

金融工程2010年年度投资策略

基于小波分析和支持向量机的 指数预测模型

秦国文

Jun. 10, 2010, 深圳



国信证券经济研究所

GUOSEN

Guosen Securities Economic Research Institute

主要观点

■ 结论与投资建议

- 1、结合小波消噪的SVM预测模型在刻画预测走势的形态上具有较好的准确度，预测每日涨跌的准确率也较高，预测指数下跌的效果较好。
- 2、基于预测数据的简单择时策略，在震荡市中效果理想。多头策略和多空策略均跑赢了大盘，这说明了模型的预测具有一定参考价值。
- 3、基于高频数据的技术指标交易策略（如MACD）在结合了小波消噪后效果十分显著。

■ 实证效果

- 1、SVM预测模型的实证例子时间为2009年5月15日至2010年5月21日（250交易日）。多头策略在1年内回报率为10.33%，多空策略31.61%，同期沪深300指数回报率为1.03%。
- 2、基于高频数据的技术指标交易策略的实证例子时间为2010年5月7日至2010年6月3日（20个交易日，5400分钟），期指高频数据为1分钟。收益率：多头18.7%、空头27%、多空策略100.3%，同期沪深300期指收益率-9%。

内容目录

1 数据挖掘

2 小波分析在金融时间序列中的应用

3 支持向量机简介及其在股票市场中的应用

4 小波分析与技术指标结合的投资策略

5 投资建议

内容目录

1 数据挖掘

2 小波分析在金融时间序列中的应用

3 支持向量机简介及其在股票市场中的应用

4 小波分析与技术指标结合的投资策略

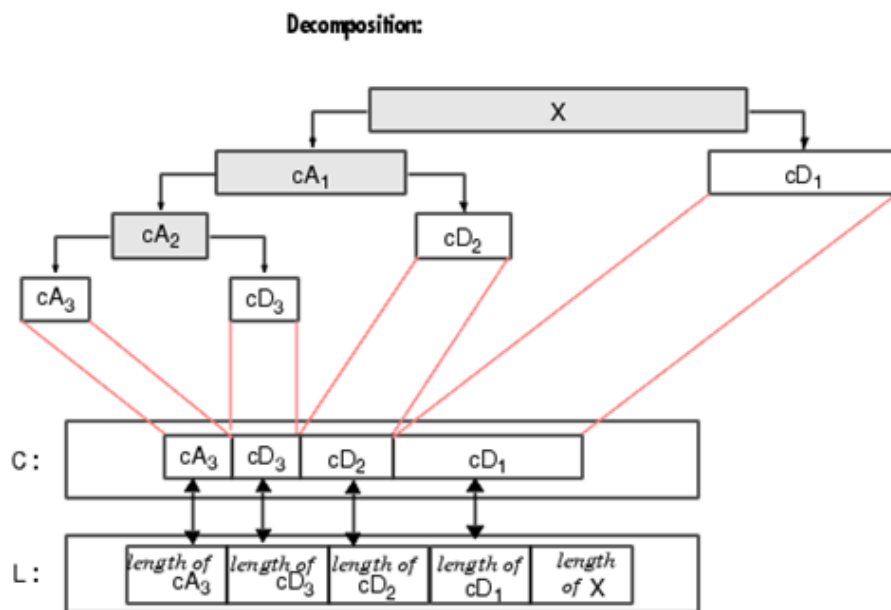
5 投资建议

2.1 小波分解流程

进行小波变换消噪的过程可以分为3个基本步骤：

1. 小波分解：选择要分解的级数 j 及所采用的小波函数，对给定的信号进行分解；
2. 细节（高频）系数的处理：对于从1到 j 的每一级，对细节系数选择不同的消噪原理进行修改或处理；
3. 信号重建：在原来得到的第 N 级离散逼近信号（低频）系数和修改后的第1级到第 j 级离散细节系数的基础上重建信号。

