

特征滞后计算的股市波动预测

姚宏亮, 李大光*, 李俊照

(合肥工业大学 计算机与信息学院, 合肥 230009)

(*通信作者电子邮箱 1065442573@qq.com)

摘要:针对股票价格波动拐点难以有效预测导致预测精度不高的问题,提出一种特征滞后程度计算的均值门限广义自回归条件异方差(LRD-TGARCH-M)模型。首先,基于股价波动与指标变化出现的 inconsistency,给出了滞后性的定义,并引入能量波动概念,从能量角度提出特征滞后程度(LD)计算模型;然后,用LD度量拐点出现之前的风险大小,将其加入到股价均值方程中,克服均值门限广义自回归条件异方差(TGARCH-M)模型对拐点预测的不足;其次,根据拐点附近波动相对剧烈,将LD加入到误差项的方差方程中,优化方差的变化,提高模型的预测精度;最后,给出了LRD-TGARCH-M模型的波动预测公式和精度分析,并在股票数据上进行实验,结果表明,与TGARCH-M模型相比,精确度提高了3.76%;与均值指数GARCH(EGARCH-M)模型相比,精确度提高了3.44%,证明了LRD-TGARCH-M模型可以提高股价走势预测精度,减小误差。

关键词:股价波动;特征滞后;能量性;波动风险;门限广义自回归条件异方差模型

中图分类号: TP399; TP182 **文献标志码:** A

Stock market volatility forecast based on calculation of characteristic hysteresis

YAO Hongliang, LI Daguang*, LI Junzhao

(School of Computer and Information, Hefei University of Technology, Hefei Anhui 230009, China)

Abstract: Focusing on the issue that the inflection points are hard to forecast in stock price volatility degrades the forecast accuracy, a kind of Lag Risk Degree Threshold Generalized Autoregressive Conditional Heteroscedastic in Mean (LRD-TGARCH-M) model was proposed. Firstly, hysteresis was defined based on the inconsistency phenomenon of stock price volatility and index volatility, and the Lag Degree (LD) calculation model was proposed through the energy volatility of the stock. Then the LD was used to measure the risk, and put into the average share price equation in order to overcome the Threshold Generalized Autoregressive Conditional Heteroscedastic in Mean (TGARCH-M) model's deficiency for predicting inflection points. Then the LD was put into the variance equation according to the drastic volatility near the inflection points, for the purpose of optimizing the change of variance and improving the forecast accuracy. Finally, the volatility forecasting formulas and accuracy analysis of the LRD-TGARCH-M algorithm were given out. The experimental results from Shanghai Stock, show that the forecast accuracy increases by 3.76% compared with the TGARCH-M model and by 3.44% compared with the Exponential Generalized Autoregressive Conditional Heteroscedastic in Mean (EGARCH-M) model, which proves the LRD-TGARCH-M model can degrade the errors in the price volatility forecast.

Key words: price volatility; characteristic hysteresis; energy characteristic; fluctuation risk; Threshold Generalized Autoregressive Conditional Heteroscedastic in Mean (TGARCH-M) model

0 引言

股票市场波动性一直是研究的热点,早期主要有自回归(Auto Regressive, AR)模型、自回归滑动平均(Auto-Regressive and Moving Average, ARMA)模型以及自回归积分滑动平均(Autoregressive Integrated Moving Average, ARIMA)模型等应用于时间序列预测^[1],但是这些模型主要运用于单变量、同方差场合的线性模型,有很大的局限性。Engle^[2]和Bollerslev^[3]对传统时间序列模型进行改进,分别提出了自回归条件异方差(Autoregressive Conditional Heteroskedasticity, ARCH)模型和广义的ARCH(Generalized ARCH, GARCH)模型应用于金融时间序列的波动性建模,取得了很大的突破。

但GARCH模型假定条件方差是滞后残差平方的函数,故残差的符号不影响波动,因此不能解释股票收益和收益变化波动之间出现的负相关现象。

近期,研究者们对GARCH模型进行了扩展,Nelson^[4]提出的指数GARCH(Exponential GARCH, EGARCH)模型,通过加入解释非对称性的附加项来解释杠杆效应;Glosten等提出的门限GARCH(Threshold GARCH, TGARCH)模型能够捕捉波动率的非对称性影响;紧接着Glosten等^[5]又提出把条件方差作为直接风险因素引进到均值方程中扩展为均值GARCH(GARCH in Mean, GARCH-M)模型;Ke等^[6]对比了GARCH-M模型、TGARCH模型和EGARCH模型,用GARCH-M模型测试了长期波动风险和回报之间的相关性,用TGARCH模型

收稿日期:2015-02-10;修回日期:2015-03-29。 基金项目:国家自然科学基金资助项目(61175051,61070131,61175033)。

作者简介:姚宏亮(1972-),男,安徽桐城人,副教授,博士,主要研究方向:人工智能、知识工程;李大光(1989-),男,安徽明光人,硕士研究生,主要研究方向:人工智能、知识工程;李俊照(1975-),男,安徽桐城人,讲师,博士研究生,主要研究方向:机器学习、人工智能。

和 EGARCH 模型模拟了股价波动的杠杆效应^[7-8],验证了股票价格波动的不对称性异方差性。文献[9-10]提出了风险价值和波动率是密切相关的。这些模型都可以描述出股票市场波动率的时变性、波动聚集性以及非对称性^[11-12],能在一定程度上描述股价的波动,但是股价波动是频繁且剧烈的,波动的拐点频繁出现,这些模型都没有针对拐点深入考虑,导致对拐点预测准确度和精度都不高,影响了整体的预测效果。

股价波动的拐点是股价经过一段时间同一走势(如上涨或下跌)之后,发生波动走势转变的点。在拐点出现之前,股价一直处于某种走势的波动变化过程,在此过程中,针对当前波动走势,会累积一定的风险性,进而引发波动的转势,因此对拐点之前股价波动变化的风险性进行计算研究,对拐点的预测有一定的参考意义。

