

基于 GARCH 模型的上证 50 的比较研究

胡后燕

(安徽大学 经济学院,安徽 合肥 230601)

摘 要:2004 年 1 月 2 日发布了上证 50 指数,它是从上海证券市场里面挑选出的在规模,流动性方面比较有代表性的 50 只股票作为样本股,利用科学的方法编制而成的指数,主要用来总体反映上海金融市场非常具有实力的一些大型企业的发展状况。通过利用 GARCH 模型族以及 VAR 的方法分析上证 50 的对数收益率,研究上证 50 指数的波动性,杠杆效应以及风险性,得出在正态分布、T 分布以及广义误差分布下的 VAR 值,然后选择合适的模型,为规范国内上证 50 风险管理提供依据。

关键词:波动性;GARCH 模型;VAR 计算;上证 50

分类号:F224 **文献标识码:**A **文章编号:**1673-1395 (2018)04-0073-07

一、VaR 及波动率介绍

(一) VaR 的描述

VaR 一般表示在一定风险状态下的价值,即在一定的置信水平下某个既定的持有时间内,某种金融序列或者金融组合可能会遭受的最大限度上的损失额度。对于 VAR 的理解需要掌握几个要点:首先是置信水平,VAR 不是一般意义上的损失额而是对应置信水平的最大损失额。其次是持有期 VAR 是对应持有期的潜在的最大损失额。最后就是正常市场, VaR 不是极端市场的最大损失额而是正常市场中出现的最大损失^[1]。从数学的角度来讲 VaR 就是针对某个损益在预期分布里面相应的置信水平下的分为点数的值,其数学表达形式为:

$$c = Pr(\Delta V \geq -VAR) = \int_{-VAR}^{+\infty} f(r) dr \text{ 或者}$$

$$1 - c = Pr(\Delta V < -VAR) = \int_{-\infty}^{-VAR} f(r) dr$$

VAR 的计算公式为:

$$VAR = V_0 \delta Z_\alpha \sqrt{T}$$

式中的 V_0 代表金融序列的初始值, δ 为方差, Z_α 表示下分位数, T 表示持有期。

(二) 波动率的描述

传统的理论一般认为市场是有效的,波动率指的是资产收益率的条件标准差,因此金融资产的波动是恒定的,金融序列的波动性基本上不会因为时间的关系改变,根据很多的实证研究结果,我们可以总结金融序列在波动性方面存在一些比较明显的特点。如,金融序列的波动性一般在均值旁边的点会存在比正态分布的状态下更高的尖峰厚尾的特点,同时在分布的尾部也会看起来比正态分布宽大,分布的峰高超过 3;金融序列的某一个变动后面会伴随着一种相应的变动,这就是所谓的波动集聚效应。波动性对有利和不利消息的反应展现出不对称的状态,也就是说在一定的条件下,不利消息对市场的影响要超过有利消息,这就是非对称性。金融序列收益率的绝对值展现出缓慢衰减的现象,即使相互之间有着较长的时间间隔,也还是相互间存在关联,这就是长期记忆性所呈现的历史对将来的影响。在一定程度上长期记忆性表明收益存在可预测性,市场上可能因为收益的可预测性产生套利行为^[2]。因此有效市场假说不成立。波动性是影响上证 50 股票交易的因素之一,笔者侧重对上证 50 股票收益率的

波动性进行研究,考量其条件异方差模型。条件异方差模型有两种类型,一种是从函数的角度去描述方差;另外一种借助随机方程来描述方差。GARCH模型就是属于第一种类型。波动率模型主要研究资产收益率是序列相不相关,还是低阶相关的,我们可以通过对金融序列的单位根检验图来进行判断。假设金融资产的收益率序列为 X_t ,我们知道波动率是刻画金融序列收益率的某种相关关系,如果给定前一时刻已经获得的信息集合是我们可以得到 X_t 的条件均值和条件方差,其关系式分别为:

$$\mu = E(X_t / F_{t-1})$$

$$\delta^2 = \text{Var}(X_t / F_{t-1}) = E[(X_t - \mu_t)^2 / F_{t-1}]$$

上述 F_{t-1} 表示已知的信息集合,它包括了金融资产收益率的函数,在某种程度上条件异方差模型就是把一个动态方程加到一个金融序列的模型里面,这又在一定程度上阐明了资产收益率的条件异方差会顺时而变的特点。

