

## 管理科学

# 基于 SVR 的石油期货价格短期预测

慕晓茜 何 佳 倪旭敏 段雨墨 陆 昊 张作泉\*

(北京交通大学理学院,北京 100044)

**摘要** 将 SVR 原理引入到石油期货价格的时间序列中,并以美原油价格进行了实证分析。结果表明:该方法能充分反映石油期货价格序列走势,对短期价格的预测具有较高的精度;并且还发现,超级参数的选择服从一定规律,即其乘积在一定范围内效果较佳。之后,还将此理论推广到多维影响因素和其他金融时间序列的预测中。

**关键词** 支持向量机回归 石油期货价格 时间序列预测 拐点预测

**中图法分类号** F830.91; **文献标志码** A

石油是人类赖以生存和各国经济发展所必须的资源,世界上恐怕没有任何一种商品可以像石油一样对世界经济、政治乃至国家安全产生如此深刻和广泛的影响。正是由于石油对世界的重要性,加之其天然的稀缺性和不可再生性,石油问题成为了国际社会上永久的话题,石油价格的走势已成为世界各国关注的焦点<sup>[1-3]</sup>。然而,由于影响石油价格和期货价格的众多因素的相互结合,使得石油期货价格的波动趋势尤为复杂,也使得石油期货价格预测变得十分困难,而支持向量机方法的出现为石油期货价格的预测提供了一种新的可能<sup>[4]</sup>。

支持向量机回归(SVR)是数据挖掘中的一个新方法,能非常成功地处理回归问题(时间序列分析)和模式识别(分类问题、判别分析)等诸多问题<sup>[5]</sup>,并可推广于预测和综合评价等领域。

现将 SVR 引入石油期货价格这个金融时间序列,对其整体趋势以及短期走势做了精确的预测;在超级参数的选择上,本文通过收集大量数据进行统计分析,发现核函数参数  $g$  以及惩罚因子  $C$  之间

有很大的相关性。此外,文中的方法还将为多维输入变量以及其他领域的预测提供了借鉴。

## 1 SVR 预测模型

Vapnik 等人根据统计学原理提出了支持向量分类机(SVM)的学习方法<sup>[6]</sup>。SVM 的基本思想主要有两点:(1)先对线性可分情况进行分析,对于线性不可分的情况,可通过核函数的方法,将低维的原始空间映射到高维的特征空间,巧妙地避开了高维空间的复杂运算,使算法的实现成为可能;(2)基于结构风险最小化理论在特征空间中寻找最佳分类面<sup>[7]</sup>,使得学习器得到全局最优化。在 Vapnik 引入  $\varepsilon$  不敏损失函数后,SVM 推广到解决非线性回归问题,称为支持向量机回归<sup>[5]</sup>。

本文采用 SVR 的原理进行预测分析即:当所给数据集是线性可分时,回归问题可转化为求解如下约束条件的优化问题:

$$\begin{aligned} \min = & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t. } & \begin{cases} y_i - [w^T x_i + b] \leq \varepsilon + \xi_i \\ [w^T x_i + b] - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l \end{cases} \end{aligned}$$

由对偶规划理论,上述凸二次规划问题可以转化为求解下述新问题:

2010 年 4 月 8 日收到 大学生创新性实验计划项目资助  
第一作者简介:慕晓茜(1989—),女,信息与计算科学。E-mail:  
07271046@bjtu.edu.cn

\* 通信作者简介:张作泉(1964—),男,副教授,博士,研究方向:计算数学金融数学。E-mail:zqzhang@bjtu.edu.cn。

$$\max w(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)^T (x_i \cdot x_j) (\alpha_j - \alpha_j^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*)$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l \end{cases}$$

由此得到支持向量回归的估计函数为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) (x_i \cdot x_j) + b.$$

而对于非线性回归问题,通过引入核函数

$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ , 将  $(x_i \cdot x_j)$  用  $K(x_i, x_j)$  代替, 得到最终的支持向量回归估计函数为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b.$$

下面给出 SVR 用于金融时间序列预测的一般步骤:

(1) 选取合适的训练集  $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$ , 其中  $x_i \in X = R^n$ ,  $y_i \in Y = R$ ,  $i = 1, 2, \dots, l$ ;

(2) 选择适当的核函数  $K(x_i, x_j)$ 、核函数参数  $\varepsilon$  以及惩罚因子  $C$ ;

(3) 构造并求解最优化问题:

$$\min_{\alpha, \alpha^*} \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)^T K(x_i, x_j) (\alpha_j - \alpha_j^*) +$$

$$\varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \sum_{j=1}^l y_j (\alpha_i - \alpha_i^*)$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l \end{cases}$$

得最优解:  $\alpha = (\alpha_1, \alpha_1^*, \alpha_2, \alpha_2^*, \dots, \alpha_l, \alpha_l^*)^T$ ;

