基于灰色支持向量机的基金波动率预测研究 *

耿立艳, 马军海
(天津大学 管理学院, 天津 300072)

摘 要: 鉴于灰色预测方法和支持向量机各自的优点,将灰色预测方法与支持向量机相结合,建立灰色支持向量机模型,并以极差替代收益的标准差度量波动率,运用新模型对深圳基金波动率进行实例分析。通过与 v-支持向量机的预测结果对比,发现所提出的模型适合于基金波动率的中短期预测。

关键词: v-支持向量回归; 灰色支持向量机; 波动率预测

中图分类号: TP18; N941.5 **文献标志码**: A **文章编号**: 1001-3695(2009)07-2471-03

doi:10.3969/j.issn.1001-3695.2009.07.019

Volatility forecasting for fund market using grey support vector machine

GENG Li-yan, MA Jun-hai
(*School of Management, Tianjin University, Tianjin 300072, China*)

Abstract: As for the advantages of grey forecasting method and support vector machine, this paper proposed a grey support vector machine model and applied it to forecasting Shenzhen fund volatility in which volatility was measured using range instead of return's standard deviation. Compared with the forecasting results of v-support vector machine, the new model by this paper was applicable to forecasting short- and middle-term fund volatility.

Key words: v-support vector regression; grey support vector machine; volatility forecasting

0 引言

波动率的预测对于金融资产的定价以及金融风险的测度和管理都有着极其重要的理论和实际意义。目前,波动率预测法可分为两类,即统计方法和人工智能方法。统计方法是一种参数方法,是在数学理论和假设基础上通过演绎推理的方法建立起来的,参数的相关形式已知,通过训练样本估计参数的值。人工智能方法主要包括人工神经网络(ANN)和支持向量机(SVM)。ANN 和 SVM 是数据驱动的非线性非参数模型,在没有任何限制参数的建模假定下,由数据确定模型的结构和参数。它们强大的非线性逼近能力能够降低因假定有误而引起的预测误差。ANN 方法建立在经验风险最小化(ERM)基础上,在实际应用中经常会遇到过拟合、局部极小值和维数灾难等问题。而基于统计学习理论的支持向量机以结构风险最小化(SRM)为原则建模,综合考虑了置信风险和经验风险,适合于小样本问题,在一定程度上克服了统计方法和 ANN 的一些缺陷,在波动率的预测中获得了比 ANN 更好的预测精度[1,2]。傅东升等人[3]以收益率的标准差度量波动率,运用支持向量机预测基金波动率,结果表明,SVM 比 BP 神经网络获得了更好的基金波动率预测值。

灰色预测方法是近年来应用比较广泛的预测方法,其独特的数据处理方式具有削弱原始数列的随机性,挖掘系统演化规律的独特功效,在数据较少的情况下即可得到较高的预测精度,而且它对一般模型具有很强的融合力和渗透力。将灰色预测方法与其他预测模型相结合,能够使预测精度得到极大的改善。唐万梅和牛东晓等人[4,5]分别提出了灰色支持向量机模型,并用于电力系统的短期预测。结果表明,新模型的预测精度优于支持向量机模型、神经网络和灰色预测模型。

本文在前人研究的基础上,将灰色预测方法引入到 v-支持向量机中,建立灰色支持向量机模型。利用新模型预测了深市基金的波动率,并将预测结果与 v-支持向量机进行了对比,分析了它们在预测基金极差波动率方面的适用性。

1 理论模型

在标准的支持向量机回归(ε-SVR)模型中,ε 不敏感系数控制着解的稀疏性和模型的泛化能力,而事先合理地确定 ε 值比较困难。为此,Scholkopf 等人[6]将一个具有明确取值范围和意义的参数 v 引入到 ε-SVR 模型中,提出了一种新型的支持向量机算法——v-支持向量回归(v-SVR)模型。v-SVR 能够自动计算 ε 的值,其解的稀疏性由参数 v 控制。

