

基于小波分析的 ARIMA 模型对上证指数的分析与预测

石鸿雁¹, 尤作军², 陈忠菊³

(1. 沈阳工业大学 理学院, 辽宁 沈阳 110000)

(2. 沈阳工业大学 研究生学院, 辽宁 沈阳 110000)

(3. 辽宁公安司法管理干部学院 公共安全系, 辽宁 沈阳 110031)

摘 要: 股票价格的预测一直受到广泛关注, 其预测方法虽然很多, 但是往往存在预测精度有限、容易陷入局部极小等问题. 为了提高股票价格预测的准确性, 提出了基于小波分析的 ARIMA 模型的股票价格预测方法, 同时利用该方法对上证指数收盘价的月平均值进行实例分析, 并与其他方法的预测结果进行了比较, 结果表明了提出方法的有效性.

关键词: 小波分析; ARIMA 模型; 股票价格预测; 上证指数

众所周知, 股票市场是一个复杂的系统, 股票价格的变化受到经济、政治及投资者心理等多种因素的影响, 各因素的影响程度、时间范围和方式也不尽相同, 且股市各因素间相互关系错综复杂, 主次关系变化不定, 数量关系难以提取及定量分析, 因此建立准确的预测模型非常困难, 使得未来股票价格的预测有一定局限性. 目前股票预测的方法很多^[1-2], 比较常用的方法有 ARIMA^[3]、神经网络^[4-5] 及支持向量机^[6-7] 等. 这些预测方法在一定程度上揭示了股市的运行规律, 但是股价系统内部结构的复杂性、外部因素的多变性限制了股市预测的准确性, 现有的分析预测方法应用效果并不理想. 如 ARIMA 模型、贝叶斯网络预测精度高, 技术成熟, 但是建模方法繁琐, 对数据的统计特征要求较高, 只能对满足一定要求的特殊序列建模; 人工智能和神经网络适合非线性模型的拟合, 但收敛速度慢, 并且容易陷入局部极小点, 影响结果的准确性; 支持向量机适合小样本的快速预测, 能得到全局最优, 但在应用中也存在一些问题, 特别是一些参数的选取将直接影响算法的性能和预测效果, 使得支持向量机的参数选择缺乏公认有效的结构化方法.

小波分析理论是根据时-频局部化的要求发展起来的, 具有自适应和数学显微镜的性质, 适合非平稳、非线性信号的处理. 小波分析理论指出, 可以将信号分解到不同的频率上. 由于分解后的信号在频率成分上比原始信号单一, 并且小波分解对信号作了平滑处理, 因此对分解后的时间序列建立模型比原始时间序列建立模型得到的结果更准确. 文献 [8] 中, 利用小波分析中的多分辨率分解, 将中国人均 GDP 依尺度分解成不同层次, 对低频部分和高频部分分别进行重构, 其中用到时间序列分析中的 AR(1) 模型, 灰色预测的 GM(1,1) 模型, 然后分别对每层建立不同模型进行预测, 最后合成得到所需的预测值. 该文中利用 1986-2000 年的

收稿日期: 2014-02-25

资助项目: 国家自然科学基金 (61074005)

人均 GDP 值, 预测 2001–2005 年的 GDP. 研究结果显示, 该方法的预测误差小于直接利用 GM(1,1) 模型, 达到了较好的效果. 文献 [9] 提出将小波分析和 ARMA 模型、神经网络结合, 利用小波消噪对原始时间序列进行滤波, 利用小波分解序列的特征和分解数据随尺度倍增而倍减的规律用于 BP 神经网络和自回归移动平均模型的建模, 实验表明该方法既可以将时间序列中的各种隐周期分离并进行有效描述, 又能很好地刻画时间序列的非线性性, 从而提高了数据预测的精度.

针对股票数据的复杂性, 及基于小波分析的预测模型的优势, 本文将小波分析理论中的分解、重构算法与时间序列分析中的 ARIMA 模型相结合, 给出一种股票价格的预测方法, 并通过上证指数的月平均收盘价验证这种预测方法的有效性. 该方法对中国股市的预测研究有一定的实际应用价值, 为投资者做出投资决策提供了参考依据.

