

Hushen 300 Index Price Forecasting Based on Wavelet Transform

Sihui Wang, Yu Fei*

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming
Email: 18807426430@163.com, *1350691353@qq.com

Received: Oct. 8th, 2014; revised: Nov. 11th, 2014; accepted: Nov. 22nd, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper based on wavelet transform and non-wavelet transform established ARMA models to fit the daily closing price of Shanghai and Shenzhen 300 index and do a short-term forecast. It also compared their normalized mean square error (NMSE). Results display that due to the characteristics of wavelet transform which are good time frequency localization and its multi-resolution features, combined forecast model for short-term forecast of Shanghai and Shenzhen 300 index is superior to single forecast model.

Keywords

Wavelet Transform, Shanghai and Shenzhen 300 Index, ARMA, Forecast

基于小波变换的沪深300 指数预测

汪思慧, 费宇*

云南财经大学, 统计与数学学院, 昆明
Email: 18807426430@163.com, *1350691353@qq.com

收稿日期: 2014年10月8日; 修回日期: 2014年11月11日; 录用日期: 2014年11月22日

*通讯作者。

摘 要

本文通过基于小波变换和未基于小波变换对沪深300指数日收盘价序列分别建立ARMA拟合模型并做短期预测,对其归一化均方误差(NMSE)进行比较,结果显示,由于小波变换良好的时频局域化特性,以及它的多分辨功能,使组合模型较之于单个预测模型对于沪深300指数的短期预测更优。

关键词

小波变换, 沪深300, ARMA, 预测

1. 引言

由于证券市场的特殊性和波动集群特征,股价变化趋势研究的重要性日益突出,小波分析和时序理论在合理的预测股价及其波动趋势、降低投资风险的同时最大化投资收益方面应用也越来越广泛。侯璐(2009) [1]基于 ARIMA 模型探讨了短期石油价格分析,说明了国际经济状况、政策波动、投机活动等都对油价波动造成相当大的影响。范雯(2011) [2]在国际石油期货价格预测一文中利用小波与卡尔曼滤波混合的方法提出了具有实时递归性且多尺度分析能力的混合预测模型,实现了周期内对目标状态的预测。

为了模型有更好的预测精度及稳定性,本文将小波变换的时频局域化特征、多分辨功能与时序理论结合引入到沪深 300 指数日收盘价格预测问题的研究中,试图通过是否小波变换的前提下,来识别不同时序模型之间的短期预测效果,从而为正确分析非线性经济系统,完善经济系统的预测提供了实用价值。

2. 小波分析理论基础

2.1. 小波变换的含义

小波变换中,“小”说的是它的衰减性;而“波”则是指它的波动性,其振幅正负相间的震荡形式。较之于傅里叶变换,小波变换是时域和频域的局部分析,它通过由粗到细地逐步观察信号并进行伸缩平移运算细分,最终达到在低频处为频率细分,而高频处是时间细分,自适应时频信号分析,从而聚焦到信号的任意细节。我们也可以把小波变换看成是带通滤波器在不同尺度下对信号做滤波。

2.2. 几种常见的小波函数

常用的小波分析中所用到的小波函数并不是唯一的,即小波函数是多样的。小波分析中常见的小波函数主要有: Haar 小波、Daubechies(dbN)小波、Symlet(symN)小波、Coiflet(CoifN)小波、Biorthogonal (biorNr.Nd)小波[3]。

而合理的选择适宜的小波函数对数据进行变换又具有重要意义,所以我们选用如下标准作为选择小波函数的依据: 1) 对称性。基于该小波对应的滤波器具有线性相位的特点,对称性小波在图像或信号处理中能有效地避免相位畸变。常用的小波函数中,仅 Biorthogonal(biorNr.Nd)小波不具有对称性; 2) 消失距。在压缩数据和特征提取方面是非常有用的; 3) 支撑长度。支撑长度越长,一般需要耗费更多的计算时间,且产生更多高幅值的小波系数。dbN 系小波具有良好的紧支撑性; 4) 相似性。选择和信号波形相似的小波,这对于压缩和消噪是有参考意义的。5) 正则性。正则性好的小波,可以在信号或图像的重构中获得较好的平滑效果,减小量化后或舍入误差的视觉影响。常用的小波函数中,仅 Biorthogonal (biorNr.Nd)小波不具有正则性[3]。

本论文将采用上述性质根据输入信号的奇异性、相似性以及分解重构后信号的误差来合适的小波函数[4]。

2.3. 基于多分辨分析小波的信号分解与重构

合适的小波函数进行小波变换的同时我们必须考虑如何对信号进行分解和重构，而在小波函数的多分辨性基础上对信号进行分解重构是必须的，MATLAB 小波工具箱提供了可视化的小波分析工具，它包含图像化的工具及命令行函数，可以实现：测试、探索小波和小波包的特性、对信号进行压缩和降噪等功能。本文主要运用 MATLAB 小波工具箱如下实现小波分解重构的函数[4]：