从股价波动与指标波动不一致性现象出发,提出一种特征滞后程度计算的广义自回归条件异方差模型。首先,给出了滞后性的定义,结合能量特性提出特征滞后程度(Lag Degree, LD)的计算模型,将特征滞后程度作为当前走势风险情况,引入到股价均值方程中,引导启发模型对拐点的判断和预测;同时,由于拐点处波动变化比较大导致拐点的预测精度较低,将特征滞后加入到随机误差项的方差方程中,优化方差的变化,并给出了将特征滞后程度作为风险因素的均值 TGARCH(Lag Risk Degree TGARCH in Mean, LRD-TGARCH-M)模型的波动预测公式和精度分析,最后在股票上证指数数据上进行了实验验证。

1 滞后性

指标是和股价密切相关的要素^[13],理论上指标的变化是随着股价的波动而同步变化,然而现实中,经常出现不同步,尤其是上涨和下跌的末期,股价走势会慢于指标的变化,这种现象就是滞后性,因此,可以通过滞后性的程度来衡量当前的风险性大小。

对于目标股票 S , 设其数据(Stock Data, SD) 为一个三元组 $SD = \{Y, R, I\}$, 其中: Y, R, I 分别表示股票收盘价、涨跌幅和指标数据。

针对这种滞后性, 本文给出如下定义。

定义1 股价走势。为了方便体现出股价波动变化的走势, 选取股价的3日移动平均值作平滑处理, 设 Y_t 为 t 时股价的3日移动平均值, 涨跌幅 $R_t = (Y_t - Y_{t-1})/Y_{t-1}$, 若连续 $n(n \geq 3)$ 日出现 $|R_t| \geq \varepsilon$ (ε 可设为给定的正值) 则称股价处于上涨或下跌走势, 否则称为股价横盘整理。

类似地, 对于任一指标 I_t 为 t 时的3日移动平均值, 若连续 $n(n \geq 3)$ 日出现 $|I_t - I_{t-1}| \geq \varepsilon'$ (ε' 可设为给定的正值) 则指标向上或向下运动; 否则指标处于水平移动。

定义2 滞后性。设有股票数据三元组 $SD = \{Y, R, I\}$, 其中: Y 为股票收盘价, R 为股票涨跌幅, I 为股票的指标值。对于收盘价 Y_t 、涨跌幅 R_t 、指标 I_t , 若连续 $N(N \geq 3)$ 日出现 $R_t \geq \theta$, 且 $I_t - I_{t-1} < \theta'$ 或者连续 N 日出现 $R_t \leq -\theta$, 且 $I_t - I_{t-1} > -\theta'$, 则股价变化落后于指标 I 的变化, 称为滞后性。

定义3 序列窗口(Series Window, SW)。股票数据三元组 $SD = \{Y, R, I\}$, 设 $SW = \{Y_i, Y_{i+1}, \dots, Y_{i+H-1}; R_i, R_{i+1}, \dots, R_{i+H-1}; I_i, I_{i+1}, \dots, I_{i+H-1}\}$ 是由 H 组连续的不同时间序列数据组成, 其中: i 是窗口开始的时间戳, $i+H-1$ 是窗口结束的时间戳, $SW \subseteq SD$, 窗口大小为 H 。

定义4 特征滞后程度(LD)。在步长为 h 的序列窗口时间内, 用 $LD_{h,I}$ 表示在窗口中因股价相对于指标 I 变化滞后的程度。设从 i 时关注股价走势, 在 $i+m$ 时出现滞后, 在 $i+m+n$ 时滞后结束, 则特征滞后程度为 $LD_{h,I} = \sum_{j=i+m}^{i+m+n} f(R_j, I_j) = \sum_{j=i+m}^{i+m+n} (R_j / e^{f_{h,i}})$, 其中 I 为出现滞后性的指标。当股价走势发生转变时, $LD_{h,I}$ 清零。

2 指标特征的能量性和滞后性

股票波动的内在能量可以通过指标的状态体现, 股价波动与内在能量的一致性, 可表现为股价波动趋势落后于指标波动趋势。

2.1 指标特征的能量性

定义5 多空比率净额 VA 。设股价收盘价 Y , 当天最高价 Y_{\max} , 当天最低价 Y_{\min} , 成交量 V , 则 $VA = ((Y - Y_{\min}) - (Y_{\max} - Y)) / (Y_{\max} - Y_{\min}) * V$ 。

成交量(Volume, VOL) 是体现股票波动内在能量的一个主要要素, 因此直接取 VA 的变化作为成交量的能量体现, 即 $E_V = (VA_t - VA_{t-1}) / VA_{t-1}$ 。

根据指标特征的变化与股价变化的关系, 本文给出指数平滑异同平均线(Moving Average Convergence/Divergence, MACD), 随机指标(KDJ 指标)的能量计算公式如下。

1) MACD 指标的能量计算。

这里的 MACD 指标, 只考虑 MACD 的能量柱和 DIF 线。设 t 时, MACD 能量柱 M_t , MACD 白线 DIF_t , 多空比率净额 VA_t , 则 t 时 MACD 指标的能量为 E_{M_t} 。

当 $M_t \geq M_{t-1}$ 时:

$$E_{M_t} = (300 - DIF_t) / 300 \times e^{(M_t - M_{\max}) / M_{\max}} \times (M_t - M_{t-1}) \times |DIF_t - M_t| \times |(VA_t - VA_{t-1}) / VA_{t-1}| \quad (1)$$

当 $M_t < M_{t-1}$ 时:

$$E_{M_t} = (300 + DIF_t) / 300 \times (2 \arctan(i) + e^{(M_t - M_{\max}) / M_{\max}}) \times (M_t - M_{t-1}) \times |DIF_t - M_t| \times |(VA_t - VA_{t-1}) / VA_{t-1}| \quad (2)$$

