

深度报告

金融工程

量化投资

2010 中期策略会专题报告

2010 年 05 月 25 日

本报告的独到之处

■利用小波变换的消噪原理去除时间序列中的细微波动，只考虑大体趋势，从而对时间序列进行平滑处理；

■首次使用支持向量机对指数进行价格预测。

专题报告

基于小波分析和支持向量机的指数预测模型

支持向量机 (support vector machine, SVM) 是数据挖掘中的一项新技术，是借助于最优化方法解决机器学习问题的新工具。它成为克服“维数灾难”和“过学习”等传统困难的有效办法，虽然他还处在飞速发展的阶段，但它的理论基础和实现途径的基本框架已经形成。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。而当我们把走势看作两种状态（涨、跌），问题便转化为分类问题，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。

本报告将从实际应用角度出发，用沪深 300 指数为代表的金融时间序列数据，研究小波理论在金融时间序列中的可操作性。主要应用小波变化的消噪功能，对时间序列数据进行优化，进一步拓展支持向量机在时间序列预测上的研究和应用。

通过对沪深 300 指数价格的预测，构造多头和多空两种择时策略。

实证效果：

实证例子的时间为 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 指数回报率会 1.03%。观察 1 年内策略的最高和最低回报发现，多头策略对下方风险控制较好，收益率不会出现大幅波动。

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，这说明了模型的预测具有一定参考价值，报告中所述策略需要长期坚持。

分析师：黄志文

电话：0755-82130833-6210

E-mail: huangzw@guosen.com.cn

SAC 执业证书编号：S0980206110185

分析师：葛新元

电话：0755-82130833-1870

E-mail: gexy@guosen.com.cn

SAC 执业证书编号：S0980200010107

独立性声明：

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

内容目录

小波分析理论简介.....	4
金融时间序列概况	4
小波理论的发展概况	4
小波理论的主要应用	5
小波降噪的基本原理	5
小波分析在金融时间序列中的应用	6
小波降噪的步骤	6
沪深 300 时间序列降噪的实证分析	7
支持向量机在股票市场的应用	10
结合小波降噪的支持向量机预测模型	10
研究对象	11
原始数据的标准化	11
核函数的选择	11
输入向量的选取方法	11
参数的选择	12
股票价格趋势预测方法	12
SVM 模型的实际运用	13
实证例子分析.....	14
预测沪深 300 指数走势	14
基于预测走势的交易策略	14
模型的改进方案和后续拓展	15
SVM 预测模型存在的问题及解决方案	15
后续拓展方向	16

图表目录

图 1: 小波分解图释 (3 层分解)	7
图 2: 含噪音的沪深 300 原始数据	8
图 3: 小波分解, 低频 cA4.....	8
图 4: 小波分解, 高频 cD4.....	9
图 5: 小波分解, 高频 cD3.....	9
图 6: 小波分解, 高频 cD2.....	9
图 7: 小波分解, 高频 cD1.....	9
图 8: 消噪后的沪深 300 数据	9
图 9: 沪深 300 指数的预测走势和真实走势 (近一年)	14
图 10: 多头策略, 考虑交易成本	14
图 11: 多头策略, 未考虑交易成本.....	14
图 12: 多空策略, 考虑交易成本	15
图 13: 多空策略, 未考虑交易成本.....	15
表 1: 各组配对收益率表现	15

小波分析理论简介

小波分析理论是目前科学界和工程界讨论和研究最多的课题之一，它包含了丰富的数学内容，又具有巨大的应用潜力。小波分析是在 Fourier 分析的基础上发展起来的，是调和分析近半个世纪以来的结晶。其基本思想是将一般函数（信号）表示为规范正交小波基的线性叠加，核心内容是小波变换。由于小波变换在时域和频域具有良好的局部化性质，能自动调整时、频窗口，以适应实际分析需要，因而已成为许多工程学科应用的有力工具。在进一步介绍小波分析理论之前，我们先了解一下金融时间序列的概念。

金融时间序列概况

金融市场是国家经济生活的核心，寻找其中的变化规律，进行有效合理的管理是各国政府及相关研究机构孜孜以求的目标之一。同时，人们对金融预测也作了大量的探索，取得了丰硕的成果。典型的金融预测是时间序列预测。金融市场的数数据绝大多数都是时间序列数据，即指这些数据是按照时间的排序取得的一系列观测值，如股票或期货价格、货币利率、外汇利率等。这些数据具有非常复杂的变化规律，而利用一定的数学方法对其进行分析和研究将有助于制定更为精确的定价和预测决策，当然对于金融投资与风险管理活动具有十分重要的意义。金融时间序列分析主要是以统计理论和方法为基础，通过模型假设、参数估计、回归分析等技术来描述其内在的规律。适当的数学工具和真实的数据使金融时间序列分析成为金融经济研究中独具魅力的一块领域，例如美国经济学家 Engle 和英国经济学家 Granger 就因其提出的 ARCH 模型和协整理论而荣获 2003 年度诺贝尔经济学奖。

一般来说，时间序列的分析可以通过时域和频域两个途径进行。但是很多金融时间序列表现出较强的非平稳性和长记忆性，这使得许多传统的单独集中于时域或频域的研究分析方法已经不再适用。而小波分析作为一种新型的信号分析方法，因其在时域和频域都具有表征信号局部特征的能力，被誉为“数学显微镜”，因而它非常适用于分析非平稳信号。小波分析是近 20 年发展起来的新兴数学分支，也是目前数学界和工程界讨论最多的话题之一。并且已经在信号和图像处理、模式识别、语音识别、地震勘测等众多学科中得到了广泛应用。相对来说，其在金融时间序列分析和建模中的应用却相对较少。但近年来，小波分析方法在金融时间序列分析中的重要地位已经越来越受到人们的关注。

小波理论的发展概况

小波这一名称首先是由法国地质学家 J.Morlet 与 A.Grossmann 在分析地质数据时引进的，Y.Meyer, S.Mallat 及 I.Daubechies 等人对小波理论的发展都做了非常重要的贡献，本报告也参考了他们发表的相关文献。至上世纪 90 年代初期经典的小波理论已经基本成熟，目前国际上的重点已转向小波的推广和应用。

