DOI:10.13529/j.cnki.enterprise.economy.2012.08.001

基于自回归条件异方差模型的上证指数收益波动分析

□孙海涛

[摘　要] 股票市场的股价波动会引起股价指数的涨跌。本文基于计量经济学的自回归条件异方差模型，对上证指数 2007 年以来的 1276 个交易日的样本数据进行实证研究。结果表明: 上证指数收益序列具有尖峰厚尾特征和波动的时段集群特征，适合利用自回归条件异方差模型进行分析及预测。研究结果为各方从不同角度把握上证指数收益波动的规律提供了股票投资理论和方法。

[关键词] 回归条件异方差; 收益率; 波动研究

[中图分类号] F832.5; F224　[文献标识码] A　[文章编号] 1006－5024(2012)08－0168－04

[基金项目] 山东省哲学社会科学项目"低碳经济视野下山东省经济可持续发展研究"(批准号: 11CJJZ12)

[作者简介] 孙海涛，青岛理工大学经贸学院副教授，研究方向为经济统计学、系统评价理论与应用。(山东 青岛 266520)

Abstract: The stock price fluctuation in stock market causes the falling and rising of stock index. Based on econometrics autoregressive conditional heteroscadicity, this paper conducted empirical study on 1276 trading days of the sample data of Shanghai index since 2007. The results show that Shanghai index return sequence features fluctuation cluster during a period of time and sharp peak and thick tail, which is suitable for applying regressive conditional heteroscadicity model to conduct analysis and prediction. The research results provided scientific theory and methods to grasping the law of shanghai index return fluctuation and then conduct stock investment.

Key words: regressive conditional heteroscadicity; yields rate; volatility research

国内外学者研究发现，某些时间序列特别是金融时间序列，常常会出现某一特征值成群出现的情况。对时间序列进行建模，其随机干扰项往往在较大幅度波动后面伴随着较大幅度的波动，在较小幅度波动后面伴随着较小幅度的波动，这就是波动的集群特征。

上海证券交易所自1990年11月26日成立以来，经过20多年的积淀和发展，已成为国内首屈一指的市场，上市公司数量、上市股票数量、市价总值、流通市值、证券成交总额、股票成交金额和国债成交金额等各项指标均居全国首位。1990年上证指数收盘价为127.61%，交易量仅为7800股，至2012年3月30日上证指数收盘价为2262.788%，交易量达到62.02亿股。2007年10月16日，上证指数创历史最高点，达到6124.044%。之后由于受全球性金融危机的冲击，上证指数一路走低，于2008年10月28日降至1664.925%。股票市场上证指数的波动，必将引起众多利益相关者的密切关注。因此，加强对股票市场收益波动的分析，具有重要的现实意义。

一、文献综述

针对上证指数收益波动问题，国内一些学者选取不同时段的样本数据从不同角度进行了有益的探讨。王真真等用 ARCH－M 模型对上证日收盘价序列进行了分析和实证研究，结果表明，上证股价指数存在条件异方差特性，并表现出非正态性; 然后用该模型对上证日收盘价进行短期预测，结果表明，该模型能够较好地反映上证股价指数的变化特征。曾慧采用 ARCH 模型及其扩展模型对上证综合指数的波动性进行实证研究，结果显示，我国股票市场收益有明显的尖峰厚尾性、波动集簇

性以及波动的信息不对称性等特点，从而表明我国股票市场与发达国家成熟市场有相似之处。许爱霞通过基于正态分布和 t 分布的 GARCH 模型对沪市行业指数的波动性进行了比较分析，实证结果表明，基于 t 分布的 GARCH 模型能更精确地描述股市的波动性。此外，文章还用 EGARCH 模型检验了市场波动的不对称性，实证结果表明沪市行业指数除地产指数外都存在明显的“杠杆效应”，市场冲击曲线也能直观说明股票市场的“杠杆效应”。张雪蓉等应用 TARCH_M 模型，并且引入迭代累计平方和法则对上证指数波动时段进行划分，针对不同波动时段分析上证指数日收益率上涨和下跌对上海股票市场非对称的影响特点。

