

ANN, GARCH下医疗器械板块预测分析

孙 晓

上海对外经贸大学, 上海
Email: zxrr19961007@163.com

收稿日期: 2021年5月27日; 录用日期: 2021年6月11日; 发布日期: 2021年6月23日

摘 要

2020年新型冠状病毒疫情的爆发,使得股票市场中某些板块产生剧烈的波动,尤其是医疗器械板块,本文对医疗器械板块的股票信息进行了神经网络估计与传统GARCH模型估计,发现神经网络对股票价格估计相比较于GARCH模型产生了极强的优势。基于传统计量假设的GARCH模型忽视了股市中的非线性关系,而且对于模型的假设比较苛刻,而神经网络对于原始股票序列的宽松假设,可以弥补这种缺陷。结果表明:神经网络能捕捉到医疗板块序列中的序列之间的非线性关系,在股票序列价格估计方面优于传统的GARCH模型。

关键词

收益率, GARCH模型, 神经网络

Prediction and Analysis of Medical Device Sector Based on ANN and GARCH

Xiao Sun

Shanghai University of International Business and Economics, SUIBE, Shanghai
Email: zxrr19961007@163.com

Received: May 27th, 2021; accepted: Jun. 11th, 2021; published: Jun. 23rd, 2021

Abstract

In 2020, the outbreak of New Coronavirus epidemic caused some violent fluctuations in some sectors of the stock market, especially the medical device sector. In this paper, we estimated the stock information of the medical device sector and estimated the traditional GARCH model, and found that the neural network has a strong advantage in the stock price estimation compared with the

GARCH model. The GARCH model based on the traditional econometric hypothesis ignores the nonlinear relationship in the stock market, and the hypothesis of the model is harsh. The loose hypothesis of neural network for the original stock sequence can make up for this defect. The results show that: the neural network can capture the nonlinear relationship between the series in the medical plate series, and it is better than the traditional GARCH model in the estimation of stock series price.

Keywords

Rate of Return, GARCH Model, Neural Network

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

股票市场作为一国经济的晴雨表，其波动往往引起公众的广泛关注。而股票市场的波动性一直是统计研究的热点问题，加之 2020 年爆发的新型冠状病毒肺炎疫情，医疗器械板块的股价波动引起本文的关注。本文选取了医疗器械板块的股票作为研究对象，对其价格进行估计预测。股市是一个复杂的非线性系统，股票价格涉及许多不确定因素，且各个因素之间的相关关系错综复杂[1]。传统计量经济学模型认为股价波动完全是随机的，但大量事实表明，股价波动存在某种规律性[2]。我们可以将股市看作确定的非线性动力系统。股票价格作为一种常见的金融时间序列，其波动影响因子之间关系呈现复杂的特征，基于传统的计量经济假设下的 GARCH 模型只能捕捉到序列之间的线性关系，而忽略了序列中复杂的非线性特征，从而导致估计结果的偏差过大，为此本文引进机器学习中的神经网络对股票价格进行预测分析，发现神经网络对于股票价格的估计预测优于传统的 GARCH 模型，进而对投资者而言，能够更好的规避股市风险、制定合理的投资计划[3]。

传统的股票时间序列预测模型是通过建立在时间序列模型之间具有弱相关关系的基础上建立的，以此来预测未来的股票价格[4]。1982 年，美国经济学家 Engle 针对金融时间序列首次提出了 ARCH 模型(自回归条件异方差模型)，由于 ARCH 模型对一个序列的预测存在滞后阶数多，参数难以估计的缺点，在此基础上，Bollerslev 提出了改进的 ARCH 模型，GARCH (广义自回归条件异方差模型)，使用较低的滞后阶数就可以很好的展现出收益的波动率特征，从而弥补了 ARCH 模型在参数估计方面的缺点，进一步刻画金融时间序列的波动性特征[1]。但是由于股票市场往往存在杠杆效应，即负面影响的冲击往往大于正面影响带来的冲击，传统的 GARCH 模型对于股票收益率序列的非对称性无法做到较好的捕捉，在此基础上，Nelson (1991)提出了对 GARCH 模型的改进 EGARCH (指数自回归条件异方差模型)，能够更好的刻画股票收益率的非对称性。