1882 年，法国数学家 Fourier 从热力学的角度提出一种新的理论即“热的解析理论”，即被后人广泛应用和称誉的 Fourier 分析方法。小波分析是在 Fourier 分析方法的基础上发展起来的，由于小波变换在时域和频域同时具有良好的局部化性质，能自动调整时、频窗口以适应实际分析的需要，从而可以聚焦到分析对象的任意细节，因而具有简单、随意、灵活的特点。

小波变换虽然是在 Fourier 变化的基础上发展起来，但是相比之下，它是一种实践尺度的局域变换，能够同时在时域和频域进行局域化分析的方法。小波分析的

思想来源于伸缩和平移。1910 年，Alfred Haar 给出了 Haar 小波的构造，但由于不光滑，理论上没有引起重视和发展。1936 年，Littlewood-Paley 建立 L-P 理论，即提出对频率进行 2^i 分划并证明其本质上不影响函数的形状和大小。1981 年，法国地址物理学家 Morlet 首先提出平移伸缩的小波公式，用于地质勘测。1986 年，Meyer 证明了一维小波基的存在，构造了第一个真正的小波基，国际上从此开始形成研究小波的热潮。1988 年，Mallat 和 Meyer 合作提出了多分辨分析的框架。同年，年轻的女数学家 I. Daubechies 在其发表在美国 Comm. Pure & Appl. Math. 的一篇文章中构造了具有紧支集的有限光滑小波函数，被视为小波分析的经典纲领文献。后来，信号分析专家 Mallat 提出了多分辨分析的理论，给出了构造正交小波基一般的方法，并以此为基础提出了著名的快速小波算法—Mallat 算法，这是小波理论突破性的成就。从此，小波分析就从理论研究走向广泛的应用研究。1990 年，我国的崔锦泰和王建构造了基于样条函数的单正交小波。1991 年，Alpert 和 Rokhlin 通过构造 $r(r \geq 2)$ 个尺度函数，形成了多小波理论的思想。1994 年 Goodman 等人基于 r 重多分辨分析，建立了多小波的基本理论框架。至此经典小波分析理论已经基本成熟，近年来高维小波理论已逐步被人们所关注。

小波理论的主要应用

由于众多金融时间序列表现为较强的非平稳和长记忆性，小波分析方法在时域和频域上都具有表现局部特征的能力，可以广泛地应用于金融时间序列的建模和相关波动率分析等。

目前，利用小波变换在时间序列方面的分析主要集中在以下几个方面：（1）利用小波变换研究时间序列的长记忆过程，并可估计长记忆参数；（2）根据小波分析的分解与重构原理对事件序列季节、趋势、周期等因素的分解、然后根据不同尺度进行分析预测、从而使非平稳时间序列的处理变得简单化；（3）利用小波变换的消噪原理去除时间序列中的细微波动，只考虑大体趋势，从而对时间序列进行平滑处理；（4）小波分析还可以检测信号的异常点，对时间序列的奇异点进行较为准确的检测和定位。以上是小波分析在时间序列分析中的主要研究范围。但主要集中于理论研究，实际应用研究较少。

本报告将从应用角度出发，用沪深 300 指数为代表的金融时间序列数据，研究小波理论在金融时间序列中的可操作性。结合之前报告《支持向量机在股票价格预测方面的应用》，主要应用小波变化的消噪功能，对时间序列数据进行优化，进一步拓展支持向量机在时间序列预测上的研究和应用。

小波消噪的基本原理

一般来说，一个含有噪声的一维信号可以用如下公式简单地表示：

$$s(t) = f(t) + \sigma \cdot \varepsilon(t) \quad t=0,1,2,\dots,n-1$$

其中， $f(t)$ 表示为真实信号， $\varepsilon(t)$ 为噪声， $s(t)$ 表示为含噪音的信号，通常简单假设 $\{\varepsilon(t)\}$ 是白噪音序列，即相互独立同分布的且具有零均值（ $E\varepsilon_t = 0$ ）和有限方差（ $E\varepsilon_t^2 = \sigma^2$ ）的随机序列。在应用中，有用信号通常表现为低频信号或较平稳的信号，噪声信号则表现为高频信号，所以消噪过程可以按一下方法进行处理。

首先，对实际信号进行小波分解，选择小波并确定分解层次为 N ，则噪声部分通常包含在高频中。然后对小波分解的高频系数进行门限阈值量化处理。最后根据小波分解的第 N 层低频系数和经过量化后的 1 至 N 层高频系数进行小波重构，达到消除噪声的目的，即抑制信号的噪声，在实际信号中恢复真实信号。

目前，小波消噪的方法大概可以分为三大类：

(1) 基于小波变换墨迹大致原理提出的交替投影 (AP) 算法, 即根据信号和噪声在小波变换尺度上的不同传播特征, 剔除由噪声产生的模极大值点, 保留信号所对应的模极大值点, 然后重构小波系数, 进行恢复信号。基本的操作过程概括如下: 首先对含噪音信号进行二进小波变换, 一般分为 4 至 5 个尺度, 求出每个尺度上的小波变换系数的模极大值, 然后从最大尺度开始, 选取阈值, 若极值点对应的幅值的绝对值小于阈值, 则去掉该点, 否则予以保留, 最后将每个尺度上保留下来的极值点利用适当的方法重构小波系数, 进行信号恢复;

(2) 根据小波变换后相邻尺度间小波系数的相关性进行取舍重构信号。基本原理是信号经过小波变换后, 其小波系数在各尺度上有较强的相关性, 尤其在信号的边缘附近, 而噪声所对应的小波系数在尺度间却没有明显的相关性。因此, 可以利用小波系数在不同尺度上对应点的相关性来区分有效信号和噪声, 达到去噪效果;

(3) 阈值消噪: 该方法首先对含噪音的信号直接进行小波变换, 得到一组小波系数, 然后选择适当的阈值, 对小波系数进行阈值处理, 一般认为, 信号对应的小波系数包含信号的重要信息, 幅度较大, 予以保留, 噪声对应的小波系数幅度较小, 可以剔除。常用阈值消噪方法有软阈值和硬阈值方法。

总结上述原理可得, 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。

小波分析在金融时间序列中的应用

金融市场中数据由于各种偶然因素的影响, 即使不存在暗箱操作, 或没有什么重要新闻、重要政策出台, 也会表现一种小幅的随机波动。这些随机波动可以看成是信号的噪声, 不具有分析和预测的价值, 而且这些随机波动往往严重地影响了进一步的分析和处理。因而在做金融事件序列的建模分析之前, 往往对数据进行预处理, 消除这些噪音。