$[C, L] = \text{wavedec}(X, N, 'wname')$ 返回信号 s 在 N 层的小波分解。 N 则必须是严格的正整数。输出分解结构包含了小波分解向量 C 和相应的记录向量 L 。由一维小波系数进行单支重构， $a = \text{wrccoef}('type', c, l, 'wname', N)$ 是基于小波分解结构 $[C, L]$ 在 N 层计算重构系数向量，以及指定的小波 'wname' 或者重构滤波器 Lo_R 和 Hi_R ，来重构一维信号的系数。

3. 实证研究

3.1. 数据准备与检验

本论文采用沪深 300 指数进行预测分析，数据来源于大智慧软件下载。

沪深 300 指数由 300 只来自沪深两市场的股票构成，是沪深证券交易所于 2005 年 4 月 8 日联合发布的用以反映 A 股市场变动全貌及运行状况，为指数化投资和指数衍生产品创新提供基础条件。其 300 只样本股选原原则是通过：1) 确定样本空间。主要需具备以下条件：上市时间超过一个季度(除非该股票上市以来日均 A 股总市值在全部沪深 A 股排名前 30 位)，非 ST、*ST 股票，非暂停上市股票，经营状况良好，近一年无重大违法违规事件，财务报告无重大问题，股价变化无明显异动或市场操纵，剔除其它经专家委员会认定的应该剔除的股票；2) 计算样本空间内股票近一年(新股上市以来)的 A 股日均成交金额和日均总市值；3) 对样本空间股票近一年的 A 股日均成交金额排名，选择排名前 50% 的股票中近一年日均 A 股总市值排名在前 300 名的股票作为样本股[5]。

其计算公式：报告期指数=报告期成份股的调整市值/基日成份股的调整市值*1000。

其中，调整市值 = $\sum(\text{市价} \times \text{调整股本数})$ 。

而调整股本是对自由流通股本通过分级靠档后获得的。自由流通股本指的是剔除了上市公司股本中的不流通股份及由于战略持股性质或其它原因导致的基本不流通股份剩下的股本部分[5]。

考虑到采用的数据是源于 2009 年 10 月~2014 年 9 月的沪深 300 指数日收盘价数据，样本大小为 1204，而小波分解尺度即(层数) k 必须满足信号长度能被 2^k 整除。因此，先选择运用不同小波函数对原序列进行 2 层的分解与重构。分解重构的误差比较如表 1：

由分解重构后的误差比较可知，分解重构误差较小的是 dbN 系小波和 symN 系小波，特别是：sym1 和 db1。又根据沪深 300 指数曲线的不平缓特性，其奇异性很明显，且我们期望得到更为准确的短期预测结果，而消失距在时域、频域方面表征信号局部的特征、检测信号的瞬态或奇异点方面是非常有用的，阶数小时域局部性越好，良好的紧支撑性则能凸显出信号的时域特性，综上考虑，选择具有较好紧支撑性、消失距阶数较低的 db1 小波函数来对原始信号进行分解和重构。

以下是 db1 小波函数对原始数据序列分解和重构后的各统计量(其中 $D1$ 、 $D2$ 分别代表分解重构后的第一层、第二层高频分量， $A1$ 、 $A2$ 则分别代表分解重构后的第一层、第二层低频分量)(表 2)：

由于低频信号象征了原始信号的波动性特点，而高频部分中则包含影响市场走向的突发事件，根据 db1 小波函数 2 层分解重构后得到的各层分量序列的统计量可知，低频信号 $A1$ 的最值、标准差、均值与

Table 1. Errors comparison
表 1. 误差比较

	小波基	Coif	Sym	Db
	小波基 1	2.68E-09	2.72E-11	2.72E-11
	小波基 2	3.32E-08	1.74E-09	1.74E-09
	小波基 3	1.85E-09	2.37E-08	2.37E-08
分解重构后误差	小波基 4	8.31E-08	1.48E-09	4.33E-09
	小波基 5	1.75E-05	5.30E-10	6.96E-09
	小波基 6	-	2.31E-09	4.44E-09
	小波基 7	-	2.30E-09	4.63E-09
	小波基 8	-	7.35E-10	1.11E-08

Table 2. The layer sequence statistics through Wavelet decomposition and reconstructed of 2-story
表 2. db1 小波函数 2 层分解重构后的各层序列统计量

