度的性能公式计算值

SSE	SST	SSR	NMSE	R^2
0.016	930,864	930,809	1.69369E-08	0.999

Table 3. Calculating value of the performance formula of fitting accuracy
表 3. 判断拟合准确度的性能公式计算值

SSE	SST	SSR	NMSE	R^2
3,059,321	122,363,978	113,620,073	0.025	0.929

4.5. VAR 模型的检验

进行 Granger 因果检验。平稳性检验能对每一个时间序列进行 AR 根检验。AR 根检验需通过 Eviews 中的 AR Roots Graph 和 AR Roots Table 功能进行判断, 若无特征根在单位圆外则序列平稳, 则通过平稳性检验。

对收集得到的数据作 AR 根图如下。如图 4 可知, 所有根的模的倒数都小于 1, 均位于单位圆内, 说明该模型是稳定的, 以此建立的 VAR 模型有效。

为了得出各变量之间是否存在互相影响的关系, 此处进行 Granger 因果关系检验, 显著性水平设为 0.05, 确定滞后期的值为 2。检验结果如表 4:

分析上表可以看出, 利率、EI、财政收入变量的 F 统计量的概率 P 值均小于显著性水平 5%, 故拒绝原假设。表明利率、EI、财政收入变量会对房地产价格产生影响。

将房地产价格设为 y , 利率为变量 x_1 , EI 为 x_2 , 财政收入为 x_3 , 运用房价、利率、EI、财政收入数据后运行 Eviews 得到的 VAR 模型如下:

$$\ln y = 0.041 \ln x_{1,t-1} - 0.048 \ln x_{1,t-2} + 1.102 \ln x_{2,t-1} - 0.905 \ln x_{1,t-2} + 0.114 \ln x_{3,t-1} - 0.033 \ln x_{3,t-2} + 0.614 \ln y_{t-1} - 0.122 \ln y_{t-2} + 0.943$$

4.6. VAR 模型预测及结果分析

运用 VAR 构建的模型, 根据 2013 年 1 月~12 月 12 个月度数据得到房地产价格预测值, 预测效果见图 5。可见预测值与实际值总体看不是很接近, 预测效果不好。

本文进一步对 VAR 模型进行预测准确度的定量分析, 其中常用的判断预测准确度的性能计算公式如表 5。

$R^2 = 0.492$ 表明在房地产价格的总变差中, 被估计的拟合方程所解释的比例为 49.2%。即因变量(房地产价格)的 49.2% 可由模型确定。说明预测效果不理想。

5. 结论

5.1. ε -TSVR 模型与 VAR 模型比较分析

本文将对 VAR 模型和 ε -TSVR 模型的预测效果进行比较。两个模型分别根据已知变量预测了 2013 年 1~12 月份的中国房地产价格, 见图 6。

图 6 中模型 1 预测值为 VAR 模型下的预测值, 模型 2 预测值为 ε -TSVR 模型下的预测值。从图 6 可以看出两个模型都可以较好的进行时间序列预测, 但是 ε -TSVR 模型预测的效果更接近实际值, 因而 ε -TSVR 模型预测效果更好。

进一步对两个模型进行预测准确度的定量分析, 其中常用的判断预测准确度的性能计算公式如下:

$$\text{MAE (平均绝对误差): } \text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

平均绝对误差利用了全部数据信息, 因此能比较客观的反映变量分布的离散程度。MAE 值越大, 表示变量分布离散程度越大; MAE 越小, 则变量分布离散程度越小。

$$\text{MPE (平均绝对百分误差): } \text{MPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

平均绝对百分误差是衡量模型好坏的一个指标。MAPE 值越大代表预测值与原始值偏差越大, 预测效果越差; MAPE 值越小, 预测值与原始值差别越小, 也即预测效果越好。

$$\text{RMSE (均方根误差): } \text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$