小波消噪的步骤

为了更有效的预测金融市场的未来趋势, 我们可以将大幅波动作为有用信号保留, 而将小的波动作为噪音消除, 先将数据进行预处理之后, 再做时间序列的建模分析。假设原始的时间序列 S_0 , 建模的基本步骤如下:

1、小波分解; 选择合适的小波函数和小波分解的层次, 计算时间序列 S_0 到第 N 层的小波分解。即首先对含噪音信号 $s(k)$ 进行小波变换, 得到一组小波系数 $Wf(j,k)$ 。根据多分辨率分析理论, 分解的层次越高, 去掉的低频成份就越多, 而低频成份主要代表有用信号。因而分解的层次越高, 去噪效果越好, 但是相应的失真程度也越大。在我们下面的沪深 300 期指实证中, 使用的是 Matlab 的小波分析工具包, 所采用的小波函数为 "db4", 分解的层数为 4 层。

2、阈值处理; 将分解得到的小波系数进行阈值处理来区分信号和噪声。阈值的确定对消噪性能有很大影响, 阈值过高会使信号失真, 阈值过低又会使得消噪不

完全。一般来说，选定阈值可以有以下几种常用准则：（1）无偏风险估计准则(rigrsure)，即一种基于 Stein 的无偏似然估计原理的自适应阈值选择方法，对每个阈值求出对应的风险值，风险最小的即为所选；（2）固定阈值准则(sqtwolog)，设 n 为小波系数向量长度，则设定阈值为 $T_r = \sqrt{2\log n}$ ；（3）混合阈值准则：用于最优预测变量阈值的选择，它是 rigrsure 准则和 sqtwolog 准则的混合。（4）最小最大阈值准则(minimax)，是根据统计学中的绩效极大估计量而设定的一种固定阈值选择方法。以上四种阈值准则中，rigrsure 准则和 sqtwolog 准则是相对比较保守的准则，它仅是部分系数置零，可以保留较多的高频信号。因而根据金融数据高频性特征，我们一般可以选择 rigrsure 准则或 sqtwolog 准则来确定阈值。在我们实证中，阈值取的是 140。

3、小波消噪及重构；一般来说，除了简单的强制消噪方法（该方法直接将小波分解结构中的高频系数置零），阈值消噪方法可分为默认阈值消噪处理和软（硬）阈值消噪处理两种，后者在实际应用中比前者更具有操作性。通过阈值选择的高频和低频系数及小波逆变换公式，计算出信号的小波重构，达到消噪的目的。

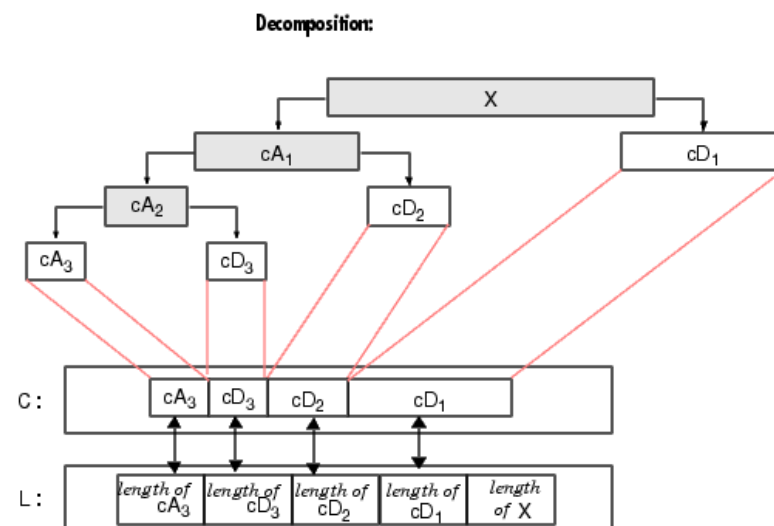
常用的小波变换重构公式为：

$$S_f(j-1, k) = S_f(j, k) * h(j, k) + W_f(j, k) * g(j, k)$$

其中 $S_f(j, k)$ 为尺度系数， $W_f(j, k)$ 为小波系数， h 和 g 分别为对应于尺度函数和小波函数的低通和高通滤波器。

下图截自 Matlab 的 help 文档，诠释的是小波分解流程，cA 为分解出的低频数据、cD 为高频数据。

图 1：小波分解图释（3 层分解）



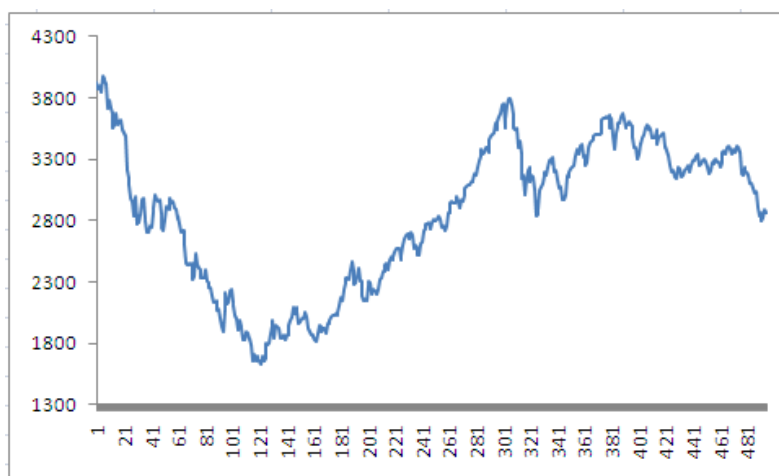
资料来源： Matlab 7.8.0

沪深 300 时间序列消噪的实证分析

首先我们对 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。原始的时间序列见下图。根据上述的理论，对此金融序列进行小波分解，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。由于理论上通常认为噪声部分包含在高频中，因此我们对小波分解的高频系数进行阈值量化处理，其中阈值处理选择 sqtwolog 阈值估计准则。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。