基于 ARMA-GARCH 模型对股票时间序列进行建模分析预测时，往往会出现预测结果不理想、精确度不高等问题[5]。1990 年，Varies、Versino 首次把人工神经网络用于经济研究领域，对金融时间序列问题进行建模，这也引起本文的兴趣。用人工神经网络对医疗器械板块股票重新进行估计，实证分析发现相比较于计量经济学分析方法，人工神经网络具有更强的预测分析能力[6]。Weigend 等把神经网络用于研究太阳黑子的年平均活动情况，通过与回归方法的比较表明神经网络的预测优于统计预测。20 世纪 90 年代以来，利用神经网络对股票价格预测方法就已经层出不穷，本文便使用了前向传播网络与 GARCH

模型分别对股票价格和收益率进行对比(对于 GARCH 模型预测出的对数收益率, 本文将其转化成价格进行比较), 发现神经网络在预测方面的确优于传统的模型[7] [8] [9]。

2. 模型简述

2.1. ARMA 模型

将任意一个时间序列经过预处理后, 变成蕴含相关信息的平稳非白噪声序列, 这时就可以用 ARMA 模型对此序列进行拟合, 提取出关键信息。ARMA 为自回归移动平均模型, 可细分为 AR 模型, MA 模型, 和 ARMA 模型三大类, 分别具有如下形式:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (1)$$

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p} \quad (2)$$

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p} \quad (3)$$

AR 模型有偏自相关函数截尾的性质; MA 模型有相关函数截尾的性质。有些因果线性时间序列有与 AR 和 MA 类似的表现, 但是不能在低阶实现偏自相关函数截尾或者相关函数截尾。ARMA 模型结合了 AR 和 MA 模型, 在对数据拟合优度相近的情况下往往可以得到更简单的模型, 而且不要求偏自相关函数截尾也不要求相关函数截尾。在式(3)中, ε_t 为独立同分布零均值白噪声序列, ε_t 与 x_{t-1}, x_{t-2}, \dots 独立。 $1 - \phi_1 x - \dots - \phi_p x^p$ 称为特征多项式, 特征多项式的根都在单位圆外, 这个条件称为可逆条件。两个多项式没有公共根, 否则统一模型可能会有不同的表示。

2.2. ARCH 模型

在经济研究领域, 经常发现有在消除确定性非平稳因素后, 残差序列大部分时段是平稳的但在某一时段会有较大波动, 某一时段会有较小波动, 具有异方差波动特征, 存在波动聚集的现象, 不满足时间序列建模的同方差的假设, 也就无法使用时间序列的分析方法分析问题, 于是 Engle (1982) 提出了 ARCH 模型(自回归条件异方差模型), 这是对波动率定义为条件标准差, 第一次提出的波动率的理论模型, 基本思想是: 1) 资产收益率的扰动序列 $a_t = r_t - E(r_t | F_{t-1})$ 是前后不相关的, 但是前后不独立; 2) a_t 的不独立性, 描述为 $Var(r_t | F_{t-1}) = Var(a_t | F_{t-1})$ 可以用 a_t^2 的滞后值的线性组合表示。

具体的 ARCH (m) 模型为:

$$a_t = \sigma_t \varepsilon_t \quad (4)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2 \quad (5)$$

其中 ε_t 是零均值单位方差的独立同分布的白噪声, $\alpha_0 > 0, \alpha_j \geq 0, j = 1, 2, \dots, m$ 。 α_j 还需要满足一些条件使得 $Var(a_t)$ 有限, 类似于 AR (p) 序列的平稳性的特征根条件。

ARCH 模型实质就是构造一个模型, 利用历史波动信息, 得到条件方差信息, 采取自回归形式刻画波动的变化。如下式所述:

$$Var(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 \quad (6)$$

对于一个时间序列来说, 不同时刻包含不同历史信息, 所以有不同的条件方差, 用 ARCH 模型可以很好的刻画时间序列的这种波动特征。

2.3. GARCH 模型

ARCH 模型实质是用残差平方序列的移动平均来拟合当期的异方差数值, 通常只适用于短期自相关

过程, 对于有长期自相关性的残差序列(ARCH 模型具有较高的滞后阶数), 使用 ARCH 模型会增加它待估参数的难度和估计精确度, 于是引入 GARCH (广义自回归条件异方差)模型。对于一个对数收益率序列 r_t , 令 $a_t = r_t - \mu_t = r_t - E(R_t | F_{t-1})$ 为其新息序列, 称 a_t 服从 GARCH(m, s) 模型, 如果 a_t 满足:

$$\begin{aligned} \alpha_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^m \alpha_i \alpha_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \end{aligned} \quad (7)$$

