doi:10.3969/j.issn.1673-4785.2011.03.014

小波变换和GM-ARMA组合模型的股指预测

吴朝阳
(康考迪亚大学 统计与数学系,蒙特利尔 H3G 2H9)

摘　要:当前在利用小波分解和其他模型建立组合模型的过程中,对小波基方程的选择和分解层数并没有一个标准,基本上是通过经验和一些实验来决定这2个因素;而且很多利用小波分解建立的组合模型并不考虑模型之间相互的影响,对各个子模型的参数估计采取各自独立的估计,从而导致预测结果不是最优. 为此,提出了先对小波基方程和分解层数这2个特征进行参数化,然后定量地对所有子模型的特征参数进行统一、综合的评估,以达到建立最佳组合模型的目的. 由于该组合模型是由小波分解、灰色模型和ARMA模型组合而成的,因此称为WGM-ARMA模型. 股指预测的实例验证了WGM-ARMA模型大幅度地降低了预测误差,说明了该组合模型的有效性、实用性和可行性.

关键词:小波分解;灰色模型;ARMA模型;GM-ARMA模型;股指预测
中图分类号:TP18　**文献标识码**:A　**文章编号**:1673-4785(2011)03-0279-04

Using wavelet transformation and a GM-ARMA model to forecast stock index

WU Zhaoyang
(The Department of Mathematics and Statistics, Concordia University, Montreal H3G 2H9, Canada)

Abstract: During the process of building a hybrid model by combining wavelet decomposition and other techniques, there is no standard in terms of selecting a wavelet base function and decomposition level. The commonly used ways are usually based on the researcher's experience or several experiments instead of a quantitative approach. In addition, many hybrid models based on wavelet decomposition do not consider the interaction between sub models. Instead of estimating the parameters in all sub models as the whole, they estimate the parameters separately, which lead to that the prediction result is not optimal. In order to solve this problem, this paper first introduced two new parameters, wavelet functions and decomposition levels, then quantitatively estimated all the parameters as a whole for the purpose of building an optimal hybrid model. For convenience, the model was called the WGM-ARMA model because it combines the wavelet decomposition, grey model, as well as autoregressive integrated moving average (ARMA) model. Experimental results show that the hybrid model significantly reduces prediction errors. As a result, it can be concluded that the model in terms of forecasting stock index is valid and useful, along with the method used to construct the optimal hybrid model.

Keywords: wavelet decomposition; grey model; ARMA model; GM-ARMA model; stock prediction

ARIMA模型(autoregressive integrated moving average model)和GM(1,1)模型作为应用广泛的时间序列模型,长期以来被许多学者用于股票价格序列的研究中[1-5]. 由于这2种模型对于时间序列的预测各有侧重,因此一些学者提出了整合这2个模型的组合模型并称之为GM-ARMA模型(grey model and autoregressive integrated moving average model)[6]. 但是GM-ARMA模型并不是一个最优的模型,因为子模型GM(1,1)模型没有经过优化,同时2个子模型在结合时,也没有考虑进行最佳的整合. 为此笔者提出了改进的GM-ARMA模型以克服上述缺点并称为RGM-ARMA(revised GM-ARMA)模型[7],实例证明RGM-ARMA的预测误差小于单一模型和GM-ARMA模型.

但是RGM-ARMA模型在预测误差上还是偏

收稿日期:2010-09-18.
通信作者:吴朝阳. E-mail: hostingca@gmail.com.

大,特别是方向预测误差超过了50%.从RGM-ARMA模型的建模方法上可以看出,继续在GM(1,1)模型和ARMA模型的组合上下功夫已经不太可能大幅度地降低预测误差,有必要引进新的建模技术到RMG-ARMA模型中来降低预测误差,其中一个可以选择的技术就是小波分解.小波分解近年来开始被用于结合其他模型进行股价时间序列的预测[8-9],其基于的经济学背景来源于Pepters,他认为股票市场的投资者可以分为长期投资者、中期投资者和短期投资者,不同投资者由于投资习惯的不同,所造成的统计特征也不同[10].基于以上分析,一个可以尝试的办法就是把股票时间序列分解成不同频率的子时间序列,并对不同的子序列采用合适的模型进行建模.本文对于代表趋势的低频序列,用灰色模型进行预测,因为灰色模型具有很好的动态捕捉趋势的能力;对于代表波动的高频序列,采用ARMA模型进行预测,因为可以猜想具有相同投资习惯的投资者的投资习惯在某种程度上是自相关的,其特征应该基本类似并保持基本平稳.