小波分解图释（3层分解）

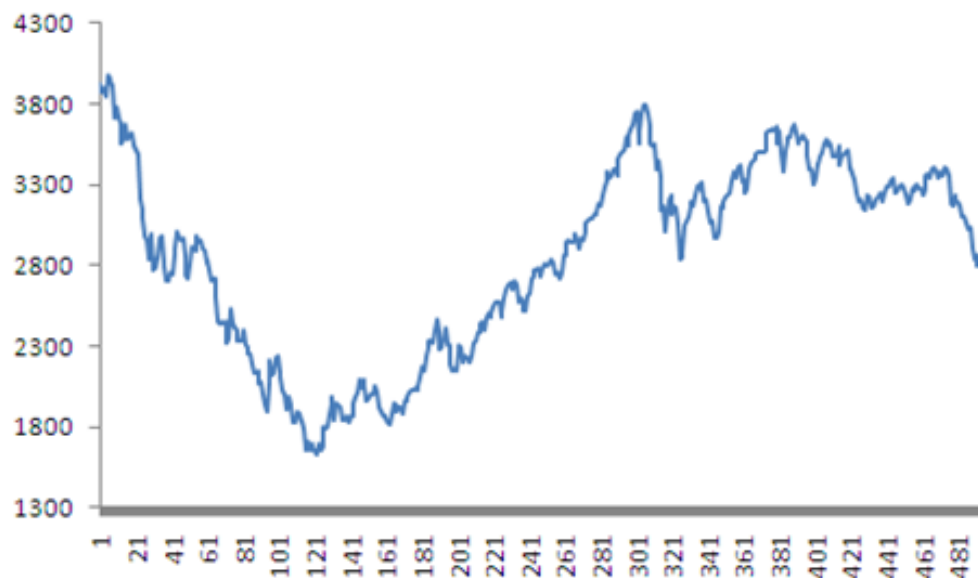


2.2 小波消噪实例分析 – 沪深300

我们对2008年5月至2010年5月共500个交易日的沪深300收盘价信号进行去噪实验。

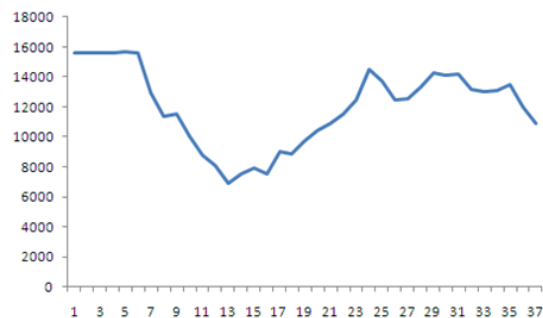
1. 选择小波Daubechies小波系(db4)并确定分解层次为4层，得到4层高半频和4层低半频序列；
2. 阈值处理选择sqtwolog阈值估计准则；
3. 最后根据小波分解的第4层低频系数和经过量化处理后的1至4层高频系数进行小波重构，达到消除噪音的目的。

