股票价格预测的最优选择模型

贺本岚

(重庆工商大学 数学与统计学院,重庆 400067)

摘　要:文章首先介绍了我国学者对股票价格指数的研究现状,并阐述了时间序列分析中两种常见的模型:自回归移动平均(ARIMA)模型和条件异方差(ARCH)模型。然后分别对上证指数近八年的346个有效收益数据进行建模,并对未来三个月的收盘价进行预测。结果表明,ARCH模型的整体预测效果优于ARIMA模型。

关键词:上证指数;时间序列;ARIMA模型;ARCH模型

中图分类号:F830.91　　**文献标识码**:A　　**文章编号**:1002-6487(2008)06-0135-02

证券市场股票的价格往往随时间变化而波动,股票的价格走势直接影响着投资者的经济利益以及不同行业的景气状况,也影响和反映着国家的宏观经济政策,被视为中国经济运行的"晴雨表",受到了人们的广泛关注。时间序列分析法是经济领域研究的主要工具之一,它采用合适的模型描述历史数据随时间变化的规律,并预测经济变量值。随着我国市场经济的发展,特别是证券市场的发展,时间序列分析方法将为投资者在证券市场上的正确投资战略决策提供有效依据。我国学者黄永兴[1]采用英国水文学家赫斯特提出的R/S分析(Rescaled range analysis)方法论证了我国证券市场为持久性随机过程,建立了沪深两市的AR—EGARCH模型;王波,张凤玲[2]利用时间序列中的ARIMA模型和人工神经网络建立了两类股票价格预测模型;张雅,肖冬荣,陈科燕,王东[3]等对股票价格分别建立了自回归模型和ARIMA模型并对股价进行了短期预测。徐绪松[4]对上海股票市场的GARCH效应进行了实证分析,得出GARCH模型可以很好地拟合沪市的波动情况。研究证明,对不同的序列、同一序列的不同时段可能会有不同的模型最适合它。

1 模型的简单介绍

1.1 Box-Jenkins 方法——ARIMA 模型

Box-Jenkins方法用变量X_t自身的滞后项以及随机误差来解释该变量,具体形式可表达成ARIMA(p,d,q)。其中p表示自回归过程阶数,d表示差分的阶数,q表示移动平均过程的阶数。

若时间序列是平稳的,可直接运用ARMA模型:

$$X_t-\varphi_1X_{t-1}-\varphi_2X_{t-2}-\cdots-\varphi_nX_{t-n}=a_t-\theta_1a_{t-1}-\theta_2a_{t-2}-\cdots-\theta_ma_{t-m}$$

若时间序列是非平稳的,则需要经过d阶差分,将非平稳时间序列转换成平稳时间序列。平稳时间序列可表示成:

$$X_t^*=(1-B)^dX_t$$

1.2 ARCH 模型

许多实际问题中,序列$\{x_t\}$(t=1,2,…,n)的随机扰动项的条件方差随着时间t的变化而变化,即序列具有变方差的特性。根据这一特点,Engle于1982年首次提出了条件异方差(ARCH)模型。

若有一随机过程$\{\varepsilon_t\}$,它的平方ε_t^2服从AR(q)过程

$$\varepsilon_t^2=\alpha_0+\alpha_1\varepsilon_{t-1}^2+\cdots+\alpha_q\varepsilon_{t-q}^2+\eta_t$$

其中,$\{\eta_t\}$独立同分布,且有$E(\eta_t)=0,D(\eta_t)=\lambda^2$,t=1,2,…,则称$\{\varepsilon_t\}$服从q阶的ARCH过程,记作$\varepsilon_t$~ARCH(q)。