其中: i 表示能量柱的第 i 轮向下或向上运动; M_{\max} 表示上一轮运动能量柱的最大值或最小值, 且当 $i = 1$ 时, $M_{\max} = M_t$ 。

2) KDJ 指标的能量计算。

KDJ 指标由 K、D、J 线组成, 设 t 时, KDJ 指标的 K、D、J 线数值 K_t, D_t, J_t , 多空比率净额 VA_t , 则 t 时 KDJ 指标的能量为 E_{K_t} 。

当 $K_t \geq D_t$ 时:

$$E_{K_t} = (100 - K_t) / 100 \times |J_t - J_{t-1}| / J_{t-1} \times (K_t - D_t) \times |(VA_t - VA_{t-1}) / VA_{t-1}| \quad (3)$$

当 $K_t < D_t$ 时:

$$E_{K_t} = (100 + K_t) / 100 \times |J_t - J_{t-1}| / J_{t-1} \times (K_t - D_t) \times |(VA_t - VA_{t-1}) / VA_{t-1}| \quad (4)$$

2.2 指标特征的滞后性

定义6 量价滞后。成交量与股价变化之间的不一致, 表现为有价无量, 或有量无价。设 t 时, 股价涨跌幅 R_t , 成交量 V_t , 最近5日的平均成交量 V_t' , 若 $|R_t| \geq \mu$, 且 $V_t / V_t' < \sigma$, 则表示有价无量; 若 $|R_t| < \mu'$, 且 $V_t / V_t' \geq \sigma'$, 表示有量无价; 都表示股价变化与量能不一致。

定义7 同比放量滞后。相同几波走势之间量能的不一,表现为相同股价处量能不足。设 t 时,股票收盘价 Y_t ,3日平均成交量 V_t , (Y_t, V_t) 是上一相同走势的相近股价范围内的股票收盘价和3日平均成交量,若 $Y_t - Y_i < \omega$,且 $V_t/V_i < 1$,则为同比放量滞后,量能不足。

定义8 MACD指标滞后。MACD能量柱的变化与股价变化的不一致,表现为股价上涨或下跌中,MACD的能量柱开始变短。设 t 时,股价涨跌幅 R_t ,能量柱 M_t ,若连续 $s(s \geq 3)$ 日出现 $R_t \geq 0$,且 $M_t - M_{t-1} < 0$,为上涨过程MACD指标滞后;若连续 s 日出现 $R_t < 0$,且 $M_t - M_{t-1} \geq 0$,为下跌过程MACD指标滞后。

定义9 KDJ指标滞后。KDJ指标J线的变化与股价变化的不一致,表现为J线不随股价上涨而向上或不随股价下跌而向下。设 t 时,股价涨跌幅 R_t ,KDJ指标的K、D、J线 K_t, D_t, J_t ,若连续 $s'(s' \geq 3)$ 日出现 $R_t \geq 0$ 且 $(J_t - J_{t-1})/J_{t-1} < 0$ 且 $|K_t - D_t| < \varepsilon$,则上涨过程KDJ指标滞后;若连续 s' 日出现 $R_t < 0$,且 $(J_t - J_{t-1})/J_{t-1} < 0$,且 $|K_t - D_t| < \varepsilon$,则下跌过程KDJ指标滞后。

2.3 股市滞后程度计算

对于股票数据 $SD = \{Y, R, I\}$,其中: $I = \{V, M, K\}$, V 表示VOL指标数据, M 表示MACD指标数据, K 表示KDJ指标数据。设定序列窗口 SW 大小为 H_{SW} ,开始时间戳为 i 。

对于股价涨跌幅 $R_t, R_{t+1}, \dots, R_{t+H_{SW}-1}$,若出现 $R_j \geq \theta$ 且 $R_{j+1} \geq \theta$ 且 $R_{j+2} \geq \theta$,则股价 j 时开始处于上涨波动中,股价在上涨波动中的滞后程度 LD 计算过程如下。

1) 若 t 时出现量价滞后,则 $LD_t = LD_{t-1} + f_{T,t}(R_t, V_t) = LD_{t-1} + R_t/e^{E_{V_t}}$ 。

2) 若 t 时出现同比放量滞后,令 $E'_{V_t} = (VA_{t+1} - VA_t)/VA_t$ (VA_t 上一波上涨的相同股价处的多空比率净额),则 $LD_t = LD_{t-1} + f_{T,t}(R_t, V'_t) = LD_{t-1} + R_t/e^{E'_{V_t}}$ 。

3) 若 t 时出现MACD指标滞后, $t+m$ 时消失,则对于 $t+x(0 \leq x \leq m)$ 时:

$$LD_{t+x} = LD_{t+x-1} + f_{T,t+x}(R_{t+x}, M_{t+x}) = LD_{t+x-1} + R_{t+x}/e^{E_{M_{t+x}}}$$

4) 若 t 时出现KDJ指标滞后, $t+n$ 时消失,则对于 $t+x(0 \leq x \leq n)$ 时:

$$LD_{t+x} = LD_{t+x-1} + f_{T,t+x}(R_{t+x}, K_{t+x}) = LD_{t+x-1} + R_{t+x}/e^{E_{K_{t+x}}}$$

直至出现 $R_k \leq -\theta$ 且 $R_{k+1} \leq -\theta$ 且 $R_{k+2} \leq -\theta$,则股价上涨走势结束, LD_k 清零。