假设一组由未知函数随机、独立产生的数据点样本集 $L=\{(x_i,y_i)\mid i=1,2,\cdots,l\}$。其中:$x_i\in R^n$ 是 n 维输入变量,$y_i\in R$ 是对应的输出值,v-支持向量回归(v-SVR)模型通过某一非线性映射函数 $\phi(x)=\{\phi(x_1),\phi(x_2),\cdots,\phi(x_l)\}$将样本映射到高维特征空间 R^{n_f},在此特征空间中构造最优决策函数 $f(x)=w^{\mathrm{T}}\phi(x)+b$, $w\in R^{n_f}$, $b\in R$。根据结构风险最小化(SRM)原则,权

收稿日期: 2008-11-04; **修回日期**: 2008-12-29 **基金项目**: 国家自然科学基金资助项目(60641006)

作者简介:耿立艳(1979-),女,天津宝坻人,博士,主要研究方向为复杂金融时序分析及预测等(gengliyan_28117@yahoo.com.cn);马军海(1965-),男,山东青岛人,教授,博导,主要研究方向为经济系统复杂性.

向量 w 和偏差量 b 可通过优化以下模型得到：

$$\min 1/2\|w\|^2 + C[v\varepsilon + 1/l\sum_{i=1}^{l}(\xi_i + \xi_i^*)]$$

$$\text{s.t.}\begin{cases}((w\phi(x_i)) + b) - y_i \leqslant \varepsilon + \xi_i \\ y_i - ((w\phi(x_i)) + b \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geqslant 0, \varepsilon \geqslant 0 \\ i = 1, 2, \cdots, l\end{cases} \tag{1}$$

其中：$C(\geqslant 0)$ 是惩罚参数，决定置信风险与经验风险间的平衡；$v(0\leqslant v\leqslant 1)$ 是异常样本占训练样本总数比例的上界与支持向量占训练样本总数比例的下界；这里的 ε 作为优化问题的变量出现。

式(1)为一凸二次优化问题。根据最优化理论，其解法是引入 Lagrange 乘子，在对偶空间建立 Lagrange 方程求解上述最优化问题，可得到原始优化问题的对偶问题：

$$\max -1/2\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i, x_j) + \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \leqslant Cv \\ \alpha_i^{(*)} \in [0, C/l], v \geqslant 0, C \geqslant 0 \\ i = 1, 2, \cdots, l\end{cases} \tag{2}$$

其中：核函数 $K(x_i, x) = \phi(x_i)^T\phi(x)$ 为满足 Mercer 条件的对称正定函数，通常可选择线性函数、多项式函数、径向基函数和 Sigmoid 函数。求解式(2)即可得到 v-SVR 预测输出决策函数：

$$f(x) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)K(x_i, x) + b \tag{3}$$

可以证明，在一定条件下，当对独立同分布样本进行拟合时，v 以概率 1 渐进等于支持向量比例和错误样本比例。因此，参数 v 直接控制了支持向量的数目。

2 灰色支持向量机模型

2.1 灰色预测模型

灰色系统理论是由邓聚龙教授提出的一门新兴的横断学科，它是在经典控制理论、现代控制理论、模糊控制理论的基础上，针对一类要求高而又难于用传统方法建模的系统发展起来的。灰色系统理论中的灰色预测就是通过原始数据的处理和灰色模型的建立，发现、掌握系统发展规律，对系统未来状态作出科学的定量预测。灰色预测模型具有以下三个特征[7]：

a)建模所需信息较少，通常只要有四个以上数据即可建模。

b)不必知道原始数据分布的先验特征，对无规律或不服从任何分布的任意光滑离散原始序列，可以通过有限次的生成即可转化成有规律序列。

c)模型的精度较高，可保持原系统的特征，能较好地反映系统的实际状况。

最常用的灰色预测模型是由一个只包含单变量的一阶微分方程构成，称之为 GM(1,1)模型。其实质是对原始数列进行一次累加生成，使生成数列呈现一定的规律性，通过建立具有微分、差分、近似指数规律兼容的方程模型，求得拟合曲线，用于对未来作出预测。

