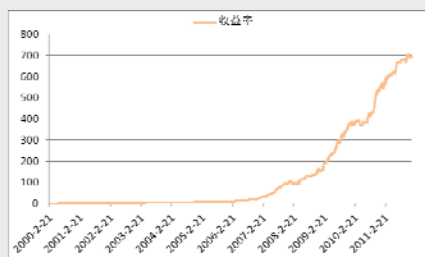


量化择时系列二

水致清则鱼自现——小波分析与支持向量机择时研究

量化择时模型收益比较图



相关研究报告

《基金仓位估算模型》
《欲寻良驹，先觅伯乐——量化选股系列报告一》
《众推之下无弱股——量化选股系列报告之二》
《寻找黑金中的黑马——采掘行业量化选股报告》
《寻觅小公司中的大机会——中小盘量化选股报告》
《春江水暖鸭先知——行业轮动研究报告》
《拨云散雾待月明——量化择时系列报告一》

证券分析师

王韧 首席策略研究员
投资咨询资格编号
S1060210090010
07551-22623376
wangren740@pingan.com.cn

研究助理

周谧 一般从业资格编号
S1060110070105
0755-22625861
Zhoumi259@pingan.com.cn

程宁巧 一般从业资格编号
S1060110080129
0755-22629753
Chengningqiao151@pingan.com.cn

请通过合法途径获取本公司研究报告，
如经由未经许可的渠道获得研究报告，
请慎重使用并注意阅读研究报告尾页
的声明内容。

■ 模型对上证指数取得较好的择时收益

本文采用小波分析加支持向量机的方法构建量化择时模型，并检验了在不同参数条件下预测模型对应交易策略的有效性。最终发现预测模型得到的交易策略对上证指数具有较好择时效果，在训练时间窗为 5 个交易日 的情况下，经过小波分析滤波后得到的预测模型单日预测正确概率可以达到 56.01%，交易成功概率达到 84.65%，每次交易扣除 1% 的交易成本后，从 2000 年 2 月 21 日到 2011 年 12 月 31 日，模型共获得 84.28 倍的收益。

■ 模型对沪深 300 指数的择时收益尚待改进

本文在使用小波分析和支持向量机的模型对沪深 300 指数进行建模后，发现无论是 指数的日数据还是周数据，都取得了明显的预测效果，预测涨跌的正确概率都显著高于 55%，正确概率的均值达到 60%。虽然如此，该模型的预测序列在对应的交易策略中并没有取得较高择时收益。总体来看，周数据交易模型收益率要高于日数据交易模型，滤波后效果最好的交易模型扣除单次 1% 的交易成本后也只创造 4.5 倍的收益。

■ 小波分析对提升模型预测效果功勋卓著

小波分析是目前进行时频分析应用最广泛的工具，在信号异常分析、信号压缩、信号滤波等方面有广泛的应用。本文在实证中发现，无论对于上证数据还是沪深 300 数据，无论对于日交易数据还是周交易数据，无论对于何种长度时间窗，在采用小波分析进行滤波之后，模型预测效果以及对应交易模型的效果都有显著提高。

■ 模型在未来会进行两个方面的改进

坦率的讲，该模型对小规模资金进行择时操作具有较好适应性，而对大资金的择时操作则显吃力，这主要是由于模型交易频率偏高所致。从对模型的实证检验的结果来看，过高的交易频率使得交易成本高企，如果资金规模较大，由于冲击成本提高将使交易成本显著超过 1%。在未来，平安金融工程将会从两个方面改进模型，一个方向是降低模型的交易频率，使得模型更适合大资金操作；另一个方向是提高模型的交易频率，使得该模型更适合投资像股指期货等高交易频率、低交易成本的品种。

正文目录

一、	量化择时的一些思考.....	4
1.1	技术分析的有效性.....	4
1.2	如何构建较好量化择时模型	5
二、	小波分析及支持向量机介绍.....	7
2.1	小波分析的原理介绍	8
2.2	支持向量机的原理介绍.....	12
三、	小波分析及支持向量机实证检验.....	16
3.1	支持向量机对原始数据的预测效果.....	16
3.2	支持向量机对经过小波滤波后数据的预测效果	22
四、	总结.....	28

图表目录

图表 1	常用技术指标的表现情况	4
图表 2	累积收益率为正的概率与预测次数的关系（单次正确概率 $p=55\%$ ）	6
图表 3	累积收益率为正的概率与预测次数的关系（单次正确概率 $p=45\%$ ）	6
图表 4	累积收益率为正的概率与预测次数的关系（单次正确概率 $p=50\%$ ）	7
图表 5	2007.11.23——2012.12.28 上证指数走势图（原始数据）	9
图表 6	2007.11.23——2012.12.28 上证指数走势图（从第二层构建原序列）	9
图表 7	2007.11.23——2012.12.28 上证指数走势图（从第四层构建原序列）	10
图表 8	常用的几种小波函数	11
图表 9	结构风险最小化	13
图表 10	支持向量机源于最优超平面	14
图表 11	支持向量机对原始数据的预测情况	17
图表 12	原始数据建模后的交易策略表现	17
图表 13	支持向量机对周数据预测效果	18
图表 14	周数据建模后的交易策略效果	19
图表 15	沪深 300 日数据建模后的预测情况	20
图表 16	沪深 300 日原始数据建模后的交易策略收益率情况	20
图表 17	沪深 300 周数据建模后的预测情况	21
图表 18	沪深 300 周数据建模的交易策略收益率	21
图表 19	上证指数与展开数据	22
图表 20	滤波之后的上证指数图象	23
图表 21	支持向量机对滤波数据预测效果	23
图表 22	滤波构建支持向量机的交易策略情况	24
图表 23	支持向量机对滤波周数据预测效果	25
图表 24	支持向量机对滤波上证周数据预测效果	25
图表 25	支持向量机对滤波沪深 300 日数据预测效果	26
图表 26	滤波沪深 300 建模后交易策略收益情况	26
图表 27	支持向量机对滤波沪深 300 周数据预测效果	27
图表 28	滤波沪深 300 建模后交易策略收益情况	28

一、 量化择时的一些思考

1.1 技术分析的有效性

在最初进行量化择时研究的时候，首先想到的方法就是技术分析方法。技术分析方法又分为形态分析法和技术指标法。

形态分析法主要根据最近一段时间股票价格的运动轨迹构成的图形判断未来市场的涨跌，有些形态分析只根据每天的单一价格（开盘价或收盘价等）路径所呈现的图形对大盘走势进行预测，有些形态分析需要根据一段时间的K线数据组成的图形对大盘进行预测。形态分析最主要的特点就是主观分析比较容易，但是由于人脑从事机械运算的能力较差，导致主观分析的样本量较少，统计出来的结果在统计意义上不具有说服力。此外由于人的主观意识作祟，每个人对技术图形的判断有所不同，这也造成了技术形态法仁者见仁，智者见智。由于形态分析法具有上述特征，使得其难以用量化的方式进行分析，目前也有少量的一些报告用形态分析法进行研究，主要以模式识别法为主。

技术指标法主要根据最近一些价格序列或者交易量序列的四则运算得到了新的时间序列对大盘进行分析，形成对大盘未来的判断。技术指标法最大好处在于指标构建比较客观，适合做程序化交易。从我们跟一些做股指期货期货以及外汇交易的投资者交流情况来看，技术指标在程序化交易中仍然有效，但是现在的技术指标法已经不再像原来那样仅仅依靠几个简单的公式就能获利了。目前期货以及外汇交易所使用的技术指标法通常都是数个技术指标的组合或者在普通技术指标基础上经过重新计算所得到的，这些指标也主要被投资者运用在了高频交易的领域，这种方法对头寸规模有较大的限制。

我们也检验了一些常用单技术指标，发现市场对这一块的信息已经基本挖掘充分，很难再靠这种简单的指标获得较高收益率。

图表1 常用技术指标的表现情况

指标名称	交易次数	成功次数	成功概率	总收益率
MACD	75	31	41.48%	26.7%
RSI	228	123	54%	46.7%
单均线	60	26	42.9%	505%
双均线	62	29	46.8%	602%
ROC	146	53	36.6%	20%
KDJ	220	81	36.7%	5%
PSY	10	3	33%	-34%
OBV	98	35	35.9%	153%
BOLL	179	69	38.3%	123%

资料来源：平安证券研究所、

上面的图表给出了大家常用的一些技术指标及其择时效果的情况，上述表格所取的行情是从2000年1月4日——2011年11月30日，共2879个交易日。可以看出单一技术指标的择时效果都不太理想，因

此我们应该考虑用多个技术指标共同得到量化择时的模型。

1.2 如何构建较好量化择时模型

■ 1.2.1 量化择时模型的两个维度

量化择时模型需要从两个维度考虑问题：第一个维度是择时模型发出信号的频率，第二个维度是择时模型发出信号的准确率，两个维度共同决定着量化择时模型的效果，并且决定着量化择时模型的适用性。

频率既是指择时模型在一定时间内发出的买卖交易信号的次数。对于基金的投资者来讲，频率高于1次/日的择时模型都是没有用的。由于股票交易的T+1机制，当天买入的股票只能在第二天卖出，因此我们的量化择时信号发出买卖交易信号的时间间隔必须要大于1日。而对于高频交易和日内交易的投资者则不存在这个问题。频率越高也意味着对单次获利的要求提高，由于每次交易都会产生交易佣金、印花税等费用，因此频率越高的量化择时模型将产生更高的交易费用，只有大部分买卖的获益是大于交易成本的，才能保证模型获利。此外频率越高意味着该模型能够容纳的资金规模越有限，时间间隔越小，在时间间隔内进行大规模的资金换手将产生更大的冲击成本，从而将低买高卖的收益抵消掉。因此在进行量化择时模型时，一定要根据资金规模的大小来决定量化择时模型的频率。准确率是量化择时模型的第二个维度，衡量模型在一段时间内给出交易信号的正确概率，正确概率是量化择时模型的根本，一个量化择时模型只要正确概率高，而交易次数以及平均每次交易的盈利不是差得离谱，都是一个好模型。当然也有一些交易模型比较有点，虽然成功概率不高，但是成功一次所带来的收益率是很高的，而每次亏损的收益率比较小，这种模型仍然可以赚钱，其主要的功劳是在于止损策略的设置，使得平均盈利收益率/平均亏损收益率具有很高的比例，这种模型的收益率也呈现阶梯型的上升。而成功概率较高的收益率模型，其收益率的图形基本是一直往上生长的，基本不回头。

