

[金融学论丛]

非参数支持向量回归和分类理论 及其在金融市场预测中的应用

Finance
Monographs

THE NONPARAMETRIC THEORY OF SUPPORT VECTOR REGRESSION AND
CLASSIFICATION WITH APPLICATIONS TO FORECASTING FOR FINANCIAL MARKETS

陈诗一/著



北京大学出版社
PEKING UNIVERSITY PRESS

F830.9/231

2008

[金融学论丛]

非参数支持向量回归和分类理论 及其在金融市场预测中的应用

Finance
Monographs

THE NONPARAMETRIC THEORY OF SUPPORT VECTOR REGRESSION AND
CLASSIFICATION WITH APPLICATIONS TO FORECASTING FOR FINANCIAL MARKETS

陈诗一/著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

非参数支持向量回归和分类理论及其在金融市场预测中的应用/陈诗一著. —北京:北京大学出版社, 2008. 4

(金融学论丛)

ISBN 978 - 7 - 301 - 13715 - 4

I. 非… II. 陈… III. 时间序列分析 - 经济模型 - 应用 - 金融市场 - 市场预测 - 中国 IV. F832. 5

中国版本图书馆 CIP 数据核字(2008)第 058500 号

书 名: 非参数支持向量回归和分类理论及其在金融市场预测中的应用

著作责任者: 陈诗一 著

责任编辑: 朱启兵

标准书号: ISBN 978 - 7 - 301 - 13715 - 4/F · 1895

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn> 电子邮箱: em@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752926

出版部 62754962

印 刷 者: 北京汇林印务有限公司

经 销 者: 新华书店

650 毫米 × 980 毫米 16 开本 16 印张 262 千字

2008 年 4 月第 1 版 2008 年 4 月第 1 次印刷

印 数: 0001—3000 册

定 价: 34.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子邮箱: fd@pup.pku.edu.cn

献给我的九十二岁高龄的爷爷
——陈印老先生

献给离开我整整二十年的奶奶
——丁成训女士

作者的话

未来充满了不确定性,无论是国家宏观经济管理者还是微观经济单位和个体在做出经济政策和投资决定的优化决策时都离不开对未来重要经济和金融变量的预测。而正确预测的获得除了依赖于决策者的良好判断外,还必须依靠各种定量预测方法。预测方法的发展可谓日新月异,多种多样。本书主要研究最新的非参数、非线性支持向量回归和分类预测方法,并对金融市场的重要价格变量进行预测,以检验该方法的出色预测能力。

本书的主要研究成果由作者的博士论文及其后续研究组成。作者博士论文的题目为“Forecasting Methods for Financial Markets Using Support Vector Regression”,主要是关于支持向量回归的理论和应用研究。回国后,作者继续这方面的研究并相继发表了一些研究成果,特别是先后两次赴德国洪堡大学,与计量经济学家和统计学家 Härdle 教授合作完成了论文“Estimation of Default probabilities with Support Vector Machines”。这些研究成果构成了本书的主要内容。

本书得以出版离不开多方面的支持,在此,作者要特别感谢张军教授、Jeong 教授和 härdle 教授的学术指导以及他们在百忙中为本书所作的序言!感谢北京大学出版社朱启兵编辑的热情工作!感谢上海市浦江人才计划、上海市重点学科建设项目(项目编号 B101)和复旦大学 985 国家哲学社会科学创新基地“中国经济国际竞争力研究”课题的资助!

相对于传统预测方法,支持向量回归和分类采用独特的结构风险最小化设计,不仅和传统方法一样最小化经验风险,以获得良好的拟合能力,而且同时最大化泛化推广能力,这是其他预测方法所没有的,正是这种特殊的结构设计从理论上赋予了支持向量算法卓越的预测推广能力。虽然支持向量算法属于数据挖掘导向的方法,但是它同样可以与经济理论结合,纳入经济理论所要求的各种重要解释变量。本书第一章简单介绍了预测的定义、发展和重要性。第二章详细讨论了支持向量回归和分类的理论特性。第三章利用通常的前馈支持向量回归来估计自回归模型,对日汇率进