其中 ε_t 为零均值单位方差的独立同分布白噪声序列, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $0 < \sum_{i=1}^m \alpha_i + \sum_{j=1}^s \beta_j < 1$, 最后一个条件是用来保证模型的 a_t 的无条件方差有限且不变, 而条件方差 σ_t^2 可以随时间 t 的变化而变化。

2.4. EGARCH 模型

由于 GARCH 模型在处理时间序列中没有考虑到股票收益率的非对称效应, Christie (1982)的研究认为, 当股票价格下降时, 资本结构当中附加在债务上的权重增加, 如果债务权重增加的信息泄漏以后, 资产持有者和购买者就会产生未来资产收益率将导致更高波动性的预期, 从而导致该资产的股票价格波动。因此, 对于股价反向冲击所产生的波动性, 大于等量正向冲击产生的波动性, 这种“利空消息”作用大于“利好消息”作用的非对称性。为了克服 GARCH 模型处理时间序列上的问题, Nelson (1991)考虑了加权的新息, 由此提出了 EGARCH 或指数(Exponential) GARCH 模型。

$$\begin{aligned} r_t &= \phi_0 + \phi_1 x_t + \alpha_t, \\ \alpha_t &= \sigma_t \varepsilon_t, \\ g(\varepsilon_t) &= \theta \varepsilon_t + \gamma (|\varepsilon_t| - E|\varepsilon_t|), \\ \ln(\sigma_t^2) &= \alpha_0 + \sum_{j=1}^s \beta_j \ln \left(\sigma_{t-j}^2 + \sum_{i=1}^m \alpha_i \frac{|\alpha_{t-i}| + \gamma_i \alpha_{t-i}}{\sigma_{t-i}} \right) \end{aligned} \quad (8)$$

其中 θ 和 γ 是实常数。 ε_t 和 $|\varepsilon_t| - E|\varepsilon_t|$ 都分别是零均值独立同分布白噪声, 分别为连续分布, 易知 $Eg(\varepsilon_t) = 0$ 。

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma E|\varepsilon_t|, & \varepsilon_t \geq 0 \\ (\theta - \gamma)\varepsilon_t - \gamma E|\varepsilon_t|, & \varepsilon_t < 0 \end{cases} \quad (9)$$

根据上式可见 $g(\varepsilon_t)$ 的分布是非对称的。当 $\varepsilon_t \sim N(0, 1)$ 时, EGARCH(m, s) 模型可以用滞后算子的形式写成:

$$\begin{aligned} \alpha_t &= \sigma_t \varepsilon_t, \\ \ln(\sigma_t^2) &= \alpha_0 + \frac{1 + \beta_2 B + \dots + \beta_s B^{s-1}}{1 - \alpha_1 B - \dots - \alpha_m B^m} g(\varepsilon_t) \end{aligned} \quad (10)$$

其中 α_0 为常数, 其中 B 是滞后算子, 多项式 $1 + \beta_2 B + \dots + \beta_s B^{s-1}$ 和 $1 - \alpha_1 B - \dots - \alpha_m B^m$ 的根都在单位圆外且两个多项式没有公因子。记 $\xi_t = \ln(\varepsilon_t^2)$, 则上式给出的 ξ_t 为一个平稳线性 ARMA($m, s-1$) 序列, 以零均值独立同分布白噪声 $g(\varepsilon_{t-1})$ 为新息; 但是, $\ln \varepsilon_t^2$ 通过 $\varepsilon_{t-1} = \frac{\alpha_{t-j}}{\sigma_{t-j}}$, 序列依赖。原始的 GARCH 模型

的 ε_t^2 直接依赖于 α_{t-j}^2 的, $\pm \alpha_{t-j}$ 对 σ_t^2 影响相同。易知, $E \ln(\sigma_t^2) = \alpha_0$ 。EGARCH 与 GARCH 模型的区别还有: 1) 使用条件方差的对数建模, 因为对数值可正可负, 这就取消了 GARCH 模型对系数必须非负的限制。

限制; 2) $g(\varepsilon_{t-j}) = g\left(\frac{\alpha_{t-j}}{\sigma_{t-j}}\right)$ 的使用使得波动率对 α_{t-j} 的依赖关系与 α_{t-j} 的正负号有关, 可以用来描述正负收益率的不同的影响。