至此,通过研究股市的特点,结合所需子模型的特点,本文最终决定引进小波分解技术来尝试降低预测误差.具体的思路为:首先用小波分解对原始股价序列进行分解,然后针对分解后的子序列的特点选择合适的模型,这里采用GM-ARMA模型和单个的ARMA模型,对子序列进行建模并预测,最后把各自子序列的预测值合成为股价的最终预测值.为阐述方便起见,称该组合模型为WGM-ARMA模型(wavelet transform, grey model and autoregressive integrated moving average model).

1　WGM-ARMA模型

在WGM-ARMA模型中,需要整合3种模型技术,这里模型的整合优化是建立WGM-ARMA模型的关键.在小波分解中,小波基方程的选择和分解层数对预测的影响很大,但是对小波基方程的选择和分解层次并没有一个标准,基本上是通过经验和一些实验来决定这2个因素.而且,很多利用小波分解建立的组合模型并不考虑模型之间相互的影响,各个子模型的参数估计是独立的,而不是互相制约和影响的,这使得组合模型并不是有机整合在一起的,从而导致预测结果不是最优.为了有机地整合小波分解、灰色模型和ARMA模型,有必要对小波基和分解层数参数化.这里用参数m代表小波基,参数k代表分解层数.

通过对小波分解的2个主要指标参数化后,就可以参考RGM-ARMA模型的建模思路了.在灰色模型GM$(1,1,\mu,v)$中的特征参数是(μ,v),在ARMA模型的特征参数是(p,q),在小波分解的特征参数是(m,k),而整合灰色模型、ARMA模型和小波分解的过程就是找到特征参数组合(μ,v,p,q,m,k)最佳组合的过程.同建立RGM-ARMA模型一样,其基本的前提条件是(μ,v)、(p,q)和(m,k)的选择必须基于相同的统计准则.显然参数(m,k)可以用TAE准则[7]选择,因此模型的整合可以依据TAE准则进行选择.

具体来说,构造WGM-ARMA模型的思想如下.

对于给定的时间序列$X=(x_1,x_2,\cdots,x_n)$,已知小波基方程m和小波分解的层数k,根据小波分解,该时间序列X可以被分解为

$$X = A_0^m + D_1^m + D_2^m + \cdots + D_k^m.$$

式中:A_0^m和$D_1^m,D_2^m,\cdots,D_k^m$分别代表时间序列通过小波基方程m,分解k层后得到的低频序列和高频序列集.A_0^m可以看作是序列X的近似,代表X的大致趋势,其特点适合用RGM-ARMA模型进行拟合;$D_1^m,D_2^m,\cdots,D_k^m$是相对高频的时间序列集,其特点为没有趋势值并有近似为0的均值和相对稳定的方差,适合用ARMA模型来拟合.这样序列X的模型就为

$$\begin{aligned}X &= \hat{X} + \varepsilon = \hat{A}_0^m + \hat{D}_1^m + \hat{D}_2^m + \cdots + \hat{D}_k^m + \varepsilon = \\ &[\mathrm{GM}(1,1,\mu,v) + \mathrm{ARMA}(p^0,q^0)]_0^m + \\ &[\mathrm{ARMA}(p^1,q^1)]_1^m + [\mathrm{ARMA}(p^2,q^2)]_2^m + \cdots + \\ &\mathrm{ARMA}(p^k,q^k)]_k^m + \varepsilon.\end{aligned}\tag{1}$$

式中:$\hat{X}$是X的预测拟合序列;$\hat{A}_0^m$是A_0^m的预测拟合序列;$\hat{D}_1^m,\hat{D}_2^m,\cdots,\hat{D}_k^m$是$D_1^m,D_2^m,\cdots,D_k^m$的预测拟合序列;$\varepsilon$是WGM-ARMA模型的残差序列.