二、理论模型的介绍

(一)ARCH模型

恩格尔在1982年提出了ARCH模型,认为时间序列数据一般存在一种特殊的异方差,即自回归条件异方差。ARCH模型的简单形式为:

$$\text{ARCH}(1): Y_t = \beta X_t + \epsilon_t$$

$$\delta_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

式中 X_t 表示研究中选取的自变量, Y_t 表示为因变量,其中的随机干扰项不具有时间序列相关性。其中ARCH模型中的残差项的方差具有序列相关性,在构成的方程式中,有一个解释变量,即前一期因变量的随机干扰项的平方 ϵ_{t-1}^2 ,该项成为ARCH项,该式表明当期残差的方差和以往多期的残差无关,仅仅与前一期残差的平方有关。模型的线性部分可以是自变量也可以是因变量 Y 的过去数值,误差项可以服从正态分布也可以服从 t 分布以及GED分布^[3]。以上模型仅仅考虑了前一期残差对当期残差平方影响,类似的可以推广到更广泛的范围,即认为当期的残差的方差依赖于过去的残差,ARCH(p)其一般形式为:

$$X_t = \delta_t \epsilon_t$$

$$\delta_t^2 = C_0 + \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-2}^2 + \cdots + \alpha_p X_{t-p}^2$$

ARCH模型认为金融序列的扰动项之间是不相关的并且相互之间也是不独立的;金融序列收益率的某一种变动往往可能会跟着一个类似的变动。我们一般用LM检验、 F 检验、 Q 检验,来检验金融

序列的ARCH效应,这里最常用的就是拉格朗日乘法法检验,检验步骤为:

$H_0: \alpha_1 = \alpha_2 = \cdots \alpha_p = 0$ 如果 $\alpha_1 = \alpha_2 = \cdots \alpha_p = 0$ 则说明不存在ARCH效应,如果其中存在一个 α_p 不是0,说明ARCH效应是存在的。然后检验统计量为 $LM = nR^2 \sim \chi^2(q)$ 这里的 n 代表样本数量, R^2 代表辅助回归项系数,如果原假设被推翻则认为该金融序列存在ARCH效应,反之如果原假设成立则样本序列不存在ARCH效应。

(二)GARCH模型

波勒斯列夫在1986年基于ARCH模型的基础上提出了GARCH模型,为特定的金融序列制作回归模型,特别适用于波动性的分析和预测。它的基本思想是在ARCH模型的基础上,再加上 δ_t^2 的自回归部分,即 δ_t^2 还是 $\{\delta_{t-1}^2 \cdots \delta_{t-p}^2\}$ 的函数。GARCH模型最简单形式为GARCH(1,1),其表达式为:

$$\delta_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \delta_{t-1}^2$$

由上式可知, t 期残差的方差由其滞后一期残差的平方和滞后一期残差方差的平方决定,将上式进行推广可以得到更一般的GARCH模型,即GARCH(p, q)模型,其方程形式为:

$$\delta_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \delta_{t-1}^2 + \beta_q \delta_{t-q}^2$$

在GARCH模型里面有几个重要的参数,即 α ,表示误差系数,它在很大程度上决定了金融序列的波动性对市场运动反应的敏感度,并且 α 数值越大,意味着波动性的反应越迅速。 α 的数值通常小于0.25; β 表示滞后系数,它在很大程度决定着波动性对市场运动产生反应的持久性, β 的数值越大,说明产生波动越持久。 β 数值通常大于0.7;上述两个参数的和 $(\alpha + \beta)$ 被持久称作持久度,它的作用就是在一定程度上决定着波动性向均值反转的速度。 $(\alpha + \beta)$ 的数值越大,说明持久性越强,向均值回复的速度就会更慢。

(三)TGARCH模型

不利的消息往往比有利的消息对金融资产的价格产生更大更强烈的影响,鉴于此Glosten, Jagannathan and Runkle在1993年提出了TGARCH模型。