股价下跌走势同理可得,横盘整理不考虑。

3 特征滞后的TGARCH-M模型

均值TGARCH(TGARCH in Mean, TGARCH-M)模型通过拟合历史股价实现预测,但对波动拐点的预测准确度和精度不高。

3.1 特征滞后的均值方程

TGARCH-M模型中均值方程为:

$$y_t = \lambda X_t + \varphi \sigma_t^2 + \varepsilon_t$$

股价的变化总是存在很大的风险因素, TGARCH-M模型将条件方差 σ_t^2 的信息作为风险因素加入均值方程中。为了提高对拐点的预测准确度,由于拐点附近风险性明显增大,本文算法将滞后性引到均值方程中,可得:

$$y_t = \lambda X_t + \varphi \times Ld_{t-1} \times \sigma_t^2 + \varepsilon_t$$

当出现指标特征滞后时 $Ld_t = 1 + \sum (R_t/e^{E_{I,t}})$,当无指标特征滞后时, $Ld_t = 1$ 。 R_t 是股价的涨跌幅, $E_{I,t}$ 是出现滞后性的指标 I 的能量值。

3.2 特征滞后的方差方程

在TGARCH-M模型中,方差方程为 $h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} + \gamma \varepsilon_{t-1}^2 d_{t-1}$ 。

进一步考虑,股价变化的拐点处波动相对剧烈,预测的精度不高,为了提高精度,考虑风险大小对随机扰动的影响,因此,将特征滞后加入到方差方程,则特征滞后的方差方程为:

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} + \gamma \times Lr_{t-1} \times \varepsilon_{t-1}^2$$

当出现指标特征滞后时 $Lr_t = d_t + R_t \times E_{I,t}$,当无指标特征滞后时 $Lr_t = d_t$ 。 R_t 是股价的涨跌幅, $E_{I,t}$ 是出现滞后性的指标 I 的能量值。

3.3 特征滞后的广义自回归条件异方差模型

本文给出LRD-TGARCH-M模型,描述如下:

$$\begin{cases} y_t = \lambda X_t + \varphi \times Ld_{t-1} \times \sigma_t^2 + \varepsilon_t \\ h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} + \gamma \times Lr_{t-1} \times \varepsilon_{t-1}^2 \end{cases} \quad (5)$$

其中, $Lr_t = d_t + R_t \times E_{I,t}$, $Ld_t = 1 + \sum (R_t/e^{E_{I,t}})$, R_t 是股价的涨跌幅, $E_{I,t}$ 是出现滞后性的指标 I 体现的能量值; $\varepsilon_t = \sqrt{h_t} v_t$ 为随机误差项, v_t 是独立同分布的白噪声过程,且 $E(v_t) = 0$, $Var(v_t) = 1$; $\sigma_t^2 = h_t$,当 $\varepsilon_{t-1} < 0$ 时 $d_{t-1} = 1$,当 $\varepsilon_{t-1} \geq 0$ 时 $d_{t-1} = 0$; $\lambda, \varphi, \gamma, \alpha_i (i = 0, 1, \dots, p), \beta_j (j = 1, 2, \dots, q)$ 为未知的参数,且 $\alpha_i \geq 0 (i = 1, 2, \dots, p), \beta_j \geq 0 (j = 1, 2, \dots, q)$, $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1, \varphi > 0, \gamma > 0, \alpha_0 > 0$ 。

3.4 LRD-TGARCH-M模型的误差分析

定理1 LRD-TGARCH(1,1)-M模型的条件期望和条件方差的解析式分别为:

$$\begin{cases} E_{t-1}(y_t) = \lambda_0 + \lambda_1 y_{t-1} + \dots + \lambda_m y_{t-m} + \varphi \times Ld_{t-1} \times ((1 + \beta_1 + \beta_1^2 + \dots) \alpha_0 + (\alpha_1 + \gamma \times Lr_{t-1}) \varepsilon_{t-1}^2 + \alpha_1 \beta_1 \varepsilon_{t-2}^2 + \alpha_1 \beta_1^2 \varepsilon_{t-3}^2 + \dots) \\ Var_{t-1}(y_t) = (1 + \beta_1 + \beta_1^2 + \dots) \alpha_0 + (\alpha_1 + \gamma \times Lr_{t-1}) \varepsilon_{t-1}^2 + \alpha_1 \beta_1 \varepsilon_{t-2}^2 + \alpha_1 \beta_1^2 \varepsilon_{t-3}^2 + \dots \end{cases} \quad (6)$$

证明 令 $T_1 \rightarrow T_2 \rightarrow T_3$ 为一个完整的股价上涨下跌过程,其中: T_1 时股价上涨开始; T_2 时上涨结束,下跌开始; T_3 时,下跌结束。根据滞后性定义,设 $i(T_1 \leq i \leq T_3)$ 时,开始出现指标特征的滞后现象,且在之后会陆续出现其他指标特征滞后,则 i 时将上涨过程分为两段,即 $T_1 \rightarrow i$ 和 $i \rightarrow T_2$ 。

$T_1 \rightarrow i$ 过程 设 t 为其中某一时刻,由于未出现指标特征滞后,有 $Ld_t = 1, Lr_t = d_t$ 。随着上涨的持续,必然会出现指标特征滞后,股价波动进入 $i \rightarrow T_2$ 过程。

$i \rightarrow T_2$ 过程 设 t 为其中某一时刻,由于出现指标特征滞后,有 $Ld_t = 1 + \sum (R_t/e^{E_{I,t}})$, $Lr_t = d_t + R_t \times E_{I,t}$,且随着上涨的持续,出现更多的指标特征滞后, Ld_t 随着滞后程度的累积,越来越大, Lr_t 则由当前时刻的特征滞后情况确定。由 $\varepsilon_t = \sqrt{h_t} v_t$ 可以得到随机误差项 ε_t 的条件方差为 $E_{t-1}(\varepsilon_t^2) =$

$h_t E_{t-1}(v_t^2) = h_t$, 根据式(5) 可得:

$$E_{t-1}(\varepsilon_t^2) = \alpha_0 + (\alpha_1 + \gamma \times Lr_{t-1})\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (7)$$