数据序列	均值(MEAN)	最大值(MAX)	最小值(MIN)	标准差(STD)
D1	0	109.075	-109.075	19.78220037
D2	0	100.2775	-100.278	24.63132712
A2	2685.622	3633.008	2126.193	412.1707567
A1	2685.622	3646.425	2103.225	412.9060849
S	2685.622	3668.83	2086.97	413.3796927

原始信号的各统计量更为接近，即低频信号 A1 更能代表沪深 300 指数的波动特点，反映原始序列的长期趋势，且保留了原始序列的大部分信息，因此最终我们考虑使用 db1 小波函数对沪深 300 指数日收盘价序列进行 1 层的分解和重构。

3.2. 基于 AR 模型的低频价格序列建模分析

接下来我们对沪深 300 指数日收盘价以及经过小波分解重构后的各高频及低频信号进行了平稳性检验。由 ADF 检验结果得到变量 A1 不平稳但 D1 是平稳的，对 A1 进行 1 阶差分后平稳性检验通过，即 $D(A1)$ 是平稳的[6]。根据一阶差分后的低频序列相关结果可以看出沪深 300 指数低频增长率的偏自相关系数是 2 阶截尾的，自相关系数也是 2 阶截尾的。因此我们考虑采用枚举的方法来确定最优的 q 和 p 值。取模型的阶数和，经过枚举法综合各组合模型相应的 AIC 和 SC 值可知以下组合为最理想的模型[7]。

模型参数估计结果如表 3:

据此得出模型:

$$A1_t = -0.6493 + 0.2279A1_{t-2} + 0.2279A1_{t-3} + A1_{t-1} + \varepsilon_t - 0.0688\varepsilon_{t-2}$$

再依据模型进行静态预测得到后 4 个月的预测值。并对模型进行残差检验，可知模型不存在序列相关，模型拟合较好。

3.3. 基于 ARMA 模型的高频价格序列建模分析

同样的对高频率序列进行序列相关检验，可知高频率序列的偏自相关系数是拖尾的，自相关系数在 1 阶

截尾。尝试对 $D1$ 序列建立 ARMA(10,1)模型依据其系数显著性选择带 AR(4)、AR(6)、AR(8)、AR(10)、MA(1)项的模型(表 4):

模型结果为:

$$D1_t = 0.0047 - 0.0750D1_{t-4} - 0.1076D1_{t-6} - 0.0675D1_{t-8} - 0.0873D1_{t-10} + \varepsilon_t - 1.0069\varepsilon_{t-1}$$

残差检验的结果表明，模型的偏相关和自相关系数均在两倍标准差内即残差为白噪声序列，模型检验通过。

3.4. 实证结果比较分析

经过小波变换后的序列应满足： $s = d1 + d2 + \dots + di + ai$ ，其中， i 指的是分解的层数(即尺度) ($i = 1, 2, \dots, n$)， di 是高频分量， ai 是低频分量， s 为原始数据。

综上得到序列的拟合值也应该满足： $\hat{s} = \hat{a}1 + \hat{d}1$ ，则由静态预测得到经 db1 小波变换后的沪深 300 指数序列模型 NMSE 为 0.005166。

对沪深 300 指数日收盘价原序列进行单位根检验得到其不平稳但其一阶差分 $D(S)$ 是平稳的。对沪深 300 指数一阶差分的序列相关检验可知，偏自相关系数是拖尾的，自相关系数也是拖尾的。

因此我们考虑运用枚举法来确定最优的 p 、 q 值。取模型的阶数 $p = 7$ 和 $q = 7$ ，先建立 ARIMA(7, 1, 7)模型。由于建立的模型中有的解释变量的系数估计值是不显著的，经过枚举法对不同的 p 、 q 进行组合建立模型，根据 AIC 和 SC 得到最优模型为[7]:

$$s_t = -1.0115 + 0.9071s_{t-5} + 0.2279s_{t-6} + s_{t-1} + \varepsilon_t - 0.9112\varepsilon_{t-5}$$

Table 3. Parameter estimate
表 3. 参数估计

变量	输出变量: da1			
	系数	标准差(Std.)	T 统计量	P 值
C	-0.64934	1.120559	-0.57948	0.5624
AR(2)	0.227882	0.15291	1.4903	0.1364
MA(2)(	-0.06883	0.157249	-0.43769	0.6617

Table 4. Parameter estimate
表 4. 参数估计

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004702	0.004017	1.170353	0.2421
AR(4)	-0.074992	0.028843	-2.60001	0.0094
AR(6)	-0.107617	0.028693	-3.75066	0.0002
AR(8)	-0.067521	0.028679	-2.35439	0.0187
AR(10)	0.087257	0.028434	3.068732	0.0022
MA(1)	-1.006921	0.006188	-162.717	0