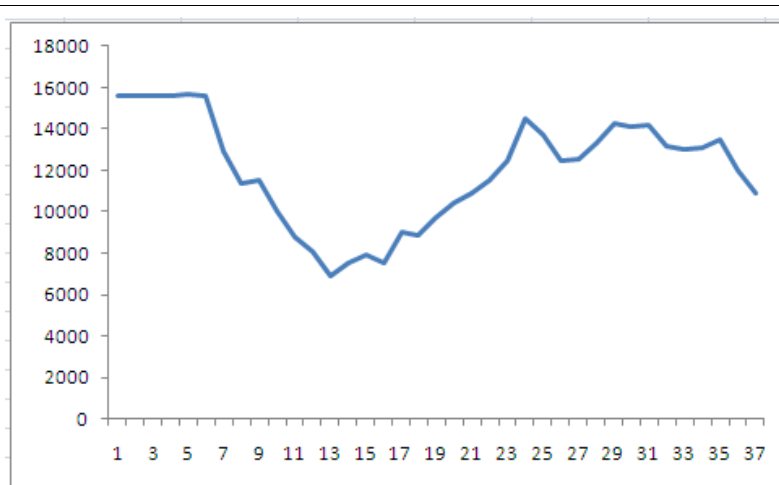
图 2：含噪音的沪深 300 原始数据



资料来源：国信证券经济研究所

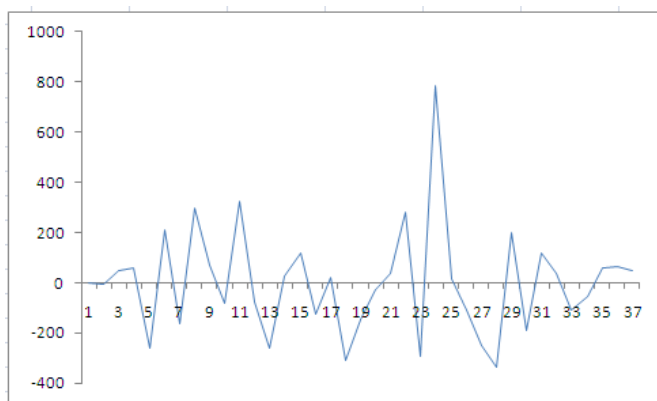
以下依次小波分解后的第 4 层低频和高频第 4、第 3、第 2 和第 1 层信号：

图 3：小波分解，低频 cA4



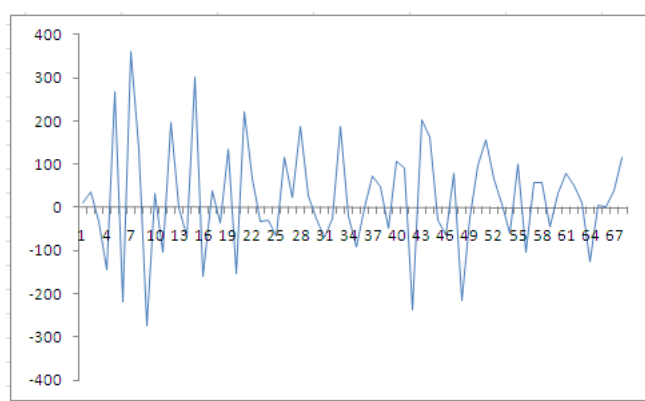
资料来源：国信证券经济研究所

图 4：小波分解，高频 cD4



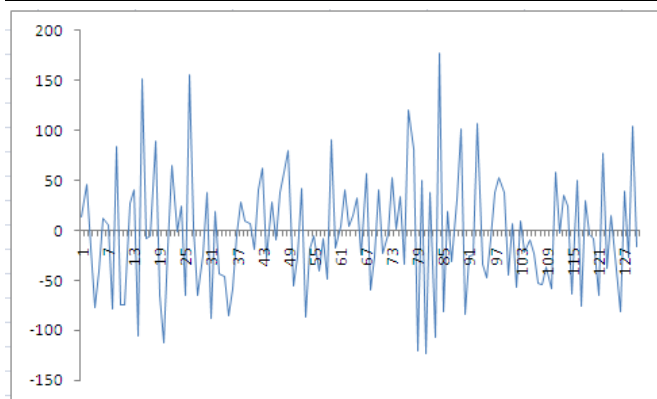
数据来源：国信证券经济研究所，

图 5：小波分解，高频 cD3



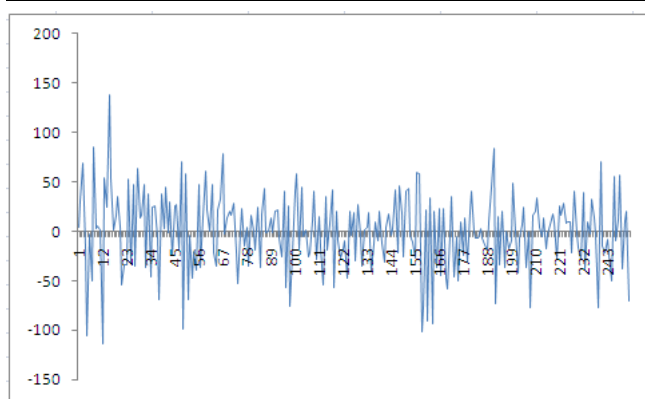
数据来源：国信证券经济研究所，

图 6：小波分解，高频 cD2



数据来源：国信证券经济研究所，

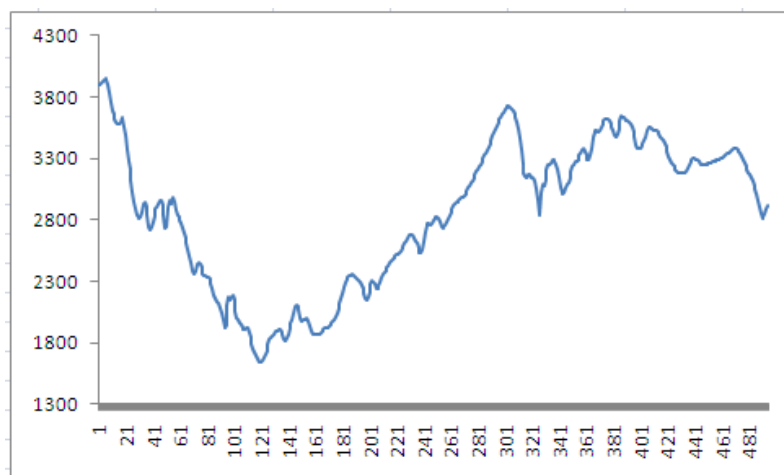
图 7：小波分解，高频 cD1



数据来源：国信证券经济研究所，

提取第 4 层的低频信号及根据固定阈值处理的高频信号，进行小波重构，得到新的消噪沪深 300 走势如下：

图 8：消噪后的沪深 300 数据



资料来源：国信证券经济研究所

将图 8 与图 2 比较，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。

支持向量机在股票市场的应用

先回顾一下在上篇关于支持向量机的综述报告《支持向量机在股票价格预测方面的应用》中的核心内容。**支持向量机的最大特点**是改变了传统的经验风险最小化原则，而是针对结构风险最小化原则提出的，因此具有很好的泛化能力。另外，支持向量机在处理非线性问题时，首先将非线性问题转化为高维空间中的线性问题，然后用一个核函数来代替高维空间中的内积运算，从而巧妙地解决了复杂计算问题，并且有效地克服了维数灾难以及局部极小问题。