其中, $h_{t-1} = \alpha_0 + (\alpha_1 + \gamma \times Lr_{t-2})\varepsilon_{t-2}^2 + \beta_1 h_{t-2}$, 依次进行无穷次迭代, 可得:

$$E_{t-1}(\varepsilon_t^2) = (1 + \beta_1 + \beta_1^2 + \dots)\alpha_0 + (\alpha_1 + \gamma \times Lr_{t-1})\varepsilon_{t-1}^2 + \alpha_1 \beta_1 \varepsilon_{t-2}^2 + \alpha_1 \beta_1^2 \varepsilon_{t-3}^2 + \dots \quad (8)$$

由条件期望公式, 股价收益率的条件期望为 $E_{t-1}(y_t) = E_{t-1}(\lambda X_t + \varphi \times Ld_{t-1} \times \sigma_t^2 + \varepsilon_t)$, 又 X_t 为 y_t 的先验知识, 取 AR(m) 模型, 即 $X_t = \lambda_0 + \lambda_1 y_{t-1} + \dots + \lambda_m y_{t-m}$, 则:

$$E_{t-1}(y_t) = \lambda_0 + \lambda_1 y_{t-1} + \dots + \lambda_m y_{t-m} + \varphi \times Ld_{t-1} \times E_{t-1}(h_t)$$

根据式(8) 可得, 股价收益率的条件期望为:

$$E_{t-1}(y_t) = \lambda_0 + \lambda_1 y_{t-1} + \dots + \lambda_m y_{t-m} + \varphi \times Ld_{t-1} \times ((1 + \beta_1 + \beta_1^2 + \dots)\alpha_0 + (\alpha_1 + \gamma \times Lr_{t-1})\varepsilon_{t-1}^2 + \alpha_1 \beta_1 \varepsilon_{t-2}^2 + \alpha_1 \beta_1^2 \varepsilon_{t-3}^2 + \dots) \quad (9)$$

根据条件方差公式 $Var_{t-1}(y_t) = E_{t-1}(y_t - E_{t-1}(y_t))^2$, 并结合式(8), 有股价收益率的条件方差为:

$$Var_{t-1}(y_t) = E_{t-1}(y_t - E_{t-1}(y_t))^2 = E_{t-1}(\varepsilon_t^2) = (1 + \beta_1 + \beta_1^2 + \dots)\alpha_0 + (\alpha_1 + \gamma \times Lr_{t-1})\varepsilon_{t-1}^2 + \alpha_1 \beta_1 \varepsilon_{t-2}^2 + \alpha_1 \beta_1^2 \varepsilon_{t-3}^2 + \dots \quad (10)$$

综上, 可得股价收益率的条件期望和条件方差方程。证毕。

由定理1 中条件期望解析式可知, TGARCH-M 模型由于加入了特征滞后的因素, 提高了拐点的预测准确度; 由条件方差解析式可知, 加入特征滞后因素, 优化方差变化, 提高精度, 使得误差减小。

3.5 LRD-TGARCH-M 模型的预测

定理2 LRD-TGARCH-M 模型的 l -步预测公式为:

$$\begin{cases} \hat{Y}_T(l) = \lambda_0 + \lambda_1 \hat{Y}_T(l-1) + \dots + \lambda_m \hat{Y}_T(l-m) + \varphi \times Ld_{T-1+l} \times h_T(l) \\ h_T(l) = \alpha_0 + \sum_{i=1}^p \alpha_i h_T(l-i) + \sum_{j=1}^q \beta_j h_T(l-j) + \gamma \times Lr_{T-1+l} \times h_T(l-1) \end{cases} \quad (11)$$

其中: 当 $l-m \leq 0$ 时, $\hat{Y}_T(l-m) = Y_{T-m+i}$; 当 $l-i \leq 0$ 时, $h_T(l-i) = \varepsilon_{T-i+i}^2$; 当 $l-j \leq 0$ 时, $h_T(l-j) = h_{T-j+i}$ 。

证明 X_t 是 Y_t 的先验知识, 令 $X_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \dots + \lambda_m Y_{t-m}$, 则式(5) 的均值方程可写为:

$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \dots + \lambda_m Y_{t-m} + \varphi \times Ld_{t-1} \times \sigma_t^2 + \varepsilon_t \quad (12)$$

令当前时刻为 T 时, 对于 $T+1$ 时, 有:

$$\begin{cases} Y_{T+1} = \lambda_0 + \lambda_1 Y_T + \dots + \lambda_m Y_{T-m+1} + \varphi \times Ld_T \times \sigma_{T+1}^2 + \varepsilon_{T+1} \\ h_{T+1} = \alpha_0 + \alpha_1 \varepsilon_T^2 + \dots + \alpha_p \varepsilon_{T-p+1}^2 + \beta_1 h_T + \dots + \beta_q h_{T-q+1} + \gamma \times Lr_T \times \varepsilon_T^2 \end{cases} \quad (13)$$

对于扰动项 ε 下标大于 T , 用 ε_t 的期望值代替, v_t 是白噪声过程, 则: $E(\varepsilon_t) = E(\sqrt{h_t} v_t) = E(\sqrt{h_t}) E(v_t) = 0$, $\sigma_{T+1}^2 = h_{T+1}$, 用预测值 $h_T(1)$ 代替, 式(13) 可进一步写成:

$$\begin{cases} \hat{Y}_T(1) = c + \lambda_1 Y_T + \dots + \lambda_m Y_{T-m+1} + \varphi \times Ld_T \times h_T(1) \\ h_T(1) = \alpha_0 + \alpha_1 \varepsilon_T^2 + \dots + \alpha_p \varepsilon_{T-p+1}^2 + \beta_1 h_T + \dots + \beta_q h_{T-q+1} + \gamma \times Lr_T \times \varepsilon_T^2 \end{cases} \quad (14)$$

