

基于小波与动态 GM(1,1)-ARIMA 模型的股价预测研究

骆 桦,陈艳飞

(浙江理工大学理学院,杭州 310018)

摘 要: 针对传统时间序列股价预测模型的时间滞后性,提出一种基于小波与动态 GM(1,1)-ARIMA 的股价预测模型。运用小波分析对股价数据进行预处理,对小波重构序列建立 ARIMA 模型,考虑预测过程中未来因素对系统的影响,建立动态 GM(1,1)模型。选取沪深 300 指数进行实证分析,结果表明所提出的小波与动态 GM(1,1)-ARIMA 模型与传统股价预测模型相比,其预测精度最高。

关键词: 股价;小波分析;动态 GM(1,1);ARIMA 模型

中图分类号: F830.91

文献标志码: A

文章编号: 1673-3851(2017)04-0575-05

0 引 言

目前,预测股价的方法很多,其中基于简单统计和传统股价技术分析的预测方法大都存在滞后性且预测效果不理想。采用小波分析、神经网络、灰色预测等进行股价预测的方法是对传统预测方法的优化,国内外学者做了大量实证研究。刘向丽等^[1]通过小波分析建立 ARMA 模型对股指期货进行分析预测,发现该模型在高频预测方面优于传统 ARMA 模型;肖燕君等^[2]运用小波对股价序列进行分解,对不同尺度信号进行预测,预测的精度比传统模型高;Lahmiri^[3]和 Jammazi 等^[4]采用小波变换对金融时间序列进行预测,Hsieh 等^[5]采用小波分析对道琼斯工业平均指数、伦敦富时 100 指数和台湾证交所市值加权指数进行研究和预测,发现基于小波分析的预测模型精度更高。然而,小波分析在分解和重构时均会产生误差,单一的小波分析在预测时会造成一定的误差。孙冰洁等^[6]运用小波与神经网络的组合模型对股价进行预测,石鸿雁等^[7]采用小波分解与重构的 ARIMA 模型对股价预测。以上研究表

明,运用小波分析对原始数据序列进行分解重构,对重构后的序列建立 ARIMA 模型,能进一步提高模型的预测精度,但由于 ARIMA 模型是线性模型,预测时会造成时间滞后性,而灰色模型在短期预测能不断更新调整,从而使得预测结果更为准确。张宇敬等^[8]运用灰色预测模型对股价进行预测,骆桦等^[9]、Kara 等^[10]运用神经网络对期权价格进行预测,发现灰色系统和神经网络对于短期价格预测具有较高精度,但对于长时期预测,模型的预测精度将降低。

考虑到各种股价预测模型的优缺点,本文提出一种基于小波与动态 GM(1,1)-ARIMA 模型的股价预测方法。采用小波分析对原始金融时间序列进行分解重构,得到高频数据和低频数据,分别运用动态 GM(1,1)模型对高频数据和 ARIMA 模型对低频数据进行拟合建模并预测,二者预测值之和为模型最终预测结果。本文以沪深 300 指数为实证分析验证所提出模型的可行性,同时比较单一的灰色模型、单一的 ARIMA 模型与本文所提出的模型对沪深 300 指数的预测精度,以验证本文所提出模型的有效性。

1 实验数据来源及选取

沪深300指数是由沪深两个证券市场中具有代表性的股票编制,能反映沪深两个市场的整体走势和市场主流投资的收益情况。本文选取沪深300指数作为研究对象,从同花顺证券交易软件下载沪深300指数2016年1月4日至2017年1月19日共257个交易日的收盘价作为实证分析的原始数据。其中2016年1月4日至2016年12月28日共242个交易日的收盘价数据为样本内数据,其余15个交易日的收盘价数据为样本外数据。沪深300指数的收盘价序列如图1所示。

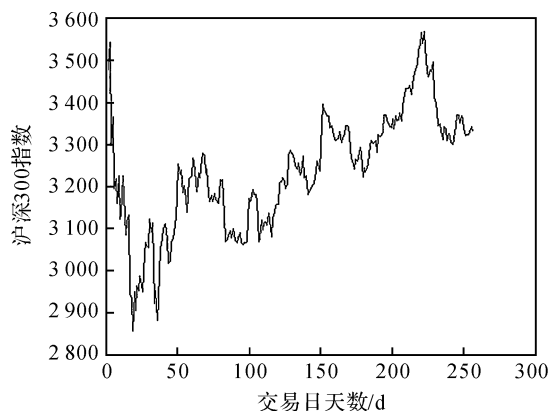


图1 沪深300指数收盘价原始数据

2 沪深300指数的小波与动态GM(1,1)-ARIMA模型

2.1 小波与动态GM(1,1)-ARIMA模型理论

2.1.1 小波分析理论

小波分析是一种时频分析方法,具有自适应性,可用于处理非线性和非平稳信号。设 $\psi(t)$ 为平方可积函数,即 $\psi(t) \in L^2(\mathbf{R})$,若其傅里叶变换满足小波函数的可容许条件: $C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty$,则称 $\psi(t)$ 为基本小波或母小波函数。小波变换的公式为:

$$WT_f(\alpha, b) = \int_{-\infty}^{+\infty} f(t) \alpha^{-\frac{1}{2}} \psi^* \left(\frac{t-b}{\alpha} \right) dt \quad (1)$$

其中: α 为尺度因子,与频率信息相关; b 为平移因子,与时间信息相关; $\psi^*(t)$ 为小波函数 $\psi(t)$ 的共轭函数。运用快速Mallat算法进行小波分解与重构,Mallat算法包含分解算法和重构算法。分解算法公式为: $c_k^i = \sum_n H(n-2k)c_n^{i-1}$, $d_k^i = \sum_n G(n-2k)d_n^{i-1}$,重构算法公式为: $c_n^{i-1} = \sum_n h(n-2k)c_k^i + \sum_n g(n-2k)d_k^i$,其中: $H(n)$ 为低通滤波器, $h(n)$ 为其对应的冲

击响应; $G(n)$ 为高通滤波器, $g(n)$ 为其对应的冲击响应; c_k^i 为低频信号, d_k^i 为高频信号。

2.1.2 动态GM(1,1)模型理论

设传统静态灰色GM(1,1)模型的原始时间序列 $\{x^{(0)}\}$ 有 n 个观察值,即 $\{x^{(0)}\} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$,通过累加生成新序列 $\{x^{(1)}\} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$,其中: $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $k=1, 2, \dots, n$ 。灰色GM(1,1)模型相应的微分方程是 $\frac{dx^{(1)}}{dt} + ax^{(1)} = u$,其中: a 为发展灰数; u 为内生控制灰数。设 A 为待估参数向量,利用最小二乘法求解,记 $A = [a, u]^T = (B^T B)^{-1} = B^T Y_n$,其中:

$$B = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{bmatrix},$$

$$Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}.$$

对微分方程进行求解,得到 $\{x^{(1)}\}$ 的时间响应方程:

$$\hat{x}^{(1)}(t+1) = \left(x^{(0)}(1) - \frac{a}{u} \right) e^{-at} + \frac{a}{u}, t = (1, 2, \dots) \quad (2)$$

将 $\{x^{(1)}\}$ 序列进行累减运算,得到原始数据序列 $x^{(0)}$ 的预测方程:

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t)$$

$$\hat{x}^{(0)}(t+1) = (1 - e^{-a}) \left(x^{(0)}(t) - \frac{u}{a} \right) e^{-at},$$

$$t = (1, 2, \dots) \quad (3)$$

当 $t < n$ 时,式(3) $\hat{x}^{(0)}(t+1)$ 的值为模型的拟合值;当 $t \geq n$ 时,式(3) $\hat{x}^{(0)}(t+1)$ 的值为模型的预测值。

动态GM(1,1)模型是基于传统静态灰色GM(1,1)模型变换得到的,动态GM(1,1)模型是对建模的基础数逐步进行更新,不断注入新信息,去除冗老旧信息,并且保持建模的原始序列元素个数不变。根据式(3)可得到 $t+1$ 时刻的预测值,为了更精确地预测 $t+2$ 时刻的值 $\hat{x}^{(0)}(t+2)$,本文将原始序列 $\{x^{(0)}\}$ 的元素进行一次更新,将序列 $\{x^{(0)}\}$ 中的第一个元素 $x^{(0)}(1)$ 去掉,剩余的元素向前移一个单

位,最后加入 $\hat{x}^{(0)}(t+1)$ 为第 n 个元素,得到动态的时间序列,并对其建立新的 GM(1,1) 模型,预测得到 $\hat{x}^{(0)}(t+2)$ 。如此往复能动态地预测出未来多期的值。

2.1.3 ARIMA 模型理论

对于随机序列 $\{x_t\}$ 定义模型 ARIMA(p, d, q), 该模型的结构为:

$$\begin{cases} \Phi(\mathbf{B})\nabla^d X_t = \Theta(\mathbf{B})\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t. \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