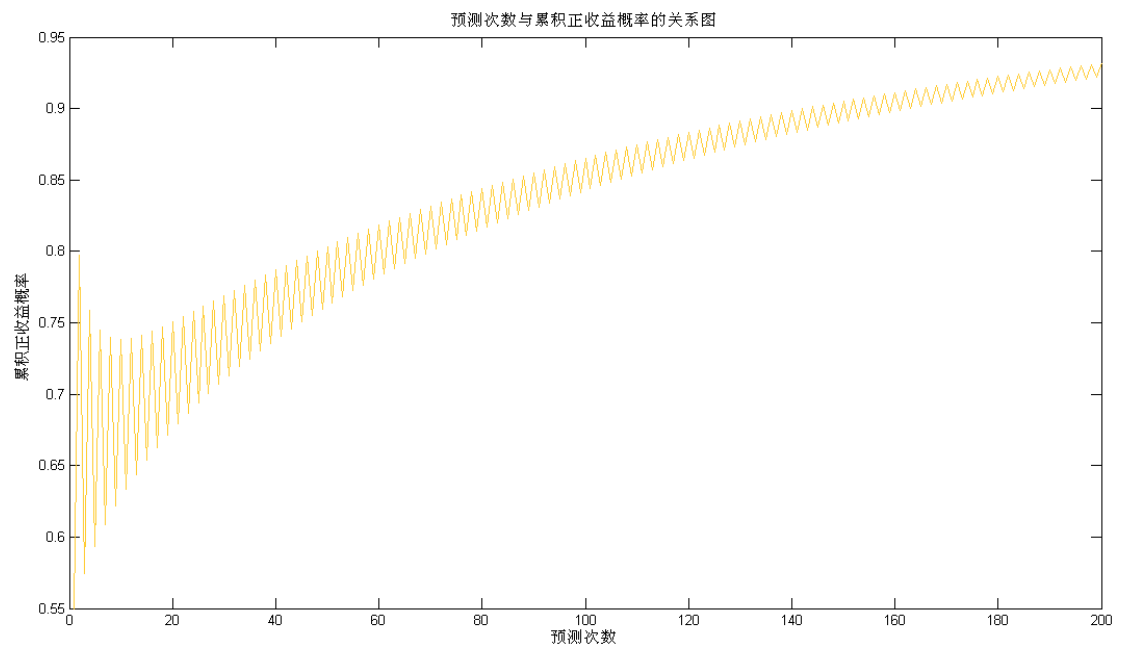
在《量化择时系列报告——拨云散雾待月明》中我们提到过，市场呈现一种类似随机游走的混沌状态，因此想对市场的长期趋势做预测几乎是不可能的，不过对短期市场趋势而言，做预测仍然是有可能的，不过由于其混沌性，我们只能保证预测趋势有一定的正确概率，而不能每次都预测准确。

■ 1.2.2 量化择时模型的需要满足的条件

由于金融工程的预测讲求一定的预测概率，而预测的次数过少的话，最终预测的总收益率波动会比较大，当预测次数增加的话，由大数定律可知预测总收益率为正的概率将增加。

我们举个例子：假定有一个预测模型，每次预测正确的概率为55%，每次预测正确收益率为1%，预测错误收益率为-1%，则预测总收益率为正的概率与预测次数的关系如下图：

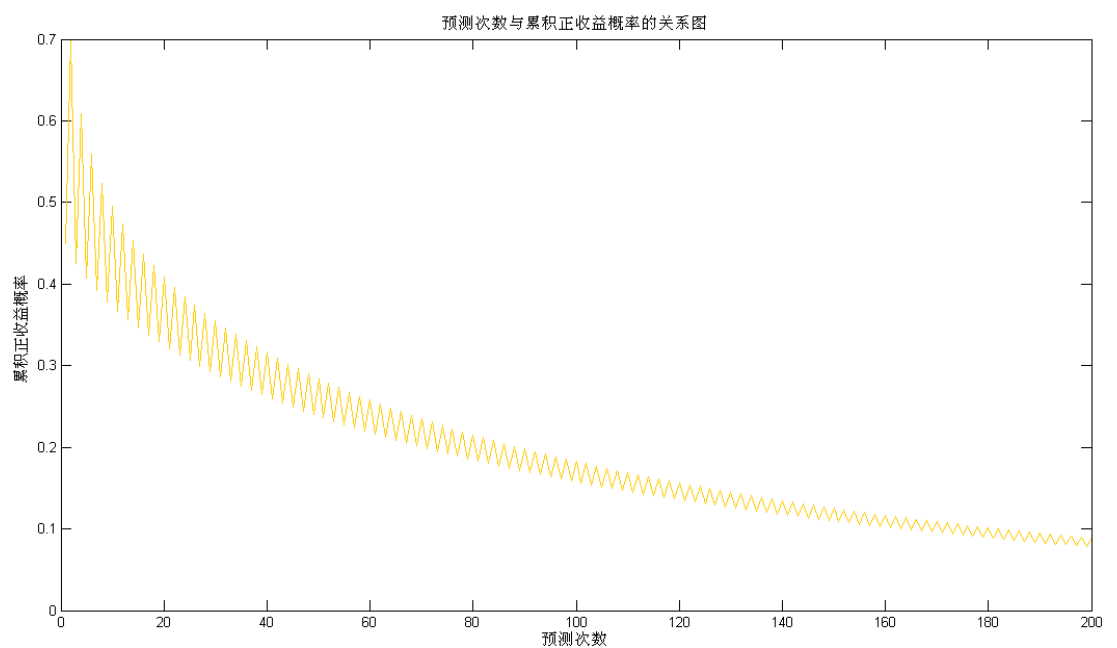
图表2 累积收益率为正的概率与预测次数的关系（单次正确概率 $p=55\%$ ）



资料来源：平安证券研究所、

另外再举个例子：假定有一个预测模型，每次预测正确的概率为45%，每次预测正确收益率为1%，预测错误收益率为-1%，则预测总收益率为正的概率与预测次数的关系如下图：

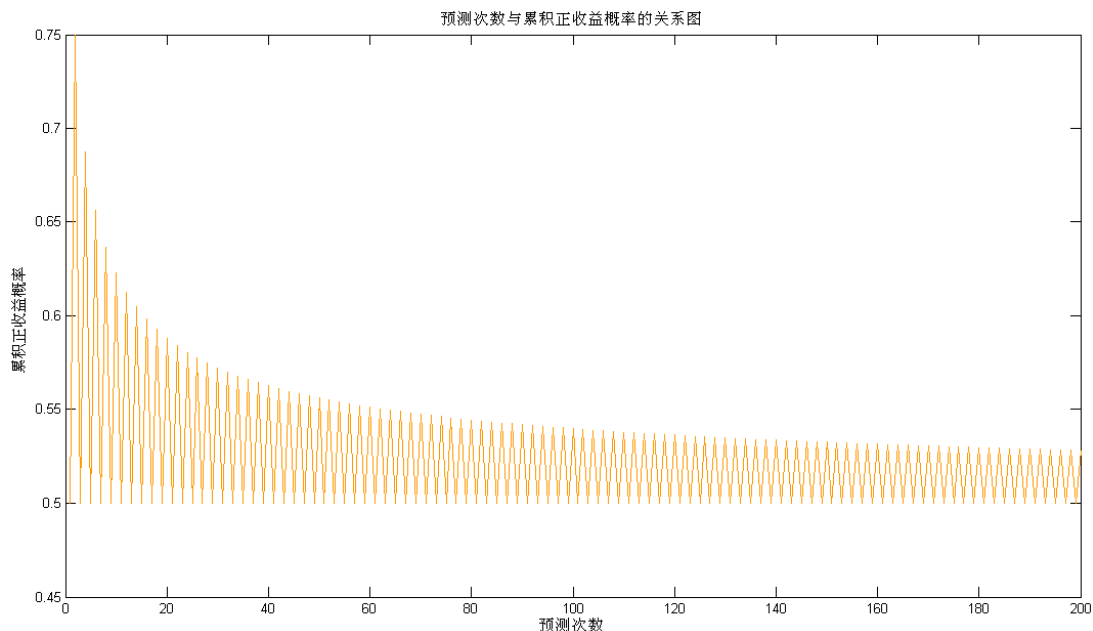
图表3 累积收益率为正的概率与预测次数的关系（单次正确概率 $p=45\%$ ）



资料来源：平安证券研究所、

最后再看例子：假定有一个预测模型，每次预测正确的概率为50%，每次预测正确收益率为1%，预测错误收益率为-1%，则预测总收益率为正的概率与预测次数的关系如下图：

图表4 累积收益率为正的概率与预测次数的关系（单次正确概率p=50%）



资料来源：平安证券研究所。

从上图可以看出，预测模型以50%的概率为界，当成功概率大于50%时，随着预测次数的增加，累积的预测收益率为正的概率接近于1，当成功概率小于50%时，随着预测次数增加，累积预测收益率为正的概率接近于0，而当成功概率等于50%时，随着预测次数增加，累积预测收益率为正的概率为50%。

因此我们在构建量化择时模型时，首要的要务是要提高模型的择时正确概率，其次要保证模型有一定的预测次数，当然由于交易成本和冲击成本的限制，该模型的交易次数不能过高。

上述模型只是上涨和下跌等幅的简化模型。假如上涨和下跌的幅度不一致，则上涨幅度和上涨概率，下跌幅度和下跌概率应该保证如下的关系才能保证该量化择时模型是个好模型：

$$\begin{cases} p(1+r_p) + q(1+r_q) \geq 1 \\ p(1+r_p) > q(1+r_q) \end{cases} \quad \text{—— (1)}$$

其中：

p 为模型预测正确的概率

q 为模型预测错误的概率

r_p 为模型预测正确时的平均超额收益率

r_q 为模型预测错误时的平均超额收益率

二、小波分析及支持向量机介绍

在量化择时报告一中，曾经采用了神经网络的方法对大盘进行预测。神经网络最大的问题就在于它

所拟合出的映射关系有可能不是全局最优的，这样得到的预测结果有可能因为过拟合导致其外推性比较差。支持向量机则是一种很好的进行预测的工具，它最大的好处在于可以得到全局最优解，这样它的预测能力和精度就有可能大幅提升。本文还引入一种工具——小波分析，它最大的特点就是可以起到平滑和过滤的作用，将价格序列中过于频率的微小变动滤掉，留下价格运动的主要趋势。