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。而我们将走势看作两种状态（涨、跌），问题便转化为分类问题，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。

自 2000 年以来，国内外对支持向量机的研究不断地增加，随着它为越来越多的人所熟知，许多相关领域的工作者都试图采用该方法解决一些实际问题，计算机领域首当其冲。在证券领域，也有很多人对于支持向量机在该领域的应用做出了很多贡献。

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。

从股票指标的选择来看，大多采用三种指标选择法，一种是采用历史上在股票市场上经常被使用的技术指标，另一种是直接利用股票的基本指标或其简单组合，第三种就是将前两者混合起来使用。没有什么科学的理由说明哪种指标的选择是最好的，因为即便是股票的技术指标也是从基本指标推导得出的，所以在 SVM 的应用中没有能够说明前者比后者更能提高预测精度的理论根据。

结合小波消噪的支持向量机预测模型

结构模型和时间序列模型是股市预测模型的两大分类。采用结构模式主要有基本面数据方法和金融资产法，前者是以影响股市的宏观经济指标或企业财务指标为影响参数。我们知道，股票的价格关键是由公司经营状况和财务状况决定，并受到宏观经济状况，制度政策等因素影响。这些因素非常适合对股票中长期预测。同时，由于我国股市发展时间较短，数据量不是很丰富，加之宏观数据，和财务数据发布的时间间隔大，最短的也需要每月公布一次，而可用月度数据非常有限，直接通过

基本面数据进行预测数据不够充分，意义不是很大。后者，根据股票和各种相关的金融资产的协同运动规律来进行预测。例如，应用这种模型，通常要有完善的金融市场，需要有良好的金融资产价格形成机制，使各种资产的风险价格化。在我国，资本市场和货币市场还很不成熟，还不能应用这种模型。还有一种是通过技术指标进行预测，由于大多数技术指标是通过日 K 线数据计算形成的，而且可用的日 K 线数据的数据量很充分，足以进行预测和检验。

如果用支持向量机的方法来分析股票或指数，指标的选取将会更加灵活方便。本文将会提出几种用支持向量机分析股票的方法，在参数的选择方面将遵循以下原则：

- 参数的选择要包含尽量多的特征，但所选特征必须充分的代表问题。
- 选择的参数要尽可能多的用到已知的信息量。
- 对于预测而言，选择的参数要尽量用到最新的数据。

研究对象

股指期货推出后，市场对其关注程度远超预期。除股指期货套利业务受到追捧外，具有较大风险偏好的投资者也希望通过期指的杠杆效应在投机业务上获利。此时对股指走势的判断显得比以往更为重要。本文的研究对象是沪深 300 指数，一般来讲，相对于个股来说，指数的波动“随机性”较弱，预测的准确率也会相对个股要高。由于参数和输入向量的选取范围很大，本人只是提出一套可行的方法，在实际的操作中，至参数和输入向量的选取等都将根据实际情况灵活进行。

对于沪深 300 指数的研究如果能够取得一定的成果，将会对股指期货投资起到非常大的指导性作用。

原始数据的标准化

一个 m 维的样本数据里，各个分量来自不同的领域，数值的取值范围各不相同，因此不可避免的情况是，不同的分量的数据处在不同的数量级上，数量级大的分量对模型的影响就较大，而数量级小的分量对模型的影响就小，小数“淹没”在了大数里而失去了意义，这相当于在构造模型之前便人为的给不同指标加上了权重，甚至是丢失了一些信息，这是十分不科学的；另外，由于计算机所能表示的数据精度有限，如果个别数据过大或者过小，也容易使计算后的数据越界而丢失信息。基于以上考虑，所选择的原始数据需要进行一定的处理，将数据标准化是在本文将要采用的策略。

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。

核函数的选择

我们参考了国内外关于支持向量机预测金融市场的报告，在以往对股票或指数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。

输入向量的选取方法

在应用 SVM 进行指数预测时，如何有效地选取输入向量的分量是决定预测模型准确性的关键。样本向量的各个分量应该选取能充分反应指数变化特征的定量指标，不加选择则会增加期望误差的上界。反之，选取指标过少难以市场的特点。

模型里所用到的输入向量是由 6 个交易基础指标计算得出，这 6 个指标分别是：开盘价、最高价、最低价、成交量、成交额、收盘价。选取上述的指标去构造出 7 个我们认为对股价具有较强相关性的输入向量：

- 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；
- 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。

参数的选择

在我们的 SVM 简单预测模型里，核函数采用的是市场公认的效果较好的 Gauss 径向基核函数，因此模型有两个参数需要调试，分别是 Gauss 核函数的参数 σ 和模型最优化求解时所用的惩罚参数 C。

在我们下面实证例子中的浦发银行实验里，这两个参数经过枚举法调试后确认了最优组合。这样的参数组合在我们测试的不同市场环境下均得到稳定的效果。

股票价格趋势预测方法

预测未来一些天的股票价格是增长还是下降的时候，除非具有非常高的正确率，否则对于实际投资并没有太大的指导意义。如果能够描绘出未来一些天股票价格的价格走势或者说是股票价格曲线的大致形状，即便预测出的股票价格与实际股票价格有一定的出入，对于股票投资来说也会有很大的指导意义。因此，本文的目标就是寻找一种方法来得到上述的预测效果。

先来简单说明一下 SVM 预测模型的运作流程：

选取 l 个数据作为训练样本，经过推导得到如下公式（详细可以参考报告《支持向量机在股票价格预测方面的应用》，我们称它为决策函数。

$$f(x) = \text{sign} \sum_{i=1}^l \alpha_i y_i \text{Ker}(x_i, x) + b \quad (*)$$