$T+2$ 时, 同理可得:

$$\begin{cases} Y_{T+2} = \lambda_0 + \lambda_1 Y_{T+1} + \dots + \lambda_m Y_{T-m+2} + \varphi \times Ld_{T+1} \times \sigma_{T+2}^2 + \varepsilon_{T+2} \\ h_{T+2} = \alpha_0 + \alpha_1 \varepsilon_{T+1}^2 + \dots + \alpha_p \varepsilon_{T-p+2}^2 + \beta_1 h_{T+1} + \dots + \beta_q h_{T-q+2} + \gamma \times Lr_{T+1} \times \varepsilon_{T+1}^2 \end{cases} \quad (15)$$

用一步预测值 $\hat{Y}_T(1)$ 代替未知项 Y_{T+1} , 并且根据一步预测值 $\hat{Y}_T(1)$ 计算得各指标数据, 结合特征滞后计算模型, 求得 $Ld_{T+1}, Lr_{T+1}; \sigma_{T+2}^2$ 由预测值 $h_T(2)$ 代替, 且 $E(\varepsilon_t^2) = E(h_t)$, 故式(15) 可写为:

$$\begin{cases} \hat{Y}_T(2) = \lambda_0 + \lambda_1 \hat{Y}_T(1) + \dots + \lambda_m Y_{T-m+2} + \varphi \times Ld_{T+1} \times h_T(2) \\ h_T(2) = \alpha_0 + \alpha_1 h_T(1) + \dots + \alpha_p \varepsilon_{T-p+2}^2 + \beta_1 h_T(1) + \dots + \beta_q h_{T-q+2} + \gamma \times Lr_{T+1} \times h_T(1) \end{cases} \quad (16)$$

依次进行迭代, 可得 l -步预测公式:

$$\begin{cases} \hat{Y}_T(l) = \lambda_0 + \lambda_1 \hat{Y}_T(l-1) + \dots + \lambda_m \hat{Y}_T(l-m) + \varphi \times Ld_{T-1+l} \times h_T(l) \\ h_T(l) = \alpha_0 + \sum_{i=1}^p \alpha_i h_T(l-i) + \sum_{j=1}^q \beta_j h_T(l-j) + \gamma \times Lr_{T-1+l} \times h_T(l-1) \end{cases} \quad (17)$$

其中: 当 $l-m \leq 0$ 时, $\hat{Y}_T(l-m) = Y_{T-m+i}$; 当 $l-i \leq 0$ 时, $h_T(l-i) = \varepsilon_{T-i+i}^2$; 当 $l-j \leq 0$ 时, $h_T(l-j) = h_{T-j+i}$ 。证毕。

3.6 LRD-TGARCH-M 模型的预测算法

算法描述如下。

输入: 股票价格数据集、相关指标数据集。

输出: 预测收益率的结果。

步骤如下。

步骤1 根据 $R_t = \ln(y_t/y_{t-1})$ 将股票收盘价转换为对数收益率, 用 $y_t = (R_t - \bar{R})/S_t$ 将其标准化。

步骤2 根据成交量、MACD、KDJ 等指标体现的能量公式, 计算得各自的能量体现 E_V, E_M, E_{KDJ} 。

步骤3 特征滞后程度的计算, 具体如下:

1) 根据 $R_j \geq \theta$ 且 $R_{j+1} \geq \theta$ 且 $R_{j+2} \geq \theta$, 判断当前股价是否为上涨; 根据 $R_j \leq -\theta$ 且 $R_{j+1} \leq -\theta$ 且 $R_{j+2} \leq -\theta$, 判断当前股价走势是否为下跌; 同时令 $Ld_j = 1$ 。

2) 对于 $t \geq j$, 判断出现滞后现象的指标 I , 则特征滞后程度 $Ld_t = 1 + \sum (R_t/e^{I \cdot t})$, 特征滞后因子 $Lr_t = d_t + R_t \times E_{I,t}$ 。

3) 直到 $(R_k \leq -\theta$ 且 $R_{k+1} \leq -\theta$ 且 $R_{k+2} \leq -\theta)$ 或 $(R_j \geq \theta$ 且 $R_{j+1} \geq \theta$ 且 $R_{j+2} \geq \theta)$, 则股价当前股价走势结束, 令 $f_{T,k} = 0$ 。

步骤4 由收益率序列和 $\{\varepsilon_t^2\}$ 的自相关函数和偏自相关函数, 确定均值方程的阶数 m , 方差方程的阶数 p, q 。

步骤5 估计 LRD-TGARCH-M 模型中均值方程的未知参数 $\lambda_0, \lambda_1, \dots, \lambda_m, \varphi$, 以及方差方程中的未知参数 $\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \gamma$ 。

步骤6 预测。

4 实验数据处理

4.1 数据选取

数据来自某软件下载的上证大盘指数, 取自2010年7月

2日至2014年2月20日共879个交易日的数据。数据包括大盘指数的日收盘价数据、日涨跌幅数据、日成交量数据、MACD指标数据、KDJ指标数据各879组。

4.2 数据的预处理

时间序列数据采用经标准化的日收盘价的对数收益率 y_t ,首先,根据 $R_t = \ln(y_t/y_{t-1})$,其中 y_t 为第 t 个交易日的收盘价,计算得日对数收益率数据,并采用 $y_t = (R_t - \bar{R})/S_t$,其中 \bar{R} 为样本平均值, S_t 为样本标准差,对日收益率序列进行标准化,消除一些干扰因素。

4.3 对比实验的选择

对比均值EGARCH(EGARCH in Mean, EGARCH-M)模型和TGARCH-M模型,并将特征滞后程度只加入到均值方程的模型(Lag Degree TGARCH-M, LD-TGARCH-M)进一步进行对比。

4.4 评价准则

对于预测的效果,可以根据观察预测与实际的折线图来判断,预测值跟实际值越吻合,则预测结果也越好,也可以根据一些数值指标来衡量平均的预测误差,平均预测误差越小,则模型越准确。