行预测。第四章基于反馈神经网络的启示,首次提出了反馈支持向量回归方法,对含有误差项的自回归共积移动平均模型进行估计,并对真实金融收益率和仿真序列的水平趋势进行预测。第五章利用第四章提出的反馈支持向量回归方法对 GARCH 模型进行估计,并对金融收益率的波动性进行预测。因为, GARCH 模型实际上服从一个自回归移动平均随机过程。所有这些预测都与基准的移动平均法、最大似然法和人工神经网络等方法进行比较以检验支持向量回归的预测能力。第六章利用支持向量分类方法来估计非线性的概率模型,以预测公司的信用风险和违约概率。支持向量分类的识别能力将与基准的 logit 模型进行对比。第七章是全书的结束语。全书所有计算都通过 MATLAB 软件来实现。

本书既适合于从事预测、统计和计量研究的研究生、科研人员 and 高等学校教师,也适合于不同层级的经济管理工作、经济分析师和企业经理人士,更适合于金融市场众多的投资单位和个人。本书理论和应用分别阐述、各自成章、脉络清晰,不同的读者可以根据自己的具体情况进行选择阅读。由于作者水平有限,成书匆匆,错误难免,欢迎读者批评指正。

序言(1)

张 军*

本书的原型主要是陈诗一博士在韩国庆北国立大学完成的博士论文。他去韩国攻读博士学位最初是我的主意。我在 2002 年秋受聘在韩国庆北国立大学讲授一个学期的“宏观经济学”和“中国经济研究专题”的英文课程,在此期间与经济系的 Jeong 教授相识。有一次我们在一起喝酒,他试探地问我,是否愿意推荐我的学生来庆北大学攻读博士学位,因为他有了一个研究项目的资助,可以为博士研究生提供三年的奖学金。

我再三考虑,决定推荐正在复旦大学攻读硕士学位的陈诗一去 Jeong 教授那里攻读计量经济学的博士学位。我做出这个决定,主要是考虑到诗一的数学基础比较好。就这样,诗一成为了 Jeong 教授指导的第一位中国留学生,也是韩国庆北国立大学计量经济学博士学位项目的第一个中国人。当然,他也是第一个获得韩国庆北国立大学计量经济学博士学位的中国人。诗一在庆北国立大学实际上只用了三年不到的时间就获得了博士学位,在韩国庆北国立大学被传为佳话。我很高兴我当时做出了这个决定。

作为陈诗一的博士学位论文,本书的选题也隶属于这一领域,它主要讨论了非参数、非线性支持向量回归和分类方法在金融时间序列预测和公司信用风险识别上的表现,并与基准的最大似然法和人工神经网络方法进行比较,以检验支持向量回归和分类理论上的出色预测能力。

在这个不确定的世界上,对未来的预测总是最重要的,无论是宏观经济管理还是微观金融市场投资的决策都离不开预测。但是,仅仅根据管理者、投资者和研究者们的判断能力,还不足以获得正确的预测,必须依靠统计方法和数量模型做出尽可能准确的定量预测。

从计量经济学的角度看,预测方法的发展大概经历了三个阶段:基于普通最小二乘法估计的结构计量预测模型、基于最大似然法估计的时间序

* 复旦大学“当代中国经济”长江特聘教授,中国经济研究中心主任。

列模型以及非参数、非线性的神经网络和支持向量回归与分类模型。陈诗一的这本书所运用的预测方法是支持向量回归和分类的方法。

支持向量回归和分类的预测能力主要来自于其独特的结构风险最小化设计,即不仅最小化经验风险(普通最小二乘法、最大似然法和人工神经网络只做到了这一点),以最大化回归样本内的拟合能力,而且还最大化分隔空间(这是其他所有预测方法没有做到的),即最大化样本外的预测泛化能力,最终获得两部分的平衡解。所以,支持向量回归和分类不仅能够像普通最小二乘法、最大似然法和人工神经网络一样有着很好的拟合能力,而且有着比它们更出色的预测推广能力。