假设条件方差方程为:

$$\delta_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \lambda_1 \epsilon_{t-1}^2 \cdot 1(\epsilon_{t-1} > 0) + \beta_1 \delta_{t-1}^2$$

式中, $1(\cdot)$ 为示性函数,即当 $\epsilon_{t-1} > 0$ 时,取值为0, $\epsilon_{t-1} \leq 0$ 即为TGARCH项。

(四)EGARCH模型

EGARCH模型是由于对称的ARCH类模型

无法准确地反映正负冲击效应而提出的改进模型,在一定程度上可以处理股价比的非对称分布问题。它的一般形式为:

$$\ln\delta_t^2=\alpha_0+\alpha_1(\frac{\epsilon_{t-1}}{\delta_{t-1}})+\lambda_1|\frac{\epsilon_{t-1}}{\delta_{t-1}}|+\beta_1\ln\delta_{t-1}^2$$

其中 $(\epsilon_{t-1}/\delta_{t-1})$ 为 ϵ_{t-1} 的标准化,也称为 EGARCH 项。只要 α_1 不为 0,则这个模型也包括了非对称效应。 $|\epsilon_{t-1}/\beta_{t-1}|$ 表示对称效应,也称为 EARCH-a 项,由于 δ_t^2 为指数形式,故称为指数 GARCH,EGARCH 的优点在于,无论 $\ln\delta_t^2$ 取值何值,

都有 $\delta_t^2=\exp(\ln\delta_t^2)>0$

(五)PGARCH 模型

PGARCH 模型是由 Ding 在 1993 年提出的,它具有一套科学模型,并能进行量化计算。

三、实证分析

偏度 $S=-0.274635<0$,峰度 $K=6.534878>3$,这说明上证 50 指数收益率比正态分布的图像偏左一些并且呈现尖峰后尾的分布状态。

表 1 数据的统计特征								
指数简称	均值	最大值	最小值	方差	偏度	峰度	JB 检验值	Prob
上证 50	0.0003	0.09233	-0.0995	0.01831	-0.27437	6.5349	1663.541	0

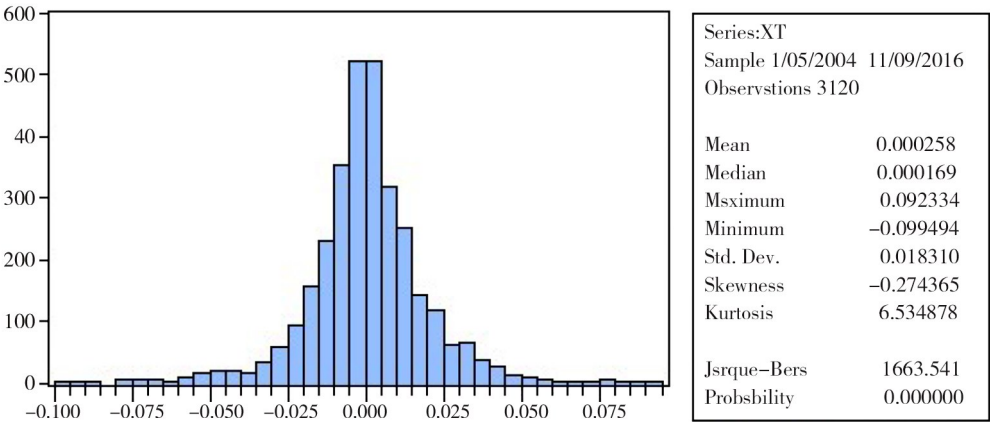


图 1 上证 50 基本统计特征

图 1 里面的后两项是总体分布的正态性检验所采取的 Jarque-Bera 检验所得到的结果,我们一般假设在 JB 检验里面样本序列一般服从正态分布,其检验的统计量为:

$$JB=\frac{n-m}{6}\{S^2+\frac{1}{4}(K-3)^2\}$$

式中 S 表示金融样本序列的偏动幅度,K 表示金融样本序列的峰高,m 代表进行样本序列估计时所采用的样本系数的个数。一般在零假设的情况下,我们可以认 JB 统计量服从 $\chi^2(2)$ 分布通过 Eviews 软件操作的结果,我们可以判读有没有必要拒绝原假设。由图 1 我们可以看到,这里的概率值近乎为 0,则说明了在至少 99% 的置信水平的条件下我们可以不接受原假设,即认为金融样本序列不服从正态分布。