对 d 阶齐次非平稳序列 $\{X_t\}$ 而言, $\{\nabla^d X_t\}$ 是一个平稳序列, 设其适合 ARMA(p, q) 模型, 即 $\Phi(\mathbf{B})\nabla^d X_t = \Theta(\mathbf{B})\varepsilon_t$, 或者 $\Phi(\mathbf{B})(1 - \mathbf{B})^d X_t = \Theta(\mathbf{B})\varepsilon_t$, 其中:

$$\Phi(\mathbf{B}) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p,$$

$$\Theta(\mathbf{B}) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q.$$

该模型称为求和自回归移动平均模型, 记为 ARIMA(p, d, q), 其中: p 为自回归阶数; d 为差分阶数; q 为移动平均阶数。

2.2 沪深 300 指数的实证分析

2.2.1 沪深 300 指数的小波分析

对数据进行小波分析时, 选取不同的小波基函数, 其信号的时频域局部化能力不同, 因此需选取合适的小波基函数。对沪深 300 指数数据采用多种小波基函数进行实验, 对比分析后本文选取实验结果最佳的 coif3 小波基函数进行尺度为 3 层的小波分解与重构。沪深 300 指数的高频重构系数与低频重构系数重构图, 如图 2 所示。设第三层的低频重构系数为 A3; 第一层到第三层的高频重构系数分别为 D1、D2 和 D3。

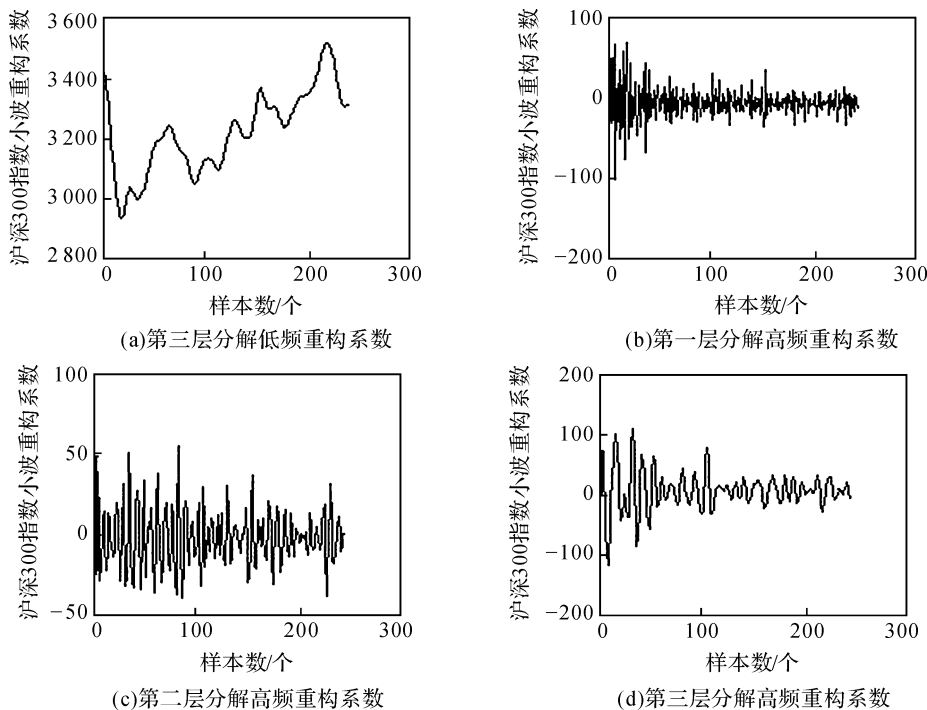


图2 沪深 300 指数的小波重构系数

从图 2 中可以看出, 与沪深 300 指数的原始价格序列图相比, 小波重构数据的日历效应减弱, 重构信号表现出低频数据趋势性明显, 高频数据周期性突出。重构后的信号对原始数据起到平滑去噪的作用。

2.2.2 建立高频数据的动态 GM(1,1) 模型

对小波重构后的细节序列(高频重构系数)D1、D2、D3 分别建立动态 GM(1,1) 模型。首先对序列的非负性进行检验, 如果序列存在负数值, 则对序列加上一个合适的常数 C 进行修正使其变成非负的时