(4) 构造决策函数

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b.$$

## 2 实证分析

### 2.1 输入变量的选择

众所周知, 影响石油期货价格的因素很多, 包括石油期货的历史价格、相关石油产品的期货价

格、汇率变动因素以及由股市反映的世界经济状况等<sup>[4]</sup>。在这诸多因素中, 石油期货的历史价格反映了以往各影响因素的变化对石油价格的影响。对于未来 10 天或一周内的石油期货价格而言, 汇率变动、世界经济情况等因素在通常情况下不会产生特别巨大的变动。所以, 石油期货的历史价格无疑是石油期货现在价格影响最大的一个因素。因此, 我们首先选取美原油期货的历史价格作为一维输入变量, 并对此种石油期货的价格进行预测, 以验证 SVR 理论的正确性。

### 2.2 核函数的选择

在 SVR 的应用中, 常见的核函数有线性核, 多项式核, RBF 核, Sigmoid 核以及多二次曲面核等<sup>[4]</sup>。但是在实践中多使用 RBF 核。因为它可以将输入空间以非线性方式映射到特征空间, 便于处理现实中以非线性方式存在的问题, 且更少遇到数值计算困难。金融时间序列中的实验表明, RBF 预测性能更好并且运行时间更短<sup>[4]</sup>。由于石油期货价格序列呈现非线性且高噪声的特征, 因此, 最终决定选取 RBF 核作为核函数进行预测。

### 2.3 预测结果

#### 2.3.1 整体趋势预测

以 2009 年 5 月 19 日至 2010 年 3 月 16 日美原油期货的价格<sup>①</sup>作为输入变量。为了放大不同价格间的细微变化, 我们均将价格扩大 50 倍进行观察研究。运用 MATLAB 软件进行编程求解, 采用穷举的方式对超级参数  $C, g$  进行搜索, 在找到最优参数后, 采用决策函数进行预测。整体预测结果如图 1 所示。

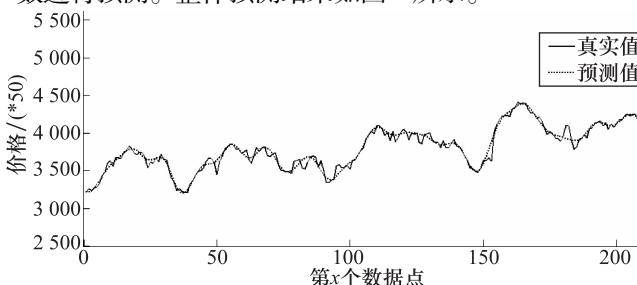


图 1 2009.5.19—2010.3.16 美原油期货价格预测值与  
真实值比较图

① 数据来源:国信期货软件

由图 1 可知,石油期货价格的预测值与真实值是十分接近的。为了更加准确地对预测效果加以评价,我们选用均方误差(mse)作为预测精确度的评价标准。

根据计算公式  $mse = \frac{1}{n} \sum_{i=1}^n (\hat{y} - y_i)^2$ , 得到本次预测的 mse 值为 1.320 932。由此可以看出,SVR 可以很好的预测石油期货的整体价格变化趋势。

### 2.3.2 短期趋势预测

短期趋势和短期内是否出现拐点的信息对于投资者来说是十分重要,所以本文着重研究了这两个问题。

首先,从总体价格走势(图 1)上来看,我们发现在第 40 个数据点左右价格开始平稳上升,从第 170 个数据点左右价格开始平稳下降。因此,我们选取从开始到第 40 个数据点(即 2009.5.19—2009.7.13)的数据以及从第 130 到第 170 个数据点(即 2009.11.17—2010.1.13)的数据分别进行此后一周的预测。预测结果如图 2,图 3 所示。