R^2 = 0.519025
AIC = 8.056301
SC = 8.081941
D.W. = 1.992522

建立带 AR(5)、MA(5)的 ARMA 模型(表 5):

残差检验的结果表明,模型的偏相关和自相关系数均在两倍标准差内即残差为白噪声序列,模型检验通过。

同样的对模型进行静态预测并计算 NMSE 为: 0.008349(表 6)。

综上所述得到,经过小波变换后的时间序列模型其 NMSE 较小,预测精度较高。体现了小波变换其较好的时频局域化特性以及它的多分辨功能在金融数据的适用性,能更好的对数据进行分解[8]。

4. 结论

股票市场作为社会主义市场经济的重要组成部分,对股票市场价格变化的准确判断,不仅可以为投资者降低投资风险的同时最大化投资收益,更可以有效地对市场进行监管,防范金融风险[9]。本论文在研究沪深 300 指数日收盘价序列的基础上把小波变换方法与 ARMA 模型相结合,构造出一个混合且较精准的沪深 300 指数日收盘价预测模型。并利用该模型对数据进行短期预测,利用小波函数的自适应性提高了预测精度。论文主要采用了如下方法得到相关的结论:

- 1) 基于小波变换对数据进行预处理得到各高频、低频序列进行分析。
 - 2) 传统统计学理论分析法采用较为成熟的 ARMA 模型,并对沪深 300 指数日收盘价 ARMA 预测模型的阶数确定、检验与分析进行了详细的说明。
 - 3) 用沪深 300 指数日收盘价历史数据分别对单个和组合模型进行训练并预测日收盘价价格,利用 NMSE 对预测结果进行误差比较,结果表明组合模型较之有更高的预测精度,整体相对误差也较小。
- 股票的预测研究是一个涉及面非常广的应用型研究课题,本论文只进行了沪深 300 指数日收盘价的短期预测工作,研究思路有待于进一步的探讨和研究[10]:

- 1) 影响沪深 300 指数日收盘价的众多影响因素中不只包含自身历史数据还包括技术指标、宏观经济等一些层面,若将这些因素有效的量化纳入预测模型中,进行综合分析,一定会提高预测精度,为投资者提供更可信的指导[5]。
- 2) 可以考虑如何利用小波分析理论更进一步的优化小波时间序列结构,样本数据的选取、预处理和数据滤波等,应该在更深与更广的范围内进行探讨。
- 3) 可以尝试运用小波函数与新兴的机器学习方法组合对比传统组合模型预测进行探讨。

基金项目

云南省省院省校教育合作咨询共建重点学科——统计学(42111217003)。

Table 5. Parameter estimate
表 5. 参数估计

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.01147	1.091511	-0.92667	0.3543
AR(5)	0.907099	0.033261	27.27236	0
MA(5)	-0.91122	0.033554	-27.1567	0

Table 6. Comparison of the predictions among two methods
表 6. 模型预测效果对比

	基于小波变换的铜期货价格 预测模型	未经小波变换的铜期货价格 预测模型
NMSE	0.005166	0.008349

参考文献 (References)

- [1] 侯璐 (2009) 基于 ARIMA 模型的石油价格短期分析预测. 硕士论文, 暨南大学, 广州.
- [2] 范雯 (2011) 国际石油价格期货预测的小波-Kalman 滤波混合方法. 硕士论文, 西安科技大学, 西安.
- [3] 辛月 (2011) 基于小波理论的期铜价格周期波动和预测模型研究. 硕士论文, 北方工业大学, 北京.
- [4] 郑纪安 (2009) 基于小波分析和神经网络的金融时间序列预测研究. 硕士论文, 厦门大学, 厦门.
- [5] 张坤 (2009) 中国股票市场投资策略的计量分析. 经济科学. 硕士论文, 江西财经大学, 南昌.
- [6] 王习涛 (2006) ARIMA 模型在期货交易预测中的应用研究. *微计算机信息*, **15**, 139-140.
- [7] 陈林, 黄章树 (2010) 基于 ARIMA 模型的期货价格分析与预测. *福州大学学报(哲学社会科学版)*, **3**, 2-4.
- [8] 徐珍, 李星野 (2012) 小波 ARMA 模型和 ARIMA 模型对上证指数的预测效果探究. *现代商业*, **30**, 3-4.
- [9] 杨丽 (2008) 小波神经网络和 ARMA 模型在股票预测中的研究与应用. 硕士论文, 北京工业大学, 北京.
- [10] Kanas, A. (2008) Modeling regime transition in stock index futures markets and forecasting implications. *Journal of Forecasting*, **27**, 649-669.