间序列, 再基于传统静态灰色 GM(1,1) 模型, 不断更新建模的原始序列 $\{x^{(0)}\}$, 最后建立动态 GM(1,1) 模型, 并将其初始预测值减去常数 C , 得到最终的预测值 $X(t+i)_1, X(t+i)_2, X(t+i)_3$ 。由于动态 GM(1,1) 模型会对原始序列进行不断更新, 每一个预测值对应的方程都会不同, 则在此列出每个预测值对应的发展灰数 a 与内生控制灰数 u 意义不大, 运用 Matlab 软件可得出三个细节序列动态灰色 GM(1,1) 模型未来 i 期的预测值 $X(t+i)_1, X(t+i)_2, X(t+i)_3$ 。

2.2.3 建立低频数据的 ARIMA 模型

将小波重构后的近似序列作为 ARIMA 模型建模的原始时间序列,首先根据序列的平稳性、序列的自相关和偏自相关系数图的特征,建立合适的 ARIMA 模型,再经过反复筛选,选取最佳的模型,最后确定为 ARIMA(2,1,1),模型的系数见表 1。

表 1 近似序列 ARIMA(2,1,1) 模型参数

模型参数	φ_1	φ_2	θ_1
系数	1.70	-0.76	1.00
系数检验值 p	0.04	0.04	0.01

由表 1 可知,ARIMA(2,1,1) 模型的回归系数检验值 p 均小于模型系数显著性检验时给定的显著性水平值 0.05,说明模型的系数显著不为 0。进一步检验模型的有效性,具体结果如图 3 所示。图 3 中残差序列的 ACF 值在 0 附近,残差序列白噪声检验的 LB 统计量概率检验 p 值为 0,则拒绝残差序列不是白噪声的原假设,得到 ARIMA(2,1,1) 模型的残差序列为白噪声序列,说明该模型可行。

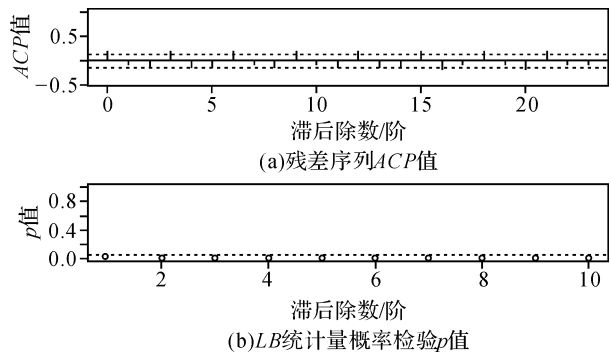


图 3 ARIMA(2,1,1) 模型的检验

2.2.4 沪深 300 指数的预测与模型的比较

沪深 300 指数未来 i 期的收盘价 $C(t+i)$ 可表示为:

$$C(t+i) = X(t+i)_1 + X(t+i)_2 + X(t+i)_3 + Y_{t+i}, i = 1, 2, \dots, 15.$$

将 2016 年 1 月 4 日至 2016 年 12 月 28 日沪深 300 指数 242 个交易日的收盘价数据用于建模,2016 年 12 月 29 日至 2017 年 1 月 19 日的 15 个数据作为模型预测结果的检验数据。为验证基于小波与动态 GM(1,1)-ARIMA 模型在股价预测中的可行性,将其与其它模型的预测结果进行比较,各个模型的预测结果见表 2。

表 2 各个模型预测值

日期	实际值	静态灰色 GM(1,1) 预测值	ARIMA 预测值	动态灰色 GM(1,1) 预测值	小波与动态 GM(1,1) -ARIMA 预测值
2016-12-29	3297.76	3414.66	3298.95	3406.82	3322.88
2016-12-30	3310.08	3416.26	3301.42	3412.66	3326.72
2017-01-03	3342.23	3417.87	3299.60	3419.37	3331.10
2017-01-04	3368.31	3419.47	3300.65	3423.74	3333.34
2017-01-05	3367.79	3421.08	3300.44	3428.94	3335.99
2017-01-06	3347.67	3422.69	3299.82	3432.44	3337.08
2017-01-09	3363.90	3424.30	3301.20	3436.25	3338.20
2017-01-10	3358.27	3425.91	3299.19	3439.45	3338.61
2017-01-11	3334.50	3427.52	3301.66	3443.44	3339.41
2017-01-12	3317.62	3429.13	3298.93	3446.32	3339.32
2017-01-13	3319.91	3430.75	3301.69	3449.36	3339.36
2017-01-16	3319.45	3432.36	3299.11	3453.50	3340.22
2017-01-17	3326.36	3433.97	3301.32	3457.14	3340.76
2017-01-18	3339.37	3435.59	3299.64	3459.73	3340.62
2017-01-19	3325.47	3437.20	3300.67	3462.71	3340.89