1 研究基础

1.1 小波分析

小波分析以其自适应和数学显微镜的特性引起人们越来越广泛的关注, 并取得了令人瞩目的进展, 在信号分析、图像处理、模式识别、语音分析、方程求解等领域都已取得了具有科学意义和应用价值的重要成果. 小波变换中的伸缩参数实质上描述了观测信号的范围, 也就是尺度, 在图像处理中称之为分辨率, 所以小波变换也可以理解为信号的多分辨率分析. 多分辨率分析 (MRA) 是由 S.Mallat 引入的, 他从空间概念上, 形象地说明了小波的多分辨率特征, 将在此之前所有的小波变换理论统一起来. 1989 年, S.Mallat 在小波多分辨率分析理论与图像处理的应用研究中受到塔式算法的启发, 提出了信号的塔式多分辨率分解与重构的快速算法, 即著名的 Mallat 算法.

小波分解和重构的 Mallat 算法如下:

1) 分解算法

$$C_{j,n} = 2^{-\frac{1}{2}} \sum_{k \in \mathbb{Z}} h_{2k-n} C_{j-1,k} \quad (1)$$

$$d_{j,n} = 2^{-\frac{1}{2}} \sum_{k \in \mathbb{Z}} g_{2k-n} C_{j-1,k} \quad (2)$$

2) 重构算法

$$C_{j-1,n} = \sum_{k \in \mathbb{Z}} h_{2k-n} C_{j,k} + \sum_{k \in \mathbb{Z}} g_{2k-n} d_{j,k} \quad (3)$$

针对股票序列的非平稳、非线性性, 本文应用小波分析的多分辨率分析将一组股票序列分解为不同频率上的趋势更为明显的序列, 由于分解后的序列在频率成分上比原始序列单一, 所以使得后续的处理过程更加精确, 为进行准确的预测打下了坚实的基础.

1.2 ARIMA 模型

对一个非平稳序列, 其数字特征, 如均值、方差和协方差等随着时间的变化而变化, 也就是说, 非平稳序列在各个时间点上的随机规律是不同的, 难以通过样本序列的一致信息去掌握序列整体上的随机性. 但在宏观经济领域的实证研究中, 多数经济时间序列都是非平稳的, 为此 Box-Jenkins 提出了以随机理论为基础的时间序列预测分析方法, 其基本模型包括: 自回归模型 (AR)、移动平均模型 (MA)、自回归移动平均模型 (ARMA) 和自回归求和移动平

均模型 (ARIMA)^[10-12]. 其中 AR、MA、ARMA 模型是处理平稳时间序列的模型, 而对于非平稳的时间序列则需要建立 ARIMA 模型, ARIMA 模型的形式如下:

考虑序列 y_t , 若其能够通过 d 次差分后变为平稳序列, 则

$$u_t = \Delta^d y_t = (1 - B)^d y_t \quad (4)$$

其中 u_t 为平稳序列, 于是可建立 ARMA(p,q) 模型:

$$u_t = c + \varphi_1 u_{t-1} + \cdots + \varphi_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (5)$$

经 d 阶差分后的 ARMA(p,q) 模型成为 ARIMA(p,d,q) 模型, 其中 p 为自回归部分的阶数, q 为移动平均部分的阶数, ε_t 为一个白噪声过程.

针对股票序列的非平稳性, 以及 ARIMA 模型在处理非平稳时间序列上的优越性, 本文选取 ARIMA 模型对小波分解后的时间序列分别建立预测模型.