二、自回归条件异方差模型的基本理论

(一) ARCH 模型

1982 年 Engle 提出了自回归条件异方差模型，该模型通常用于对回归或自回归模型的随机干扰项进行建模，以便能够更加充分地反映随机干扰项在主体模型中的作用，最终使模型的随机残差项成为白噪声。一般而言，对于一个平稳的随机变量 x_t，其自回归模型 AR(p) 可以表述为:

$$x_t = = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots \phi_p x_{t-p} + \varepsilon_t$$

其中，随机误差项 ε_t 服从条件均值为零，条件方差为 σ^2_t 的正态分布。

若 ε_t 的条件方差可以表述为 $\sigma^2_t = E(\varepsilon^2_t) = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \cdots + \alpha_q \varepsilon^2_{t-q} + \eta$，其中 η_t 为白噪声，且满足 $E(\eta_t) = 0$、$D(\eta_t) = \lambda^2$。α_i 取值非负且 $\sum \alpha_i < 1$，则 ε_t 的条件方差称为自回归条件异方差模型，即 $\varepsilon_t \sim ARCH(q)$。

由自回归条件异方差模型可以看出：首先，ε_t 的方差由 ε^2_{t-1}、…、ε^2_{t-q} 决定。当 ε_{t-1} 很大时 σ^2_t 也很大，说明 x_t 在 t－1 期的一个大的波动可能导致它在 t 期的大波动。其次，滞后期 q 的值决定随机变量 x_t 的一个波动所持续的时间长短。q 值越大，波动持续的时间就越长。因此，在金融时间序列中一个大的误差 ε_t 可能伴随另一个大的误差，而一个小的误差可能伴随另一个小的误差，并且这种变化会经常出现时段的集群特征。

(二) GARCH 模型

为解决 ARCH(q)模型中 q 值很大且难以确定问题，1986 年 Bollerslev 提出广义自回归条件异方差模型，该模型通常也用于对回归或自回归模型的随机干扰项进行建模。其中一个简单而应用广泛的模型是 GARCH(1，1)，它是在 ARCH(1) 模型的基础上加入自回归项 σ^2_{t-1} 构成的。GARCH(1，1)模型的形式为:

$$\sigma^2_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \lambda \sigma^2_{t-1}$$

其中 ε_{t-1} 为 ARCH 项，σ_{t-1} 为 GARCH 项。并要求 α_0、α_1、λ 为非负，且$0 \le \alpha_1 + \lambda < 1$。

(三) EGARCH 模型

EGARCH 模型也称指数 GARCH 模型，它是由 Nelson 于 1991 年提出的。

模型的条件方差为:

$$\log(\sigma^2_t) = \alpha_0 + \lambda \log(\sigma^2_{t-1}) + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

等式左边是条件方差的对数，这意味着杠杆影响是指数的，而不是二次的，所以，条件方差的预测值一定是非负的。杠杆效应的存在能够通过 $\gamma < 0$的假设得到检验。

当 $\gamma < 0$时，利好消息($u_t > 0$)和利坏消息($u_t < 0$)对条件方差的冲击为: 利好消息有一个 $\alpha + \gamma$ 倍的冲击; 利坏消息有一个 $\alpha + \gamma \times (-1)$倍的冲击。如果 $\gamma \neq 0$，则信息的作用是非对称的。

三、样本数据

本文选取 2007 年 1 月至 2012 年 3 月共 1276 个交易日的上证指数收盘价数据，最小值 1706. 703，最大值为 6092. 507。上证指数原序列以 SZ 表示，取对数一阶差分后的序列为收益率，以 r 表示。依据这些数据得到上证指数收益率 r 的变化趋势图(图 1)和收益率的统计特征图(图 2)。