均方根误差也是衡量样本的离散程度的一个指标。RMSE 值越大, 说明离散程度越大; RMSE 值越小, 离散程度则越小。

据公式计算 VAR 和 ε -TSVR 模型下相应的值, 见表 6。

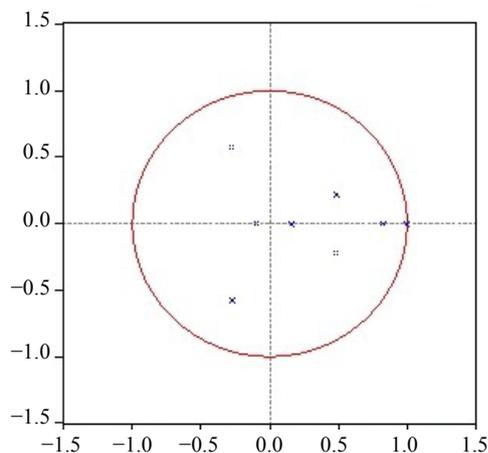


Figure 4. AR root graph

图 4. AR 根图

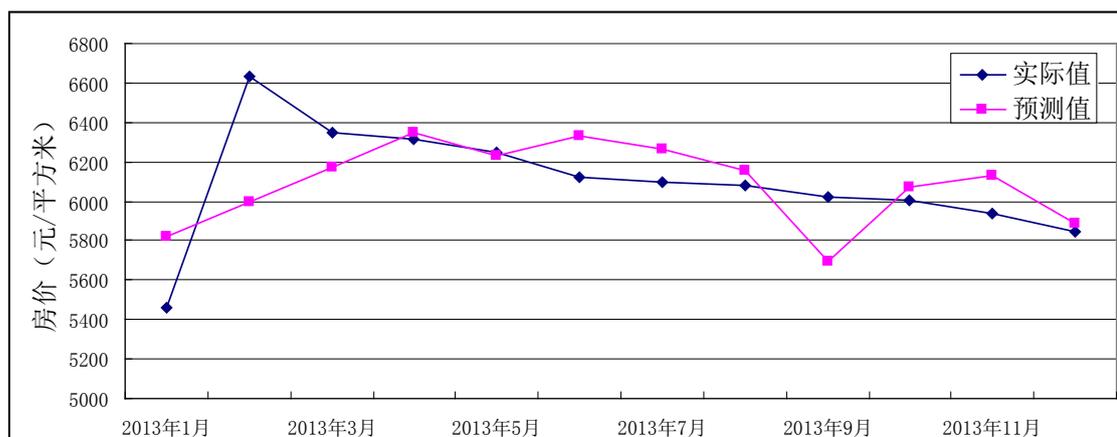


Figure 5. Price data of actual value and the predictive value comparison of VAR model

图 5. VAR 模型下的预测值和实际值的比较

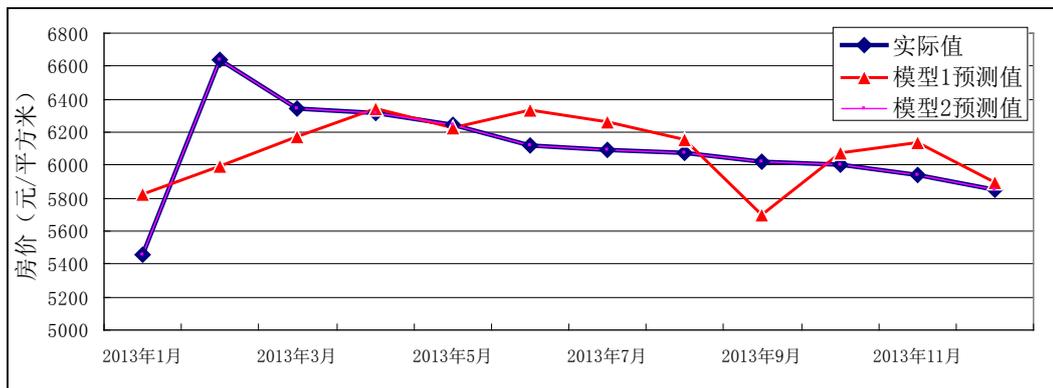


Figure 6. Comparison between the actual value and the predictive value of monthly house price based on VAR model and ε -TSVR model in 2013