此外对于序列 ε_t 拟合 GARCH 模型有一个基本要求即需为零均值, 纯随机, 异方差序列。但有时 $f(t, x_{t-1}, x_{t-2}, \dots)$ 不能充分提取原序列相关信息, ε_t 就不具有纯随机性。这时需要考虑 AR-GARCH 模型, 具有如下形式:

$$\begin{aligned} x_t &= f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t, \\ \varepsilon_t &= \sum_{k=1}^m \beta_k \varepsilon_{t-k} + v_t, \\ v_t &= \sqrt{h_t} e_t, \\ h_t &= \omega + \sum_{i=1}^p \eta_i h_{t-i} + \sum_{j=1}^q \lambda_j v_{t-j}^2 \end{aligned} \tag{11}$$

这就是先对 ε_t 拟合自回归模型, 再考虑 μ_t 的方差齐性, 如具有异方差, 对此拟合 GARCH 模型, 称为 AR(m)-GARCH(p, q)模型。

2.5. 神经网络模型

人工神经网络(Artificial Neural Network, ANN)简称神经网络(ANN), 是基于生物学中神经网络的基本原理, 在理解和抽象了人脑结构和外界刺激响应机制后, 以网络拓扑知识为理论基础, 模拟人脑的神经系统对复杂信息的处理机制的一种数学模型。该模型具有高度的非线性, 能够进行复杂的逻辑操作和非线性关系实现的系统。ANN 是由大量处理单元经广泛互连而组成的人工网络, 用来模拟脑神经系统的结构和功能。这些处理单元我们称作人工神经元。人工神经网络便可看成是以人工神经元为节点, 用有向加权弧连接起来的有向图。在有向图中, 人工神经元就是对生物神经元的模拟, 而有向弧则是轴突—突触—树突对的模拟。

机器学习中的神经网络结构由输入层、隐含层、输出层组成。假设有 N 个任意样本集 (x_j, y_j) , $j=1, 2, \dots, N$ 其中

$$\begin{aligned} x_i &= [x_{j1}, x_{j2}, \dots, x_{jm}]^T \in R^n \\ y_i &= [y_{j1}, y_{j2}, \dots, y_{jm}]^T \in R^n \end{aligned} \tag{12}$$

通过神经网络可以自动找寻 x 与 y 之间的非线性映射关系。对于具有 L 个隐含结点的单隐层神经网络的输出 t_j 可以表示为

$$\sum_{i=1}^L \theta_i g(\omega_i^T x_j + b_i) = t_j, \quad j=1, 2, \dots, N \tag{13}$$

式(13)中, g 为激活函数, 一般为 σ 函数。 $\omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{im}]^T$ 为第 i 个隐层单元的输入权重, i 为第 i 个隐层单元的偏置, $\theta_i = [\theta_{i1}, \theta_{i2}, \dots, \theta_{im}]^T$ 为第 i 个隐层单元的输出权重。 $t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T$ 为神经网络的输出值。上式的矩阵形式为

$$HB = H(\omega_1, \omega_w, \dots, \omega_L, b_1, b_2, \dots, b_L, x_1, x_2, \dots, x_N)B = T \tag{14}$$

其中,

$$H = \begin{bmatrix} g(\omega_1^T x_1 + b_1) & \cdots & g(\omega_L^T x_1 + b_L) \\ \vdots & & \vdots \\ g(\omega_1^T x_N + b_N) & \cdots & g(\omega_L^T x_N + b_L) \end{bmatrix}_{N \times L}$$

$$B = \begin{bmatrix} \theta_1^T \\ \vdots \\ \theta_L^T \end{bmatrix}_{L \times m}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

单隐层神经网络对输入的 N 个样本集 $(x_j, y_j), j=1, 2, \dots, N$ 进行自我学习, 通过最小化输出值与是机器的误差平方和估计参数 $\hat{\omega}_i, \hat{b}_i, \hat{\theta}_i$, 得到 x 与 y 之间的非线性映射关系:

$$\min_{\omega, b, \beta} \|Y - T\| = \min_{\omega, b, \beta} \|Y - H(\omega, b)B\| \quad (15)$$

式(11)等价于最小化损失函数:

$$E = \sum_{j=1}^N \left(\sum_{i=1}^L y_i - \theta_i g(\omega_i^T x_j + b_i) \right)^2 \quad (16)$$