紧接着利用 QQ 图做验证,见图 2。

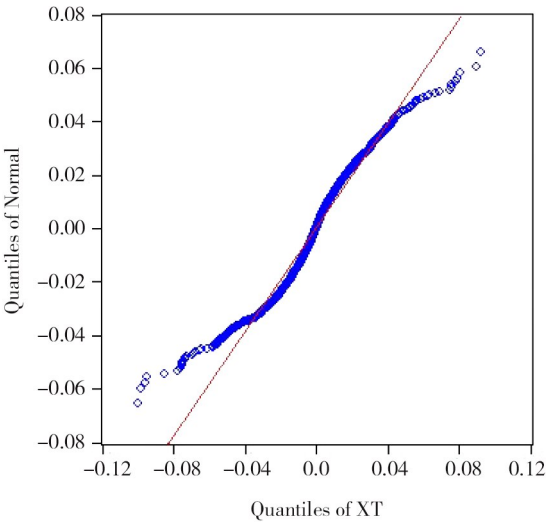


图 2 上证 50 指数对数收益率序列的 Qusntile-Qusntile

通过图 2 我们可以看到,金融序列样本的值在一起呈现出一条弯曲的线,与图中倾斜 45°的直线相偏离,即此金融序列不是正态分布。

表 2 上证 50 对数收益率序列的单位根检验图

Null Hypothesis:XT has a unit root		
Exogenous:Constant		
Lag Length:0 (Automatic-based on SIC,maxlag=28)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-55.47237	0.0001
Test critical values: 1% level	-3.432257	
5% level	-2.862268	
10% level	-2.567202	

由此可以知道上证 50 对数收益率的单位根相较于临界值小,说明这个金融序列较为稳定,继续检验相关和偏自相关性如下:

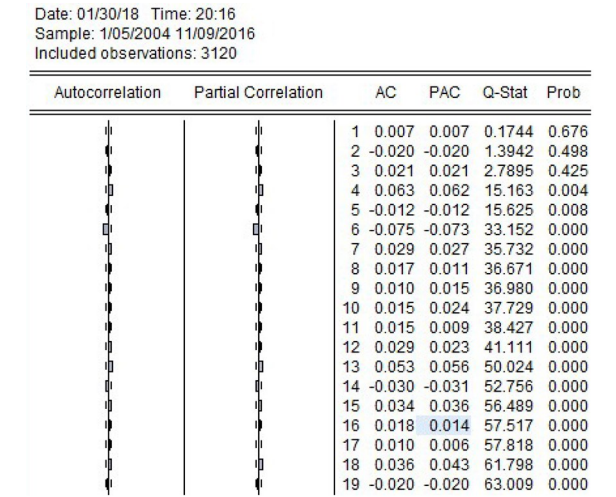


图 3 上证 50 指数对数收益率自相关图

根据图 3,我们可以利用上证 50 的三阶滞后量来求自相关函数值和偏自相关函数值,得知上证 50 指数的对数收益率之间的关联度不明显,但是在更高阶之后关联度呈现明显的现象。

(二)检验条件异方差性

进行条件异方差检验就是判断金融序列是否存在 ARCH 效应,可以通过上面的结果知道样本平方回报率存在着自相关性,紧接着用两种方式检查是不是存在条件异方差。

1.检验残差图法

由图 4,可以得知残差序列存在波动集聚效应,认为金融序列存在条件异方差。

2.检验 ARCH 效应

一般根据 ARCH 效应的存在与否来判断

(一)平稳性检验

检验上证 50 指数对数收益率的单位根得如下结果,见表 2。

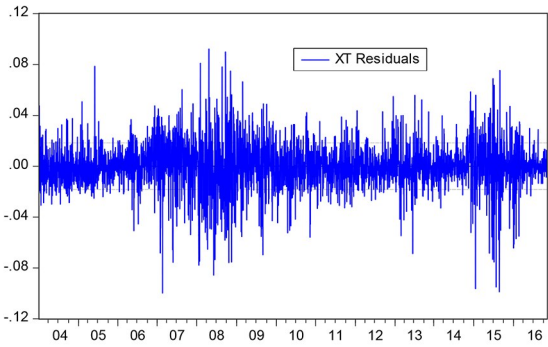


图 4 残差图

随机扰动项是否存在 GARCH 效应,利用拉格朗日乘法法检验 ARCH 效应。假设模型随机误差项 $\epsilon_t \sim \text{ARCH}(q)$ 建立辅助回归方程:

$$ht=\alpha_0+\alpha_0\epsilon_{t-1}^2+\cdots+\alpha_q\epsilon_{t-q}^2$$

若检验的原假设 $H_0=\alpha_1=\alpha_2=\cdots=\alpha_q=0$,则说明不存在 ARCH 效应,若回归系数至少存在一个不为 0,说明序列存在 ARCH 效应不接受原假设。检验统计量为 $LM=nR^2\sim\chi^2(q)$ 这里的 n 表示样本个数, R^2 表示回归系数。如果原假设被推翻则认为金融序列有 ARCH 效应,反之样本序列不存在 ARCH 效应。其结果见表 3。

由上面的分析可以看出最小二乘法的检验结果中的统计量 Obs * R-squared 的值及其概率,及统计量 F 的值及其概率,概率值都将近为零,由此认为金融序列存在 ARCH 效应。^[4]

(三)建立 GARCH 模型对其进行参数估计

笔者将对 3 个模型进行参数估计,计算出标准差,由标准差得到 VAR 的值。

1.GARCH(1,1)的模型

$$\delta_t^2=\alpha_0+\alpha_1\epsilon_{t-1}^2+\gamma_1\delta_{t-1}^2$$

表 3 ARCH-LM 检验结果输出

Heteroskedasticity Test: ARCH				
Test Equation:				
Dependent VaRiable: RESID^2				
Method: Least Squares				
Date: 01/30/18 Time: 20:43				
Sample (adjusted): 1/15/2004 11/09/2016				
Included observations: 3112 after adjustments				
VaRiable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000143	1.76E-05	8.124240	0.0000
RESID^2(-1)	0.068546	0.017931	3.822691	0.0001
RESID^2(-2)	0.091299	0.017903	5.099708	0.0000
RESID^2(-3)	0.109572	0.017945	6.106160	0.0000
RESID^2(-4)	0.082290	0.018047	4.559862	0.0000
RESID^2(-5)	0.024197	0.018046	1.340816	0.1801
RESID^2(-6)	0.059262	0.017944	3.302682	0.0010
RESID^2(-7)	0.088536	0.017901	4.945917	0.0000
RESID^2(-8)	0.047748	0.017911	2.665889	0.0077
R-squared	0.088635	Mean dependent VaR		0.000334
Adjusted R-squared	0.086285	S.D. dependent VaR		0.000789
S.E. of regression	0.000754	Akaike info criterion		-11.53969
Sum squared resid	0.001764	Schwarz criterion		-11.52221
Log likelihood	17964.76	Hannan-Quinn criter.		-11.53341
F-statistic	37.72287	Durbin-Watson stat		2.000932
Prob(F-statistic)	0.000000			

需要对这里的 $(\alpha_0 \alpha_1 \gamma_1)$ 进行参数估计。

2. EGARCH(1,1)的模型

$$\ln \delta_t^2 = \alpha_0 + \alpha_1 (\epsilon_{t-1} / \delta_{t-1}) + \gamma_1 | \epsilon_{t-1} / \delta_{t-1} | + \beta_1 \ln \delta_{t-1}^2$$

需要对这里的 $(\alpha_0 \alpha_1 \gamma_1 \beta_1)$ 4 个参数进行估计。

3. PGARCH(1,1)的模型

$$\delta_t^2 = \alpha_0 + \beta_1 \delta_{t-1}^2 + \alpha_1 (| \mu_{t-1} | - \gamma_u \mu_{t-1})^2$$