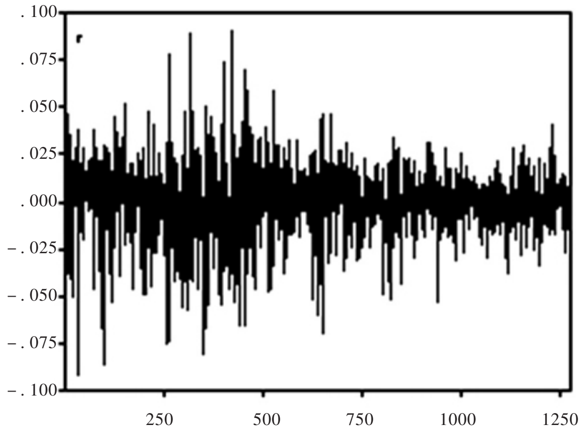

图 1

由图 1 可以看出，在样本区间内，上证指数收益率的波动较大，变化很剧烈，在有的时间范围内波动较小，而在有的时间范围内波动较大，表现出波动的集群特征。对收益率序列进行单位根检验发现，该序列为一平稳时间序列。由图 2 可以看出，收益率偏态系数为 －0. 355736，说明样本呈现左偏分布；峰态系数为

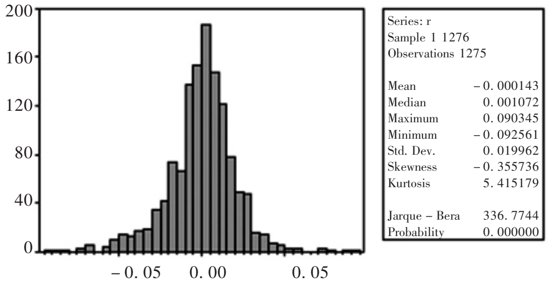

图 2

5.415179，明显大于正态分布的峰度值 3，说明样本的分布具有尖峰厚尾的特征；由于 Jarque-Bera 值为 336.7744，相应的概率为 0，可以判定收益率序列不服从正态分布。因此，对该序列的分析应适用 ARMA(p,q) 模型。根据反复实验，本文认为收益率序列适合建立 ARMA(5,5) 模型采用 OLS 估计得到 ARMA(5,5) 模型为：

$$r_t = 0.295r_{t-1} - 0.739r_{t-2} + 0.884r_{t-3} - 0.2982r_{t-4} + 0.798r_{t-5} - 0.301\varepsilon_{t-1} + 0.742\varepsilon_{t-2} - 0.872\varepsilon_{t-3} + 0.338\varepsilon_{t-4} - 0.816\varepsilon_{t-5}$$

对 ARMA(5,5) 模型残差序列进行拉格朗日乘数检验发现，残差序列不存在序列自相关性，说明对上证指数收盘价取对数一阶差分后的收益率序列所建立的 ARMA(5,5) 模型是合理的。对模型残差序列进行 ARCH 效应检验结果 LM = 20.76789，相应的显著性概率为 0，小于显著水平为 5% 的概率，认为残差序列存在自回归条件异方差，应该考虑在 ARMA(5,5) 模型的基础上建立 ARCH 模型，以便更加准确地反映上证指数收益率波动的本质特征。

表 1

方程	变量	回归系数	标准误差	z-统计量	显著性概率
均值方程	AR(1)	0.876490	0.091556	9.573278	0.0000
	AR(2)	-0.924402	0.158251	-5.841380	0.0000
	AR(3)	-0.482656	0.217995	-2.214074	0.0268
	AR(4)	0.537426	0.148911	3.609040	0.0003
	AR(5)	-0.821523	0.084588	-9.712043	0.0000
	MA(1)	-0.865063	0.083084	-10.41189	0.0000
	MA(2)	0.885734	0.143126	6.188478	0.0000
	MA(3)	0.546916	0.195371	2.799366	0.0051
	MA(4)	-0.589131	0.134633	-4.375828	0.0000
	MA(5)	0.854204	0.077005	11.09289	0.0000
方差方程	C	0.000139	1.36E-05	10.20862	0.0000
	RESID(-1)^2	0.074084	0.024474	3.027098	0.0025
	RESID(-2)^2	0.043039	0.027613	1.558650	0.1191
	RESID(-3)^2	0.161225	0.027639	5.833343	0.0000
	RESID(-4)^2	0.046541	0.027639	1.683908	0.0922
	RESID(-5)^2	0.067125	0.020459	3.281003	0.0010
	RESID(-6)^2	0.105293	0.031584	3.333769	0.0009
	RESID(-7)^2	0.104369	0.031921	3.269640	0.0011
	RESID(-8)^2	0.067052	0.027548	2.433992	0.0149