2.1 小波分析的原理介绍

■ 小波分析的应用背景

小波分析最早的应用领域是在数字信号处理领域。数字信号处理的数学方法分为时域、频域和时频域三个方面。

时域分析主要涉及对时域信号本身进行分析或利用数理统计的方法来获取信号中的某些特征信息（峰值、均值、相关函数等）。时域分析最大的缺陷就是对序列的分析过于简单，只能从序列中提取较少的特征，而大部分的信息则抓不住。

频域分析方法就是利用离散傅里叶变换将离散的时间信号序列转换到频率域中去，得到信号的频谱，也常被叫做频谱分析。离散傅里叶变换理论上将一个一信号展成无数个周期信号之和，但是在计算机里是无法处理无穷项级数和的，最终都要化成有限级数的和，这样离散傅里叶变换就便于了计算机的处理。1965年 J.W.Cooley 和 J.W.Tukey 提出了快速傅里叶变换，将傅里叶变换的运算时间减少了几个数量级，使得傅里叶在信号处理等各个领域有了大规模的应用。不过福利叶变换具有以下几个明显的缺陷：

（1）傅里叶变换只适用于分析平稳的信号，对于非平稳信号则无能为力。由于傅里叶分析的每一个分量函数都是一个固定周期的时间上无限延拓的信号，在某个时间点上的信号变化需要在所有分信号的系数上有所体现，而由于离散傅里叶变换采用有限加和的方法将导致信号的变化不能很好反映出来。

（2）为了得到一个时域信号的频域特征，必须使用信号在时域中的全部信息，甚至将来信息。

（3）如果一个信号只在某一时刻的一个小邻域内发生了变化，那么信号的整个频谱都要受到影响。此外，在频谱的变化上，我们无法得知邻域内发生的信号变化的剧烈程度。

（4）当一个信号的频率变化比较剧烈时，此时在不同的时域区间下需要采用不同频率的信号进行分析，而傅里叶分析中每个分量信号的频率是固定的，在分析此种信号时，每个信号分量都比较乏力。

时频分析将信号转换到时间——频率二维平面上，处理后的信号不仅具有频率信息，而且具有时间信息。时频分析的产品正是针对频域分析的这些缺陷所产生的。在时频分析中最著名的方法就是小波分析法。小波变换是一种时间——尺度分析法，尺度与频率相对应。它具有以下特点：

（1）在高频范围内时间分辨率高；在低频范围内频率分辨率高。既适合于分析平稳信号，也适合与分析非平稳信号。

（2）有快速算法——Mallat 算法。Mallat 算法在小波分析中的地位相当于快速傅里叶分析在傅里叶分析中的地位。Mallat 算法可以通过普通的四则运算就能实现信号的小波分解和重构，这比通过计算积分进行小波变换方便多了。有了 Mallat 算法之后，小波分析方法迅速流行起来。

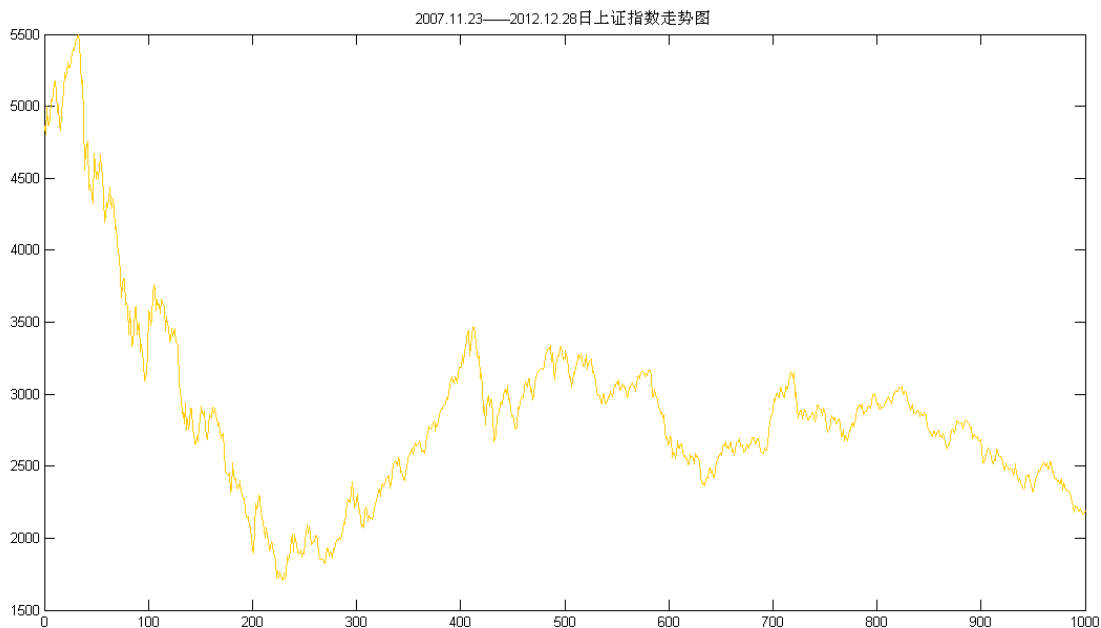
（3）利用离散小波变化可以将信号分解到各个尺度上。

小波变换应用邻域非常广泛，目前在数据压缩、图像处理、机械故障诊断、信号降噪、边缘检测、参数辨识、函数学习、语音识别与合成等方面都有应用。

目前也有许多学者在研究小波分析在金融领域中的应用，应用最广泛的一块就是对金融资产的价格

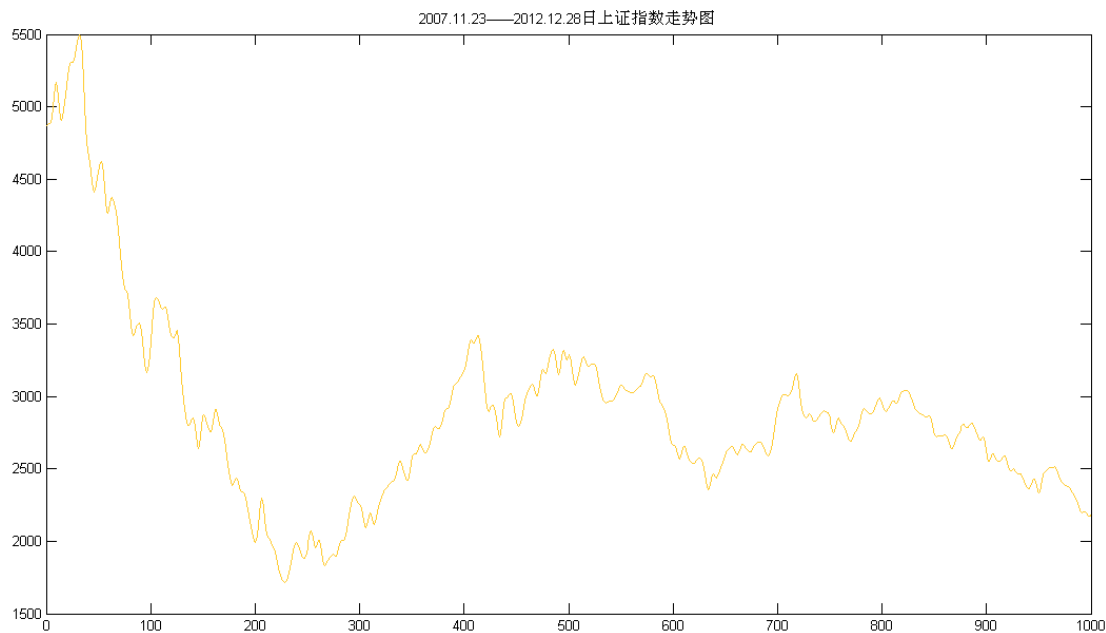
序列进行小波分析。金融资产的价格序列可以视作信号源产生的时域信号，该信号通常是非平稳的序列，并且其变动的速度和频率在不同时间区间是不同的，这种情形下的信号用小波分析是再好不过的了。以上证指数序列为例，将上证指数每个交易日的收盘价连起来，形成上证指数价格走势，该图从图像上类似于布朗运动，并且有很多“小毛刺”。这些小毛刺对我们进行大盘的预测起着干扰的作用，因此我们需要用小波分析的方法滤掉这些小毛刺，使得市场的主要趋势呈现出现，方便我们对大盘的预测。

图表5 2007.11.23——2012.12.28日上证指数走势图（原始数据）

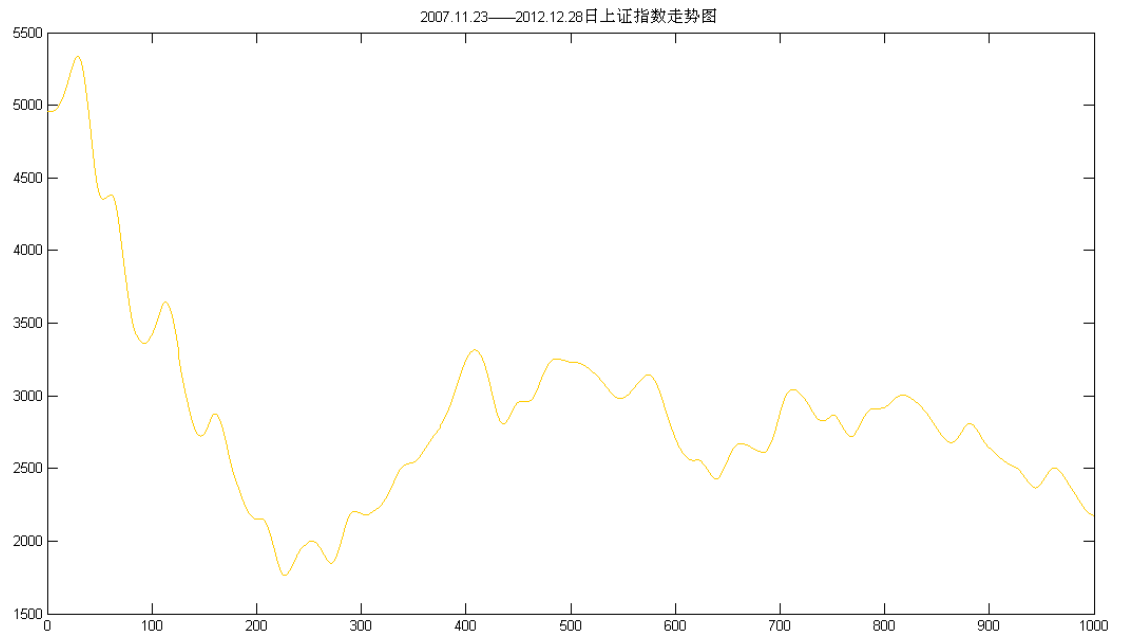


资料来源：平安证券研究所、

图表6 2007.11.23——2012.12.28日上证指数走势图（从第二层构建原序列）



资料来源：平安证券研究所、

图表7 2007.11.23——2012.12.28日上证指数走势图（从第四层构建原序列）

资料来源：平安证券研究所。

■ 小波分析的数学原理

小波的定义：设函数 $\psi \in L^2 \cap L^1$ 且 $\hat{\psi}(0) = 0$ ，则按如下方式生成的函数族 $\{\psi_{a,b}\}$