含噪音的沪深300原始数据

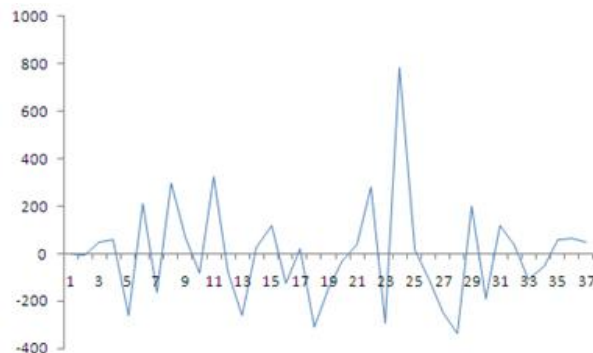


2.3 小波消噪实例分析过程

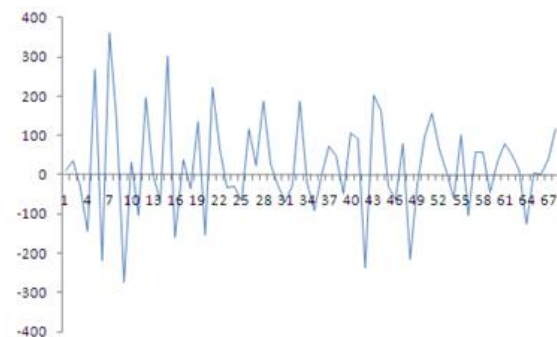
小波分析, 低频cA4



小波分析, 高频cD4



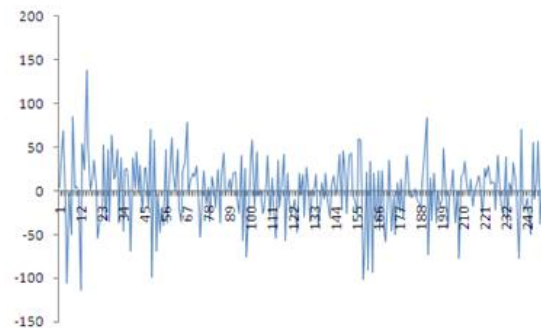
小波分析, 高频cD3



小波分析, 高频cD2



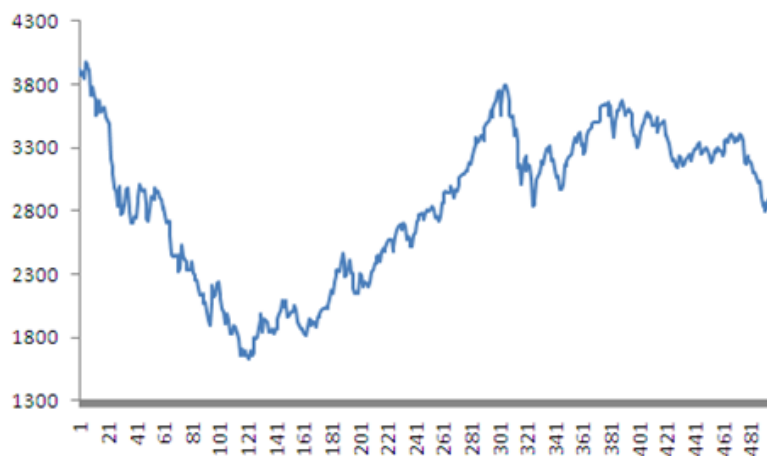
小波分析, 高频cD1



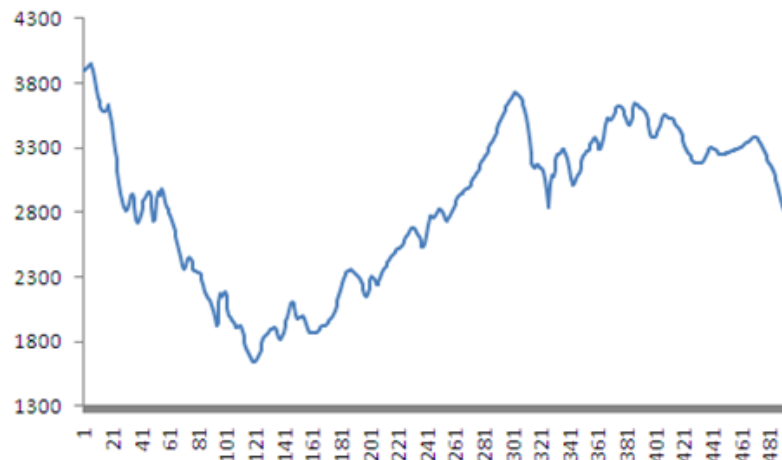
2.4 消噪后的沪深300

大部分的小幅波动被去除，而指数走势的主要特征均被保留，得到一组消噪的时间序列。由于小幅干扰的噪音已被消除，用支持向量机对其进行建模预测走势时，准确率将提升，因为消噪后的时间序列更能反映序列的主要特征。