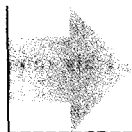
当然,支持向量回归和分类还有其他各种优点,比如不需要关于模型形式和变量概率分布的先验假定、同时适合大样本和小样本数据、能够获得全局唯一最优解、能够自动捕捉变量间的非线性依赖关系,等等。在这本书的第二章,陈诗一详细介绍了这种新的非参数支持向量算法的理论。本书利用蒙特卡罗仿真数据和真实的金融数据来检验支持向量算法的这种强大预测能力是否成立。由于支持向量算法刚刚才出现十年左右,而且主要是计算机信息工程领域的方法,把它和经济和金融领域的预测结合起来的研究还不是很多,所以该研究具有相当的挑战性。

本书在第三章利用一般的前馈支持向量回归对日汇率序列进行预测,它的预测框架是自回归共积 ARI 模型。其中,作者特别对自 2005 年 7 月 21 日开始实行的以市场供求为基础、有管理的人民币浮动汇率进行了预测。

特别值得一提的是,第四章中作者在现有文献前馈支持向量回归的基础上开发出了一种新的反馈支持向量回归算法,引入一个从输出层到输入层的全局反馈回路,反馈回路的出现可以帮助模型捕捉变量的更多动态特征,对提高前馈支持向量回归的学习能力有着深远的影响。这种新的反馈支持向量回归的提出算是本书的最大贡献。

陈诗一的博士导师 Jeong 教授是研究非参数回归和时间序列的计量经济学家。根据我的了解,Jeong 教授最近所做的这些基于半参数、非参数方法的汇率预测属于非常前沿的时间序列的计量技术。在这一领域,德国洪堡大学的 Härdle 教授也是权威性的人物。因为这样,Härdle 教授、Jeong 教授以及陈诗一开始在这一领域有了进一步合作研究的持续计划。

2007 年 12 月 10 日于上海



Preface (2)

Kiho Jeong *

It is my great pleasure to write this preface for Dr. Shiyi Chen's book. Dr. Chen was one of my students at the PhD graduate program of Kyungpook National University (KNU), the Republic of Korea, from 2003 to 2005. I will replace the preface with my personal memories on Dr. Chen.

Dr. Chen entered the graduate school in the department of economics at KNU with the Excellent International Students Scholarship of KNU. In all courses in the field of statistics, econometrics and machine learning, he had demonstrated an outstanding understanding of the basics and core analytical skills of quantitative methods, with the highest class performance.

As a PhD student, he majored econometrics with focus on forecasting. To my knowledge, he was the first student at Kyungpook National University who had completed PhD degree within three years. His dissertation was to develop new forecasting methods based on Support Vector Machine, one of the most promising among recently developed statistical techniques, and to apply the new methods to financial forecasting such as exchange rates, stock prices and their volatilities. In applications to real data, his new forecasting methods showed the outstanding forecasting ability, outperforming the competing forecasting methods.

After finishing his Ph. D. course, I and Dr. Chen kept in touch each other by either his visit to my university or my visit to Fudan University where Dr. Chen works now. In summer of 2007, we visited Humboldt University at Berlin, Germany, together for research there. In Berlin, there was one episode re-

* Professor of Econometrics, School of Economics and Trade, Kyungpook National University, Republic of Korea.

garding to Dr. Chen's diligence. We stayed at the same guest house. One day, he told me that a lady working at the guest house had never changed his room since his arrival at Berlin. I delivered his complain to the guest house' manager. A few days later, the manager said to me that he asked the lady why she did not change Dr. Chen's room and her answer was that she could not change his room because she always found him sleep in the bed about at 11 am. In Berlin, Dr. Chen always worked at his office from noon to 7 am in the morning of next day and then returned back to the guest house to sleep until noon. So the poor lady could not have any chance to change his room. I have never seen persons as diligent as Dr. Chen.

Now I am very glad to see another evidence of Dr. Chen's diligence, this book, which contains his research results so far. Considering Dr. Chen's potential research capability and diligence, I am sure he will again ask me to write a preface of his another book within a few years. I wish readers of this book share with my joy about the author.