需要对这里的 $(\alpha_0 \alpha_1 \gamma_1 \delta)$ 5 个参数进行估计。

表 4 正态分布假定下模型的参数估计

Model	α_0	α_1	β_1	γ_1	δ
GARCH(1,1)-n	2.28E-06 ***	0.061574 ***	0.932424 ***		
	(4.753962)	(12.43680)	(185.7200)		
EGARCH(1,1)-n	-0.183772 ***	0.136244 ***	-0.010821	0.989967 ***	
	(-9.415620)	(14.24234)	(-2.111729)	(496.4941)	
PGARCH(1,1)-n	3.97E-05	0.071052 ***	0.064394	0.935913 ***	1.329709 ***
	(1.381998)	(13.41182)	(1.810157)	(188.3995)	(7.945128)

注: 上述括号内的数字表示为 t 的估计统计量, *** 表示在 1% 的置信水平下显著; ** 表示在 5% 的置信水平下显著; * 表示在 10% 的置信水平下显著。

表 5 在 T 分布的假定下模型的参数估计

Model	α_0	α_1	β_1	γ_1	δ
GARCH(1,1)-t	2.03E-06***	0.064275***	0.934188***		
	(2.652126)	(7.088173)	(113.2650)		
EGARCH(1,1)-t	-0.164912***	0.139180***	-0.005720	0.992176***	
	(-5.543197)	(8.294243)	(-0.611710)	(338.6631)	
PGARCH(1,1)-t	6.05E-05	0.073052***	0.036752	0.939506***	1.177421***
	(0.849190)	(7.905627)	(0.564340)	(120.5209)	(4.467984)

表 6 在 GED 分布假定下模型的参数估计

Model	α_0	α_1	β_1	γ_1	δ
GARCH(1,1)-GED	2.14E-06***	0.062428***	0.932282***		
	(2.714216)	(7.151676)	(106.6055)		
EGARCH(1,1)-GED	-0.175459***	0.137645***	-0.007964	0.991087***	
	(-5.451540)	(8.170967)	(-0.858525)	(310.8698)	
PGARCH(1,1)-GED	4.93E-05	0.072172***	0.050400	0.936963***	1.252632***
	(0.812350)	(7.750120)	(0.786433)	(111.5842)	(4.483836)

上述表 4、表 5、表 6 是由 Eviews 得到的在正态分布、T 分布、广义误差分布下的参数估计的结果，我们可以知道 3 个模型在 1% 的置信水平下都是比较显著的，拟合的效果较好。在 GARCH、EGARCH、PGARCH 模型下 γ 都比零大，这在一定程度上说明有非对称效应的存在，即不好的消息会

比好的消息对市场产生更大的影响。根据检验可以发现 3 个模型的参数都是显著的，而在对残差项进行 ARCH 效应检验的时候，我们发现异方差现象不是很明显，这说明了上述模型可以比较好地刻画上证 50 的对数收益率的异方差现象。

表 7 正态分布下的 VAR 的计算结果

Model	置信水平	VAR 最小值	VAR 最大值	VAR 均值	失败天数	失败频率%
GARCH(1,1)	95%	30.3751	160.147	65.30175	2	0.667
EGARCH(1,1)	95%	27.14649	160.5639	65.04542	2	0.667
PGARCH(1,1)	95%	28.85692	163.6202	65.19082	2	0.667
GARCH(1,1)	99%	42.89332	226.147	92.21681	0	0
EGARCH(1,1)	99%	38.33414	226.7357	91.85202	0	0
EGARCH(1,1)	99%	40.74947	231.0374	92.05735	0	0

表 8 T 分布下 VAR 的计算结果

Model	置信水平	VAR 最小值	VAR 最大值	VAR 均值	失败天数	失败率%
GARCH(1,1)-t	95%	20.26644	164.0348	66.80074	1	0.333
EGARCH(1,1)-t	95%	26.82275	166.1019	66.62941	1	0.333
PGARCH(1,1)-t	95%	28.01325	168.7751	66.73087	1	0.333
GARCH(1,1)-t	99%	42.7962	231.9421	94.45503	0	0
EGARCH(1,1)-t	99%	37.92687	234.8651	94.21277	0	0
PGARCH(1,1)-t	99%	39.61122	238.6449	94.35624	0	0