ARCH(q)模型可以表示为:

$$\varepsilon_t=\sqrt{h_t}\cdot v_t$$

其中,$\{v_t\}$独立同分布,$E(v_t)=0,D(v_t)=1$;h_t可以表示为

$$h_t=\alpha_0+\alpha_1\varepsilon_{t-1}^2+\cdots+\alpha_q\varepsilon_{t-q}^2$$

2 实证分析

2.1 样本数据

本文采用上证指数从2000年1月7日至2006年12月29日的每周交易日的收盘价(共346个交易周)资料,选用SPSS统计软件对模型进行估计,用x_t表示从2000年1月7日算起的第t个交易周的收盘价(t=1,…,346)资料,分别采用Box-Jenkins方法和ARCH方法建立模型。

2.2 采用Box-Jenkins法建立时间序列模型

2.2.1 模型识别

首先根据ARIMA序列自相关函数及偏自相关函数的截尾、拖尾性质作模型识别(Identification or tentative specification)。

(1)上证综合指数序列的特性

通过绘制序列图(见图1),可以看出:$\{x_t\}$序列没有明显的周期性,表现出二次曲线波动现象,可以初步断定序列是非平稳的。

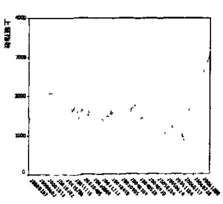

图1 上证综合指数序列图

因此,可对原始数据作二阶差分:

$W_t=\nabla^2 x_t=(1-B)^2 x_t$

此时，序列达到平稳。

(2)绘制 W_t 的自相关函数和偏自相关函数图。

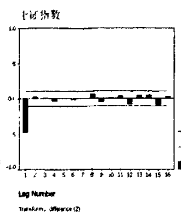

图 2 W_t 自相关函数图

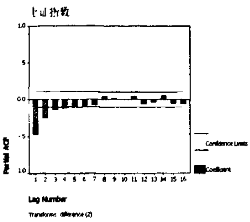

图 3 W_t 偏自相关函数图

W_t 自相关函数 2 阶截尾，偏自相关函数拖尾。故初步判断 W_t 适合 MA(2)模型。

2.2.2 模型定阶

在 MA 模型的情况下，通常采用统计学中方差分析的思想，通过假设检验来确定阶数。也就是以 MA(n)模型作为零假设，而以 MA(n+1)模型作为对立假设，然后增大 n 值，直到检验结果不显著为止。

(1) 对 W_t 拟合 MA(n)模型(n=1,2,…,6)，得残差平方和图(见图 4)。

由图 4 可知，模型阶数 n 从 1 升至 3，残差方差波动较小；n 继续升至 4，残差方差非但不减少，反而增加；继续升高模型阶数，在 n=6 时，残差方差回落但略有增加。

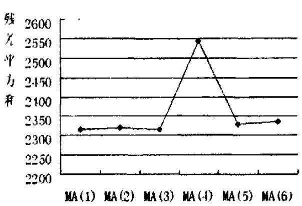

图 4 列差平方和

(2)F 检验定阶

根据图 4，将 MA(1)逐步升阶至 MA(4)，用 F 检验准则来判定阶数升高之后的模型与低一阶的模型之间是否存在显著性差异(如表 1)。建立原假设为：

$H_0:\theta_n=0$

记 Q_0 为 MA(n) 模型的残差平方和，Q_1 为 MA(n−1) 模型的残差平方和，则

$$F=\frac{(Q_1-Q_0)/S}{Q_0/(N-\gamma)}\sim F_{0.05}(S,N-\gamma)$$

其中，N 为样本长度，γ 是 MA(n)模型参数总个数。

表 1 模型拟合检验

	MA(1)	MA(2)	MA(3)	MA(4)
残差平方和	2315.2692	2319.3244	2315.7009	2543.1535
F	-0.5945	0.5305	-30.2298	30.9563
df	(1,340)	(1,339)	(1,338)	(1,337)
F_α	3.868962	3.869033	3.869118	3.869204