2.2 灰色支持向量机模型的建立

GM(1,1)模型是灰色系统理论的基本模型，灰色系统理论将任一随机过程当做是在一定幅区和一定时区变化的灰色过程，其随机量当做是在一定范围内变化的灰色量；而无规律的离散时空数列是潜在的有规律序列的一种表现，通过生成变换可将无规律序列变成有规律序列。因此，用灰色系统理论的思想、方法对原始观测数据进行必要处理，将会大大改善模型的性能。

灰色支持向量机(GSVR)模型的基本思想是，首先利用 GM(1,1)模型的建模特性对原始数列进行一次累加生成处理，削弱原始数列中的随机波动性，挖掘出数列中蕴涵的深层规律性；然后利用 v-SVR 对规律性较强的新数列建立模型并进行预测；最后将预测结果进行累减还原得到原始数列的预测值。

运用 GSVR 模型进行预测的具体步骤如下：

a)设 $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(k)\}$ 为原始非负时间序列，对其进行一次累加生成，得到新序列 $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(k)\}$，$X^{(1)}(k) = \sum_{i=1}^{k}X^{(0)}(i)$，$k = 1, 2, \cdots, n$。如果原始数列中的数有正有负时，要先作非负处理：找出序列的最小负数，将其绝对值加到各个数据上；而后再对非负数列作一次累加生成，求出预测值后再减去最小负数的绝对值。

b)将新数列 $X^{(1)}$ 归一化到[0,1]区间，确定参数 v 和惩罚参数 C 的值，选择核函数并确定核参数的值。根据样本数据中的训练集数据，建立 v 支持向量机的预测函数 $f(x) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)K(x_i, x) + b$。

c)利用训练好的 v-SVR 模型对新的累加序列 $X^{(1)}$ 进行预测，得到其预测值 $\hat{X}^{(1)}$，然后对 $\hat{X}^{(1)}$ 进行累减还原，得到原始数列 $X^{(0)}$ 的预测值 $\hat{X}^{(0)}$：

$$\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k);\ k = 0, 1, 2, \cdots, n \tag{4}$$

3 实例分析

3.1 数据的选取

选取深圳基金指数 2003 年 1 月 21 日～2006 年 11 月 27 日的数据作为样本，共 930 个观测值。

历史波动率是用金融资产的历史数据来估算其波动率，基本假设是相信过去所实现的波动率会延续到未来，且不会产生大幅变动。因此，可将用过去数据所计算而得的波动率视为未来的波动率。在金融经济学中，波动率常用资产收益的标准差来度量。Parkinson[8] 已证明，利用极差估计波动率的效率要高于通常的样本标准差。这里采用极差来度量金融资产的波动率，即

$$R_t = (\log P_{t,\text{high}} - \log P_{t,\text{low}}) \times 100 \tag{5}$$

其中：$P_{t,\text{high}}$ 和 $P_{t,\text{low}}$ 分别为 t 时刻指数的最高价和最低价。由式(5)可看出，极差的计算利用了观测值的最高价和最低价数据，能够捕捉到更多的金融数据信息，更有效地度量金融资产的波动率。