含噪音的沪深300原始数据



消除噪音的沪深300数据



内容目录

1 数据挖掘

2 小波分析在金融时间序列中的应用

3 支持向量机简介及其在股票市场中的应用

4 小波分析与技术指标结合的投资策略

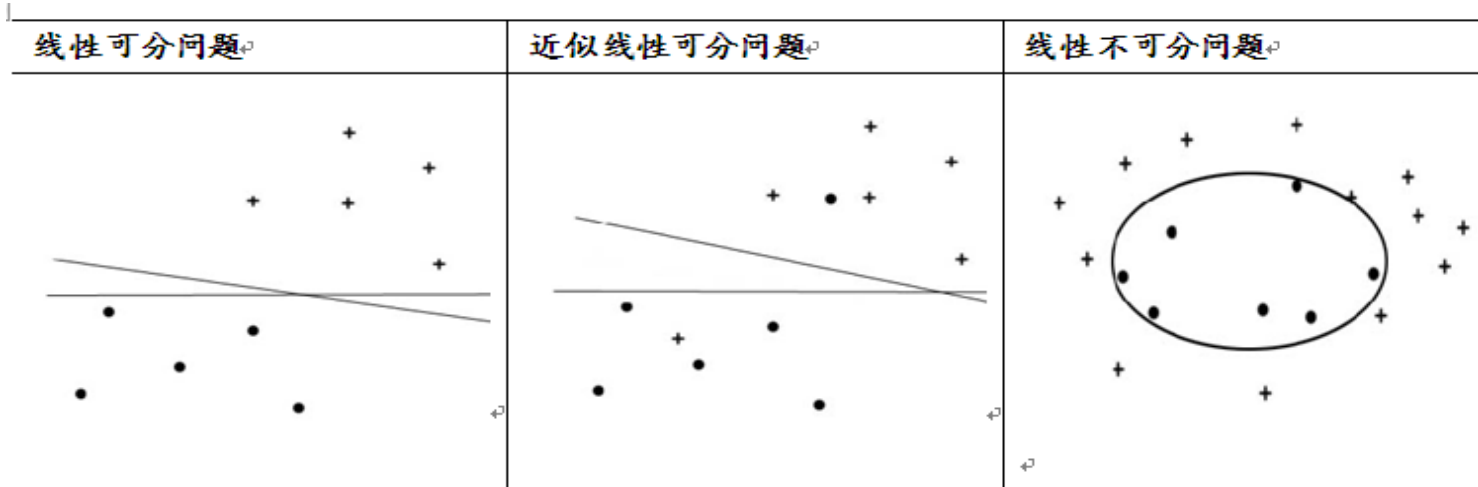
5 投资建议

3.2 支持向量机的分类问题

•支持向量机目前主要用来解决分类问题（模式识别，判别分析）和回归问题。而股市行为预测通常为预测股市数据的走势和预测股市数据的未来数值。一般来说，预测股市的行为通常为预测股市数据的走势和预测股市数据的未来数值。而当我们

将走势看作两种状态（涨、跌），问题便转化为分类的问题，一类是涨一类是跌。而预测股市未来的价格是指为典型的回归问题。我们有理由相信支持向量机可以对股市进行预测。

分类问题的三种类型



3.3 支持向量机的最优分类方法

非线性分类面

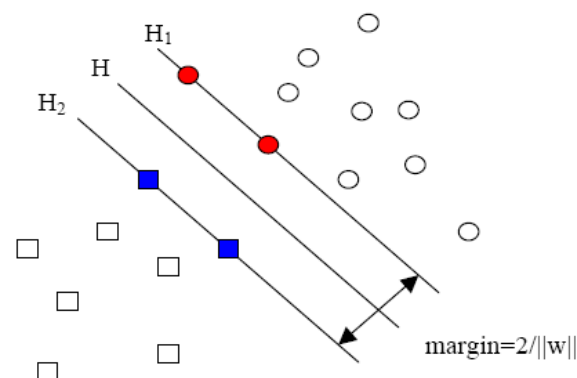
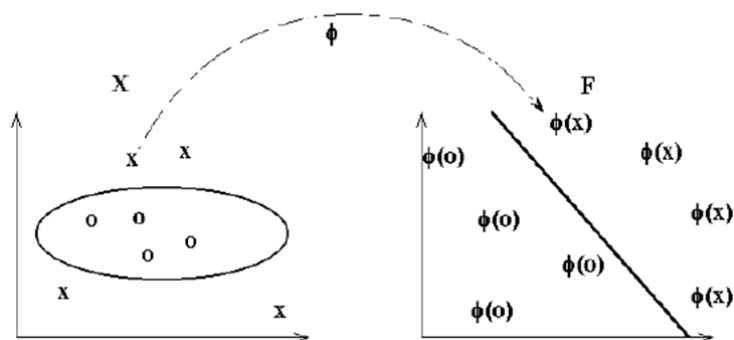


图2 线性可分情况下的最优分类线

◆SVM 是从线性可分情况下的最优分类面发展而来的,基本思想可用图2的两维情况说明。右图中,方形点和圆形点代表两类样本,H 为分类线,H1, H2分别为过各类中离分类线最近的样本且平行于分类线的直线,它们之间的距离叫做**分类间隔**(margin)。

