

数理统计在股票收益中的应用研究

陈锐

湛江科技学院经济与金融学院 广东湛江 524000

摘要: 数理统计能够发现事物随机现象的统计规律, 利用好数理统计知识, 能够更好的规避风险, 从而在实际生活中, 具有广泛的应用。文章基于数理统计中的ARCH模型, 以贵州茅台股票收盘价数据, 来实证说明数理统计在金融数据研究中的应用, 同时也实证说明GARCH模型能够很好的拟合股票数据。

关键词: 数理统计; GARCH模型; 贵州茅台

前言:

统计作为大学众多专业的基础课程, 其知识渗透到众多课程领域, 重要意义不言而喻。从局部到整体, 从描述到推断, 数据的背后往往蕴含着规律, 而这个规律往往需要运用数理统计知识去发掘, 从而运用于生活解决实际问题。文章通过选取贵州茅台数据, 基于ARCH模型进行研究, 最后分析得出以GARCH模型^[1, 2]来拟合股票收益率数据效果非常好, 并通过各项检验, 从而来实证说明数理统计在金融股票数据研究中的运用, 同时也实证说明GARCH模型能够很好的拟合股票数据。

一、数据预处理

文章选取贵州茅台2017年的股票收盘价(SP)数据, 通过EViews软件对数据预处理, 绘制出股票收盘价动态曲线, 如图1所示。

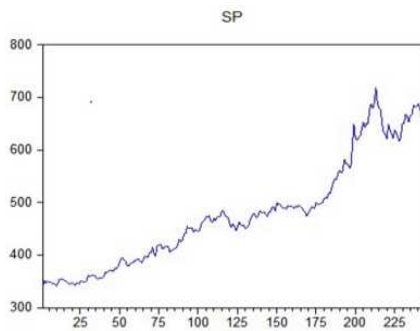


图1 贵州茅台2017年收盘价曲线图

通过曲线图1, 我们可以看出贵州茅台在2017年的收盘价(SP)随时间呈现明显的上升趋势, 从图形中我们可以看出, 在2017年9月之前, 曲线上升比较平稳, 呈直线上升; 在2017年9月之后, 上升的趋势明显加快, 趋势明显增强。然后通过EViews软件绘制自相关分析图2, 可以看出收盘价(SP)序列趋势明显。

作者简介: 陈锐(1994—), 男, 土家族, 湖北巴东人, 湛江科技学院助教, 硕士, 研究方向: 金融风险测度。

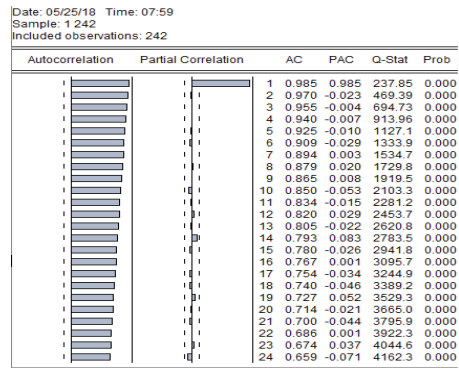


图2 收盘价(SP)自相关-偏自相关分析图

我们对贵州茅台收盘价(SP)时间序列进行一阶差分处理, 得到如图3所示。

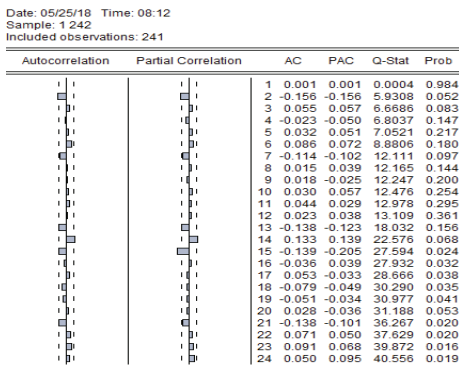


图3 贵州茅台收盘价1阶差分自相关-偏自相关图

通过图3我们可以看出, 一阶差分后我们去掉了趋势, 在显著性水平 $\alpha=0.05$ 时的置信带中, 大部分自相关系数都落入在随机区间, 所以我们考虑用一阶自回归模型, 即:

$$y_t = a + \phi_1 y_{t-1} + \varepsilon_t \quad (1)$$

通过观察图3, 残差独立性检验的相伴概率对应Prob下第24行为0.019, 远小于给定显著性水平95%, 说明残差序列不独立。通过一阶差分序列平方的自相关-偏自相关图4, 能够看出自相关拖尾, 偏自相关拖尾, 因此我们考虑对残差序列的平方做模型, 即考虑