为了找到最佳的参数组合,需要对参数$(m,k,\mu,v,p^0,q^0,p^1,q^1,\cdots,p^k,q^k)$设定上下限,并对连续参数进行相应的离散化处理,最终得到一个离散的参数空间:

$$\begin{gathered}\mu \in [l,L],v \in [r,R],p^0 \in [0,P^0],\\ q^0 \in [0,Q^0],\cdots,p^k \in [0,P^k],q^k \in [0,Q^k],\\ m \in [0,M],k \in [0,K].\end{gathered}\tag{2}$$

这样对于每个参数组合$(m,k,\mu,v,p^0,q^0,p^1,q^1,\cdots,p^k,q^k)$,都可以通过式(1)建立模型,用该模型计算出拟合序列$\hat{X}=(\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_n)$,并计算出对应的总绝对值误差$\sigma_{\mathrm{TAE}}$(total absolute error)为

$$\sigma_{\mathrm{TAE}} = \sum_{j=1}^{n} | x_j - \hat{x}_j |.\tag{3}$$

而最终的WGM-ARMA模型满足:

$$\begin{gathered}\min(\sigma_{\mathrm{TAE}});\\ \mu \in [l,L],v \in [r,R],p^0 \in [0,P^0],\\ q^0 \in [0,Q^0],\cdots,p^k \in [0,P^k],q^k \in [0,Q^k],\\ m \in [0,M],k \in [0,K].\end{gathered}\tag{4}$$

2　实例研究

用实例说明 WGM-ARMA 模型的建模过程,并用 WGM-ARMA 模型来进行股指的预测. 本节所用数据及策略和文献[7]完全一样,总数据一共为 152 个,开始的 126 个用于模型的建立,随后的 26 个用于模型的评测.

首先对参数($m,k,\mu,v,p^0,q^0,p^1,q^1,\cdots,p^k,q^k$)设定上下限. 对于代表小波基方程的参数 m,由于用 Matlab 进行编程,因此对 Matlab 提供的小波基方程进行如下编号,编号列在了表 1 中.

表 1　小波基方程的编号

Table 1　Wavelet base function number

1	2	…	11	12	…	18	19
Haar	Db1	…	Db10	Sym2	…	Sym8	Coif1
…	23	24	…	38	39	…	53
…	Coif5	Bior1.1	…	Bior6.8	Rbio1.1	…	Rbio6.8

这里可以根据需要随时添加新的小波基方程. 对于分解层数 k,由于随着分解层数的增加,计算量也会大幅度地增加,因此设 $k=2$. 对于特征参数组合($\mu,v,p^0,q^0,p^1,q^1,p^2,q^2$),采用和 RGM-ARMA 模型的设置一样,为

$$\mu \in [0,0.1,\cdots,1], v \in [5,6,\cdots,9],$$
$$p^0,q^0,p^1,q^1,p^2,q^2 \in [0,3].$$

到此,特征参数空间就定义完毕,对于每个参数组合($m,k,\mu,v,p^0,q^0,p^1,q^1,p^2,q^2$)都可以基于 126 个历史数据 $X=(x_1,x_2,\cdots,x_{126})$ 通过式(1)建立模型,并得到拟合序列 $\hat{X}$ 和残差序列 ε,并通过式(3)计算出相应的 σ_{TAE}. 并且 WGM-ARMA 模型满足式(4).

对于本例的建模数据 $X=(x_1,x_2,\cdots,x_{126})$ 和任意的一个参数组合($k=1,m=47,\mu=0.6,v=7,p^0=1,q^0=0,p^1=2,q^1=3$),其具体的 σ_{TAE} 计算过程如下.

由于 $m=9,k=1$,说明首先要用小波基方程“Db8”,通过小波变换把原始序列 X 分解 1 层,得到低频序列 A_9^0 和高频序列 D_9^1,为了方便书写,让 $A_9^0=(a_1,a_2,\cdots,a_{126})$,$D_9^1=(d_1,d_2,\cdots,d_{126})$. 由于 $u=0.6,v=7,p^0=1,q^0=0$,因此对于低频序列 A_9^0 用参数($\mu=0.6,v=7,p^0=1,q^0=0$)构建 GM-ARMA 模型,并得到其拟合序列 $\hat{A}_9^0=(\hat{a}_1,\hat{a}_2,\cdots,\hat{a}_{126})$,关于建模过程请参考文献[7]. 对于高频序列 D_9^1,由于(p^1,q^1)=(2,3),说明需要对 D_9^1 构建 ARMA(2,3)模型,得到其拟合序列 $\hat{D}_9^1=(\hat{d}_1,\hat{d}_2,\cdots,\hat{d}_{126})$.