图 6. VAR 模型和 ε -TSVR 模型下 2013 年月度的房价预测值与实际值比较

Table 4. Results of Granger causality test

表 4. Granger 因果检验结果

原假设	观察值	F 统计量	P 值	结论
利率不是影响房价的 Granger 原因	106	7.8279	0.002	拒绝
EI 不是影响房价的 Granger 原因	106	16.4442	0.000	拒绝
财政收入不是影响房价的 Granger 原因	106	27.2396	0.000	拒绝

Table 5. Calculating value of the performance formula of fitting accuracy

表 5. 判断拟合准确度的性能公式计算值

SSE	SST	SSR	NMSE	R^2
799,939	930,809	457,770	0.859	0.492

Table 6. The results of the indicators of ε -TSVR model and VAR model

表 6. VAR 和 ε -TSVR 各指标结果

	模型	MAE	MPE	RMSE
房价 I	ε -TSVR	0.025558	6.18508E-06	0.02977
房价 II	VAR	115.966	0.026	169.887

从表中数据的对比可以看出在中国房地产价格的预测中, ε -TSVR 模型的 MAE, MPE, RMSE 值均远远小于 VAR 模型, ε -TSVR 模型对于在房地产价格的预测误差更小, 说明 ε -TSVR 比 VAR 模型拟合的效果更好。

5.2. 总结

本文在影响中国房地产价格的主要因素: 利率、流动性过剩和国家财政收入的基础上, 选取 2005 年 1 月至 2012 年 12 月的单月样本数据, 分别采用向量自回归 VAR 模型和支持向量回归机 ε -TSVR 模型, 对中国 2013 年 1 月~12 月房地产价格进行预测。研究结果表明: ε -TSVR 模型的平均绝对误差(MAE)、平均绝对百分比误差(MPE)、均方根误差(RMSE)值均小于 VAR 模型, 说明 ε -TSVR 模型比 VAR 模型具有更好的预测效果, 预测结果更接近于实际值。

支持向量机是在统计学理论的基础上发展而来的借助于最优化方法来解决机器学习问题机器学习新方法, VAR 模型和 ϵ -TSVR 模型作为新型的支持向量回归机, 都能很好的进行时间序列预测, 但 ϵ -TSVR 模型在预测方面比 VAR 模型效果更好。

致 谢

本文是在云南财经大学统计与数学学院向其凤老师细心指导和建议下完成, 在此对向其凤老师表示衷心的感谢!

参考文献 (References)

- [1] 王彬 (2007) 房地产价格影响因素分析. 硕士论文, 北京交通大学, 北京.
- [2] 罗玉波 (2011) 房价影响因素分析:分位数回归方法. *统计与决策*, **6**, 158-159.
- [3] 李勇, 王有贵 (2011) 基于状态空间模型的中国房价变动的影响因素研究. *南方经济*, **2**, 38-45.
- [4] 李芳, 李秋娟 (2014) 人民币汇率与房地产价格的互动关系. *国际金融研究*, **3**, 86-96.
- [5] 李成, 马国校 (2007) VAR 模型在我国银行同业拆借市场中的应用研究. *金融研究*, **5**, 62-77.
- [6] 张卫平 (2012) 中国通货膨胀预测: 基于 AR 和 VAR 模型比较. *统计与决策*, **4**, 11-15.
- [7] 刘晓曙, 郑振龙 (2007) 商业银行 VAR 模型预测能力的验证. *当代财经*, **8**, 39-43.
- [8] 彭显刚, 王洪森 (2014) 基于竞争 ISPO 双胞支持向量回归短期负荷预测. *电力系统及其自动化学报*, **10**, 46-68.
- [9] 曹慧 (2014) 基于支持向量回归机的中国物价波动影响因素探究. 硕士论文, 浙江工业大学, 杭州.
- [10] 鲍漪澜 (2013) 基于支持向量机的金融时间序列分析预测算法研究. 博士论文, 大连海事大学, 大连.