所谓最优分类线就是要求分类线不但能将两类正确分开(训练错误率为0),而且使分类间隔最大。

使用Kernal函数推广到高维空间,最优分类线就变为**最优分类面**。

◆我们要做的是**什么**? 找到一个超平面(最优分类面),使得它能够尽可能多的将两类数据点正确的分开,同时使分开的两类数据点距离分类面最远。

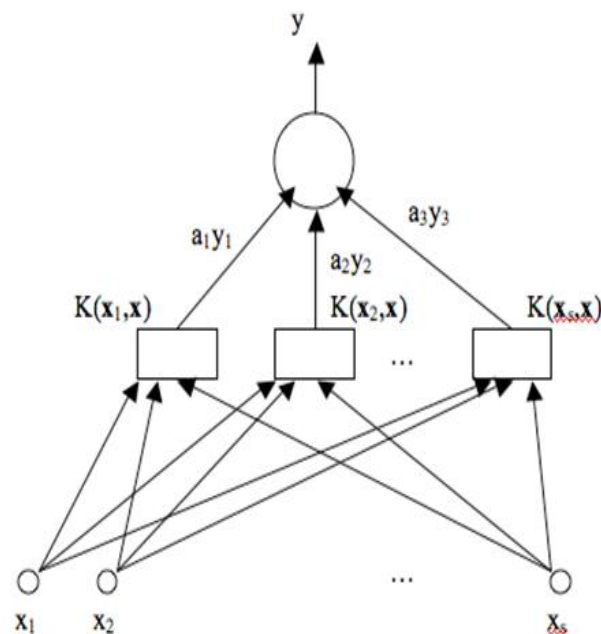
3.4 支持向量机在股票市场上的应用

通过对前人工作的参考和总结，我们发现历史上在进行这方面研究时主要需进行如下几个步骤，即：

- 股票指标的选取和原始数据的处理；
- 数据的标准化；
- 核函数及其参数的选取；
- 计算结果的分析与比较。

输入向量的选取方法：

- **VolumeIndex** :- 今日总手与过去M天平均总手的比值；
- **CloseIndex** :- 今日收盘价与过去M天平均收盘价的比值；
- **TradeAmountIndex** :- 今日交易额与过去M天平均交易额的比值；
- **TurnoverIndex** :- 今日换手率与过去M天平均换手率的比值；
- **MeanIndex** :- 过去M天的平均股价；
- **RateChangeIndex** :- 过去M天的涨跌幅；
- **HighIndex** :- 过去M天当中的最高股价；
- **LowIndex** :- 过去M天当中的最低股价。
- **M**为预测天数。



3.5 支持向量机在股票市场上的应用

◆为了避免小数“淹没”在了大数里而失去了意义，我们需要对输入参数采用标准化。

本文采用的是统计标准化，即：

$$x'_i = \frac{x_i - \bar{x}}{\sqrt{\text{var}(x)}}, \quad i = 1, 2, \dots, l,$$

其中 \bar{x} 和 $\sqrt{\text{var}(x)}$ 分别为变量 x 的平均值和标准差。

◆常见的核函数有：

线性核函数 $K(x, y) = (x \cdot y)$ ；

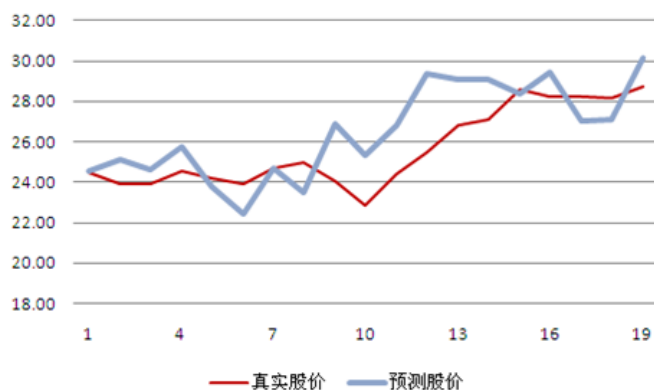
多项式核函数 $K(x, y) = [(x \cdot y) + 1]^d$ ；

径向基核函数 $K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$ ；

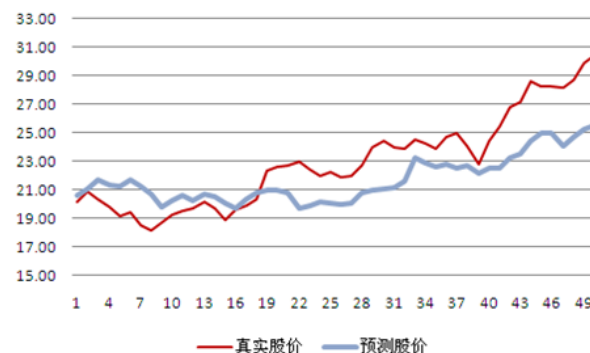
◆我们参考了国内外关于支持向量机预测金融市场的报告，在以往对股票数据研究中一般采用的是**Gauss核函数**。在本报告的初始系列里，我们暂且不为核函数的选择进行详细的讨论，而是按早前人的经验采用**Gauss核函数**。