若 x_0 为其中一个输入参数，将 x_0 代入上述公式，如果 $f(x_0) = 1$ 则认为 x_0 属于正类，否则如果 $f(x_0) = -1$ 则认为 x_0 属于负类。

通过观察我们注意到，假如把上述公式中的正负判断函数 sign 去掉，令

$$g(x) = \sum_{i=1}^l \alpha_i y_i \text{Ker}(x_i, x) + b \quad (**)$$

显然 $g(x)$ 比 $f(x)$ 含有更多的信息量，在一般的分类问题中，这些信息都被忽略掉了。从模型的理论角度来考虑， $g(x)$ 实际上体现的是预测点与训练所到的超平面

的距离。所以当我们所要研究的样本是股票数据，而且预测的目标是股票价格的涨跌幅时，这些信息便可以很好的利用起来。

我们假设 \bar{x} 为一个预测用样本（输入参数），用决策函数（*）可以预测出对应的未来某一天（假设为第 m 天）的股票价格是涨还是跌。若使用公式（**）则可以得到股价变化的幅度。

$$z = g(\bar{x}) = \sum_{i=1}^l \alpha_i y_i \text{Ker}(x_i, \bar{x}) + b$$

假如第 m 天的股票价格实际变化为 y ，而 \hat{y} 是预测得到的变化值，那么假设存在映射 φ ，使得 z 与 \hat{y} 存在一一对应关系，即：

$$\hat{y} = \varphi(z).$$

这样便可以通过 \bar{x} 预测第 m 天的股票价格相当于当天的涨跌变化幅度。

SVM 模型的实际运用

- 1) 利用股票原始数据选取并构造预测指标，标准化指标向量，从而形成向量 $x_i (i = 1, 2, \dots)$ 。
- 2) 定义 N 为训练数据的长度， M 为预测数据的长度。
- 3) 令 $y_i = \text{第}(i + M \text{ 天的收盘价} - \text{第 } i \text{ 天的收盘价})$ 。
- 4) 选取 $N(x_1, x_2, \dots, x_N)$ 个数据作为训练数据，选取适当的核函数对这 N 个数据进行训练，得到 $\alpha_1, \alpha_2, \dots, \alpha_N$ 和 b ；
- 5) 计算 $z_i = g(x_i) = \sum_{j=1}^N \alpha_j \bar{y}_j \text{Ker}(x_j, x_i) + b, i = 1, 2, \dots, N$ 。
- 6) 选取这 N 个数据随后的 $M(x_{N+1}, x_{N+2}, \dots, x_{N+M})$ 个数据作为预测数据，用该 M 个数据，计算 $z_i = g(x_i) = \sum_{j=1}^N \alpha_j \bar{y}_j \text{Ker}(x_j, x_i) + b$ ，（其中 $\bar{y}_i = \text{sign}(y_i), i = N+1, N+2, \dots, N+M$ ）。
- 7) 令

$$\delta_i = \begin{cases} (\text{第 } i \text{ 天的收盘价} + y_i) - \text{第 } i + (M - 1) \text{ 天的收盘价}, & i = N + 1 \\ ((\text{第 } i \text{ 天的收盘价} + y_i) - (\text{第 } i - 1 \text{ 天的收盘价} + y_{i-1})), & i = N + 2, \dots, N + M \end{cases}$$
 则 $\delta_i (i = N+1, N+2, \dots, N+M)$ 就代表了第 $i+M$ 天相对于第 $i+M-1$ 天的实际涨跌幅度。
- 8) 令

$$\hat{\delta}_i = \begin{cases} (\text{第 } i \text{ 天的收盘价} + \hat{y}_i) - \text{第 } i + (M - 1) \text{ 天的收盘价}, & i = N + 1 \\ ((\text{第 } i \text{ 天的收盘价} + \hat{y}_i) - (\text{第 } i - 1 \text{ 天的收盘价} + \hat{y}_{i-1})), & i = N + 2, \dots, N + M \end{cases}$$
 则 $\hat{\delta}_i (i = N+1, N+2, \dots, N+M)$ 就代表了第 $i+M$ 天相对于第 $i+M-1$ 天的实际涨跌幅度。

我们可以采用一种滚动的预测模式。简单的把 M 的值变小，实证中我们使 $M=5$ ，即用一个训练集预测 5 个交易日的走势，用循环的模式不断把训练集的时间段往后推移，让训练集数据的时间可以尽可能的贴近预测数据的时间。

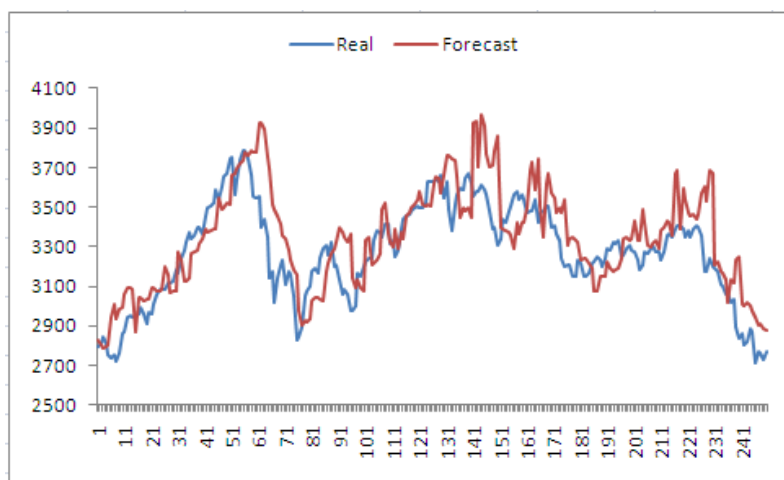
我们将利用上述计算得出的结果来绘制沪深 300 指数的价格曲线，通过对比真实与预测行情的走势，直观的来检验模型预测结果是否具有一定的说服力。