由表 2 可知,与传统灰色 GM(1,1)、ARIMA 和单一的动态灰色 GM(1,1) 模型相比,小波与动态 GM(1,1)-ARIMA 模型的预测值更接近沪深 300 指数的实际值,可见小波与动态 GM(1,1)-ARIMA 模型对股价预测的准确性最高。

选用平均预测误差(MFE)、平均绝对百分误差(MAPE)和平均绝对误差(MAD)三个指标度量各模型的预测精度。MFE 能很好地衡量模型无偏性,MFE 趋近于 0,说明预测模型无偏;MAPE 的值小

于 10,说明模型的预测精度高;MAD 能较好地反应预测精度。各误差的计算公式如下:

$$MFE = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)}{n}, MAPE = \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right|,$$

$$MAD = \frac{\sum_{i=1}^n |x_i - \hat{x}_i|}{n}.$$

各个模型的精度比较结果见表 3。

表 3 模型预测误差

模型	传统灰色 GM(1,1)	ARIMA	动态灰色 GM(1,1)	小波与动态 GM(1,1)-ARIMA
MFE	-90.01	35.63	102.21	-0.39
MAPE	0.41	0.16	0.46	0.08
MAD	90.01	35.79	102.21	18.23

由表 3 可知,小波与动态 GM(1,1)-ARIMA 模型的平均绝对误差(MAD)值为 18.23,平均绝对百分误差(MAPE)值为 0.08,平均预测误差(MFE)值为-0.39。在四个模型中,基于小波与动态 GM(1,1)-ARIMA 模型的三个误差值最小,说明该模型预测精度最高。

3 结 语

针对传统 ARIMA 模型和灰色 GM(1,1)对股价预测精度较低的问题,提出基于小波与动态 GM(1,1)-ARIMA 模型的股价预测模型。运用小波对建模数据进行预处理,有利于提高预测模型的精度。选取具有代表性的沪深 300 指数进行实证分析,得出本文所提出的预测模型的精度高于单一灰色 GM(1,1)模型、ARIMA 模型,且其预测结果与实际值吻合程度高,能为投资者投资、决策提供实际参考价值。

参考文献:

- [1] 刘向丽,王旭朋. 基于小波分析的股指期货高频预测研究[J]. 系统工程理论与实践,2015,35(6):1425-1432.
- [2] 肖燕君,张华,任若恩. 基于小波多尺度分析的股票价格组合预测方法[J]. 工业工程,2011,14(6):133-137.
- [3] LAHMIRI S. Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks[J]. Journal of King Saud University-Computer and Information Sciences, 2013, 26(2): 218-227.
- [4] JAMMAZI R, ALOUI C. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling[J]. Energy Economics, 2012, 34(3):828-841.
- [5] HSIEH T J, HSIAO H F, YEH W C. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm [J]. Applied Soft Computing, 2011, 11(2):2510-2525.
- [6] 孙冰洁,唐瑞,左毅. 小波分析下的神经网络股票预测研究[J]. 计算机与数字工程,2016,44(6):1031-1034.
- [7] 石鸿雁,尤作军,陈忠菊. 基于小波分析的 ARIMA 模型对上证指数分析与预测[J]. 数学的实践与认识, 2014, 44(23):66-72.
- [8] 张宇敬,李倩,蔡虎. 基于灰色预测模型的股价预测研究[J]. 金融教学与研究,2013(6):50-52.
- [9] 骆桦,刘兴. 基于主成分分析的神经网络算法对期权价格预测研究[J]. 浙江理工大学学报,2017,37(1):122-126.
- [10] KARA Y, MELEK A B, BAYKANCÖK. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange [J]. Expert Systems with Applications, 2011, 38(5):5311-5319.

The Research of Stock Price Prediction Based on the Wavelet Dynamic GM (1,1)-ARIMA Model

LUO Hua, CHEN Yanfei

(School of Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Aiming at the time lag of the traditional time series stock price forecasting model, a stock price forecasting model based on wavelet and dynamic GM (1,1)-ARIMA model is proposed. The data were pretreated by wavelet analysis. Based on wavelet reconstruction sequences, the ARIMA model was established and the dynamic GM(1,1) model considering the influence of future factors on the system was established. We do the empirical analysis on Shanghai and shenzhen 300 index, the results show that compared with the traditional stock price forecasting model, the proposed GM (1,1)-ARIMA model has the highest prediction accuracy.

Key words: stock price; wavelet analysis; dynamic GM (1,1); ARIMA model

(责任编辑:康 锋)