四、上证指数自回归条件异方差模型估计

(一) ARCH 模型估计

在 ARMA(5,5) 模型的基础上，经反复模拟并依据赤池信息准则，建立滞后期为 8 的 ARCH(8) 模型，模型估计结果如表 1。

由表 1 可确定 ARCH(8) 模型为：

均值方程：

$$r_t = 0.876r_{t-1} - 0.924r_{t-2} - 0.483r_{t-3} + 0.537r_{t-4} - 0.822r_{t-5} - 0.865\varepsilon_{t-1} + 0.886\varepsilon_{t-2} + 0.547\varepsilon_{t-3} - 0.589\varepsilon_{t-4} + 0.854\varepsilon_{t-5}$$

ARCH(8) 方程：

$$\sigma^2_t = 0.0001 + 0.074\varepsilon^2_{t-1} + 0.043\varepsilon^2_{t-2} + 0.161\varepsilon^2_{t-3} + 0.047\varepsilon^2_{t-4} + 0.067\varepsilon^2_{t-5} + 0.105\varepsilon^2_{t-6} + 0.104\varepsilon^2_{t-7} + 0.067\varepsilon^2_{t-8}$$

对 ARCH(8) 模型残差序列滞后 1 期进行 ARCH 效应检验，结果 LM = 0.040355，相应的显著性概率为 0.8408，远大于显著水平为 5% 的概率，则残差序列不存在自回归条件异方差。

(二) GARCH(1,1)模型估计

因 ARCH 模型中的滞后项太多，给实际建模带来很多困难，因此考虑建立 GARCH(1,1)模型。模型 GARCH(1,1)估计的均值方程和条件方差方程如表 2。

表 2

方程	变量	回归系数	标准误差	z-统计量	显著性概率
均值方程	AR(1)	-1.115090	0.025376	-43.94285	0.0000
	AR(2)	-0.808437	0.028530	-28.33612	0.0000
	AR(3)	-1.047889	0.027667	-37.87532	0.0000
	AR(4)	-0.420928	0.025868	-16.27221	0.0000
	MA(1)	1.116817	0.002858	390.7245	0.0000
	MA(2)	0.795979	0.004443	179.1469	0.0000
	MA(3)	1.068309	0.006571	162.5699	0.0000
	MA(4)	0.436835	0.005188	84.20160	0.0000
方差方程	C	2.14E-06	9.05E-07	2.366920	0.0179
	RESID(-1)^2	0.042946	0.007099	6.049214	0.0000
	GARCH(-1)	0.951133	0.007849	121.1842	0.0000

由表 2 可确定 GARCH(1,1)模型为：

均值方程：

$$r_t = -1.115r_{t-1} - 0.808r_{t-2} - 1.048r_{t-3} - 0.421r_{t-4} + 1.117\varepsilon_{t-1} + 0.796\varepsilon_{t-2} + 1.068\varepsilon^2_{t-3} + 0.437\varepsilon^2_{t-4}$$

GARCH(1,1) 方程：

$$\sigma^2_t = 0.00000214 + 0.043\varepsilon^2_{t-1} + 0.951\sigma^2_{t-1}$$

对残差序列滞后 1 期进行 ARCH 效应检验，结果 LM = 0.18956，相应的显著性概率为 0.6633，远大于显著水平为 5% 的概率，可认为残差序列不存在自回归条件异方差，模型设定正确。