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad \text{其中 } b \in R, a \in R - \{0\} \quad \text{—— (2)}$$

叫分析小波或连续小波， ψ 叫基本小波或母小波。

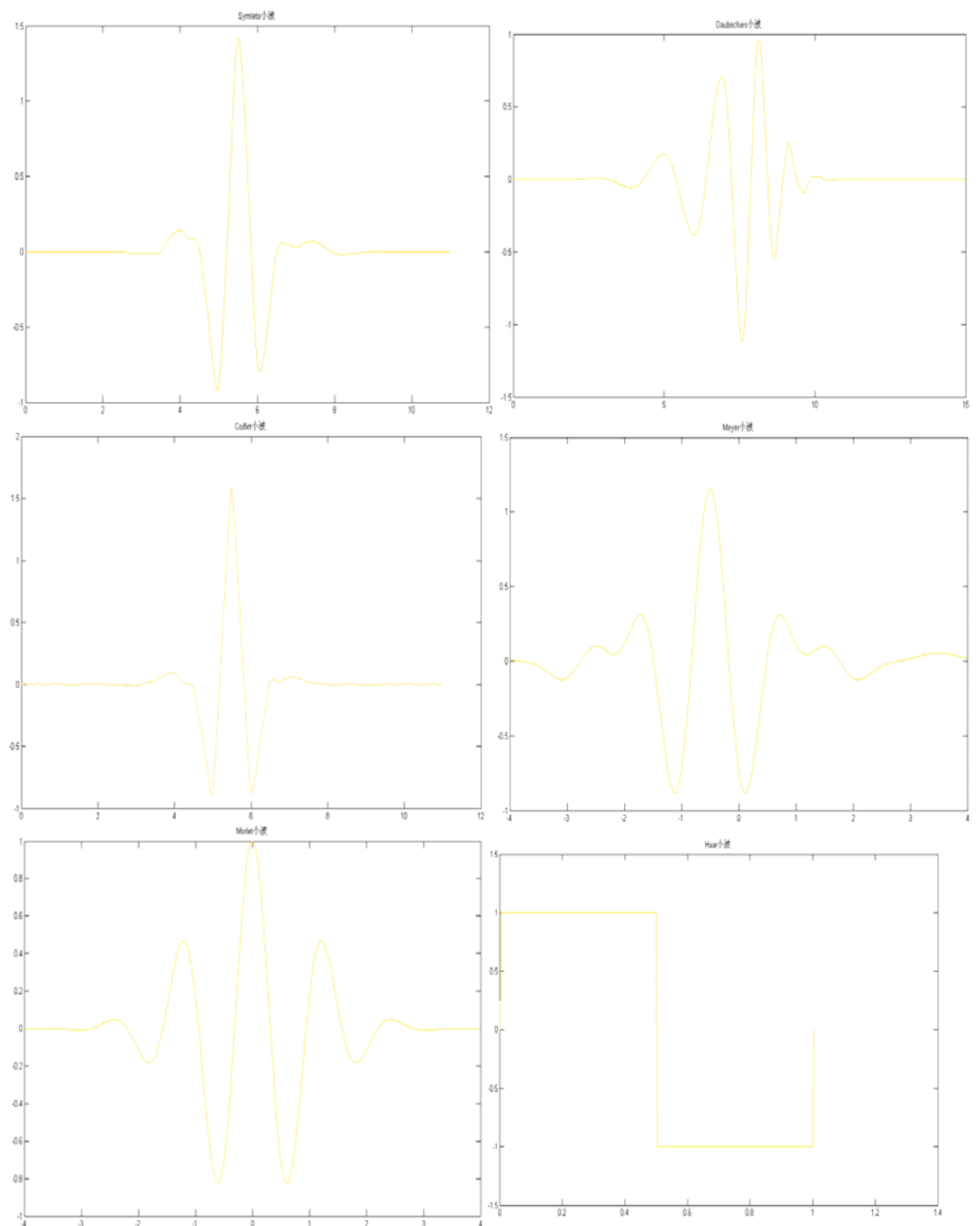
小波具有两个特点：

（1）“小”，即它在时域里都具有紧支集或近似紧支集（任何一个开时间域，都可以用有限个小波进行分析）。

（2）“波”，即小波的波动性，小波的数值具有正负交替的特性。

采用具有上述两种性质的有限定义域的函数，我们就可以在任意的连续时域空间上，对信号进行频谱分析。常见的小波函数有以下几种：

图表8 常用的几种小波函数



资料来源：平安证券研究所、

小波变化定义: 设 ψ 是基本小波, $\{\psi_{a,b}\}$ 是式 (2) 定义的连续小波, 则信号 $f \in L^2$ 的连续小波变换 (CWT) 定义为

$$(W_{\psi} f)(a, b) = \langle f, \psi_{a,b} \rangle = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (3)$$

其中, “ $\langle \rangle$ ” 表示内积, 基本小波 ψ 满足如下允许性条件:

$$C_{\psi} = \int_{\mathbb{R}} |\omega|^{-1} |\hat{\psi}(\omega)|^2 d\omega < \infty \quad \text{--- (4)}$$

其中， $\hat{\psi}$ 为 ψ 的傅里叶变换， $W_{\psi} f(a, b)$ 成为小波系数。

小波的反变换定义为：

$$f(t) = \frac{1}{C_{\psi}} \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} W_{\psi} f(a, b) \left[\frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \right] db \right\} \frac{da}{a^2} \quad \text{--- (5)}$$

上述介绍的是连续小波变换，而真正在实际运用当中所用的是离散 $\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right)$ 小波。

其中 a 是尺度， b 是偏移量。在离散小波变换中，通常把尺度 a 和偏移 b 取作幂级数形式，即

$a = a_0^j$ ， $b = ka_0^j b_0$ ， $k \in \mathbb{Z}$ 。对应的离散小波为

$$\psi_{j,k}(t) = a_0^{-\frac{j}{2}} \psi(a_0^{-j} t - kb_0)$$

信号 $f(t)$ 的离散小波变换系数为

$$C_{j,k} = \int_{-\infty}^{+\infty} f(t) \bar{\psi}_{j,k}(t) dt$$

重构公式为

$$f(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} C_{j,k} \psi_{j,k}(t)$$

2.2 支持向量机的原理介绍

■ 支持向量机背景介绍

支持向量机（SVM）是建立在统计学习理论基础之上的新数据挖掘方法，在解决小样本，非线性和高维模式识别问题中表现出许多特有的优势，它的理论依据主要来源于结构风险最小化原理和VC维数。支持向量机和二分类模式识别问题，也有人用支持向量机进行函数拟合，根据这两种用法，支持向量机又分为支持向量分类机和支持向量回归机。

目前常用的另外一种数据挖掘的方法是神经网络方法，与神经网络相比，支持向量机有坚实的统计学基础，它具有以下优点：

（1）以结构风险最小原理为基础，减小推广错误的上界，具有很好的推广性能，解决了神经网络的过适应现象。

（2）问题的求解等价于线性约束的凸二次型规划问题，具有全局最优解，解决了神经网络的局部极小问题。

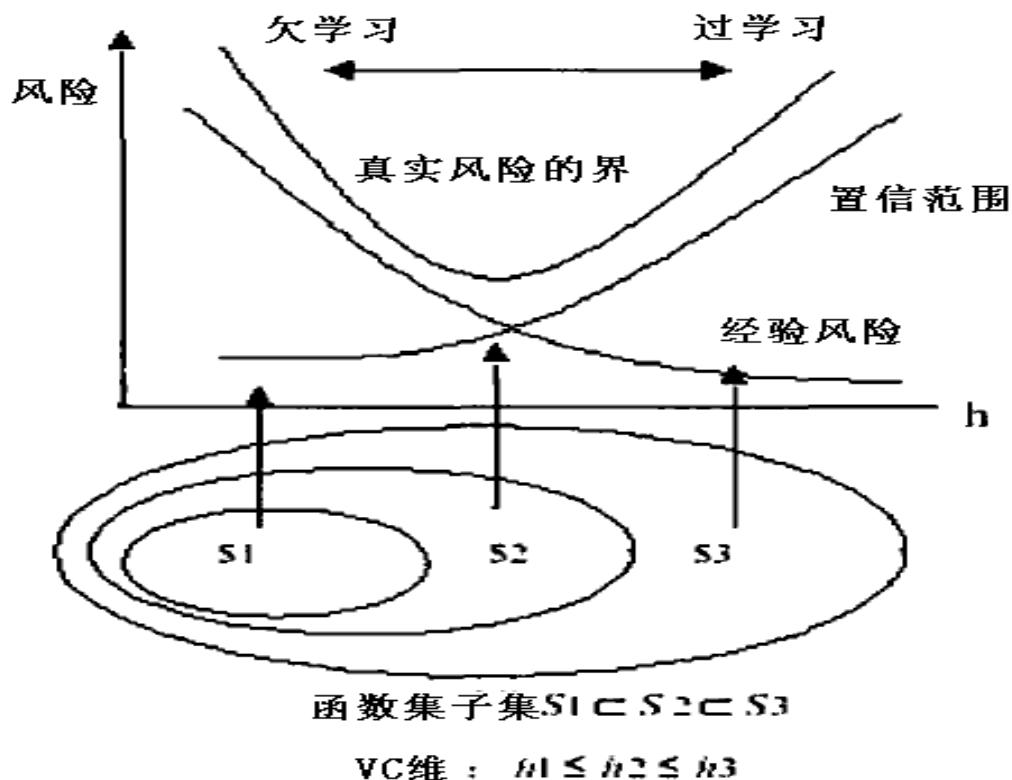
（3）把原问题映射到高维空间，通过在高维空间构造线性分类函数来实现原问题的划分，引入核函数，解决了维数灾难问题。