3.6 个股走势预测

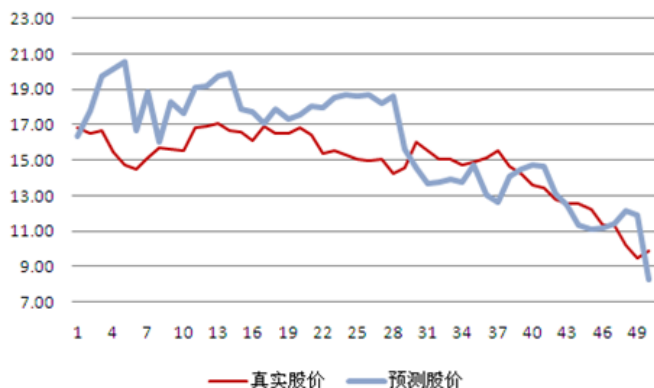
浦发银行的预测走势1，07-8-6至07-8-31（20交易日）



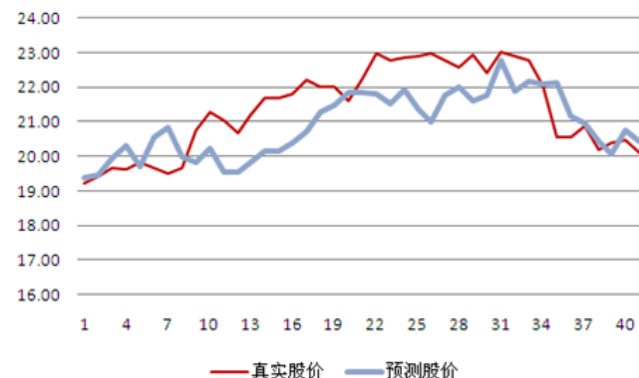
浦发银行的预测走势2，07-6-27至07-9-4（50交易日）



浦发银行的预测走势3，08-7-10至08-9-18（50交易日）

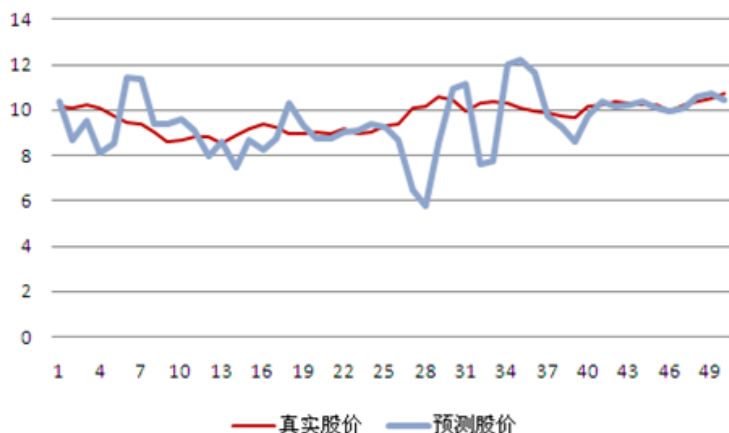


浦发银行的预测走势4，10-2-1至10-4-27（45交易日）
5日平均价



3.7 个股预测结果统计

常山股份的预测走势，10-1-14至10-3-31（50交易日）



中信证券的预测走势，10-2-24至10-5-6（50交易日）



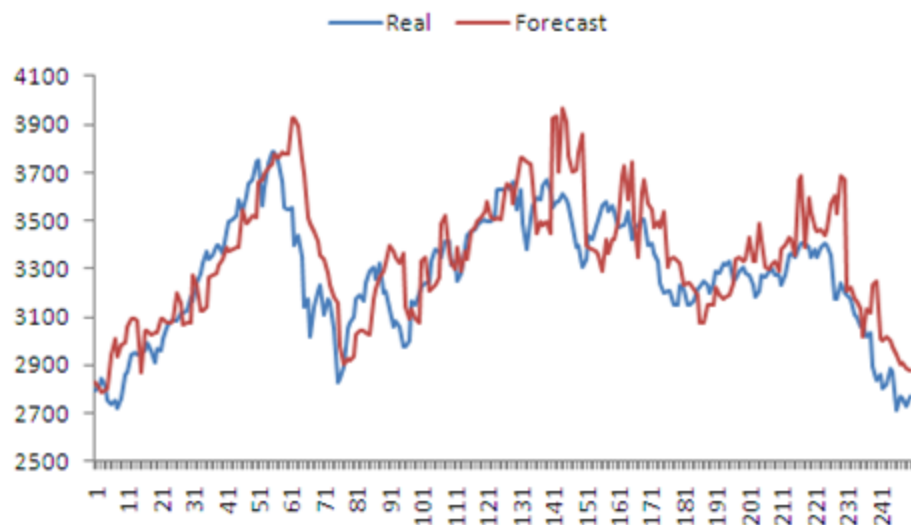
以上的实证结果表明：

1. SVM简单预测模型在当所预测的未来若干天的股票价格趋于平稳上升或者下降状态时，该模型预测的效果最佳，用作预测股票价格的大体走势可信度较高。
2. 然而在股票价格突然快速上涨或者下跌时，模型的预测往往不能立即跟上真实行情的变化速度，导致模拟走势与实际情况存在差异。
3. 实证例子还表明在牛市和熊市中，预测效果比震荡市中的效果更为准确。原因是震荡市中所要预测的实际股票价格时涨时跌，预测的曲线存在很多的急转弯，影响了其正确性。但一般来讲，从长远来看（50交易日），该模型还是能够描述出实际股票价格的大体走势，只是如果当预测走势的趋势不确定时，说明实际股票数据可能变化较小或震荡可能性较大，因此可以考虑不对该股票进行投资或者抛出该股票。

3.9 SVM预测沪深300

◆用50个交易日为训练集预测5个交易日的
方法，绘制了下面的近一年沪深300预测图