ARCH模型。

Date: 05/27/18 Time: 08:34
Sample: 1 242
Included observations: 241

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1		0.472	0.472	54.247	0.000
2		0.100	-0.158	56.687	0.000
3		0.051	0.093	57.316	0.000
4		0.180	0.172	65.359	0.000
5		0.179	0.012	73.292	0.000
6		0.111	0.032	76.363	0.000
7		0.125	0.102	80.300	0.000
8		0.130	0.012	84.549	0.000
9		0.100	0.011	87.060	0.000
10		0.140	0.117	92.029	0.000
11		0.136	-0.003	96.726	0.000
12		0.102	0.011	99.391	0.000
13		0.128	0.102	103.63	0.000
14		0.229	0.136	117.16	0.000
15		0.291	0.124	139.11	0.000
16		0.209	0.028	150.48	0.000
17		0.076	-0.048	151.98	0.000
18		0.116	0.093	155.54	0.000
19		0.141	-0.023	160.76	0.000
20		0.067	-0.085	161.96	0.000
21		0.097	0.110	164.45	0.000
22		0.094	-0.061	166.80	0.000
23		0.118	0.035	170.53	0.000
24		0.057	-0.039	171.40	0.000

图4 一阶差分序列平方自相关-偏自相关图

二、ARCH模型引入

对于通常的回归模型

$$y_t = x_t' \beta + \varepsilon_t \quad (2)$$

如果随机扰动项的平方 ε_t^2 服从 $AR(q)$ 过程，即

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \eta_t, \quad t=1,2,\dots \quad (3)$$

式中， η_t 独立同分布，并满足 $E(\eta_t) = 0, D(\eta_t) = \lambda^2$ ，则称模型 (3) 是自回归条件异方差模型，简记为 ARCH 模型。称序列 ε_t 服从 q 阶的 ARCH 过程，记作 $\varepsilon_t \sim ARCH(q)$ 。

ARCH 模型通常用于对主题模型的随机扰动项进行建模，以更充分地提取残差中的信息，使最终的模型残差项 η_t 成为白噪声。因此，对于 $AR(p)$ 模型

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (4)$$

如果 $\varepsilon_t \sim ARCH(q)$ ，则序列 y_t 可以用 $AR(p) - ARCH(q)$ 模型描述。

为方便研究并与其他拓展形式相联系， $ARCH(q)$ 模型又可以表示为：

$$\varepsilon_t = \sqrt{h_t} \cdot v_t \quad (5)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad (6)$$

式中， v_t 独立同分布，且 $E(v_t) = 0, D(v_t) = 1; \alpha_0 > 0, \alpha_i \geq 0 (i=1,2,\dots,q)$ ，且 $\sum_{i=1}^q \alpha_i < 1$ (保证 ARCH 过程平稳)。

对于任意时刻 t ， ε_t 的条件期望为：

$$E(\varepsilon_t | \varepsilon_{t-1}, \dots) = \sqrt{h_t} \cdot E(v_t) = 0 \quad (7)$$

条件方差为：

$$E(\varepsilon_t^2 | \varepsilon_{t-1}, \dots) = h_t \cdot E(v_t^2) = h_t \quad (8)$$

式 (8) 反映了序列条件方差随时间而变化的性质^[3, 4]。

三、ARCH 效应检验

序列是否存在 ARCH 效应，最常用的检验方法

是拉格朗日乘法，即 LM 检验。若模型随机扰动项 $\varepsilon_t \sim ARCH(q)$ ，则可以建立辅助回归方程

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \quad (9)$$

检验序列是否存在 ARCH 效应，即检验式 (9) 中的所有回归系数是否同时为 0。若所有回归系数同时为 0 的概率较大，则序列不存在 ARCH 效应；若所有的回归系数同时为 0 的概率很小，或至少有一个系数显著不为 0，则序列存在 ARCH 效应。检验的原假设和备择假设为：

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0 \quad H_1 : \exists \alpha_i \neq 0 \quad (1 \leq i \leq q)$$