平均绝对误差(Mean Absolute Error, MAE)的计算如下:

$$MAE = \frac{1}{n} \sum_{t=T+1}^{T+n} |y_t - y'_t|$$

均方误差(Mean Squared Error, MSE)的计算如下:

$$MSE = \sqrt{\frac{1}{n} \sum_{t=T+1}^{T+n} (y_t - y'_t)^2}$$

其中: n 是预测个数, T 是初始估计的样本长度, y_t 是真实值, y'_t 是预测值。

5 实验结果分析

5.1 模型的参数估计结果对比分析

根据样本数据集的特点,对样本数据进行分析 and 检验,选取LRD-TGARCH(1,1)-M模型进行实验,在实验的样本数据集上,对样本区间进行模拟,如图1所示。

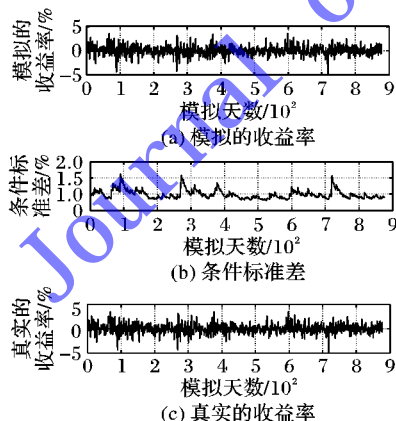


图1 LRD-TGARCH-M模型对样本区间的模拟效果

模型的参数估计值如表1所示。

从LRD-TGARCH-M模型算法对实验数据集模拟的结果图可以看出,模拟的效果与实际情况较为吻合,在波动较大的地方,算法模拟的波动也比较大,且波动拐点处波动相对更剧烈,也与实际相吻合。这说明波动大,收益率不稳定,则面临的风险性也大,LRD-TGARCH-M模型算法正是从特征滞后程

度的角度对波动拐点处进行深入分析,进而更准确地判断拐点的到来。

表1 LRD-TGARCH-M模型估计的参数值

系数	估计值	标准误差	t-统计量
c	-0.7467	0.033404	-1.7728
λ_1	-0.0178	0.037120	-0.4917
ϕ	0.7690	0.029937	1.7461
α_0	0.0742	0.016429	4.2325
α_1	0.0464	0.017730	3.9427
β_1	0.0611	0.041605	4.1819
γ	0.8672	0.015759	88.3505

5.2 模型的预测结果对比

通过对历史数据的模拟,得到各参数估计值,将各参数代入预测公式,同时将分析所得的特征滞后程度代入到均值方程和方差方程中,得到的最终方程为:

$$\begin{cases} y_t = -0.7467 - 0.0178y_{t-1} + 0.769 \times 0.895 \times \sigma_t + \varepsilon_t \\ h_t = 0.0742 + 0.0464\varepsilon_{t-1}^2 + 0.0611h_{t-1} + 0.8672 \times 0.732 \times \varepsilon_{t-1}^2 \end{cases}$$

LRD-TGARCH-M模型算法与TGARCH-M模型算法、EGARCH-M模型算法的对比如图2~3所示。

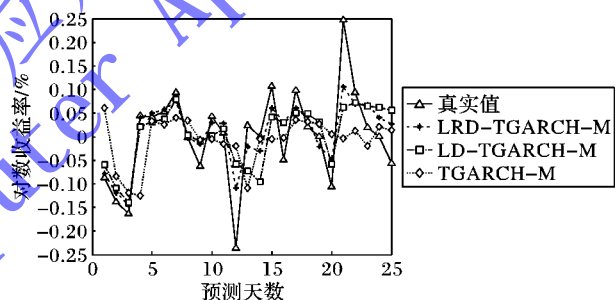


图2 与TGARCH-M模型算法的实验对比

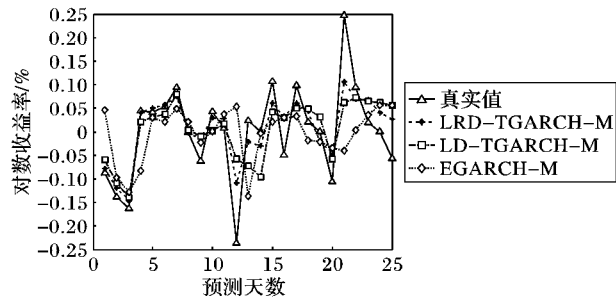


图3 与EGARCH-M模型算法的实验对比

5.3 结果分析

结果分析如表2所示。

表2 误差结果

算法	MAE	MSE	算法	MAE	MSE
TGARCH-M	0.0771	0.1067	LD-TGARCH-M	0.0538	0.0718
EGARCH-M	0.0739	0.0978	LRD-TGARCH-M	0.0395	0.0534

根据与对比算法的结果可以看出,LD-TGARCH-M模型算法和LRD-TGARCH-M模型算法的预测效果总体上是优于对比实验的,且LRD-TGARCH-M模型算法的预测效果是优于LD-TGARCH-M模型算法的。在对比图中,LRD-TGARCH-M模型算法对数据中第1,4,13,16点相对于对比模型都有比

较好的预测效果,而这些点正是股价走势的拐点,在这些点之前的股价走势中,都出现了指标特征滞后的现象,第1点出现了MACD指标滞后,第4点出现了量价滞后,第13点出现了KDJ指标滞后,第16点出现了KDJ指标滞后。正是由于出现了指标特征滞后,体现出了当前波动走势已累积一定的风险。

从预测结果的误差分析可以看出,LRD-TGARCH-M模型算法的误差值最小,相对于TGARCH-M模型和EGARCH-M模型,LRD-TGARCH-M模型算法指标特征滞后的风险对波动走势进行引导和启发,LRD-TGARCH-M模型算法也是优于LD-TGARCH-M算法的,因为还考虑到了当前波动风险对随机误差的影响,使得预测效果更好,因此,LRD-TGARCH-M模型算法能有效地减少误差。