2 建模过程及实例分析

为了提高模型的预测精度, 本文首先利用小波分析将原始序列进行分解与重构, 然后对低频序列和高频序列分别建立 ARIMA 模型, 得到低频序列和高频序列的预测结果, 从而得到原始序列的预测结果. 具体情况如下:

2.1 基于小波分析的 ARIMA 模型的建立

考虑小波分析及一般时间序列模型的建模过程, 结合股票序列的特点, 本文的建模过程为:

- 1) 小波分解. 选取合适的小波基函数, 利用 Mallat 算法对序列进行分解与重构, 得到近似序列 a_1 和细节序列 d_1 ;
- 2) 序列平稳化处理及平稳性检验. 对序列 a_1 进行 ADF 单位根检验以检验序列的平稳性, 如果序列不满足平稳性条件, 则通过差分变换等一系列变换使得序列满足平稳性条件;
- 3) 模型识别. 通过计算描述平稳化处理后的序列 a_1 的统计量 (如自相关函数和偏自相关函数), 来估计 ARIMA 模型的阶数 p 和 q , 并用 AIC 准则和 SC 准则确定阶数 p 和 q ;
- 4) 模型估计. 估计模型的未知参数, 并检验模型的平稳性及可逆性;
- 5) 模型诊断检验. 对 (4) 中所估计模型的残差进行白噪声检验, 以确定模型是否可行;
- 6) 模型预测. 利用 (4) 得到的模型对序列 a_1 进行预测, 得到 a_1 的预测值. 重复 (2)-(6) 可得到序列 d_1 的预测值, 从而得到原始序列的预测值.

建模过程如图 1 所示.

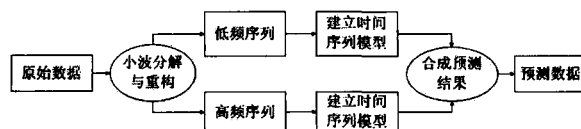


图 1 基于小波分析的 ARIMA 模型建模过程

2.2 实例分析

为检验上述方法建立模型的预测效果, 本文以上证指数收盘价的历史数据 (数据来源于上海证券交易所官方网站) 作为观察对象进行分析, 对 1990 年 12 月 19 日到 2013 年 3 月 29

日的上证指数交易日的收盘价, 按月求平均值后的 268 个数据 (记为序列 a, 如图 2 所示) 利用 Eviews6.0 软件进行分析, 该数据分为两部分: 1990 年 12 月到 2013 年 1 月数据用于估计模型的参数, 2013 年 2 月和 2013 年 3 月数据用于模型预测效果分析.

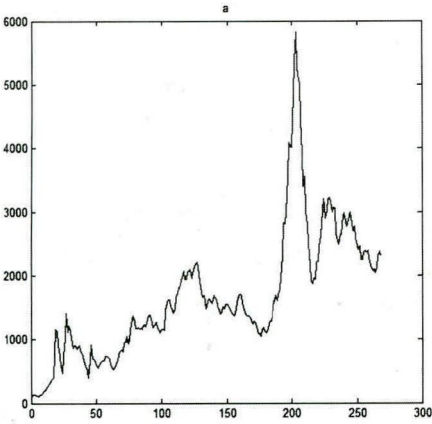


图 2 上证指数折线图

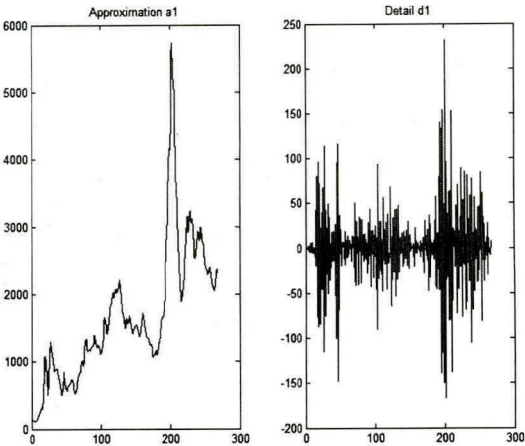


图 3 小波分解后与重构后的趋势序列 a1 和细节序列 d1

1) 小波分解

通过前期大量的试验对比分析, 本文采用 db4 作为小波基, 首先对序列 a 进行单尺度分解, 然后直接对小波系数进行重构, 得到近似序列 a1 和细节序列 d1(如图 3 所示), 之后对序列 a1、d1 分别建立 ARIMA 模型, 以下的建模过程以 a1 为例进行介绍.