根据上表可知表明 MA(1)、MA(2)无显著性差异，根据建模的“约简”性原则，适合的模型为 MA(1)模型。即 W_t 适合拟合 MA(1) 模型。

(3)参数估计

用 SPSS 软件对模型各参数进行估计，得如下模型：

$W_t=a_t+0.92294a_{t-1}$

即

$(1-B)^2x_t=(1+0.9229B)a_t$

(4)模型的适应性检验

根据残差序列的自相关和偏自相关图可知，残差序列接近于白噪声序列，是随机分布的，可以认为 ARIMA(0,2,1)模型的识别是合适的。

2.2.3 对 ARIMA(0,2,1)模型进行条件期望预测

根据所建立的 ARIMA(0,2,1)模型对 2007 年 1 月至 3 月的上证指数进行预测，可得到如下预测方程：

$\hat{x}_t(l)=E(x_{t+l})$

$=E(2x_{t+l-1}+x_{t+l-2}+a_t+0.9229a_{t-1})=2\hat{x}_t(l-1)+\hat{x}_t(l-2)$

得到预测数据如表 2。

表 2 ARIMA(0,2,1)模型预测结果

日期	实际值	预测值	误差	相对误差(%)
20070105	2641.33	2720.786	-79.4561	3.01
20070112	2668.11	2682.876	-14.7663	0.55
20070119	2832.21	2708.956	123.2543	4.35
20070126	2882.56	2878.904	3.65649	0.13
20070202	2673.21	2929.427	-256.217	9.58
20070209	2730.39	2707.921	22.46917	0.82
20070216	2998.47	2766.167	232.3031	7.75
20070302	2831.53	3045.268	-213.738	7.55
20070309	2937.91	2868.188	69.72233	2.37
20070316	2930.48	2977.876	-47.3956	1.62
20070323	3074.29	2968.197	106.0931	3.45
20070330	3183.98	3117.041	66.93948	2.10
平均相对误差				3.61

注：绝对误差=实际值-拟合值，相对误差=绝对误差/实际值。

平均相对误差 $\bar{e}=\frac{1}{n}\sum_{i=1}^{12}\frac{|x_t-\hat{x}_t|}{x_t}=3.61\%$。可见预测效果良好。

2.3 采用 ARCH 方法建立模型

2.3.1 检验是否存在异方差

根据残差序列 e_t 的自相关系数图(见图 5)，残差序列存在某种程度的自相关，股价指数序列存在异方差。

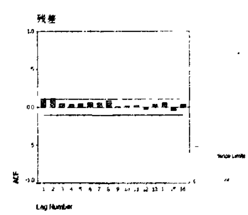

图 5 残差序列的自相关

2.3.2 ARCH 效应检验

序列是否存在 ARCH 效应，最常用的检验方法是拉格朗日乘数法，即 LM 检验。

经过反复筛选，建立模型：

$x_t=x_{t-1}+e_t$

$e_t=\sqrt{h_t}*v_t$

$h_t=1564.813+0.01257e_{t-2}^2$

ARCH 检验统计量

LM=9.261

以 α=0.05 查 χ^2 分布表，得到 $\chi^2_{0.05}(1)=3.814$，$\chi^2_{0.05}(2)=5.991$。这表明，在 5%的显著性水平下，序列 e^2 存在 ARCH 效应，可用 ARCH 模型描述序列的变化。

2.3.3 预测

采用上述模型对 2007 年 1 月至 3 月的上证指数进行预测，得到预测数据如表 3。

基于高频数据的上海股票市场隐性交易成本的实证

蒋 冠[a],田存志[b]

(云南大学 a.金融系;b.数学博士后流动站,昆明 650091)

摘 要:文章基于 Roll(1984)提出的序列协方差方法研究了上海股票市场的隐性交易成本。实证结果表明,在沪市,不同股票的隐性交易成本不一样;隐性交易成本在日内大致表现出"L"型变化特征。