(三) EGARCH(1,1)模型估计

根据掌握的数据资料建立 EGARCH(1, 1) 模型，得到相应的估计模型的均值方程和条件方差方程如表 3。

表 3

方程	变量	回归系数	标准误差	z-统计量	显著性概率
均值方程	AR(1)	-0.303145	0.025598	-11.84259	0.0000
	AR(2)	0.393119	0.100790	3.900384	0.0001
	AR(3)	-0.198496	0.056471	-3.514987	0.0004
	AR(4)	0.542186	0.108822	4.982329	0.0000
	AR(5)	0.485699	0.024253	20.02622	0.0000
	MA(1)	0.306384	0.001327	230.9616	0.0000
	MA(2)	-0.401525	0.101800	-3.944239	0.0001
	MA(3)	0.233586	0.056154	4.159737	0.0000
	MA(4)	-0.517737	0.109191	-4.741594	0.0000
	MA(5)	-0.494337	0.008292	-59.61931	0.0000
方差方程	C	-0.174889	0.037205	-4.700724	0.0000
	ABS(RESID(-1)/@SQRT(GARCH(-1)))	0.107975	0.015234	7.087803	0.0000
	RESID(-1)/@SQRT(GARCH(-1))	-0.030161	0.006991	-4.314284	0.0000
	LOG(GARCH(-1))	0.988472	0.004022	245.7504	0.0000

由表 3 可确定 EGARCH(1,1)模型为:

均值方程:

$$r_t = -0.303r_{t-1} + 0.393r_{t-2} - 0.198r_{t-3} + 0.542r_{t-4} + 0.486r_{t-5} + 0.306\varepsilon_{t-1} - 0.402\varepsilon_{t-2} + 0.234\varepsilon_{t-3} - 0.518\varepsilon_{t-4} + 0.494\varepsilon_{t-5}$$

EGARCH(1,1)方程:

$$\log(\sigma^2_t) = -0.175 + 0.988\log(\sigma^2_{t-1}) + 0.108\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - 0.030\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

对残差序列滞后1期进行 ARCH 效应检验结果 LM = 0.115592，相应概率为0.7339，远大于显著水平为 5% 的概率，可认为残差序列不存在自回归条件异方差。

EGARCH(1,1)模型中，$\alpha = 0.108$，表明利好消息有一个 0.108 倍的冲击；$\gamma = -0.03$，说明信息的作用是不对称的，杠杆效应显著，利坏消息有一个 0.138 倍的冲击。根据估计出的 EGARCH(1,1)模型结果，绘制相应信息曲线如图 3 所示。

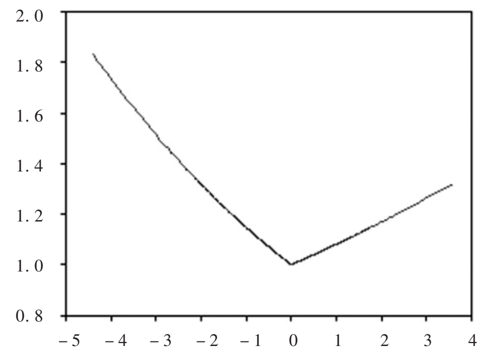

图 3　TARCH(1,1)估计模型的信息冲击线

在图 3 的信息曲线上，信息冲击小于 0 时代表负冲击，线段变化比较陡峭，而在正冲击时线段变化相对比较平缓。图 3 的信息曲线说明负冲击使得序列波动的变化更大。

五、结论

本文利用 2007 年 1 月至 2012 年 3 月共 1276 个交易日上证指数收盘价的样本，借助自回归条件异方差模型分析发现，我国股票市场与西方发达国家的成熟市场有着共同特性，均表现出尖峰厚尾特征、波动时段集群特征、杠杆效应等，但由于我国的股票市场产生较晚，并发展于计划经济向市场经济转型时期，因而又具有与其他股票市场的不同特点。

根据对上证指数的实证分析得出以下结论:

1. 上证指数属于非平稳的时间数列，但取对数后的一阶差分后变为平稳数列，即上证指数属于一阶单整数列;

2. 收益率序列表明，收益率具有尖峰厚尾特征和波动时段集群特征，适合利用自回归条件异方差模型进行分析及预测;

3. 由于收益率序列的统计特征 Jarque - Bera 值为 336.7744，相应的概率为 0，因而收益率序列不服从正态分布;

4. 在 EGARCH(1, 1) 模型中 γ 为负并在统计上是显著的，这表明在样本期间沪市的股票收盘价格指数中存在杠杆效应，即利坏消息比利好消息对股票综合指数的波动影响更大。

参考文献:

[1]易丹辉. 数据分析与 EViews 应用[M]. 北京: 中国统计出版社,2002.

[2]张晓峒. EViews 使用指南与案例[M]. 北京: 机械工业出版社,2008.

[责任编辑: 李小玉]