式(15)无解析解, 采用梯度下降法可得到模型参数估计。

3. GARCH 模型的建立

3.1. 样本选取及处理

本文选取 2019 年 12 月 2 日到 2020 年 3 月 16 日我国医疗器械板块股票的相关数据, 进行筛选得出医疗器械板块股票的日收盘价 P , 数据来源于聚宽量化分析网站。分别在 GARCH 模型与神经网络模型中进行预测分析。在神经网络模型中将 2019 年 12 月 2 日至 2020 年 2 月 15 日的数据作为训练数据, 2020 年 2 月 17 日至 2020 年 3 月 16 日数据为检验数据, 从而进行日收盘价的预测。在 GARCH 模型中将 2019 年 12 月 5 日至 2020 年 3 月 16 日的数据作为实验数据, 从而进行收益率的预测。运用 R 统计建模软件, 对最终筛选出的日收盘价 P 做时间序列图, 见图 1。由时序图可以看出, 医疗器械板块股票日收盘价随着时间的波动而波动性, 是一个非平稳的时间序列。对其进行了 ADF 检验, 检验的 P 值为 0.6021, 大于 0.05 的显著性水平, 我们有理由不拒绝原假设, 即序列是非平稳的, 这与时序图展示的结果是一致的。而在建立时间序列模型时, 假设序列是平稳的, 所以需要对原始数据进行差分处理使其达到建模的要求。又加之在研究金融时间序列时, 人们往往关注的是收益率而并非收盘价, 所以我们考虑对医疗器械板块股票日收盘价取对数, 同时做差分转化成对数收益率。再对新建立的对数收益率进行平稳性检验, 看其是否符合建模的要求。对数收益率的公式如下:

$$DR = \ln p_t - \ln p_{t-1}$$

其中, p_t 表示时刻 t 的收盘价, p_{t-1} 表示 $t-1$ 时刻的收盘价, DR 表示对数收益率。

对新建立的对数收益率序列做时间序列图, 结果如图 1 所示。

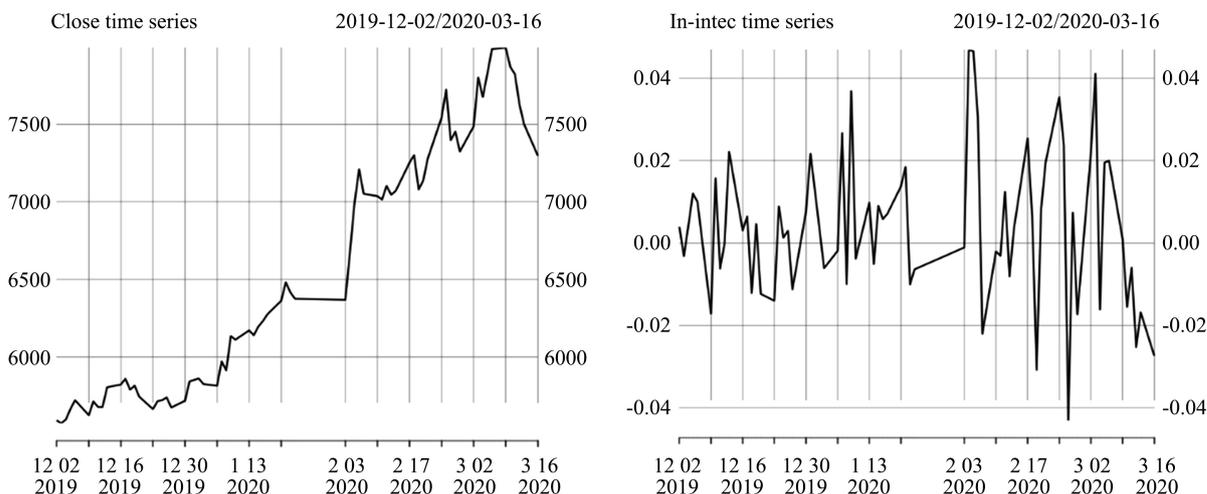


Figure 1. Medical devices sector time-series chart and yield sequence chart
图 1. 医疗器械板块时序图与收益率序列图