December 10, 2007



Preface (3)

Wolfgang Karl Härdle *

Support Vector Machines (SVM) is a modern statistical tool for classification and regression studies. This non-linear statistical learning technique has been shown to be superior to traditional approaches such as linear discrimination analysis or neural networks. The application of SVMs has, so far, mainly only been carried out in the field of pattern recognition and the classification of, for example, hand-written letters and numbers.

This book however opens this dynamic non-linear statistical learning technique up to a full new set of applications where one can expect profitable results, particularly in the field of financial econometrics where one typically applies linear estimation techniques of finite dimensions. One important field in finance econometrics is the analysis of exchange rates. In chapter 3 Chen uses feed forward SVM regression for non-linear ARI models to forecast exchange rates fluctuations. In chapter 4 he extends this concept to non-linear ARMA models and then finally in chapter 5 he applies it to GARCH models. The chapter 6 is related to support vector classification and its application to analyzing the default risk of German companies.

This work brings the notion of SVMs into a totally new and exciting domain and Chen's fresh and innovative concepts are certain to become bench marks for future work in this ever important field.

November 29, 2007

* Professor of Econometrics and Statistics, Center for Applied Statistics and Economics, Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin, German.

符号和缩写

符 号

b	阈值
d	多项式核函数参数;共积的阶数
e_i	噪声项、残差项;预测误差
$f(x)$	决策函数,或样本回归方程,或样本分类函数
$g(x)$	潜在决策函数
h_t	真实波动性
i, t	观察值
l	特征空间维数
n, T	样本长度、样本容量
p	自回归滞后阶数
q	移动平均滞后阶数
r	相关系数(第三章);SVC 高斯核参数(第六章)
t	t 统计量;学生氏 t 分布
u_t^2	真实波动性代理
$\{(x_i, y_i)\}_{i=1}^n$	训练集
x_i	投入变量,或解释变量
y_i, y_t	第三章至第五章为收益率滞后项和噪声项,第六章为会计比率 1 维产出,或被解释变量
$\bar{y}$	第三章至第五章为收益率,第六章为二分变量 收益率平均值
$\hat{y}_i, \hat{y}_t$	真实值 y_i, y_t 的拟合值或预测值
z^{-l}	l 阶滞后算子
C	惩罚参数

E	期望算子
E_t, I_t	价格变量, 比如汇率、证券指数等
$\hat{E}_t, \hat{I}_t$	真实值 E_t 或 I_t 的拟合值或预测值
H_0	零假设、原假设
H	特征空间, 或 Hilbert 空间
$I(1)$	一阶共积
$K(x_i, x)$	核函数
K	核矩阵
L	拉格朗日函数
$N(0, 1)$	标准正态分布
Q	Box-Pierce Q 统计量, Ljung-Box Q 统计量
R^1	1 维实数空间
R^m	m 维实数空间
R^2	决定系数
ε	SVR 中 ε -不敏感损失函数的自由参数
ε_i	误差项
$\alpha, \alpha', \alpha^{(')}$	拉格朗日乘子(向量表示)
$\alpha_i, \alpha'_i, \alpha_i^{(')}$	拉格朗日乘子(标量表示)
$\xi, \xi', \xi^{(')}$	松弛变量(向量表示)
$\xi_i, \xi'_i, \xi_i^{(')}$	松弛变量(标量表示)
$\varphi_i(x), \varphi(x)$	特征层非线性转换函数(标量和向量表示)
ν	SVR 自由参数
w_i, w	线性权重(标量和向量表示)
Σ	方差协方差矩阵
σ^2	SVR 高斯核参数
Δ	增长率
ϕ_1, ϕ_2, θ_1	线性 ARMA 模型参数
$\phi_1, \kappa, \delta_1, \alpha_1$	线性 GARCH(1, 1)模型参数