实证例子分析

预测沪深 300 指数走势

我们选择了应用 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%。

图 9：沪深 300 指数的预测走势和真实走势（近一年）



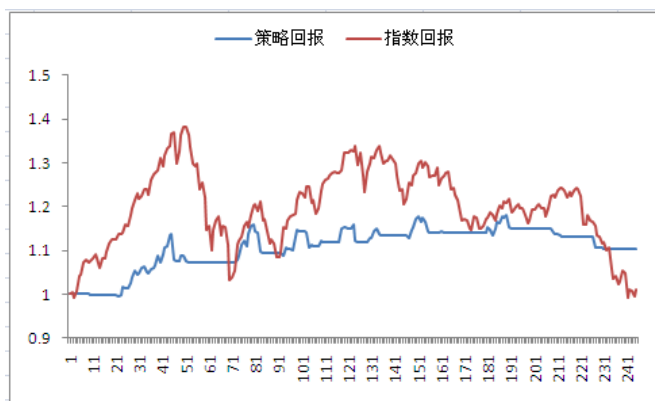
资料来源：国信证券经济研究所

基于预测走势的交易策略

由于预测走势存在长期稳定的滞后现象。运用类似于移动平均线金叉和死叉的择时策略，我们比较一下择时策略与被动跟踪指数在近一年的回报。

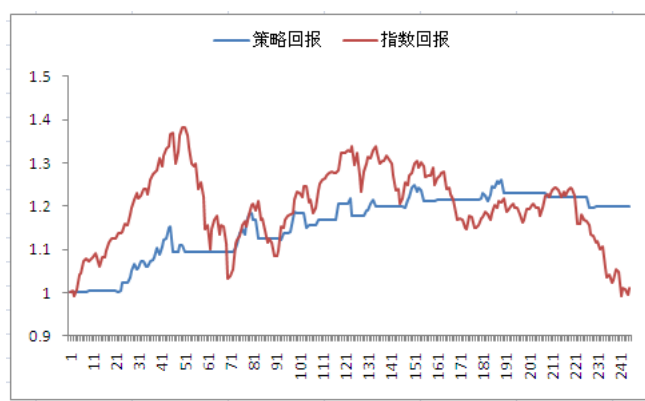
择时策略具体方法：第 n 天出现真实走势上穿预测走势（金叉），则按当日的收盘价买入；真实走势下穿预测走势（死叉）时按次日开盘价卖出。图 10、11 为多头策略，即出现买入信号持有，触发卖出信号时空仓等待机会。图 12、13 为多空策略，即出现买入信号持有多头，触发卖出信号时转多头为空头，获取指数下跌带来的收益。

图 10：多头策略，考虑交易成本



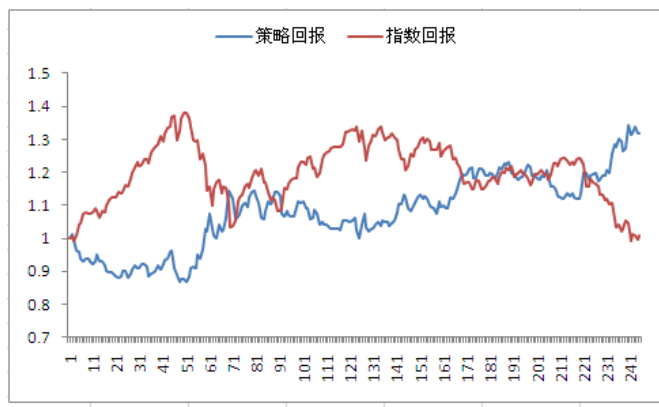
数据来源：国信证券经济研究所，

图 11：多头策略，未考虑交易成本



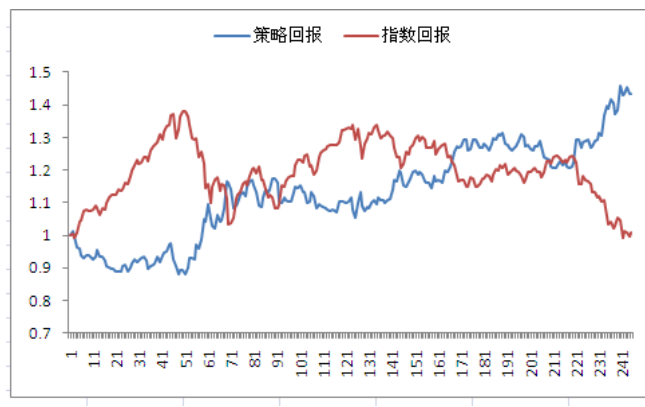
数据来源：国信证券经济研究所，

图 12: 多空策略, 考虑交易成本



数据来源: 国信证券经济研究所,

图 13: 多空策略, 未考虑交易成本



数据来源: 国信证券经济研究所,

下表统计了上述案例的结果, 实证例子的时间为 2009 年 5 月 15 日至 2010 年 5 月 21 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 指数回报率会 1.03%。观察测试周期内策略的最高和最低回报, 多头策略对下方风险控制较好, 收益率不会出现上下大幅波动。

表 1: 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	最低汇报
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%。

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。

模型的改进方案和后续拓展

SVM 预测模型存在的问题及解决方案

支持向量机因其广泛的适应能力和学习能力在非线性的预测方面有之广泛的应用前景。股票市场预测就是一个非线性的建模问题, 但是支持向量机在股市

预测中还缺乏系统的研究，本报告在这方面做出了一些尝试，对支持向量机方法进行了进一步的挖掘，并将其应用于指数价格的预测，取得了基本令人满意的结果。

尽管如此，支持向量机在股票市场预测中还有许多问题值得研究：

1. 对于一些情况预测模型泛化能力差，即拟合的非常好而预测的非常差。造成这一现象可能是训练集规模选择不当，进行短期预测，训练集过大会掩盖短期趋势，训练集过小很难把握运动趋势。

解决方法：需要通过大量实证，从统计学的角度切入，寻找最恰当的训练集规模。

2. 奇异点问题。

由于我国股市发展还不完善，投机性和政策性特征明显，经常出现暴涨暴跌的情况，造成股价运行中出现了很多的奇异点。由于造成奇异点的因素很多而且难以量化，仅靠支持向量机本身难以解决这个问题。

解决方法：可以考虑用傅立叶变换或小波变换提取出噪声和长周期因素。

3. 模型的优化。

从应用经验上讲，几种常用的核函数已经足够，但对于股市，未必是最适合的核函数，因此，为实际问题构造适当的核函数也许是最好的选择。

解决方法：最好的方法可能还是通过累积的经验来选择最适合实际情况的核函数。或对熟悉常用的核函数用排除法处理。

4. 输入向量的选择。

股市的数据量非常庞大，各种指标层出不穷，这些数据和指标都有它的实际意义，都反映了一定的股市信息。但是要把这些都作为输入量是不现实的，选择哪些数据作为输入量可以获得最好的模型并没有可靠的结论。