形。通过反复观察，我们发现预测走势有滞
后真实走势的现象，两者相关系数为0.78，
预测每日涨跌的准确率为68.5%。

沪深300指数的预测走势和真实走势



多头策略



多空策略



3.10 实证结论

以上的实证结果表明：

1. 结合小波消噪的SVM预测模型在刻画预测走势的形态上具有较好的准确度，预测每日涨跌的准确率也较高，预测指数下跌的效果较好。在指数价格出现震荡的时期，模型的预测往往不能立即跟上真实行情的变化速度，导致模拟走势与实际情况存在差异。
2. 基于预测数据的简单择时策略，在震荡市中效果理想。多头策略可以很好的规避市场下跌造成的损失，保留收益。而多空策略中做空获得的收益要高于做多所获收益。表明模型在这一时间段里预测下跌比预测上涨要准确。

两种交易策略均跑赢了大盘，这说明了两个问题，一、模型的预测具有一定参考价值；二、策略需要长期坚持。

各种策略收益率表现

	策略收益率	指数收益率	交易次数	最高回报	最低回报
多头策略	10.33%	1.03%	22	38.18%	-0.95%
多空策略	31.61%%	1.03%	44	34.23%	-13.39%

数据来源：国信证券经济研究所，考虑交易成本，单边费率0.5%。

内容目录

1

数据挖掘

2

小波分析在金融时间序列中的应用

3

支持向量机简介及其在股票市场中的应用

4

小波分析与技术指标结合的投资策略

5

投资建议

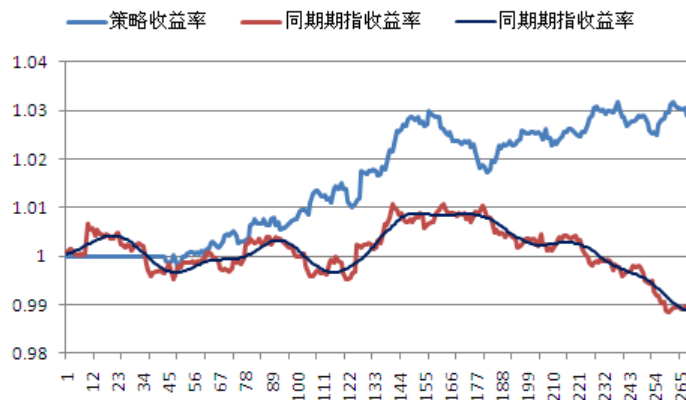
4.1 基于小波分析的高频数据技术指标交易策略

- 运用小波去噪后，消除了高频时间序列的敏感波动，保留了股价走势的大趋势。
- 此时用技术指标，如：MACD，RSI，KD等，效果明显。
- 技术指标的关键在于参数的选择，我们用穷举法用历史数据对参数进行了优化，发现在未来的一段时间内，效果较佳。