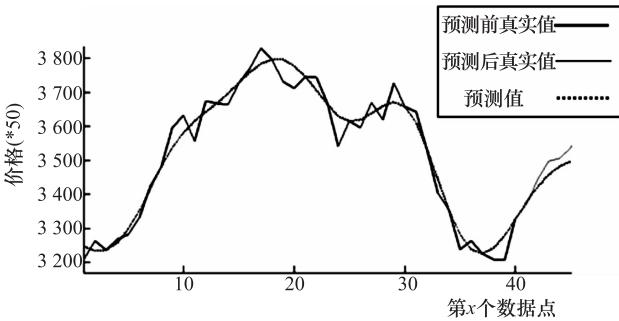


图 2 由 2009.5.19—2009.7.15 的数据得到的未来一周预测图

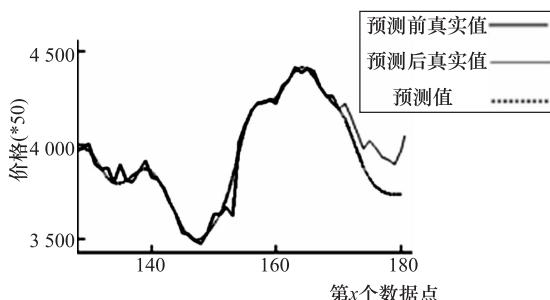


图 3 由 2009.11.17—2010.1.13 的数据得到的未来一周预测图

从图 2,图 3 可以看出,短期预测趋势与实际数据趋势能很好地吻合,因此也说明 SVR 对于短期趋势的预测是比较准确的。

下面我们来看一下拐点的出现对 SVR 的预测性能是否有所影响。从总体价格走势(图 1)上来看,我们发现在第 55 及第 113 个数据点附近均出现拐点,因此,我们选取从第 30 到第 55 个数据点(即 2009.6.28—2009.8.5)的数据以及从第 110 到第 113 个数据点(即 2009.10.23—2009.11.19)的数据分别进行此后一周的预测,预测结果如图 4,图 5 所示。

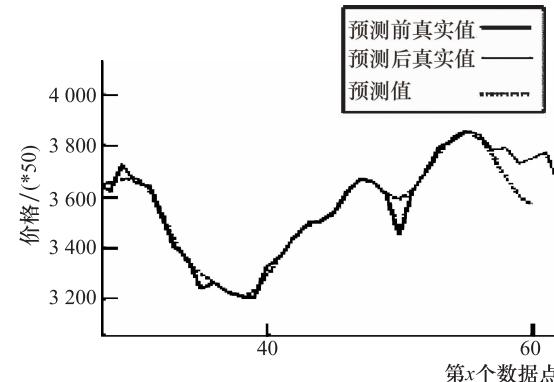


图 4 由 2009.6.28—2009.8.5 的数据得到的未来一周预测图

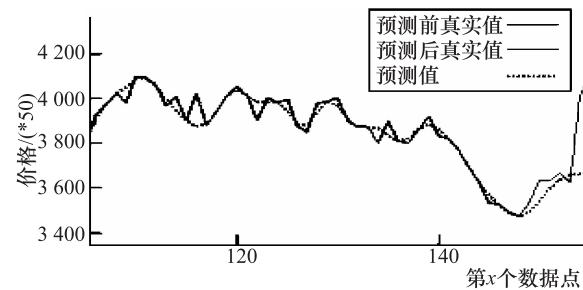


图 5 由 2009.10.23—2009.11.19 的数据得到的未来一周预测图

从图 4,图 5 可以看出,拐点的出现对 SVR 的预测性能并没有太大影响,其短期预测趋势仍与实际数据趋势很好地吻合,更加说明了 SVR 对于短期趋势预测的准确性。

### 2.4 参数分析

根据林生梁资料<sup>[8]</sup>得知核函数参数  $g$  与惩罚因子  $C$  有非常大的相关性。通过用不同的 307 组数

据进行预测,之后经过计算,我们发现  $r = \frac{\lg_2 C}{\lg_2(1/g)}$  以较大的概率落在 1.50—2.50 之间,具体分布如图 6 所示。

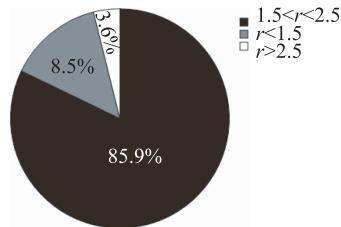


图 6 参数分布图

因此可以认为  $C, g$  满足关系式:

$$1.50 \leq r = \frac{\lg_2 C}{\lg_2(1/g)} \leq 2.50.$$

利用上式,可在穷举得到最优的  $C$  之后,由一定的范围内搜索  $g$ ,以减少程序的运行时间。

## 2.5 SVR 预测的推广

由于 SVR 预测的短期精确性,可将此方法应用到其他金融产品的价格预测上。现以“浦发银行”股票 2008.2.20—2009.6.25 的价格为例,进行未来一周的预测,效果如图 7 所示。由此可见,SVR 的方法不仅局限于石油期货一个领域。