表 9 广义误差分布下 VAR 的计算结果

Model	置信水平	VAR 最小值	VAR 最大值	VAR 均值	失败天数	失败率%
GARCH(1,1)－GED	95%	29.99193	160.4723	65.23697	2	0.667
EGARCH(1,1)－GED	95%	26.49018283	161.9967565	65.0529637	2	0.667
PGARCH(1,1)－GED	95%	27.95044126	164.563891	65.06739729	2	0.667
GARCH(1,1)－GED	99%	47.96504	257.5824	104.7156	0	0
EGARCH(1,1)－GED	99%	42.52091152	260.0302835	104.4202424	0	0
PGARCH(1,1)－GED	99%	44.93281244	264.5510452	104.6016101	0	0

(四)VAR 的计算和回测检验

利用 Eviews 软件得到金融序列的条件异方差,根据条件异方差算出标准差,然后将其带入 VaR 的公式里面,我们可以得到在不同分布不同置信水平下的 VaR 值。计算出所有金融序列 VAR 值的平均值,根据计算的结果进行回测检验,一个简单的做法就是向后构建一个检验样本。比如样本数为 300 天,根据 2016 年 11 月 10 日到 2018 年 1 月 26 日每日交易收盘价,逐日计算头寸损益,并将计算结果按升序排列,即可得到一个关于损益的排序结果。然后把金融序列的 VAR 均值与头寸损益的绝对值进行比较,如果 VAR 均值大于头寸损益,说明检验失败;如果 VAR 均值小于头寸损益,说明检验成功。统计出失败天数和失败频率就可以得到上面的 3 个表,即表 5、表 6、表 7。

(五)模型的效果分析和比较

通过以上的 3 个计算 VAR 结果的表格,(见表 7,表 8,表 9)我们可以看到上证 50 对数收益率在 T 分布下失败天数和失败率最小,这说明在 T 分布条件下,金融序列的尾部特征得到了较好的表现,但是在 T 分布下金融序列的 VAR 值高于正态分布和广义误差分布,从侧面反映出其风险被高估,正态分布和广义误差分布的估计效果也不错,失败天数和失败率一致。

综上,可以发现在同一个分布条件下,GARCH 模型、EGARCH 模型、PGARCH 模型的 VAR 值在同一个置信区间相差不大,在失败率和失败天数上正态分布和广义误差分布上基本一致,并且在正态分布和广义误差分布下 VAR 的值在相同的置信区间下,GARCH 模型下的 VAR 值要高于 EGARCH 和 PGARCH 值,说明 EGARCH 模型和 PGARCH

模型的拟合效果比 GARCH 模型的拟合效果要好。这同时证明上证 50 指数存在显著的杠杆效应,不好信息给股市带来影响力比好的信息给股市带来的影响力还要大。

四、结论

第一,上证 50 对数收益率呈现出左偏,尖峰后尾的特点,并且有强烈的波动聚集,存在自回归的条件异方差,从检验结果可以看出历史的波动对将来的波动变化的影响变得逐渐减小。

第二,上证 50 的收益率序列存在比较显著的杠杆效应,也就是在相同的条件下,股市中的利空消息给市场带来的影响比利好消息带来的影响更大,也就是所谓的非对称效应。

第三,我们根据 GARCH 模型族对金融序列进行一系列的检验,得到了在正态分布、T 分布以及广义误差分布下的参数,由参数可以得到模型在 99% 的置信区间内比较显著,从整体上来说正态分布的结果较为保守,而 T 分布和广义误差分布的效果要略胜一筹。

参考文献:

[1]赖艳丽.基于 GARCH 模型族在沪深 300 中的比较研究[J].上海管理科学,2012(4).

[2]范剑青,姚琦伟.非线性时间序列:建模、预报及应用[M].陈敏,译.北京:高等教育出版社,2005.

[3]彭作祥.金融时序计量建模分析[M].成都:西南财经大学出版社,2006.

[4]高铁梅.计量经济分析方法与建模[M].北京:清华大学出版社 2006.

责任编辑 胡号寰 E-mail:huhao2@126.com