解决方法：我们认为与输出结果有稳定关系的输入向量理论上来说能提供更准确的预测。比如对个股、行业或者大盘的资金净流量，和与股价密切相关的技术指标，如 MACD, KDJ, RSI, BIAS 等。

后续拓展方向

小波理论的另一广泛应用是在信号的奇异性检测上。信号中的奇异点及不规则的突变部分往往携带比较重要的信息，是信号的重要特征之一。因此如果能检测出这些突变点的位置及突变程度将具有重要意义，可用于模式识别等重要领域。长期以来，Fourier 变换时研究信号奇异性的主要工具，但由于 Fourier 变换单纯进行频率分析，不具有空间变量的定位功能，它只能确定一个信号奇异性的整体性质。而小波函数具有“自适应”及“变焦”特性，能进行空间局部化分析，对突变信号的突变性即突变点的位置及突变度大小的分析更加有效。

股指期货的推出象征着 T+0 时代的归来。由于高频数据含有更丰富的信息量，我们认为通过研究与股指存在密切相关的高频指标，将其作为模型的输入向量，运用 SVM 预测模型对沪深 300 做出盘中走势的预测，可以给予股指期货投资者在投机交易业务上一些指导性建议。

国信证券投资评级

类别	级别	定义
股票 投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数±10%之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业 投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10%之间
	中性	预计 6 个月内，行业指数表现介于市场指数±5%之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

免责声明

本报告信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。本报告版权归国信证券所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登。

国信证券经济研究所研究团队(含联系人)

宏观			策略			交通运输		
周炳林	0755-82130638		黄学军	021-60933142		郑 武	0755- 82130422	
林松立	010-66026312		崔 嵘	021-60933159		陈建生	0755- 82133766	
			廖 喆	021-60933162		岳 鑫	0755- 82133766	
						高 健	0755-82130678	
银行			房地产			机械		
邱志承	021- 60875167		方 焱	0755-82130648		余爱斌	0755-82133400	
黄 飙	0755-82133476		区瑞明	0755-82130678		黄海培	021-60933150	
谈 煜	010- 66025229		黄道立	0755- 82133397		陈 玲	0755-82130646	
						杨 森	0755-82133343	
						李筱筠	010-66026326	
汽车及零配件			钢铁			商业贸易		
李 君	021-60933156		郑 东	010- 66026308		孙菲菲	0755-82133400	
左 涛	021-60933164		秦 波	010-66026317		吴美玉	010-66026319	
						祝 彬	0755-82131528	
基础化工			医药			石油与石化		
张栋梁	0755-82130532		贺平鸽	0755-82133396		李 晨	021-60875160	
陈爱华	0755-82133397		丁 丹	0755- 82139908		严蓓娜	021-60933165	
邱 斌	0755-82130532		陈 栋	021-60933147				
电力设备与新能源			传媒			有色金属		
皮家银	021-60933160		陈财茂	021-60933163		彭 波	0755-82133909	
						谢鸿鹤	0755-82130646	
电力与公用事业			非银行金融			通信		
徐颖真	021-60875162		邵子钦	0755- 82130468		严 平	021-60875165	
谢达成	021-60933161		田 良	0755-82130513		程 峰	021-60933167	
			童成敦	0755-82130513				
造纸			家电			计算机		
李世新	0755-82130565		王念春	0755-82130407		段迎晟	0755- 82130761	
邵 达	0755-82132098							
电子元器件			纺织服装			农业		
段迎晟	0755- 82130761		方军平	021-60933158		张 如	021-60933151	
高耀华								
旅游			食品饮料			建材		
廖绪发	021-60875168		黄 茂	0755-82138922		杨 昕	021-60933168	
刘智景	021-60933148							
煤炭			建筑			固定收益		
李 然	010-66026322		邱 波	0755-82133390		李怀定	021-60933152	
陈 健	010-66022025		李遵庆	0755-82130532		高 宇	0755- 82130876	
苏绍许	021-60933144					侯慧娣	021-60875161	
						张 旭	010-66026340	
						蔺晓熠	021-60933146	
						刘子宁	021-60933145	
指数与产品设计			投资基金			量化投资		
焦 健	0755-82133928		杨 涛	0755-82133339		葛新元	0755-82133332	
王军清	0755-82133297		彭怡萍	0755-82133528		董艺婷	021-60933155	
彭甘霖	0755-82133259		刘舒宇	0755-82133568		林晓明	021-60933154	
阳 瑾	0755-82133538		康 亢	010-66026337		赵斯尘	021-60875174	
周 琦	0755-82133568					程景佳	021-60933166	
赵学昂	0755-66025232							
交易策略								
戴 军	0755-82133129							
秦国文	0755-82133528							
徐左乾	0755-82133090							
黄志文	0755-82133928							

国信证券机构销售团队

华北区（机构销售一部）			华东区（机构销售二部）			华南区（机构销售三部）		
王立法	010-66026352 13910524551 wanglf@guosen.com.cn		盛建平	021-60875169 15821778133 shengjp@guosen.com.cn		万成水	0755-82133147 13923406013 wancs@guosen.com.cn	
王晓建	010-66026342 13701099132 wangxj@guosen.com.cn		马小丹	021-60875172 13801832154 maxd@guosen.com.cn		魏 宁	0755-82133492 13823515980 weining@guosen.com.cn	
焦 骥	010-66026343 13601094018 jiaojian@guosen.com.cn		郑 毅	021-60875171 13795229060 zhengyi@guosen.com.cn		邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn	
李 锐	010-66025249 13691229417 lirui2@guosen.com.cn		黄胜蓝	021-60875166 13761873797 huangsl@guosen.com.cn		林 莉	0755-82133197 13824397011 linli2@guosen.com.cn	
徐文琪	010-66026341 13811271758 xuwq@guosen.com.cn		刘 塑	021-60875177 13817906789 liusu@guosen.com.cn		王昊文	0755-82130818 18925287888 wanghaow@guosen.com.cn	
			叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn		甘 墨	0755-82133456 15013851021 ganmo@guosen.com.cn	
			孔华强	021-60875170 13681669123 konghq@guosen.com.cn		段莉娟	0755-82130509 18675575010 duanlj@guosen.com.cn	
						黎 敏	0755-82130681 13902482885 limin1@guosen.com.cn	
						徐 冉	13632580795 xuran1@guosen.com.cn	