3.2 输入变量的选取

以移动平均的形式计算出基金的波动率，即以当期极差和最近四期的极差滚动计算出波动率 λ_t，得到包含最近五天信息的波动率数据。具体计算式为

$$\lambda_t = 1/5\sum_{k=0}^{4}R_{t-k} \tag{6}$$

GSVR 模型的输入变量由四阶滞后波动率组成,输出变量为当期的波动率。也就是说,利用前四期的波动率来预测当期的波动率。

3.3　网络学习及预测

分别利用 GSVR 模型和 v-SVR 模型对基金波动率进行预测,以验证 GSVR 模型的有效性。将归一化的样本数据分成两部分,前 740 个样本用于训练模型,后 190 个样本用于检验模型的预测性能。选取径向基函数 $K(x_i,x)=\exp(-\|x_i-x\|^2/\sigma^2)$ 作为核函数,这是因为基金的动态性是非线性的,非线性核函数可获得比线性核函数更好的性能。这样有三个参数 (C,σ^2,v) 需要确定。研究表明,惩罚参数 C 和核参数 σ^2 对 v-支持向量机的泛化能力有重要的影响,而参数 v 对模型的泛化能力影响不大。C 值能够平衡置信风险与经验风险,在确定的数据空间中,C 值过小会对训练数据造成欠学习现象,而 C 值过大则容易对训练数据造成过学习现象。两者都将降低模型的泛化能力,进而影响预测精度。核参数 σ^2 反映了训练样本数据的分布或范围特性,确定了局部邻域的宽度,它对模型泛化性能的影响方向与 C 正好相反。根据以往经验,将各参数设为 $C=100$, $v=0.01$, $\sigma^2=1$。为了使预测结果具有可比性,v-SVR 模型求解中的参数取值与 GSVR 模型的相同。

3.4　预测性能评价指标

选用经过异方差调整的指标 HRMSE 和 HMAE 以及 LL 三种指标来评价模型的预测性能,分别定义如下:

$$\mathrm{HRMSE}=\Big[T^{-1}\sum_{t=1}^{T}(1-\hat{\sigma}_t/R_t)^2\Big]^{0.5} \tag{7}$$

$$\mathrm{HMAE}=T^{-1}\sum_{t=1}^{T}\left|1-\hat{\sigma}_t/R_t\right| \tag{8}$$

$$LL=T^{-1}\sum_{t=1}^{T}\big[\ln(\hat{\sigma}_t)-\ln(R_t)\big]^2 \tag{9}$$

其中:$\hat{\sigma}_t$ 表示预测波动率;R_t 表示事后波动率。上述各指标的值越小,表明预测误差越小,预测就越准确。

3.5　数值结果分析

为了详细比较两种模型的预测效果,将整个预测区间分成各子区间,分别计算各子区间的预测性能评价指标,计算结果见表 1。根据 HRMSE 和 HMAE 值,在 40 步以内,GSVR 模型对基金波动率的预测误差小于 v-SVR 模型的对应值,在 40 步以外,v-SVR 模型的预测误差又逐渐小于 GSVR 模型的预测误差;根据 LL 值,GSVR 模型 35 步内的预测误差小于 v-SVR 模型的预测误差,而 35 步以外的预测误差又大于 v-SVR 模型的对应值。以上结果充分说明 GSVR 模型的中短期波动率预测性能优于 v-SVR 模型,而其长期的波动率预测性能又稍逊于 v-SVR 模型。这主要是由于 GSVR 模型发挥了 GM(1,1) 模型的一次累加生成优势,削弱了原始极差波动率的随机波动性,挖掘出数列中蕴涵的深层信息,增强了规律性,使预测精度得到提高。但同时累加生成处理也可能令原始数据的一部分有用信息丢失,进而导致 GSVR 模型的长期预测精度有所降低。由此,对基金极差波动率而言,GSVR 模型适合于刻画波动率的中短期变动趋势,而 v-SVR 模型适合于刻画波动率的长期变动趋势。

图 1、2 分别为深市基金波动率向前 35 和 190 步预测值的曲线图。由两图可看出,35 步以内,GSVR 模型比 v-SVR 模型更好地预测了基金波动率的变动趋势,各步预测值围绕事后波动率的均值变动;35 步以外,GSVR 模型预测的波动率变动趋势逐渐向下偏离了事后波动率的变动趋势,其预测值逐渐小于事后波动率的均值,而 v-SVR 模型较好地预测了事后波动率的变动趋势。