检验统计量

$$LM = nR^2 \sim \chi^2(q) \quad (10)$$

式中， n 是计算辅助回归式 (10) 时的样本数据个数； R^2 是辅助回归式 (10) 的决定系数 (采用最小二成估计)。

给定显著性水平 α 和自由度 q ，如果 $LM > \chi_a^2(q)$ ，则拒绝 H_0 ，认为序列存在 ARCH 效应；如果 $LM \leq \chi_a^2(q)$ ，则不能拒绝 H_0 ，说明序列不存在 ARCH 效应。

如图 5、图 6，通过对残差序列 ARCH 效应检验，我们发现，阶数为 83 时 χ^2 检验的相伴概率 p 值为 0.0408，小于显著性水平 0.05，拒绝原假设，说明残差序列存在滞后阶数为 83 的 ARCH 效应；而阶数为 84 时 χ^2 检验的相伴概率 p 值为 0.0522，大于显著性水平 0.05，不能拒绝原假设，说明残差序列不存在滞后阶数为 84 的 ARCH 效应。

Heteroskedasticity Test: ARCH

F-statistic	1.856432	Prob. F(83,74)	0.0036
Obs*R-squared	106.7382	Prob. Chi-Square(83)	0.0408

图5 滞后阶数为 83 的 ARCH 效应检验

Heteroskedasticity Test: ARCH

F-statistic	1.785217	Prob. F(84,72)	0.0062
Obs*R-squared	106.0715	Prob. Chi-Square(84)	0.0522

图6 滞后阶数为 84 的 ARCH 效应检验

四、GARCH 模型

若直接考虑最简单的 ARCH 模型，它的滞后阶数要考虑到 83 阶，检验通不过，所以我们考虑构造 GARCH (1, 1) 模型。

与 ARCH 模型一样，GARCH (1, 1) 模型通常也用于对回归或自回归模型的随机扰动项进行建模。若式 (6) 可以写成下面的形式：

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \theta_1 h_{t-1} + \dots + \theta_p h_{t-p} = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \theta_j h_{t-j} \quad (11)$$

则称序列服从 GARCH (p, q) 过程。

GARCH (1, 1) 模型：

$$\begin{cases} \varepsilon_t = \sqrt{h_t} \cdot v_t \\ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \theta_1 h_{t-1} \end{cases} \quad (12)$$

参数的约束:

$$\begin{cases} (1) \alpha_0 > 0, \alpha_1 \geq 0, \theta_1 \geq 0 \\ (2) \alpha_1 + \theta_1 < 1 \end{cases} \quad (13)$$

五、建立模型

我们利用EViews软件,得到如图7,通过图7,我们能够得出收盘价序列(SP)的条件均值模型(1)中的 $a = -0.4062$, $\phi_1 = 1.0039 = -0.402$ 即:

$$SP = -0.4062 + 1.0039SP(-1) + \varepsilon_t \quad (14)$$

同样残差序列平方模型(12)中的

$$\alpha_0 = 1.6516 \quad \varepsilon_1 = 0.1871 \quad \theta_1 = 0.8059, \text{ 即}$$

$$h_t = 1.6516 + 0.1871\varepsilon_{t-1}^2 + 0.8059h_{t-1} \quad (15)$$

我们可以看出。

模型的显著性:通过图3.7中SP(-1)对应的系数检验的结果P值为 $0.000 < 0.05$,说明拟合的效果非常显著;我们再看残差序列平方模型的参数估计值,RESID(-1)²与GARCH(-1)对应的系数检验的P值分别为0.0010、0.0000,都小于显著性水平0.05,说明拟合结构显著。

模型的合理性:由参数的约束条件(13),有:

$$\begin{cases} \alpha_0 = 1.6516 > 0, \alpha_1 = 0.1871 \geq 0, \beta_1 = 0.8059 \geq 0 \\ \alpha_1 + \beta_1 = 0.1871 + 0.8059 = 0.993 < 1. \end{cases}$$

即拟合的模型满足合理性。

其他性检验:由图3.7,残差序列平方GARCH(1,1)模型检验的拟合优度 $R^2 = 0.9926$,以及调整后的拟合优度值 $\bar{R}^2 = 0.9926$,拟合的效果都非常的好,模型(15)中F检验 $P = 0.000 < 0.05$,模型显著,同时D-W检验值为1.9679,非常接近2.0,说明残差序列模型是无自相关。