这样原始序列 X 基于式(1)建立的模型的拟合序列就为 $\hat{X}=(\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_{126})$,其中:

$$\hat{x}_j = \hat{a}_j + \hat{d}_j;$$

总绝对值误差 σ_{TAE} 为

$$\sigma_{\mathrm{TAE}} = \sum_{j=1}^{126} | x_j - \hat{x}_j |.$$

对于每个参数组合都可以利用式(1)建立相应的模型并求出 σ_{TAE},其中最小的 σ_{TAE} 的参数组合就是最优的参数组合,对应的模型就是所要求的 WGM-ARMA 模型.

以上算法可以用 Matlab 编程实现,最终计算出来的最小 σ_{TAE} 的参数组合为

$$k=1, m=47, \mu=1, v=6,$$
$$p^0=3, q^0=2, p^1=3, q^1=3.$$

该参数说明,首先要用小波基方程“Rbio3.3”($m=47$),通过小波分解把原始序列 X 分解 1 层($k=1$),得到低频序列 $A^0=(a_1^0,a_2^0,\cdots,a_{126}^0)$ 和高频序列 $D^1=(d_1^1,d_2^1,\cdots,d_{126}^1)$,对于低频序列 A^0 构建 GM-ARMA 模型($\mu=1,v=6,p^0=3,q^0=2$),对于高频序列 D^1 构建 ARMA(3,3)模型($p^1=3,q^1=3$). 这样序列 X 的拟合序列 $\hat{X}$ 可以表示为

$$\hat{X} = \hat{A}^0 + \hat{D}^1 = [\mathrm{GM}(1,1,1,6) + \mathrm{ARMA}(2,3)] + [\mathrm{ARMA}(3,3)].$$

效仿上面的例子,可以得到低频序列 A^0 在第 127 点上的预测值 $\hat{a}_{127}$ 和高频序列 D^1 在第 127 点上的预测值 $\hat{d}_{127}$,并由此得到序列 X 在 127 个点的预测值为

$$\hat{x}_{127} = \hat{a}_{127} + \hat{d}_{127} = 9\,248.1.$$

式中:点 $\hat{a}_{127}$ 的计算公式为

$$\hat{a}_{127} = (1-\mathrm{e}^{a}) \cdot \left(a_{121} - \frac{b}{a}\right)\mathrm{e}^{-a(7-1)} + \sum_{k=1}^{3} \varphi_k^0 (a_{123+k} - \hat{a}_{123+k}) + \tau_{127}^0 + \sum_{k=1}^{3} \theta_k^0 \tau_{123+k}^0;$$

点 $\hat{d}_{127}$ 的计算公式为

$$\sum_{k=1}^{3} \varphi_k^0 (a_{123+k} - \hat{a}_{123+k}) + \tau_{127}^0 + \sum_{k=1}^{3} \theta_k^0 \tau_{123+k}^0.$$

因此,根据 WGM-ARMA 模型,TSX 指数第 127 个点的预测值,也就是 2009 年 1 月 2 日的日线收盘价的预测值为 9 248.1. 对于第 128 个数据,要用开始的 127 个数据建立 WGM-ARMA 模型来预测,以此类推,可以得到从 2009 年 1 月 2 日—2009 年 2 月 6 日的全部 26 个预测数据. 并计算出平均绝对值误差 σ_{MAPE}(mean absolute percent error)为

$$\sigma_{\mathrm{MAPE}} = \frac{1}{26}\sum_{j=127}^{152}\left|\frac{x_j - \hat{x}_j}{x_j}\right|100\% = 1.36\%.$$

由于股市判断方向也重要,这里也计算出了方向错误率 σ_{DIR}(directional errors)为

$$\sigma_{\mathrm{DIR}} = \frac{1}{12}\sum_{j=127}^{152} d_j \cdot 100\% = 42.31\%.$$

式中:

$$d_j = \begin{cases} 0, & (x_j - x_{j-1})(\hat{x}_j - \hat{x}_{j-1}) > 0; \\ 1, & 其他. \end{cases}$$

为了比较 WGM-ARMA 模型和其他模型,这里也用相同数据计算了其他模型的误差率,表 2 为这些模型的误差比较结果.