由对数收益率的时序图可以看出序列出现了许多异常的峰值。当新冠疫情爆发之后的波动十分剧烈，也表现出了明显的波动聚集性，认为使用 GARCH 模型来刻画对数收益率的波动性是合理的。为验证可视化结果的真实性，即其是否真正处于平稳状态，同样做了 ADF 检验，检验的 P 值为 0.09756，发现其并非为平稳时间序列，因此对对数收益率进行一阶差分后进行模型的假设，根据 ADF 检验结果表明，差分后的对数收益率序列，见图 2 满足平稳时间序列的假设。

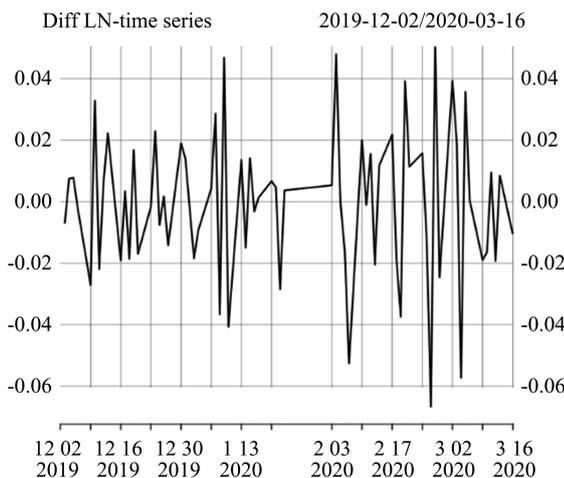


Figure 2. Rate of return difference series
图 2. 收益率差分序列

描述性统计

在对使用差分后对数收益率进行建模分析之前，本文对对数收益率序列进行了整体的描述统计，结果见表 1。

根据描述性统计，JB 检验的 P-value = 0.6392，我们有理由不拒绝服从正态分布的原假设，所以差分后的对数收益率近似的满足正态假定下的数据结构，对于之后的建模分析，在建模时，可以选择使用正态分布来对差分后的对数收益率序列进行预测分析。

Table 1. Descriptive statistical analysis table
表 1. 描述性统计分析表

统计量	数值	统计量	数值
最小值	-5.044	最大值	6.687
均值	-0.010	中位数	-0.254
峰度	-0.001	偏度	0.271
J-B 统计量	0.895	Q (1)	13.248
Q (6)	17.737	Q (12)	19.718

3.2. 建模分析

建立 GARCH 模型之前, 首先要对序列进行 ARCH 效应的检验, 若序列本身并不存在 ARCH 效应, 则没有建立 GARCH 模型的必要。为了符号上的方便, 记为均值方程的残差, 则可以用平方序列来检验条件异方差性, 即所谓的 ARCH 效应。Ljung-Box 检验的 P 值为 0.02053, 小于 0.05 的显著性水平, 所以我们有理由拒绝不存在 ARCH 效应的原假设, 所以存在 ARCH 效应, 可以进行 GARCH 建模分析。对平方序列做偏自相关图(见图 3)进行 ARCH 模型的定阶。根据偏自相关图, 选用低阶的 GARCH 模型, 如 GARCH(1, 1)模型、GARCH(1, 2)模型、GARCH(2, 1)模型等, 便可以进行预测估计。

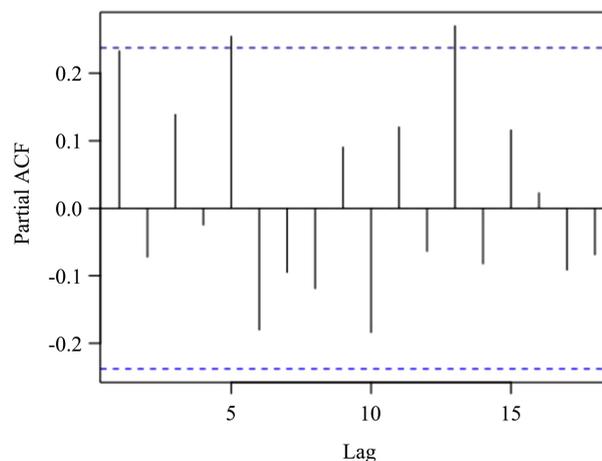


Figure 3. The partial autocorrelogram of a_t^2

图 3. a_t^2 的偏自相关图

经过大量的回归检验及尝试对比, 在对原始对数收益率序列建立 EGARCH(1, 1)模型时, 发现存在两个参数检验通不过的现象, 而且根据赤池信息统计量最小的原则, 最终对波动率的模型拟合选择了 ARMA(1, 2)-GARCH(1, 1)模型, 拟合后的 GARCH 模型为:

$$\begin{aligned}
 r_t &= 0.1044 - 0.9350r_{t-1} + \alpha_t + 0.2099\alpha_{t-1} + 1.000\alpha_{t-2} \\
 \sigma_t^2 &= 0.9151 + 0.2155\alpha_{t-1} + 0.4788\sigma_{t-1}^2 \\
 (\log \text{likelihood} &= -131.4971, \text{AIC} = 4.0734)
 \end{aligned} \tag{17}$$