关键词:隐性交易成本;非对称信息;序列协方差方法

中图分类号:F830.91 **文献标识码**:A **文章编号**:1002-6487(2008)06-0137-03

1 问题的提出

证券市场的交易成本的研究一直是金融市场微观结构理论研究的热点问题之一。Demsetz (1968) 将证券市场的交易成本分解为两部分:第一部分称之为显性交易成本,这主要包括委托费、佣金、印花税;第二部分称之为隐性交易成本,是证券市场信息非对称和市场供需不均衡时投资者买卖证券的风险补偿,一般用买卖报价价差来衡量。在报价驱动交易市场(,或做市商市场)中,买卖价差是可以直接观察到的,因而隐性交易成本可以很容易地估计出来,而在指令驱动交易市场(或竞价市场)中,买卖价差在大多数时候是不可观测和不可获得的,如何度量买卖价差乃至隐性交易成本是一个难于回答的问题。因此金融学界一直在努力探索指令驱动市场中的隐性交易成本的计量。

对于如何估算证券市场的隐性交易成本,文献中归纳起

表 3 ARCH 模型预测结果

日期	实际值	预测值	绝对误差	相对误差(%)
20070105	2641.33	2715.79	-74.46	2.82
20070112	2668.11	2695.63	-27.52	1.03
20070119	2832.21	2707.85	124.36	4.39
20070126	2882.56	2871.88	10.68	0.37
20070202	2673.21	2926.19	-252.98	9.46
20070209	2730.39	2713.17	17.22	0.63
20070216	2998.47	2776.39	222.08	7.41
20070302	2831.53	3038.54	-207.01	7.31
20070309	2937.91	2881.21	56.70	1.93
20070316	2930.48	2981.67	-51.19	1.75
20070323	3074.29	2971.80	102.49	3.33
20070330	3183.98	3113.86	70.12	2.20
平均相对误差			3.55	

2.4 模型比较

由以上两模型分析可以看出,ARCH 模型平均相对误差较 ARIMA 模型小,预测效果较好。两种方法都是根据相同股票的历史数据对未来的股价进行预测,而且都得到了较好的预测结果。但是相比较而言,ARCH 比 ARIMA 模型更加准确一些。这主要是由于股票数据的产生是随机的、不确定的,并且大多数数据都不含有白噪声,存在非线性性和异方差。而 ARIMA 模型对于股票价格变化这种非线性行为的分析、预测存在着先天的不足和缺陷,只考虑时间序列本身的特性进行预测,却没有考虑到股市本身受不可预测的许多复杂因素影响,如政治因素、政府的宏观调控等。这些因素在 ARIMA 模型中只能以随机扰动项表示,而在预期的期望值中却无法表现出来。另外,这种传统的多元回归模型均为齐方差性模型,且假定模型残差的均值为零、方差为常数,但实际上我国股价指数序列往往存在异方差现象。

3 结论

本文运用 ARIMA 模型和 ARCH 模型对我国上证指数进行了实证研究。结果表明,由于金融时间序列,尤其是股价指数存在异方差,ARCH 模型的拟合效果优于 ARIMA 模型的拟合效果。ARCH 模型能显著地描述股价波动的不确定性,有效地处理随机误差项的异方差问题,是一种对不确定性进行描述的有效方法。

从以上的分析,可以看出上海股票市场总体趋于上升趋势,但市场收盘价的波动较大。

参考文献:

[1]黄永兴.我国股价指数的时间序列模型[J].安徽工业大学学报,2002,(4).

[2]王波,张凤玲.神经网络与时间序列模型在股票预测中的比较[J].武汉理工大学学报, 2005,(6).

[3]张雅,肖冬荣,陈科燕,王东.时间序列模型对深证成指的预测分析[J].统计与决策,2003,(6).

[4]徐绪松.我国上海股票市场 GARCH 效应实证研究[J].武汉大学学报,2002,(6).

(责任编辑/亦 民)