表 2 采用不同模型预测股指的误差

Table 2 Prediction errors based on different models

模 型	σ_{TAE}	$\sigma_{\mathrm{MAPE}}/10^3$	$\sigma_{\mathrm{DIR}}/10^3$
ARIMA	4 500.15	0.195	5.769
GM(1,1,μ,v)	3 990.67	0.173	5.769
RGM-ARMA	3 852.97	0.169	5.384
WGM-ARMA	3 122.09	0.136	4.231

从表 2 可以看出,WGM-ARMA 模型明显优于单个的 ARIMA 模型、GM(1,1,μ,v)模型,也优于 RGM-ARMA 模型.特别是方向误差低于 50%,这说明了引进小波分解到 RGM-ARAM 模型中是一个很好的办法,同时也说明了直接在原始序列上建模不能发挥出灰色模型和 ARMA 模型的优势,必须排除干扰后才能将这些模型的优势发挥出来,以达到减少预测误差的目的.

3 结束语

本文在分析股票运动特征的基础上,提出了基于小波分解、灰色模型和 ARMA 模型的组合模型,实例证明该模型在各个指标上都明显优于本文中用到的单个和其他的组合模型.更为重要的是,本文提出了一个全新的构建组合模型的思路,利用这个思路,可以对其他基于小波分解的组合模型进行有效的优化,进而构建出更多更精确的预测模型.尽管 WGM-ARMA 模型的初衷是用来预测股票,但是 WGM-ARMA 模型也可以用于其他时间序列的预测,例如 GDP 和销售量的预测.

参考文献:

[1]朱宁,徐标,仝殿波.上证指数的时间序列预测模型[J].桂林电子工业学院学报,2006,26(2):124-128.
ZHU Ning, XU Biao, TONG Dianbo. Time-series prediction model of Shanghai composite index[J]. Journal of Guilin University of Electronic Technology, 2006, 26(2): 124-128.

[2]POTERBA J M, SUMMERS L H. The persistence of volatility and stock market fluctuations[J]. American Economic Review, 1986, 76(5): 1143-1151.

[3]FRENCH K R, SCHWERT G W, STAMBAUGH R F. Expected stock returns and volatility[J]. Journal of Financial Economics, 1987, 19(1): 3-29.

[4]李国平,于广青,陈森发.中国股票价格灰色预测研究综述[J].东南大学学报:哲学社会科学版,2005,7(2):28-30,126.
LI Guoping, YU Guangqing, CHEN Senfa. A review of research on stock price gray forecast in China[J]. Journal of Southeast University: Philosophy and Social Science, 2005, 7(2): 28-30, 126.

[5]郭宁,向凤红.灰色理论和神经元网络在证券市场中的应用[J].自动化技术与应用,2008,27(10):1-3.
GUO Ning, XIANG Fenghong. Application of grey model and neural network in the stock-market[J]. Techniques of Automation and Applications, 2008, 27(10): 1-3.

[6]吴庚申,梁平,龙新峰.基于 GM-ARMA 的年电力负荷组合模型[J].湖北电力,2005,29(2):21-23.
WU Gengshen, LIANG Ping, LONG Xinfeng. An annual electric consumption combined model based on GM-ARMA [J]. Hubei Electric Power, 2005, 29(2): 21-23.

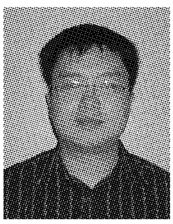
[7]吴朝阳.改进的灰色模型与 ARMA 模型的股指预测[J].智能系统学报,2010,5(3):277-281.
WU Zhaoyang. Forecasting stock index based on revised grey model and ARMA model[J]. CAAI Transactions on Intelligent Systems, 2010, 5(3): 277-281.

[8] 郑纪安.基于小波分析和神经网络的金融时间序列预测研究[D].厦门:厦门大学,2009:45-61.
ZHENG Ji'an. The research on forecasting of financial time series baesd on wavelet analysis and neural network[D]. Xiamen: Xianmen University, 2009: 45-61.

[9] 曲文龙,李海燕,刘永伟,等.基于小波和支持向量机的多尺度时间序列预测[J].计算机工程与应用,2007,43(29):182-185.
QU Wenlong, LI Haiyan, LIU Yongwei, et al. Research on multi-scale prediction of time series based on wavelet and support vector machines[J]. Computer Engineering and Applications, 2007, 43(29): 182-185.

[10]PETERS E E. Fractal market analysis: applying chaos theory to investment and economics[M]. New York, USA: John Wiley and Sons Inc, 1996: 39-50.

作者简介:



吴朝阳,男,1975 年生,工程师,主要研究方向为灰色系统理论、时间序列和小波变换的股票预测.