6 结语

本文提出一种特征滞后程度计算的广义自回归条件方差模型(LRD-TGARCH-M),该模型算法从能量角度给出特征滞后程度的计算模型;将特征滞后程度作为风险情况,加入到均值和方差方程中,启发模型对拐点的判断和预测。从实验对比中可以看出,LRD-TGARCH-M模型算法在股价拐点处的走势预测方面有更好的预测效果。

参考文献:

- [1] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control [M]. Hoboken: John Wiley and Sons, 2008: 784.
- [2] ENGLE R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation [J]. *Econometrica*, 1982, 50(4): 987-1007.
- [3] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity [J]. *Journal of Econometrics*, 1986, 31(3): 307-327.
- [4] NELSON D B. Conditional heteroskedasticity in asset returns: a new approach [J]. *Econometrica*, 1991, 59(2): 347-370.
- [5] GLOSTEN L R, JAGANNATHAN R, RUNKLE D E. On the relation between the expected value and the volatility of the nominal excess return on stocks [J]. *The Journal of Finance*, 1993, 48(5): 1779-1801.
- [6] KE J, CHEN Z, ZHANG R. Comparative analysis of international stock market volatility with ARCH model [C]// FITME'08: Proceedings of the 2008 International Seminar on Future Information Technology and Management Engineering. Washington, DC: IEEE Computer Society, 2008: 494-497.
- [7] ZHANG S. A study on the weekday effect and leverage effect on CSI-300 index futures volatility - according to expanded conditional autoregressive range model of application [C]// ICMSE 2012: Proceedings of the 2012 International Conference on Management Science & Engineering. Piscataway: IEEE, 2012: 1522-1527.
- [8] FANG L, GUO B, ZENG Y. The compare of GARCH models' predicted ability: a semiparametric method [J]. *The Journal of Quantitative and Technical Economics*, 2010(4): 148-161. (方立兵, 郭炳仲, 曾勇. GARCH族模型的预测能力比较: 一种半参数方法 [J]. *数量经济技术经济研究*, 2010(4): 148-161.)
- [9] WANG C, LI G, ZHAO X. VaR-GARCH model using simulated annealing algorithm [J]. *Journal of Systems Engineering*, 2003, 18(1): 1-7. (王春峰, 李刚, 赵欣. 基于模拟退火算法的 VaR-GARCH 模型 [J]. *系统工程学报*, 2003, 18(1): 1-7.)
- [10] XIONG Z, ZHANG J. A comparative study on the performance of the value-at-risk using realized volatility and GARCH model [C]// MASS 2010: Proceedings of the 2010 International Conference on Management and Service Science. Piscataway: IEEE, 2010: 1-4.
- [11] CHEN F. The volatility risk of stock price and its defusing [C]// CSO 2009: Proceedings of the 2009 International Joint Conference on Computational Sciences and Optimization. Piscataway: IEEE, 2009: 512-516.
- [12] YANG J, ZHANG R. Measurement and analysis on the risk of China stock market under different economic systems [C]// BIFE 2012: Proceedings of the Fifth International Conference on Business Intelligence and Financial Engineering. Piscataway: IEEE, 2012: 214-218.
- [13] ZHANG R, WEI Y, JIN W. HS300 index volatility forecast based on Markov regime-switching EGARCH model [J]. *Journal of Systems Engineering*, 2011, 26(5): 628-635. (张锐, 魏宇, 金炜东. 基于 MRS-EGARCH 模型的沪深 300 指数波动预测 [J]. *系统工程学报*, 2011, 26(5): 628-635.)
- [8] KORTUEM G, KAWSAR F, FITTON D, *et al.* Smart objects as building blocks for the Internet of things [J]. *IEEE Internet Computing*, 2009, 14(1): 44-51.
- [9] HONG S, KIM D, HA M, *et al.* SNAIL: an IP-based wireless sensor network approach to the Internet of things [J]. *IEEE Wireless Communications*, 2009, 17(6): 34-42.
- [10] LIU X, WU M. Vehicular CPS: an application of IoT in vehicular networks [J]. *Journal of Computer Applications*, 2012, 32(4): 900-940. (刘小洋, 伍民友. 车联网: 物联网在城市交通网络中的应用 [J]. *计算机应用*, 2012, 32(4): 900-940.)
- [11] XUE X, WANG Q, ZHANG F. Key technologies and application evolution of Internet of things [J]. *Journal of Computer Applications*, 2013, 33(10): 2701-2706. (薛小平, 王翥, 张芳. 物联网核心技术及应用演进 [J]. *计算机应用*, 2013, 33(10): 2701-2706.)
- [12] ATZORIA L, IERAB A, MORABITO G. The Internet of things: a survey [J]. *Computer Networks*, 2010, 54(15): 2787-2805.
- [13] TANG B, HE C, CAO X. Topology of wireless sensor networks for mechanical vibration monitoring [J]. *Journal of Vibration Measurement & Diagnosis*, 2010, 30(4): 357-361. (汤宝平, 贺超, 曹小佳. 面向机械振动监测的无线传感器网络结构 [J]. *振动、测试与诊断*, 2010, 30(4): 357-361.)
- [14] LI S H, CHIANG H D. Nonlinear predictors and hybrid corrector for fast continuation power flow [J]. *IET Generation, Transmission & Distribution*, 2008, 2(3): 341-354.
- [15] YIN F. The research and design of remote data transmission system basing on GPRS [D]. Changsha: Hunan University, 2009: 18-20. (尹锋. 基于 GPRS 的远程数据传输系统研究和设计 [D]. 长沙: 湖南大学, 2009: 18-20.)

(上接第 2076 页)