缩 写

ACF	自相关函数值	AD	摊销和折旧
ADF	扩大迪基-富勒检验	AIC	Akaike 信息标准
ANN	人工神经网络	AP	应付账款
APT	套期定价理论	AR	应收账款
AR	精度比率;自回归模型	ARCH	自回归条件异方差
ARI	自回归共积模型	ARIMA	自回归共积移动平均模型
AUD	澳大利亚元	CA	流动资产
CAD	加拿大元	CAP	累积精度轮廓
CAPM	资本资产定价模型	CD	下降趋势正确预测度
CHER	人民币名义汇率	CHF	瑞士法郎
CIER	平均条件信息量比率	CL	流动负债
CLRM	古典线性回归模型	CM	嵌套检验统计量之三
CU	上升趋势正确预测度	DA	判别分析
DA	方向精度	DJC	道·琼斯证券指数
EBIT	息税前利润	EBITDA	息税折旧摊销前利润率
EUR	欧元	exp	指数
GA	遗传算法	GARCH	广义自回归条件异方差
GBP	英镑	GLS	广义最小二乘
GRNN	广义回归神经网络	HLN	嵌套检验统计量之二
INR	印度卢比	INV	库存资产
IQR	四分位数间距离	IV	工具变量
JB	正态分布检验统计量	KKT	库恩塔克条件
KRW	韩国元	LAD	最小绝对偏差估计
log	自然对数	logit	逻辑斯蒂回归模型
MA	移动平均	MAE	预测误差绝对值平均值
MEU	最大期望效用	MLE	最大似然估计
MLP	多层感应器神经网络	MSE	均方预测误差
NI	净收入	nsup	支持向量数目
NMSE	标准化均方预测误差、泛化误差	NZD	新西兰元

NYSE	纽约证交所综合指数	NW	Nadaraya-Watson 估计量
OE	嵌套检验统计量之一	OF	本金
OI	业务收入、营业收入	OLS	普通最小二乘
PACF	偏自相关函数值	PAV	Pool Adjacent Violator 算法
PD	违约概率	PHP	菲律宾比索
PP	菲利普斯-费龙检验	QA	速动资产
QP	二次规划	RBF	径向基函数神经网络
RMB	人民币	ROA	资产利润率
SBC	Schwartz 贝叶斯标准	SHCI	上证综合指数
sign	预测方向准确度指标	SSR	残差平方和
SVM	支持向量算法、支持向量机	SVC	支持向量分类
SVR	支持向量回归	TA	总资产
TL	总负债	WC	周转资金
ε -SVR	ε -支持向量回归	ν -SVR	ν -支持向量回归

第一章 预测概述	/1
第一节 预测的重要性	/1
第二节 什么是预测?	/7
第三节 预测方法的发展	/12
第四节 预测与决策	/19
第二章 支持向量回归和分类理论	/22
第一节 支持向量算法	/22
第二节 支持向量回归	/24
第三节 支持向量分类	/32
第四节 蒙特卡罗仿真	/39
附录	/46
第三章 汇率预测:基于前馈 SVR 的非线性 ARI 模型	/59
第一节 介绍	/59
第二节 数据收集和处理	/62
第三节 实证模型设定	/64
第四节 预测方案和评估标准	/71
第五节 预测结果比较分析	/74
第六节 人民币汇率预测	/87
第七节 结论	/92
第四章 金融收益率水平预测:基于反馈 SVR 的非线性 ARIMA 模型	/97
第一节 介绍	/97
第二节 反馈 SVR 机制设计	/101
第三节 金融收益率定义	/104
第四节 固定预测评估	/106
第五节 递归预测评估	/118

目录

非参数支持向量回归和分类理论及其在金融市场预测中的应用

第六节 中国证券指数和汇率收益率水平预测	/133
第七节 结论	/139
第五章 金融收益率波动性预测:基于反馈 SVR 的非线性 GARCH 模型	/145
第一节 介绍	/145
第二节 实证模型和预测方案	/149
第三节 蒙特卡罗仿真	/154
第四节 真实数据检验	/159
第五节 中国金融波动性预测案例	/167
第六节 结论	/173
第六章 公司信用风险预测:基于 SVC 的非线性概率模型	/180
第一节 介绍	/180
第二节 数据描述和处理	/184
第三节 预测分析框架	/195
第四节 实证分析	/201
第五节 CAPM 检验案例	/215
第六节 结论	/219
第七章 结束语	/226
词汇表	/230
后记	/236