结构风险最小化原理：在统计学习理论中，真实的风险是由两部分内容来刻画的，一是经验风险，它代表一个分类函数在给定样本上的误差；二是置信风险，代表了我们可以多大程度上可以信任分类器在未知文本上分类的结果。我们进行统计学习时，所得到的最优函数追求的不是经验风险最优化，而是结构风险最优化。

在机器学习方法都把经验风险最小化作为努力的目标，但后来发现很多分类函数能够在样本集上轻

易达到100%的正确率，在真实分类时却一塌糊涂（即所谓的推广能力差，或泛化能力差）。此时的情况便是选择了一个足够复杂的分类函数（它的VC维很高），能够精确的记住每一个样本，但对样本之外的数据一律分类错误。之所以出现这样的情况主要是因为在小样本的学习中，如果样本数量相对与总样本量过小，以至于这个小样本不能很好地反应所有样本的性态，这样就导致了在进行外推时效果很差。此时需要知道我们可以在多大程度上相信分类函数对未知样本分类的能力，这个需要用置信风险来衡量。

图表9 结构风险最小化



资料来源：平安证券研究所。

VC 维数：学习算法的推广能力是学习机器的重要方面，它建立在 VC 维的基础之上。VC 维是目前为止对函数学习性能的最好的描述指标。

VC 维的定义如下：设 F 是一个 n 维向量 X 到 $\{0,1\}$ 的函数族，则 F 的 VC 维定义为 X 子集 E 的最大元素数， E 满足：对任意 $S \subseteq E$ ，总存在 $f \in F$ ，使得当 $x \in S$ 时， $f(x)=1$ ， $x \notin S$ 但 $x \in E$ 时， $f(x)=0$ 。

上述的数学定义可以用直观的描述：若存在一个 n 个样本的样本集能够被一个函数集中的函数按照所有可能的 2^n 种形式分为两类，则称函数集能够把样本数为 n 的样本打散。VC 维就是函数集中的函数能够打散的最大样本集的样本数目。

VC 维反映函数能够打散的最大样本集的样本数目。如果任意数目的样本都有函数将它们打散，则函数集的 VC 维为无穷大。VC 维反映函数集的学习能力，VC 维越大说明学习机越复杂。根据 VC 维，我们就能得出一个机器学习机的结构风险的上界公式：

对函数集 $f(x, w)$ ，损失函数 $L(y, f(x, w))$ 取值为 0 或 1，则对所有的风险函数和该损失函数，经验风险和实际风险至少以概率 $1-\eta$ 满足

$$R(w) \leq R_{emp}(w) + \sqrt{\frac{h(\ln n(2n/h) - n(\eta/4))}{n}} \quad \text{—— (6)}$$

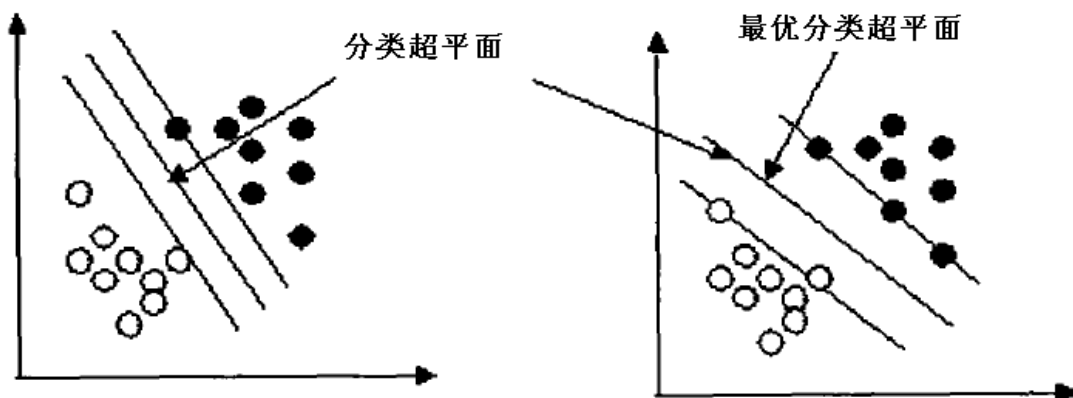
其中 n 是样本数， h 是学习机器的 VC 维，上述公式给出了置信概率与实际风险上界之间的关系。

■ 支持向量机的数学原理

➢ 支持向量分类机

支持向量分类机主要用于解决分类问题。根据给定的训练集 $T = \{(x_1, y_1), \dots, (x_n, y_n)\} \in (X \times Y)^n$ ，其中 $x_i \in X = R^n$ ， $y_i \in Y = \{-1, 1\}$ ， $i = 1, 2, \dots, n$ ，寻找 $X = R^n$ 上的一个实值函数 $g(x)$ ，使之可以用决策函数 $f(x) = \text{sgn}(g(x))$ 推断每一个 x 相对应的 y ，也就是找到一个把 R^n 上的点分成两部分的规则。

图表10 支持向量机源于最优超平面



资料来源：平安证券研究所。

给定训练样本 $T = \{(x_1, y_1), \dots, (x_n, y_n)\} \in (X \times Y)^n$ ，如果训练集中的所有向量均能被某超平面正确划分，并且距离平面最近的异类向量之间的距离最大(即边缘最大化)，则该超平面为最优超平面。(如图 10) 其中距离超平面最近的异类向量被称为支持向量(Support Vector, SV)，一组支持向量可以唯一确定一个超平面。SVM 是从线性可分情况下的最优分类面发展而来，其超平面记为 $(w \cdot x) + b = 0$ ，我们对它归一化，使得对线性可分的样本集满足：

$$y_i(w_i \cdot x + b) - 1 \geq 0, i = 1, 2, \dots, n$$

支持向量与超平面之间的距离为 $1/\|w\|$ ，异类支持向量之间的距离为 $2/\|w\|$ ，因此问题转化为求 (7) 的最小值，即

$$\rho(w) = \frac{1}{2}(w' \cdot w) \quad \text{—— (7)}$$

另外，考虑到可能存在一些样本不能被分类超平面正确划分，因此引入松弛变量

$$\xi_i \geq 0, i = 1, 2, \dots, n \quad \text{—— (8)}$$

超平面的约束变为：

$$y_i(w_i \cdot x + b) \geq 1 - \xi_i, i = 1, 2, \dots, n \quad \text{—— (9)}$$

引入约束条件(8)、(9)之后，转化为求下式的极小值

$$\rho(w) = 0.5 \times (w' \cdot w) + C \sum_{i=1}^n \xi_i \quad \text{--- (10)}$$

最终(10)式及其约束条件转化为以下对偶问题：

$$\begin{aligned} \min \quad & 0.5 \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^n a_i \quad \text{--- (11)} \\ & \sum_{i=1}^n y_i a_i = 0, \quad i = 1, 2, \dots, n \\ & 0 \leq a_i \leq C \quad i = 1, 2, \dots, n \end{aligned}$$

该问题为一个线性约束的凸一次规划问题，有唯一解。其中大多数为零解，只有很少一部分是非零解，这部分非零解为支持向量，记作 a^* ，判别函数为

$$f(x) = \text{sgn} \left(\sum_{SV} a_i^* \cdot y_i \cdot (x_i' \cdot x_i) + b^* \right) \quad \text{--- (12)}$$

在线性支持向量机算法中，数据是以内积形式 $\langle x_i, x_j \rangle$ 表示。而在非线性情况下，需要用一个非线性映射将输入空间映射到某一特征空间，记为 $\Phi: R^n \rightarrow H$ 。然后在特征空间中利用和线性支持向量机一样的方法求解。如果存在一种核函数 K ，使得 $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ ，就可以在特征空间中进行计算，而且并不需要知道具体的映射 Φ 。

这样利用核函数就可以将线性支持向量机中的内积运算用核函数代替，相应地变为

$$\begin{aligned} \min \quad & W(a) = 0.5 \cdot a' H a + e' a \quad \text{--- (13)} \\ & 0 \leq a \leq C, a' y = 0 \end{aligned}$$

其中， $H_{ij} = y_i y_j K(x_i, x_j)$ 是一个 $n \times n$ 对称阵。 $e = (1, 1, \dots, 1)'$ ， $y = (y_1, y_2, \dots, y_n)'$ ， $C = (c, c, \dots, c)'$ 。

其中核函数通常有以下两种：

多项式核函数 $K(x, x') = [(x \cdot x') + c]^q \quad c \geq 0$

Guass 径向基核函数 $K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{\sigma^2}\right)$

➤ 支持向量回归机

支持向量分类机主要用于模式识别，而支持向量回归机主要用于进行预测。支持向量回归机的数学原理也是基于二次规划。

假定有样本集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ， $x_i \in R^n, y_i \in R, i = 1, 2, \dots, n$ ，则回归问题为

$$\begin{aligned} \min_{\xi, \xi^*, w} \quad & [0.5 \|w\|^2 + C(\sum_{i=1}^n \xi_i + \sum_{j=1}^n \xi_j^*)] \quad \text{--- (14)} \\ & y_i - w \cdot x_i - b \leq \varepsilon + \xi_i, \quad i = 1, 2, \dots, n \\ & w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^*, \quad i = 1, 2, \dots, n \\ & \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{aligned}$$

该问题的对偶问题为：

$$\begin{aligned} \min_{a, a^*} \quad & \varepsilon \sum_{i=1}^n (a_i + a_i^*) - \sum_{i=1}^n y_i (a_i - a_i^*) + 0.5 \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*)(a_j - a_j^*) K(x_i, x_j) \quad (15) \\ & \sum_{i=1}^n a_i = \sum_{j=1}^n a_j^* \\ & 0 \leq a_i \leq C \quad i = 1, 2, \dots, n \\ & 0 \leq a_i^* \leq C \quad i = 1, 2, \dots, n \end{aligned}$$

支持向量回归机是一种非常好的非线性函数拟合的工具，适合于用小样本数据进行预测。

三、小波分析及支持向量机实证检验

本章将用小波分析和支持向量机构建大盘指数的预测模型。首先抛开小波分析，直接用支持向量机建立预测模型，然后再用小波分析法对原始行情数据进行滤波之后的行情建立支持向量机的预测模型，可以发现模型的预测精度有了显著的提高，最终本文决定将小波分析与支持向量机结合起来，以期得到更优的预测效果。