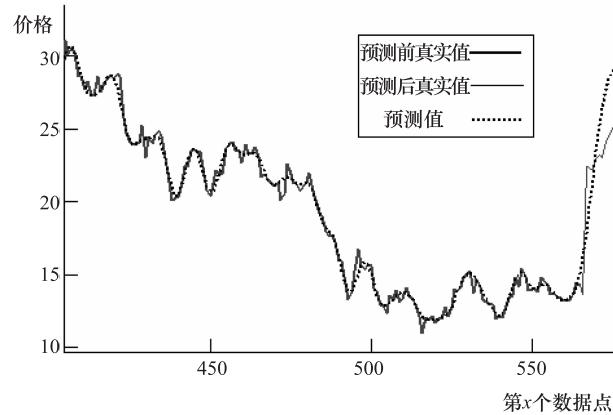


图 7 由 2008.2.20—2009.6.25 的数据得到的未来一周预测图

## 3 讨论与结论

### 3.1 讨论

由于影响石油期货价格的因素很多,只用单一

变量进行输入还是存在较大误差,因此可考虑综合各影响因素后进行多维输入。

对于石油期货的历史价格,将每日收盘原始价格序列转化成相对百分比差价,可以使转换后的数据分布更对称,更接近正态分布,继而可以提高预测性能。因此,根据祝金荣资料<sup>[4]</sup>得出,采用滞后 1 至 4 天的相对百分比价格差(RDP)和 7 天移动平均调整后的收盘价(EMA7)作为第一类输入变量是比较有代表性的,其具体公式如表 1 所示。

表 1 第一类输入变量

变量	计算公式
EMA7	$p(i) - \text{EMA}_7(i)$
RDP—1	$(p(i) - p(i-1)) \times 100/p(i-1)$
RDP—2	$(p(i) - p(i-2)) \times 100/p(i-2)$
RDP—3	$(p(i) - p(i-3)) \times 100/p(i-3)$
RDP—4	$(p(i) - p(i-4)) \times 100/p(i-4)$

注: $p(i)$  为第  $i$  天收盘价,  $\text{EMA}_7(i)$  为第  $i$  天的七天指数移动平均值

除去期货的历史价格外,影响石油期货价格的因素主要有以下三类,分别为:经济环境、汇率变动以及相关能源期货价格。在每一大类中选取三个小类(如表 2 所示),得到 9 个相关因素。之后,利用计量经济学中单位根检验、Granger 检验及协整检验的相关理论,对影响石油期货价格的这 9 类因素进行了相关性分析。最终得到的第二类输入变量为纳斯达克指数,英镑/美元汇率以及天然气价格。

表 2 石油期货价格影响因素表

大类因素	经济环境	汇率变动	相关能源期货价格
道琼斯指数	美元指数	美燃油	
小类因素	纳斯达克指数	英镑/美元	原油内
	标普 500 指数	美元/日元	天然气

因此,可考虑选用以上分析得出的 8 个影响因素作为输入变量,代替之前的 1 维输入,以得到更加精确的结果。

### 3.2 结论

本文尝试将 SVR 引入石油期货价格的金融时间序列中,对其整体趋势以及短期走势做了比较精

确的预测,并且发现,期货价格短期拐点的出现并没有影响预测结果的准确性。此外,本文通过对大量数据进行分析,得到核函数参数  $g$  与惩罚因子  $C$  的关系式  $1.50 \leq r = \frac{\lg_2 C}{\lg_2(1/g)} \leq 2.50$ 。在模型的推广中,尝试将 SVR 的方法引入股票价格的预测中,也得到了较好的效果。但是由于石油期货价格的影响因素十分复杂,本文只分析了价格的影响因素,并没有对多维变量的输入结果进行分析,这也是今后需要继续研究的内容。

### 参 考 文 献

- 1 Farooq M, Mahdi N. Forecasting output using oil prices: a cascaded artificial neural network approach. *Journal of Economics and Business*, 2006;58(2):168—180

- 2 Gori F, Ludovisi D, Cerritelli P F. Forecast of oil price and consumption in the short term under three scenarios: parabolic, linear and chaotic behaviour. *Energy*, 2007;32(7):1291—1296
- 3 Mirmirani S, Li H C. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. *Advances in Econometrics*, 2004;19:203—223
- 4 祝金荣. 石油期货价格预测. 北京:冶金工业出版社,2008:54—75
- 5 邓乃扬,田英杰. 数据挖掘中的新方法—支持向量机. 北京:科学出版社,2005
- 6 Vapnik V N. *Statistical Learning Theory*. New York: Wiley, 1998
- 7 李新战,赵震宇. 支持向量机基础及其应用前景. *科技信息*, 2009;17:39—40
- 8 林升梁,刘志. 基于 RBF 核函数的支持向量机参数选择. *浙江工业大学学报*, 2007;4:163—167

## Petroleum Futures Prices' Short-term Forecasting Based on SVR

MU Xiao-qian, HE Jia, NI Xu-min, DUAN Yu-mo, LU Hao, ZHANG Zuo-quan\*

(Beijing Jiaotong University, Beijing 100044, P. R. China)

**[Abstract]** The support vector regression (SVR) principle is introduced into the petroleum futures forward price and has carried on the empirical analysis by the US crude price. The results indicate that this method can fully reflect the petroleum forward price sequence trend and has the high precision to short term price's forecast. Moreover, super parameter's choice obedience certain rule, namely the majority of its product falls within a certain range. This theory is extended to the multidimensional impact of this factor and other financial time series also.

**[Key words]** support vector regression      petroleum futures prices      time series prediction      turning point prediction