2) 序列平稳化处理及平稳性检验

实验过程中发现直接对序列 a1 建模时, 残差序列是自相关的, 为了消除残差的自相关性, 对序列 a1 的增长率进行分析, 令

$$y_t = \frac{a1_t - a1_{t-1}}{a1_t}, \quad t = 2, 3, \dots, n$$

(6)

为了消除序列 y_t 的趋势并同时减小序列 y_t 的波动, 对序列 y_t 进行一阶差分处理, 令

$$u_t = y_t - y_{t-1}, \quad t = 3, 4, \dots, n$$

(7)

对经过式 (6) 与式 (7) 处理后的序列 u_t 进行 ADF 检验如图 4, 该检验结果表明, ADF 检验的 t 统计量 = -9.507092 均小于检验水平 1%、5%、10% 的临界值, 因此拒绝存在单位根原假设, 即序列 u_t 是平稳的, 由此确定 ARIMA 模型中的 $d=1$.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.507092	0.0000
Test critical values: 1% level	-3.455685	
5% level	-2.872586	
10% level	-2.572730	

图 4 序列 u_t 进行 ADF 检验结果

3) 模型识别

为了对 u_t 建立 ARIMA(p,d,q) 模型, 首先考察序列 u_t 的自相关、偏自相关函数和 Q 统计量 (如图 5 所示). 由图 5 看出自相关函数和偏自相关函数均呈现出拖尾性, 因此可以对序

列 u_t 建立 ARMA 模型, 经过反复筛选, 确定 $p=4, q=8$.

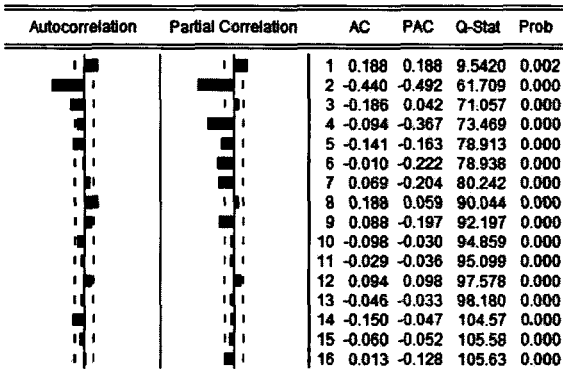


图 5 序列 u_t 的自相关、偏自相关函数

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000127	0.000109	-1.159941	0.2472
AR(2)	-0.116432	0.087670	-1.328070	0.1854
AR(4)	-0.139145	0.063485	-1.666500	0.0999
MA(1)	0.201231	0.062700	3.209445	0.0015
MA(2)	-0.722772	0.061783	-11.69853	0.0000
MA(3)	-0.512036	0.065355	-7.834643	0.0000
MA(8)	0.053019	0.048343	1.096724	0.2738
R-squared	0.467842	Mean dependent var	0.000473	
Adjusted R-squared	0.455221	S.D. dependent var	0.077471	
S.E. of regression	0.057181	Akaike info criterion	-2.658841	
Sum squared resid	0.627222	Schwarz criterion	-2.762776	
Log likelihood	378.6233	Hannan-Quinn criter.	-2.820102	
F-statistic	37.07042	Durbin-Watson stat	2.105127	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.40+.46i	.40-.46i	.40+.46i	.40-.46i
Inverted MA Roots	.99	.50	.22+.52i	.22-.52i
	-.39+.55i	-.39+.55i	-.73+.31i	-.73-.31i

图 6 模型 ARIMA(4,1,8) 的估计结果

4) 模型估计

为了检验模型的平稳性及模型的可逆性, 需要进行模型估计, 模型估计的结果如图 6 所示. 根据图 6 显示结果, AR 过程和 MA 过程滞后多项式的倒数根的模都在单位元内, 从而表明 ARIMA 模型是平稳的也是可逆的, 同时得到了模型中各参数的估计值.