3.1 支持向量机对原始数据的预测效果

本文分别选用了上证指数和沪深 300 指数作为进行预测的标的指数。其中上证指数的检验样本从 2000-1-4 至 2011-12-31 日共 2901 个交易日，沪深 300 指数从 2005-1-4 至 2011-12-31 日共 1701 个交易日。上证指数共有 600 个交易日，沪深 300 指数共有 353 周。

用支持向量机进行预测需要经历以下一些步骤：

- 1、选定支持向量机的输入参数：目前采用的输入参数主要包括，最近 M 日最高价、最近 M 日最低价、当日成交额在最近 M 日成交额中占比、过去 M 日涨跌幅、最近 M 日均价、MACD、RSI，共 7 个指标，上述 7 个指标从价量以及上涨下跌动能等各方面反映市场的状况，以期预测提供尽量多的素材。
- 2、对输入参数进行标准化处理：在用支持向量机进行预测时需要注意一点，各个输入参数的数量级之间不能相差过大，若是相差过大则会导致最终的最优公式中某些优化参数的值过大或过小，导致输入参数跟实际情况有误差的时候，该误差被错误的放大或缩小，导致最终预测的结果不准。
- 3、支持向量机的构建：确定需要采用的核函数，以及进行预测时输入样本的个数。本文检验了不同长度的样本时间窗作为输入样本，并且检验了各个时间窗下得收益率情况。
- 4、得到支持向量机的预测结果：构建好支持向量机之后，计算得到了预测结果。
- 5、分析支持向量机的最终结果。

■ 3.1.1 上证指数预测结果

我们选取上证指数的时间范围为 2000-1-4 至 2011-12-31，共 2901 个交易日。由于市场指数呈现非线性的混沌现象，所以在进行市场预测的时候应该尽量用近期的一些数据进行预测，而距离较长时间的数据应该少用。

本问将对支持向量机预测下一日涨跌正确概率以及由此得到的交易策略效果进行检验。

图表11 支持向量机对原始数据的预测情况

时间窗长度	交易次数	成功次数	成功概率
5	2871	1496	0.5211
10	2866	1522	0.5311
15	2861	1569	0.5484
20	2856	1551	0.5431
25	2851	1537	0.5391
30	2846	1523	0.5351
35	2841	1508	0.5308
40	2836	1490	0.5254
45	2831	1503	0.5309
50	2826	1475	0.5219

资料来源：平安证券研究所。

从图表 11 可以看出用支持向量机对原始数据进行学习进而预测市场是有一定的效果，下面我们检验不同时间窗情况下支持向量机的预测效果是否能够稳定的超过 50%，我们采用的检验式为：

$$t = \frac{n \text{ 个样本概率的均值} - 0.5}{n \text{ 个样本概率的方差} / \sqrt{n}} \quad \text{—— (16)}$$

最后得到的 t 值为 7.7822，而 t(9) 的 95% 的分位数临界值为 1.9，因此我们可以认为采用支持向量机之后预测的正确概率是显著大于 50% 的。

上面是每一次预测的正确概率，我们要将它转换成一个策略的成功概率，这里采用这样的策略。若当天的指数头寸为空，而模型预测下一个交易日会上涨，这时在下一个交易日开盘的时候满仓买入指数，如果模型预测下一个交易日仍然是上涨，则继续持有该头寸；当到了某日，模型预测下一天的市场是下跌，这时清空持有的指数头寸。这样我们就认为完成了一次买卖交易。在样本期内，可能有无数次这样的交易，我们也统计了不同时间窗下得交易次数、成功概率、总收益率情况。

图表12 原始数据建模后的交易策略表现

时间窗长度	交易次数	成功次数	成功概率	总收益率	总收益率
				(无交易成本)	(单边成本 0.5%)
5	260	199	0.7654	768.8900	59.1300
10	290	220	0.7586	251.0900	13.4000
15	240	176	0.7333	117.3731	10.0963
20	303	217	0.7162	128.1315	5.3814
25	327	239	0.7309	111.1044	3.3813

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
30	367	263	0.7166	78.974265	1.0840
35	401	277	0.6908	79.2621	0.4675
40	438	296	0.6758	80.1433	0.0230
45	445	293	0.6584	60.7036	-0.2795
50	463	314	0.6782	107.8247	0.0743

资料来源：平安证券研究所、

从图表 12 可以看出，时间窗越短的情况下，虽然在预测每天的涨跌准确率上并不占优势，但是在构建交易策略上却表现出很大优势。在对比了短时间窗和长时间窗的交易序列之后，发现它们间有以下几个区别：

- 1、短时间窗的预测模型交易的频率更低，而长时间窗的预测模型交易的频率更高，这导致长时间窗的预测模型要支付更多的手续费。
- 2、短时间窗的预测模型买入指数后平均持仓时间长于长时间窗模型，而长时间窗模型则经常会出现持有期限为 1 天的交易，这说明短时间窗模型的预测结果比长时间窗更加稳定。
- 3、短时间窗的预测模型的交易成功概率要高于长时间窗模型。

这些现象说明在构建支持向量机的预测模型时，并不是学习的样本越多越好，样本越多反而因为过于“混沌”使得规律过于散漫。模型的预测是以日为单位，更多关心的是短期趋势，因此应用较新的数据去进行预测。

下面我们选取 2000-1-4 至 2011-12-31 共 600 周的数据进行分析，看看更长一点的周期用支持向量机能否得到较好的结果。首先来看支持向量机对未经滤波的周数据的预测效果：

图表13 支持向量机对周数据预测效果

时间窗长度	预测次数	成功次数	成功概率
3	571	343	0.6007
5	569	332	0.5835
8	566	343	0.6060
12	562	328	0.5836
15	559	333	0.5957
20	554	344	0.6209
22	552	327	0.5924
26	548	313	0.5712
30	544	320	0.5882

资料来源：平安证券研究所、

从图表 13 可以看出支持向量机对于周数据的预测准确概率大大提升，正确概率均值在 59%左右，这比上证指数日数据的正确概率大大提升，说明采用周数据中的数据含有的干扰性的噪音更少，有利于提高向量机的预测效果。下面我们检验一下原始周数据构建的支持向量机得到的交易策略的效果

图表14 周数据建模后的交易策略效果

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
3	50	20	0.4000	2.9878	1.4384
5	46	18	0.3913	1.6772	0.6954
8	41	19	0.4634	3.0082	1.6749
12	40	12	0.3000	3.0412	1.7255
15	41	14	0.3415	2.7502	1.5030
20	50	18	0.3600	3.4023	1.6875
22	57	21	0.3684	2.6855	1.0933
26	62	22	0.3548	2.3622	0.8145
30	65	22	0.3385	2.2466	0.6997

资料来源：平安证券研究所。

从图表 14 可以看出，虽然在预测单次效果较好，但是作为交易策略策略来讲，无论是交易的成功概率还是交易的收益率结果都很一般。出现这种现象我们给出的解释是虽然预测效果好，但是一旦预测错误，投资者将遭受更大的损失，导致前面预测正确所积累的收益率很大程度被侵蚀掉。

■ 3.1.2 沪深 300 指数预测结果

前面检验的都是未经滤波的上证指数数据构建的量化择时模型的检验，下面再检验支持向量机对沪深 300 的原始数据的预测效果及交易策略的收益率情况。

沪深 300 指数由于开始时间较晚，我们选其起始日 2005 年 1 月 4 日至 2011 年 12 月 31 日，共 1701 个交易日，或 353 周的数据。

图表15 沪深300日数据建模后的预测情况

时间窗长度	预测次数	成功次数	成功概率
5	1669	911	0.5458
10	1664	925	0.5559
15	1659	938	0.5654
20	1654	939	0.5677
25	1649	929	0.5634
30	1644	919	0.5590
35	1639	910	0.5552
40	1634	893	0.5465
45	1629	899	0.5519
50	1624	884	0.5443

资料来源：平安证券研究所、

从图表 15 可以看出支持向量机模型对日数据预测的正确概率较高，根据统计检验也可以得出支持向量机预测成功概率显著高于 50%。因此支持向量机对预测沪深 300 指数下一日的涨跌具有一定的作用。下面根据向量机的预测模型构建交易策略，并检验预测模型的收益率情况。

图表16 沪深300日原始数据建模后的交易策略收益率情况

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
5	136	54	0.3971	0.8787	-0.5191
10	146	53	0.3630	1.5166	-0.4166
15	121	47	0.3884	3.1492	0.2419
20	156	62	0.3974	2.5582	-0.2515
25	175	77	0.4400	2.1552	-0.4524
30	182	82	0.4505	1.9896	-0.5163
35	194	85	0.4381	1.6792	-0.6163
40	222	100	0.4505	1.5068	-0.7294
45	224	103	0.4598	1.1649	-0.7711
50	238	107	0.4496	0.7853	-0.8364

资料来源：平安证券研究所、

从图表 16 可以看出，支持向量机采用上证指数的交易策略取得的收益率不高，虽然无交易成本下的

收益率都为正，但是扣除了交易成本之后所得到的收益率都为负。之所以出现单日预测收益率显著而交易策略收益率低，我们认为应该是预测收益率的特征造成的，这一点我们在以后报告中将会分析。对此的改进方向我们认为可以从两个角度入手：一是调整支持向量机模型，提高单日的预测准确概率或者使得预测序列适合用上证指数所说的交易模型。另一个角度是保持原有的预测序列，而调整交易策略使得该交易策略能取得更高的收益率。

下面检验一下沪深 300 周数据的预测和交易策略效果：

图表17 沪深300周数据建模后的预测情况

时间窗长度	预测次数	成功次数	成功概率
3	323	196	0.6068
5	321	184	0.5732
8	318	190	0.5975
12	314	192	0.6115
15	311	197	0.6334
20	306	186	0.6078
22	304	184	0.6053
26	300	177	0.5900
30	296	175	0.5912

资料来源：平安证券研究所、

从图表 17 可以看出，支持向量机对原始的周数据预测效果较好，平均成功概率可以达到 60%左右，远远超过对日数据的预测效果，其中 12——22 周的预测效果最好，而其它长度的时间窗预测效果则一般。下面检验其交易策略的收益率情况：

图表18 沪深300周数据建模的交易策略收益率

时间窗长度	交易次数	成功次数	成功概率	总收益率	总收益率
3	31	12	0.3871	3.3985	2.2570
5	27	10	0.3704	3.5443	1.7222
8	24	10	0.4167	4.4833	3.3500
12	22	12	0.5455	5.6135	4.3627
15	18	11	0.6111	5.8730	4.8083
20	28	13	0.4643	4.1467	2.9264
22	28	11	0.3929	4.4396	3.1503
26	28	9	0.3214	3.6481	2.5409
30	24	10	0.4167	3.8828	2.8664