在模型的输出的回归结果中，各项系数都是十分显著的，并且给出了标准化残差的 Ljung-Box 白噪声检验结果，滞后 10 的 p 值为 0.7244，承认白噪声； a_1^2 的滞后 10 的 Ljung-Box 白噪声检验结果 p 值为 0.5764 在 5% 的显著性水平下也可以承认白噪声。所以 ARMA(1, 2)-GARCH(1, 1) 模型对波动率的拟合效果是比较好的。

3.3. 预测

在对模型进行选择之后，将 2019 年 12 月 2 日至 2020 年 2 月 17 日的数据做为实验数据进行模型估计，预测后 12 期的收益率变化，见图 4。

在使用 GRCH 模型进行预测分析时，同样是选择以后 12 天的进行对比，从图中可以看出，GARCH 模型与以后 12 天的估计有一个明显的上升趋势，并且会在第 7 天左右达到一个较平稳的状态，而实际情况在之后 12 天中存在明显的上下波动情况，并不是会在某一天趋于平稳的状态。见图 5。

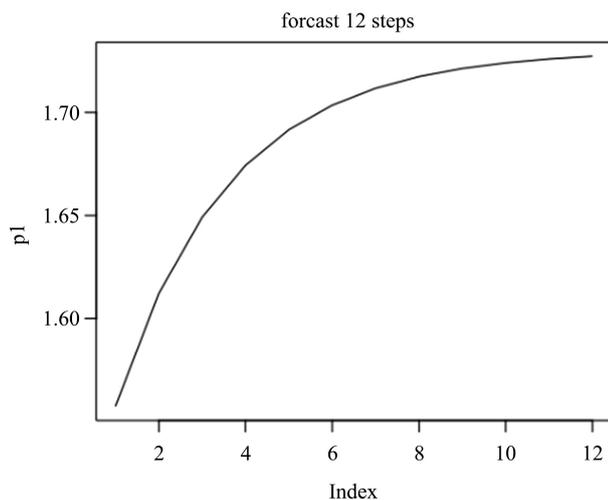


Figure 4. 12 steps ahead of the forecast

图 4. 超前 12 步预测

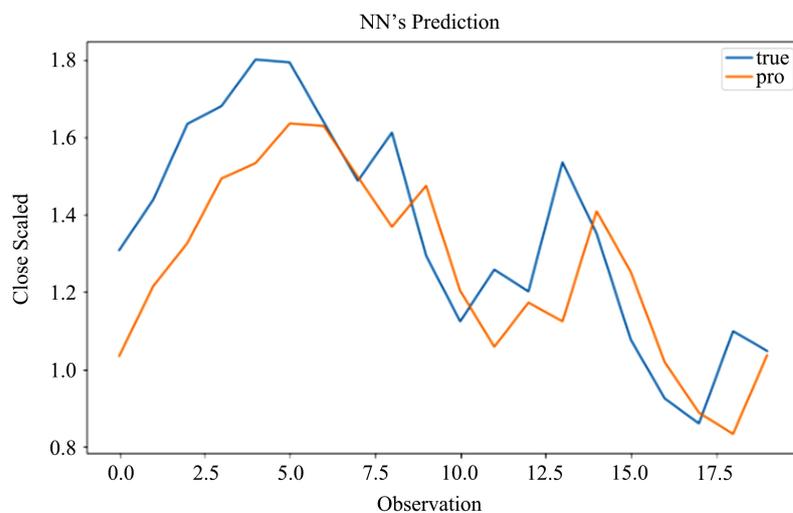


Figure 5. ANN model prediction results

图 5. ANN 模型预测结果

4. ANN 模型的建立

在神经网络模型中将 2019 年 12 月 2 日至 2020 年 2 月 15 日的数据作为训练数据, 2020 年 2 月 17 日至 2020 年 3 月 16 日数据为检验数据, 从而进行日收盘价的预测。采用 MinMaxScaler 函数对数据进行归一化处理, 将归一化后的取值设定在 $[-1, 1]$, 从而避免数据进入饱和区。创建一个序列模型, 将医疗器械板块的日收盘价作为输入到第一层, 输出维度为 15, 选用线性整流函数 Relu (校正线性单位), 损失函数选用平均绝对离差函数, 为防止过拟合, 选择早停法(Early Stopping), 在 5 个周期内都没有出现提高则结束训练。最终训练结果用训练数据集、测试数据集的拟合优度来展示。利用上述假设, 利用梯度下降法进行的 ANN 预测的结果如下。

ANN 的训练数据得到的拟合优度为 0.960, 经过一系列参数的设定, 最终测试数据得到的拟合优度为 0.497。模型的训练数据集的误差达到很好的效果, 在参数的选择方面, 为避免过拟合现象, 测试误差相比较而言偏大, 但相比传统的 GARCH 模型, ANN 模型的预测效果是优于 GARCH 模型的。序列的预测结果见图 5。