5) 模型诊断检验

为了检验模型的参数估计是否正确, 对模型的残差序列进行白噪声检验, 即检验残差序列的随机性, 结果如图 7 所示. 图 7 表明, 模型估计结果的残差序列的样本自相关函数都在 95% 的置信区间以内, 各阶滞后的自相关函数相应的概率值 P 都大于 0.052, 即大于检验水平 0.05, 因此不能拒绝原假设, 即可认为模型 ARIMA(4,1,8) 的估计结果的残差序列满足随机性假设, 从而可以利用 ARIMA(4,1,8) 模型对序列 a_1 进行预测.

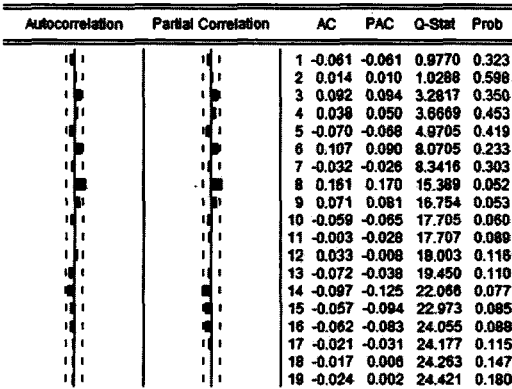


图 7 模型 ARIMA(4,1,8) 的残差相关图

6) 模型预测

通过上述过程的分析, 对序列 a1 建立如下 ARIMA(4,1,8) 模型:

$$\begin{cases} y_t = \frac{a1_t - a1_{t-1}}{a1_t} \\ u_t = y_t - y_{t-1} \\ u_t = -0.000127 - 0.116432u_{t-2} - 0.139145u_{t-4} + \varepsilon_t + \\ 0.20123\varepsilon_{t-1} - 0.72272\varepsilon_{t-2} - 0.512036\varepsilon_{t-3} + 0.053019\varepsilon_{t-8} \end{cases} \quad (8)$$

与序列 a1 建模过程类似, 确立序列 d1 的模型为 ARIMA(1,1,1). 对序列 a1 和序列 d1 分别利用 ARIMA(4,1,8) 和 ARIMA(1,1,1) 得到的 a1 在 2013 年 2 月到 2013 年 3 月的预测值与 d1 在 2013 年 2 月到 2013 年 3 月的预测值, 将其合成得到上证指数 2013 年 2 月到 2013 年 3 月的预测值. 为了检验本模型预测结果的准确性, 分别与其他模型的预测结果进行了比较, 结果如表 1 所示.

表 1 不同模型预测效果比较			
2013 年 2 月实际值 2380.363			
	预测值	绝对误差	相对误差
小波 ARIMA	2449.571	69.208	2.91%
ARIMA(1,2,4)	2485.216	104.853	4.40%
ARCH	2291.785	88.578	3.72%
2013 年 3 月实际值 2306.167			
	预测值	绝对误差	相对误差
小波 ARIMA	2438.009	131.812	5.72%
ARIMA(1,2,4)	2486.584	180.387	7.82%
ARCH	2272.940	33.227	1.44%

通过表 1 可以看出, 利用基于小波分析的 ARIMA 模型对 2013 年 2 月和 2013 年 3 月预测值分别为 2449.571 和 2438.009, 与实际值的相对误差分别是 2.91%和 5.72%. 而利用 ARIMA 模型和 ARCH 模型对 2013 年 2 月的预测值分别为 2485.216 和 2291.785, 相对误差分别为 4.40%和 3.72%, 显然基于小波分析的 ARIMA 模型对 2 月的预测值比 ARIMA 模型和 ARCH 模型更加精确. 但是对于 2013 年 3 月, ARIMA 模型和 ARCH 模型的预测值分别为 2486.584 和 2272.940, 与实际值的相对误差分别为 7.82%和 1.44%. 可以明显看出, 基于小波分析的 ARIMA 模型对于上证指数滞后一期的预测结果比直接利用 ARIMA 模型或者 ARCH 模型的预测结果更加精确, 但是对于滞后两期的预测结果还有待提高.