资料来源：平安证券研究所、

从图表 18 看出，周数据的交易成功概率整体要显著高于日数据的交易成功概率，又由于周数据的交易次数远远低于日数据的交易次数，因此周数据的总收益率以及扣除交易成本之后的收益率都要远远高于日数据的交易模型的收益率。由此可以看出用支持向量机建立沪深 300 指数的交易模型时，更适合用长周期的数据。

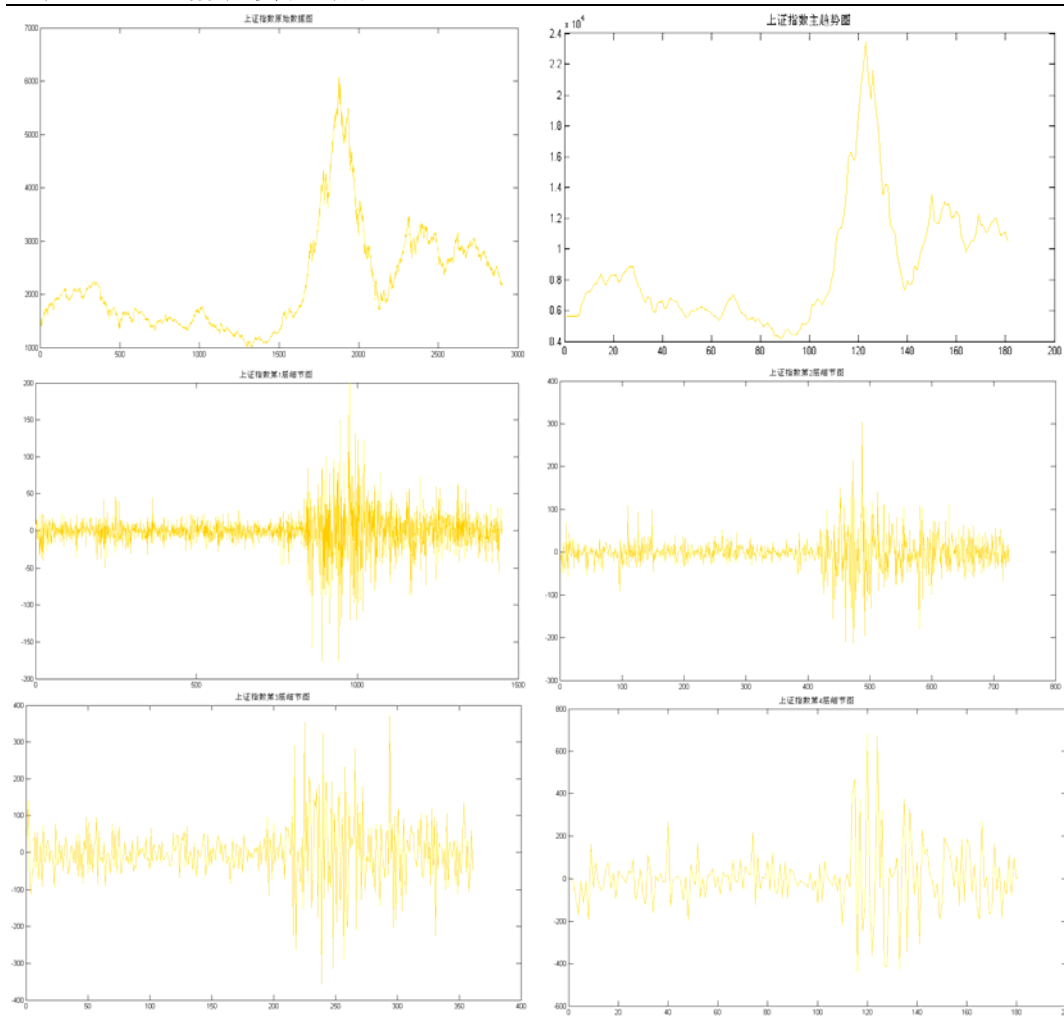
3.2 支持向量机对经过小波滤波后数据的预测效果

小波分析是一种非常优秀的时频分析工具，在数字信号处理等领域有广泛的应用，我们可以把股指也看成是一种特殊的信号，因此可以用小波分析对其进行分析。由于股票指数价格序列有许多微小的高频扰动，我们可以用小波分析对高频的信号进行滤波。首先将信号按小波尺度指标分层展开，再把小波高频信号的小波系数设为 0，最后再将分层小波信号重构，得到滤掉高频信号的股票指数。

■ 小波滤波之后的大盘数据情况

以上证指数为例，我们首先用常用的 Symlets 小波将上证指数分四层展开，得到四个细节图象，一个主趋势图象。

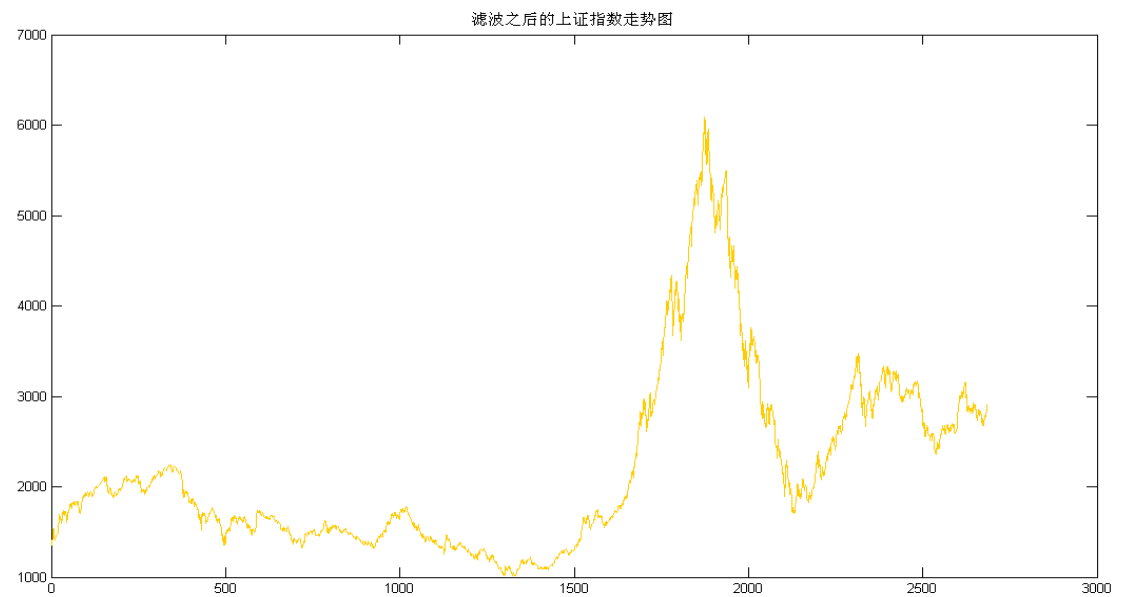
图表19 上证指数与展开数据



资料来源：平安证券研究所。

第一层中的细节部分含有许多更高频的噪音信息，我们采用 `rigsure` 方法选择好阈值，采用软约束方法处理这些频率高于阈值的高频信号，最终再根据重构主趋势以及处理过后的第一层细节，和第二、三、四层细节重构得到滤波之后的上证指数图象。

图表20 滤波之后的上证指数图象



资料来源：平安证券研究所。

■ 小波滤波之后的上证指数表现

上证指数我们选取时间为 2000-1-4 至 2011-12-31，共 2901 个交易日或 600 周。我们按照 3.2.1 节中介绍的方法对上证数据进行小波分析，将上证指数进行滤波。再将滤波后的数据进行处理，得到滤波之后的分类数据训练集。

首先来看看上证指数日数据的建模效果：

图表21 支持向量机对滤波数据预测效果

时间窗长度	预测次数	成功次数	成功概率
5	2871	1608	0.5601
10	2866	1610	0.5618
15	2861	1639	0.5729
20	2856	1652	0.5784
25	2851	1643	0.5763
30	2846	1668	0.5861
35	2841	1659	0.5839
40	2836	1650	0.5818
45	2831	1647	0.5818
50	2826	1649	0.5835

资料来源：平安证券研究所。

从图表 21 可以看出数据经过滤波之后，其预测效果相对于未滤波后的数据有明显的改善。经过滤波

之后，不同时间窗的模型在预测单日涨跌的成功概率都有显著提高，平均的成功概率在 58%左右。这说明了经过小波滤波处理之后，支持向量机对上证数据的单日预测效果有显著的改善。

下面再检验一下用滤波之后学习得到的支持向量机来构建交易策略，该交易策略的收益率情况为：

图表22 滤波构建支持向量机的交易策略情况

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
5	215	182	0.8465	694.0269	84.2849
10	223	157	0.7040	122.7893	11.7660
15	229	156	0.6812	70.7235	6.3651
20	268	183	0.6828	110.9364	6.9113
25	296	206	0.6959	84.0165	3.5195
30	318	215	0.6761	83.1734	2.5514
35	339	221	0.6519	72.7078	1.5145
40	363	243	0.6694	83.6418	1.2739
45	364	240	0.6593	54.8013	0.4500
50	389	249	0.6401	62.7166	0.2924

资料来源：平安证券研究所、

从图表 22 可以看出，随着训练时间窗的增大，交易策略的交易次数随之增大，交易成本也跟着攀升。另外，时间窗的增大也会导致成功概率降低，无交易和扣除交易成本之后的收益率也随之降低。因此建议在建模预测市场的时候使用小波分析对市场进行滤波以去掉市场中的一些噪音。

我们也能从图表 22 看出，经过滤波之后的交易策略交易成功概率也大幅上升，达到 84%的成功概率，这是一个非常可观的成功概率，同时扣除交易成本之后，12 年的收益率也有高达 84 倍，这是一个非常理想的交易系统。

下面再对周交易数据进行检验，看看滤波之后支持向量机对其的预测情况以及据此建立的交易策略的收益率情况。首先检验支持向量机对每周涨跌的预测情况。

图表23 支持向量机对滤波周数据预测效果

时间窗长度	预测次数	成功次数	成功概率
3	571	350	0.6130
5	569	339	0.5958
8	566	343	0.6060
12	562	329	0.5854
15	559	333	0.5957
20	554	345	0.6227
22	552	332	0.6014
26	548	314	0.5730
30	544	320	0.5882