由图可知, ANN 模型应用到医疗器械板块股票上的验证中, 模型对于序列的波动性质拟合效果存在一定的滞后效果, 本文的时间跨度区间仅仅选择了冠状病毒爆发的近 100 天, 样本的大小也限制了模型的估计结果, 但总体与传统的 GARCH 模型相比, 充分考虑了序列的各个方面的特性, 有效的抓住了非线性相关关系。为检验 ANN 模型的鲁棒性, 选择了同时段的汽车板块的股票日收盘价作为输入。模型的预测结果见图 2。由图 2 可知, ANN 模型对选取的汽车板块股票测试中表现的依然良好, 说明模型具有一定的鲁棒性。

5. 结论

针对股票市场中的非线性特征, 选用 ANN 的模型算法与传统的 GARCH 模型对股票价格进行预测, 结果表明选用 ANN 算法估计优于传统的 GARCH 模型估计, 尽管使用最简单的 ANN 算法, 并且由于数据的跨度较小, 会产生一定的过拟合现象, 但与传统的 GARCH 模型相对比, 还是展现出了极强的预测能力。尤其是在趋势的预测上, 使用 ANN 算法能够展现出股票日后趋势的大体走势。而传统的 GARCH 模型, 对于时间跨度较短、波动较大的序列, 没有展现出很好的估计效果, 估计误差较大, 在预估趋势方面也存在一定的差异。

致 谢

感谢老师对课题研究 with 论文撰写的指导, 对于该研究的实证模拟比较困难, 在这一过程中, 老师和同学都帮助了我很多, 通过一步一步地悉心指导, 帮助我完成了论文的撰写。此外, 也感谢论文评阅老师们的辛苦工作。

参考文献

- [1] 王蒋凤, 吴群英. 基于 GARCH 族模型对中国股市波动的分析与预测[J]. 经济研究导刊, 2011(34): 74-77+234.
- [2] 吴玉霞, 温欣. 基于 ARIMA 模型的短期股票价格预测[J]. 统计与决策, 2016(23): 83-86.
- [3] 付燕, 栗锋. ARMA 模型在我国体育股票价格预测中的应用[J]. 统计与决策, 2012(21): 101-103.
- [4] 鲁万波, 于翠婷, 王敏. 基于非参数条件自回归极差模型的中国股市波动性预测[J]. 数理统计与管理, 2018, 37(3): 544-553.
- [5] 陈晨, 刘光武, 陈涛, 温仲清. 伏牛山区栓皮栎天然次生林地位指数 ANN 模型构建[J]. 西北林学院学报, 2019, 34(1): 206-210+223.
- [6] 刘阳, 李艳丽, 陆贵斌. 基于信息更新 NN-GARCH 模型的统计套利研究[J]. 统计与决策, 2016(2): 169-171.

- [7] 欧邦才. 基于 BP 神经网络的经济预测方法[J]. 南京工程学院学报(自然科学版), 2004(2): 11-14.
- [8] 潘水洋, 刘俊玮, 王一鸣. 基于神经网络的股票收益率预测研究[J]. 浙江大学学报(理学版), 2019, 46(5): 550-555.
- [9] 陈卓雷, 蒋寒迪. 基于 GARCH-BP 模型的股指预测及实证分析[J]. 当代财经, 2006(6): 41-44.