Dependent Variable: SP
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 05/25/18 Time: 08:27
Sample (adjusted): 2 242
Included observations: 241 after adjustments
Convergence achieved after 65 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.406228	2.541225	-0.159855	0.8730
SP(-1)	1.003948	0.005655	177.5340	0.0000
Variance Equation				
C	1.651553	1.185512	1.393113	0.1636
RESID(-1)^2	0.187132	0.056684	3.301336	0.0010
GARCH(-1)	0.805907	0.057124	14.10794	0.0000
R-squared	0.992608	Mean dependent var	477.1124	
Adjusted R-squared	0.992577	S.D. dependent var	102.0739	
S.E. of regression	8.794231	Akaike info criterion	6.801205	
Sum squared resid	18483.90	Schwarz criterion	6.873503	
Log likelihood	-814.5452	Hannan-Quinn criter.	6.830333	
F-statistic	8023.484	Durbin-Watson stat	1.967946	
Prob(F-statistic)	0.000000			

图7 GARCH模型参数估计与检验结果

故我们通过对贵州茅台2017年一年的收盘价(SP)数据进行分析,拟合出GARCH(1,1)模型:

$$\begin{cases} SP = -0.4062 + 1.0039SP(-1) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} \cdot v_t \\ h_t = 1.6516 + 0.1871\varepsilon_{t-1}^2 + 0.8059h_{t-1} \end{cases} \quad (16)$$

六、模型的预测

通过构造的GARCH(1,1)模型,我们预测了贵州茅台2017年股票收盘价(SP)的曲线图,如图8中左上图中所示的蓝线,而两侧的红线则表示在2倍的标准差下的预测区间,图8中的左下图蓝线是预测的方差 h_t 的图形;同时,我们通过图8右侧图表可以看出,平均绝对百分误差MAPE=1.185811,说明我们通过拟合的GARCH(1,1)模型来预测的贵州茅台2017年股票收盘价(SP)精度是非常高的,即可信度是非常高的。

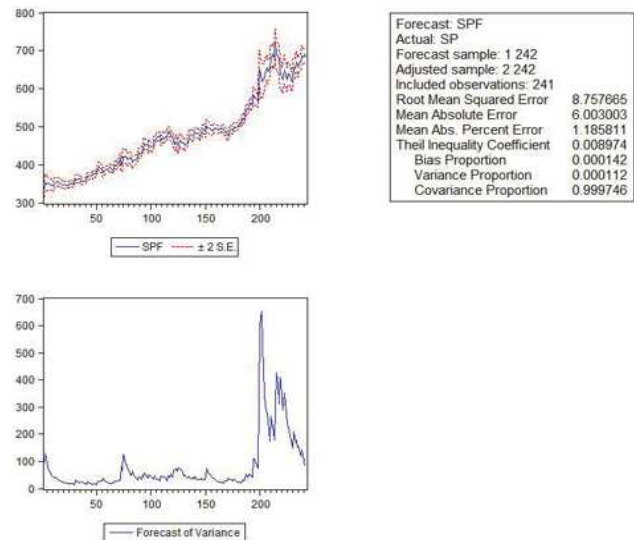


图8 收盘价预测图及残差图

七、总结

文章基于贵州茅台股票数据建立GARCH模型,通过模型估计、检验和预测,得出GARCH模型能够很好的拟合原始数据。从而证实数理统计知识作为研究随机现象统计规律的一门学科的重要性的广泛性;也证实ARCH类模型对金融股票数据具有很好的拟合效果。

参考文献:

- [1]冀南南,杨天兴.基于GARCH族模型股票收益的波动性研究[J].中国管理信息化,2021,24(23):128-129.
- [2]廖欣昱,李喜梅,古雨禾,张天立.基于GARCH族模型的深证指数波动性研究[J].中国商论,2021(08):94-97.DOI:10.19699/j.cnki.issn2096-0298.2021.08.094.
- [3]张世英,柯珂.ARCH模型体系[J].系统工程学报,2002(03):236-245.
- [4]钱争鸣.ARCH族计量模型在金融市场研究中的应用[J].厦门大学学报(哲学社会科学版),2000(03):126-129.