资料来源：平安证券研究所、

从图表 23 可以看出，支持向量机在预测滤波周数据方面，准确率比日数据正确率有显著的提升，但是相对于未滤波周数据构建的支持向量机，效果并没有显著的提升。接着检验滤波周数据模型对应交易策略的收益率情况：

图表24 支持向量机对滤波上证周数据预测效果

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
3	20	50	0.4000	2.9878	1.4384
5	15	38	0.3947	1.6353	0.8076
8	18	38	0.4737	3.5234	2.1202
12	21	45	0.4667	2.8433	1.4649
15	20	45	0.4444	3.7874	2.0784
20	25	46	0.5435	4.9960	2.8224
22	23	47	0.4894	3.8685	2.0661
26	25	57	0.4386	2.3786	0.9171
30	23	52	0.4423	2.2009	0.9092

资料来源：平安证券研究所、

从图表 24 可以看出，支持向量机对滤波后的周数据进行建模后得到的交易策略相对于未进行滤波的

周数据的模型有了较大的改进。不过图表 24 也反应了其它一些信息，对于滤波得到的交易策略，效果最好的时间窗在 8—22 周的这些模型，处于这个区间两侧的交易策略效果一般。

■ 滤波后沪深 300 指数预测结果

沪深 300 指数的滤波方法与上证指数的方法相同，在此就不再赘述。沪深 300 指数从 2005 年 1 月 4 日至 2011 年 12 月 31 日，共 1701 个交易日，或 353 周的数据，根据滤波后的数据建立支持向量机的训练集。

首先对沪深 300 的日数据进行检验：

图表 25 支持向量机对滤波沪深 300 日数据预测效果

时间窗长度	预测次数	成功次数	成功概率
5	1669	945	0.5662
10	1664	965	0.5799
15	1659	991	0.5973
20	1654	986	0.5961
25	1649	982	0.5955
30	1644	979	0.5955
35	1639	977	0.5961
40	1634	979	0.5991
45	1629	966	0.5930
50	1624	964	0.5936

资料来源：平安证券研究所、

从图表 25 可以看出经过滤波之后，支持向量机对沪深 300 指数的预测效果显著地提升，平均正确概率在 59%左右。而未滤波的时候，平均预测正确概率为 55%左右，提升了 4 个百分点。

下面检验在滤波后的交易策略的收益率情况。

图表 26 滤波沪深 300 建模后交易策略收益情况

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
5	108	52	0.4815	3.4865	0.5345
10	115	46	0.4000	5.0422	0.9297
15	110	56	0.5091	11.8764	3.0056

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
20	124	57	0.4597	9.1520	1.9723
25	149	80	0.5369	9.7829	1.4573
30	153	81	0.5294	12.5928	1.9838
35	175	91	0.5200	11.5563	1.2073
40	190	102	0.5368	9.1567	0.5346
45	189	99	0.5238	9.5393	0.6085
50	195	102	0.5231	10.3098	0.6263

资料来源：平安证券研究所、

经过滤波之后，沪深 300 模型的交易成功概率较未滤波时明显提高，无交易成本和扣除交易成本之后的收益率都远远高于未滤波时的情况。无交易成本时候的收益率最高有 12.6 倍，但是扣除成本之后最高收益只有 3 倍，这主要是由于交易频繁以及我们设置的交易成本较高所导致的。如果将交易对象变为股指期货之后，相信收益率会有显著的提高。

下面再检验一下滤波后的沪深 300 周数据构建的交易模型收益率情况：

图表27 支持向量机对滤波沪深300周数据预测效果

时间窗长度	预测次数	成功次数	成功概率
3	323	201	0.6223
5	321	190	0.5919
8	318	194	0.6101
12	314	190	0.6051
15	311	192	0.6174
20	306	194	0.6340
22	304	188	0.6184
26	300	183	0.6100
30	296	184	0.6216

资料来源：平安证券研究所、

由图表 27 可以看出，支持向量机在预测滤波之后的周数据上也较日数据以及未滤波的周数据有明显的提高，平均的成功概率为 61%。这说明无论的上证指数和沪深 300 指数的日数据还是周数据，滤波之后的模型效果都要比原始数据的预测效果好。

图表28 滤波沪深300建模后交易策略收益情况

时间窗长度	交易次数	成功次数	成功概率	总收益率 (无交易成本)	总收益率 (单边成本 0.5%)
3	29	13	0.4483	3.8020	2.6316
5	22	9	0.4091	3.2090	2.4040
8	22	9	0.4091	3.7191	2.8151
12	22	12	0.5455	5.9471	3.6359
15	22	13	0.5909	5.8219	4.5364
20	26	13	0.5000	5.2195	3.8432
22	27	12	0.4444	4.8197	3.4842
26	25	10	0.4000	4.1922	3.0736
30	24	10	0.4167	3.8828	2.8664

资料来源：平安证券研究所。

由图表 28 可知，滤波之后的沪深 300 周数据建立的模型收益率较日数据的模型有大幅提高。虽然周数据模型的交易成功概率并没有显著提升多少，但是由于周数据模型交易频率大幅下降，这使得该模型的收益率大部分保存下来。以每次交易成本 1%为计，多交易 100 次会导致收益率只相当于原来的 37%，多交易 200 次会导致收益率只相当于原来的 13.7%，这就是频繁交易对收益率的损伤效果。

四、 总结

本文将小波分析和支持向量机结合起来得到了一种量化择时模型。小波分析在信号分析、异常检测、信号压缩、信号消噪中都有广泛的应用。本文利用小波分析在信号消噪方面的功能，对含有高频扰动信号的上证指数及沪深 300 指数的原始序列进行滤波，得到了较为平滑的股指序列。平滑之后的股指序列能够去除高频扰动信号的干扰，使得股票指数的趋势能够更好的呈现出来，也使得支持向量机的预测效果更加的显著。支持向量机是模式识别中的一种新方法，主要用于分类研究，根据是否引入核函数，我们可以解决线性和非线性的分类问题。支持向量机分为支持向量分类机和支持向量回归机，在本文中主要采用支持向量分类机。未来的股票市场无非就是“涨”和“跌”两种情况，我们可以根据目前的 market 情况，将它们分成对应未来上涨的情况以及对应未来下跌的情况。这样根据对未来市场状况的分类，我们就能够预测未来市场的涨跌。

两种方法相结合产生了一种量化择时模型，也派生出一种交易策略。上述交易策略对上证指数显示出了强大的适应能力，以 5 日为学习时间窗的预测模型得到的交易策略，最终创造出 12 年扣除交易成本达 84.28 倍的投资收益，年化投资收益率为 44.7%，交易成功概率高达 84.65%。上述预测模型对沪深 300 指数也表现出较好的预测效果，但是在构建沪深 300 的交易模型时却遇到了挫折，无论是原始数据还是滤波数据，无论是日数据还是周数据，它们构建的交易策略模型都不能产生类似上证指数的交易效果。我们未来可以从改进支持向量机模型的参数以及改变交易策略两方面入手进行改进。

总体来讲上述量化择时模型换仓的频率较高，和大资金做股票的要求还有一定的差距，未来我们会考虑降低调仓的频率，使得收益率不被高换手率所带来的高交易成本侵蚀，也使模型更符合股票型基金的需求。另一方面，该模型可以将交易数据的频率提高到秒、分级别，得到高频交易的模型，目前期货交易的交易成本较低，并且具有 T+0 机制，是高交易频率低交易成本的品种。如果能将该量化择时模型用于上述品种的交易，相信会有令人印象深刻的效果。

未来平安金工小组将从上述两个方面研究量化择时模型，希望能为股票投资者和期货投资者同时奉上精美和趁手的模型！

平安证券综合研究所投资评级：

股票投资评级：

强烈推荐（预计 6 个月内，股价表现强于沪深 300 指数 20%以上）

推 荐（预计 6 个月内，股价表现强于沪深 300 指数 10%至 20%之间）

中 性（预计 6 个月内，股价表现相对沪深 300 指数在 $\pm 10\%$ 之间）

回 避（预计 6 个月内，股价表现弱于沪深 300 指数 10%以上）

行业投资评级：

强烈推荐（预计 6 个月内，行业指数表现强于沪深 300 指数 10%以上）

推 荐（预计 6 个月内，行业指数表现强于沪深 300 指数 5%至 10%之间）

中 性（预计 6 个月内，行业指数表现相对沪深 300 指数在 $\pm 5\%$ 之间）

回 避（预计 6 个月内，行业指数表现弱于沪深 300 指数 5%以上）

公司声明及风险提示：

负责撰写此报告的分析师（一人或多人）就本研究报告确认：本人具有中国证券业协会授予的证券投资咨询执业资格。

本公司研究报告是针对与公司签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本公司研究报告仅提供给上述特定客户，并不面向公众发布。未经书面授权刊载或者转发的，本公司将采取维权措施追究其侵权责任。

证券市场是一个风险无时不在的市场。您在进行证券交易时存在赢利的可能，也存在亏损的风险。请您务必对此有清醒的认识，认真考虑是否进行证券交易。

市场有风险，投资需谨慎。

免责条款：

此报告旨在发给平安证券有限责任公司（以下简称“平安证券”）的特定客户及其他专业人士。未经平安证券事先书面明文批准，不得更改或以任何方式传送、复印或派发此报告的材料、内容及其复印本予任何其它人。

此报告所载资料的来源及观点的出处皆被平安证券认为可靠，但平安证券不能担保其准确性或完整性，报告中的信息或所表达观点不构成所述证券买卖的出价或询价，报告内容仅供参考。平安证券不对因使用此报告的材料而引致的损失而负上任何责任，除非法律法规有明确规定。客户并不能尽依靠此报告而取代行使独立判断。

平安证券可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。报告所载资料、意见及推测仅反映分析员于发出此报告日期当日的判断，可随时更改。此报告所指的证券价格、价值及收入可跌可升。为免生疑问，此报告所载观点并不代表平安证券有限责任公司的立场。

平安证券在法律许可的情况下可能参与此报告所提及的发行商的投资银行业务或投资其发行的证券。

平安证券有限责任公司 2010 版权所有。保留一切权利。

中国平安 PINGAN

平安证券综合研究所

地址：深圳市福田区金田路大中华国际交易广场 8 层

邮编：518048

电话：4008866338

传真：(0755) 8244 9257