表 1　样本外预测效果

预测步数	HRMSE		HMAE		LL	
	GSVR	v-SVR	GSVR	v-SVR	GSVR	v-SVR
3	0.331 1	0.439 1	0.322 1	0.438 5	0.134 4	0.209 7
5	0.276	0.364 5	0.257 6	0.337 3	0.089 5	0.139 7
10	0.577	0.685 7	0.459 2	0.522 1	0.214	0.247 4
15	0.907 7	1.007 1	0.602 7	0.657 9	0.331 7	0.362 3
20	0.809 7	0.884 3	0.538	0.557 5	0.279 4	0.291 1
25	0.742 2	0.812 3	0.494 5	0.521	0.242 6	0.258
30	0.700 5	0.761 1	0.473 9	0.490 4	0.278	0.291 6
35	0.729 6	0.772 9	0.498 2	0.516	0.290 5	0.298 5
40	0.728 6	0.758 5	0.509 7	0.516 1	0.314 5	0.310 6
45	0.770 1	0.757 6	0.555 4	0.530 6	0.353 4	0.324 8
50	0.753 7	0.744 5	0.545 9	0.528 7	0.342 5	0.318
100	0.718 7	0.672 4	0.541 2	0.495 6	0.392 6	0.334 2
190	0.614 4	0.618 1	0.479 1	0.474 6	0.438	0.300 5

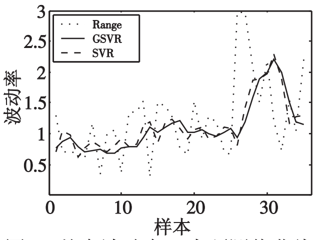

图1　基金波动率35步预测值曲线

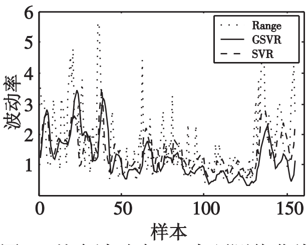

图2　基金波动率190步预测值曲线

4　结束语

支持向量机是一种新型的神经网络模型,具有坚实的理论基础和严格的理论分析,综合考虑了置信风险和经验风险,具有良好的鲁棒性和较高的预测精度。而灰色预测方法独特的数据生成方式可以削弱数据序列的随机性,增强数据的规律性。本文将灰色预测方法引入到 v-支持向量机模型中,充分利用两者的优势对深圳基金的极差波动率进行了预测。通过与 v-支持向量机和灰色预测模型的对比,证明了新模型是一种有效的波动率预测方法,适合于基金波动率的中短期预测。

参考文献:

[1]　PÉREZ-CRUZ F, AFONSO-RODRIGUEZ J A, GINER J. Estimating GARCH models using support vector machines[J]. Quantitative Finance, 2003, 3(3):163-172.

[2]　GAVRISHCHAKA V V, GANGULI S B. Volatility forecasting from multiscale and high-dimensional market data[J]. Neurocomputing, 2003, 55(1-2):285-305.

[3]　傅东升,曹丽娟. SVM 与 BP 网络对基金波动的预测效果比较分析[J]. 世界经济情况,2007(8):45-50.

[4]　唐万梅. 基于灰色支持向量机的新型预测模型[J]. 系统工程学报,2006,21(4):410-413.

[5]　牛晓东,谷志红,王会青,等. 基于灰色支持向量机的季节型负荷预测方法[J]. 华东电力,2007,35(6):1-5.

[6]　SCHÖLKOPF B, SMOLA A J, WILLIAMSON R C, *et al*. New support vector algorithms[J]. Neural Computation, 2000, 12(5):1207-1245.

[7]　傅立. 灰色系统理论及其应用[M]. 北京:科技技术文献出版社,1992.

[8]　PARKINSON M. The extreme value method for estimating the variance of the rate of return[J]. Journal of Business, 1980, 53(1):61-65.