3 结论

本文利用基于小波分析的 ARIMA 模型对上证指数进行了分析预测, 并与直接利用 ARIMA 模型和 ARCH 模型的预测效果进行比较, 可以看到由于小波分析的引入提高了模型短期预测的精度, 拟合结果在一定程度上反应了上证指数的短期走势. 因此可以利用该模型对股票市场进行短期的预测分析, 为进一步分析未来股市的变化提供了理论依据, 也为投资者投资决策提供了参考. 但是随着时间的推移, 模型预测精度会越来越低, 因此对股票价格的长期预

测还有待进一步研究.

参考文献

- [1] Pai Pingfeng, Lin Chihsheng. A hybrid ARIMA and support vector machines model in stock price forecasting[J]. Omega, 2005(33): 497-505.
- [2] Wang Jujie, Wang Jianzhou, Zhang Zhegeorge, Guo Shupo. Stock index forecasting based on a hybrid model[J]. Omega, 2012(40): 758-766.
- [3] 白营闪. 基于 ARIMA 模型对上证指数的预测 [J]. 科学技术与工程, 2009, 9(16): 4885-4888.
- [4] 王文波, 费浦生, 羿旭明. 基于 EMD 与神经网络的中国股票市场预测 [J]. 系统工程理论与实践, 2010, 30(6): 1027-1032.
- [5] 杨一文, 刘贵忠, 蔡毓. 基于模糊神经网络和 R/S 分析的股票市场多步预测 [J]. 系统工程理论与实践, 2003(3): 70-76.
- [6] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究 [J]. 电子科技大学学报, 2011, 40(1): 2-7.
- [7] 徐国祥, 杨振建. PCA-GA-SVM 模型的构建及应用研究 — 沪深 300 指数预测精度实证分析 [J]. 数量经济技术经济研究, 2011(2): 135-147.
- [8] 刘晏玲, 胡芬, 付恩成. 基于小波分析的中国人均 GDP 分析与预测 [J]. 长江大学学报 (自然科学版), 2008, 3(5): 135-137.
- [9] 佟伟民, 李一军, 单永正. 基于小波分析的时间序列数据挖掘 [J]. 计算机工程, 2008, 1(34): 26-29.
- [10] 蒋金良, 林广民. 基于 ARIMA 模型的自动站风速预测 [J]. 控制理论与应用, 2008, 25(2): 374-376.
- [11] Ni Lihua, Chen Xiaorong, Huang Qian. ARIMA model for traffic flow prediction based on wavelet analysis [C]. 2010 2nd International Conference on Information Science and Engineering (ICISE), IEEE Conference Publications, 2010: 1028-1031.
- [12] 朱帮助, 林建. 基于 ARIMA 和 LSSVM 的非线性集成预测模型 [J]. 数学的实践与认识, 2009, 39(12): 34-40.

Analysis and Prediction of Shanghai Composite Index by ARIMA Model Based on Wavelet Analysis

SHI Hong-yan¹, YOU Zuo-jun², CHEN Zhong-ju³

(1. School of Science, Shenyang University of Technology, Shenyang 110000, China)

(2. Graduate School, Shenyang University of Technology, Shenyang 110000, China)

(3. Liaoning Administrators College of Public and Justilc, Public Security Department, Shenyang 110031, China)

Abstract: Stock price prediction attracts the extensive attention, although there are many forecasting method. But they often has problems such as limited prediction accuracy, easy to fall into local minimum and so on .In order to improve the accuracy of stock price forecasting. The modified model of ARIMA based on wavelet analysis of stock price forecasting methods is established. Then using this model analysis the monthly average closing price of the Shanghai composite index. And compared the prediction results with other methods. The results show the effectiveness of the proposed method.

Keywords: wavelet analysis; ARIMA Model; stock price prediction; shanghai composite index