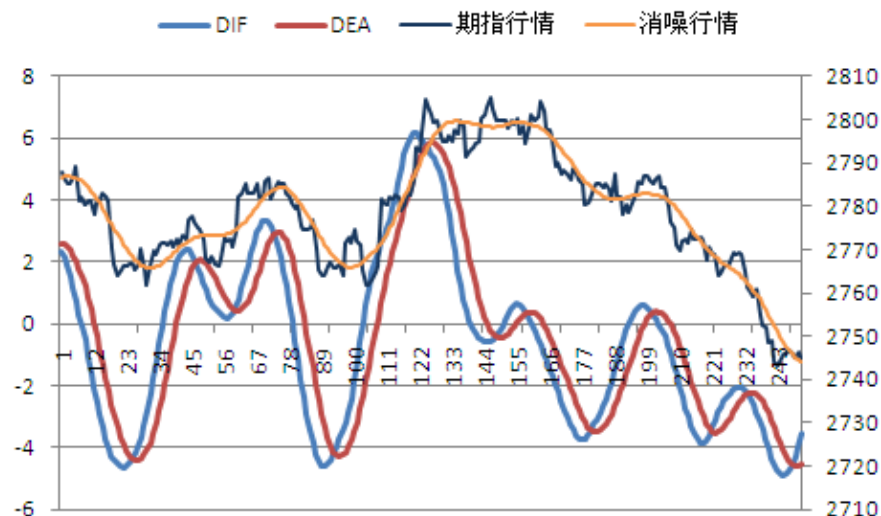
MACD:

- $DIF: EMA(CLOSE, SHORT) - EMA(CLOSE, LONG);$
- $DEA: EMA(DIF, MID);$ 。 DIF上穿DEA，买入。DIF下穿DEA, 卖出。

6月3日多空策略收益率



MACD择时



4.2 策略效果比较

•交易时间：2010年5月7日至2010年6月3日，合计**20个交易日**。

•高频数据频率：1分钟。

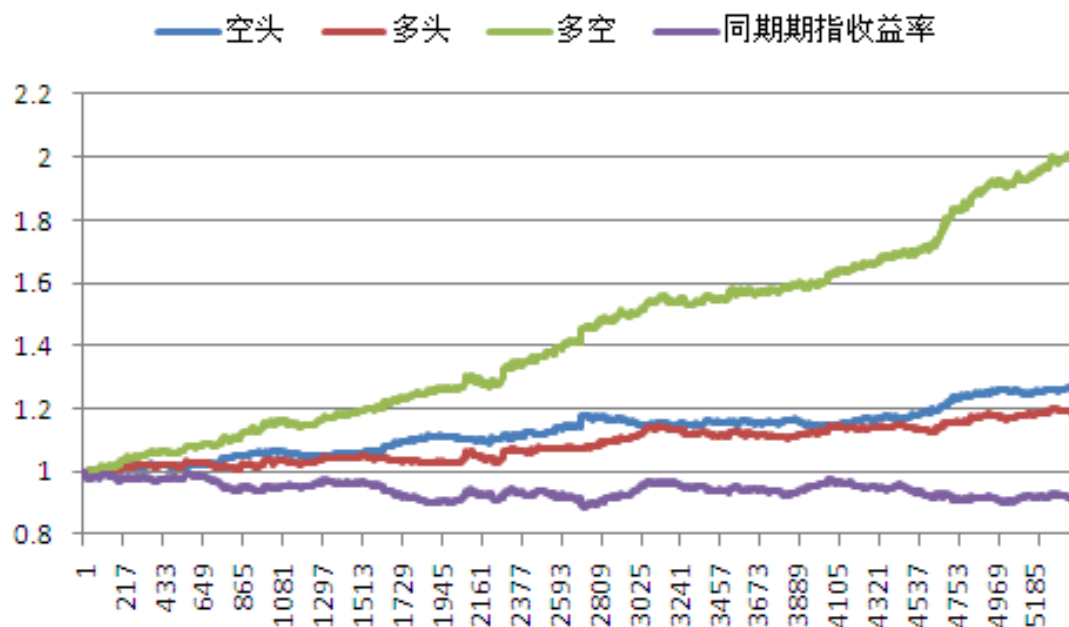
•交易次数：多头、空头各140次，多空策略280次。

•手续费：万分之一

•滑价：2个指数点位。

•策略回报：多头18.7%、空头27%、多空策略100.3%。

MACD高频交易策略



内容目录

1

数据挖掘

2

小波分析在金融时间序列中的应用

3

支持向量机简介及其在股票市场中的应用

4

小波分析与技术指标结合的投资策略

5

投资建议



国信证券经济研究所

Guosen Securities Economic Research Institute

全球视野 本土智慧
GLOBAL VIEW LOCAL WISDOM

