FINC 520: Time Series
Problem Set #4

Qiushi Huang
Marius Ring

May 14, 2015

1 Question 1

1.1 part (a)

We can rewrite the GARCH(1,1) as

$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 - \beta v_{t-1} + v_t \tag{1}$$

$$v_t = \epsilon_t^2 - \sigma_t^2 = (z_t^2 - 1)\sigma_2^2 \tag{2}$$

$$z_t \ \sim \text{I.I.D. N}(0, 1) \tag{3}$$

We find that

$$\mu \equiv \mathbb{E}\epsilon_t^2 = \frac{\omega}{1 - \alpha - \beta} \tag{4}$$

Thus we may write

$$\epsilon_t^2 - \mu = (\alpha + \beta)(\epsilon_{t-1}^2 - \mu) - \beta v_{t-1} + v_t \tag{5}$$

Then we use the Yule-Walker (not White-Walker as in Game of Thrones!!) method to find the autocorrelations for ϵ_t^2, which will be the same as those of y_t^2 under our assumption. The first autocorrelation:

$$\rho_1 = \frac{\alpha(1 - \alpha\beta - \beta^2)}{1 - 2\alpha\beta - \beta^2} \tag{6}$$

Which provides us with all the subsequent ones due to p=q=1 in the GARCH(p,q):

$$\rho_k = (\alpha + \beta)^{k-1}\rho_1 \tag{7}$$

1.2 part(b)

The heteroskedasticity consistent estimator for the asymptotic covariance matrix is

$$\hat{V} = \hat{Q}^{-1}\Omega\hat{Q}^{-1} \tag{8}$$

where

$$\hat{\Omega} = \frac{1}{T}\sum \hat{\epsilon}_t^2 y_{t-1}^2 \tag{9}$$

$$\hat{Q} = \frac{1}{T}\sum y_{t-1}^2 \tag{10}$$

We will have that

$$\sqrt{T}(\hat{\phi} - \phi) \to N(0, Q^{-1}\Omega Q^{-1}) \text{ in distribution.} \tag{11}$$

Where we can, asymptotically, replace $Q^{-1}\Omega Q^{-1}$ with $\hat{V}$.

$$\Omega = \text{plim}\frac{1}{T}\sum \hat{\epsilon}_t^2 y_{t-1}^2 = \mathbb{E}\epsilon_t^2\epsilon_{t-1}^2 = \rho_1\gamma_0 + \mu^2 \tag{12}$$

$$Q = \text{plim}\frac{1}{T}\sum y_{t-1}^2 = \mathbb{E}\epsilon_{t-1}^2 = \mu \tag{13}$$

$$V = \frac{\rho_1\gamma_0 + \mu^2}{\mu^2} \tag{14}$$

where

$$\gamma_0 = 3\omega^2(1 + \alpha + \beta)[(1 - \alpha - \beta)(1 - \beta^2 - 2\alpha\beta - 3\alpha^2)]^{-1} \tag{15}$$

$$\mu = \frac{\omega}{1 - \alpha - \beta} \tag{16}$$

$$\rho_1 = \frac{\alpha(1 - \alpha\beta - \beta^2)}{1 - 2\alpha\beta - \beta^2} \tag{17}$$

So in the case where $\alpha = \beta = 0$ we have that $\rho_1 \to 0$ and $V \to 1$.

Otherwise, it follows from the terms, and our assumptions on ω, α, and β that $\rho_1 > 0$, and thus

$$V > 1 \tag{18}$$

1.3 part (c)

Parameters	Rejections
A	0.047
B	0.06
C	0.088
D	0.158

$$\hat{\phi}_T = \left(\sum_{t=2}^{T} y_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} y_t y_{t-1} \tag{19}$$

The asymptotic variance of $\sqrt{T}\hat{\phi}_T$ will just be 1 under our H_0, and the assumption that error terms are IID Normal. We see this in part (d): just let $\rho_1 = 0$.

$$\tag{20}$$

In A and B $V = 1$ and the IID Normal assumptions on the error term hold, so we would expect to see that that we reject 5% of the time, which we also do.

IN C and D $V > 1$, thus using the IID Normal assumption on the error term, we would be using a variance that is too low, and thus use a test statistic that is too high, and thus reject too often. i.e. **heteroskedasticity that is unaccounted for increases the size of the test beyond our confidence level.**

1.4 part (d) information criterions

The 'approximate' log likelihood function for the AR(p) estimation, assuming that the error terms are IID, is this:

$$logL = -\frac{T-p-1}{2}\log(2\pi) - \frac{T-p-1}{2}\log(\sigma^2) - \frac{1}{2}\sum_{t=p+1}^{T}\frac{\epsilon_t^2}{\sigma^2} \tag{21}$$

$$\epsilon_t = y_t - \sum_{j=1}^{p}\phi_j y_{t-j} \tag{22}$$

$$\hat{\sigma}_p^2 = \frac{1}{T-p-1}\sum_{t=p+1}^{T}\hat{\epsilon}_t^2 = \frac{1}{T-p-1}\sum_{t=p+1}^{T}\left(y_t - \sum_{j=1}^{p}\hat{\phi}_j y_{t-j}\right)^2 \tag{23}$$

Thus for our Information Criterion comparisons we use:

$$\log \hat{L}_p = -\frac{T}{2}\log(\hat{\sigma}_p^2) \tag{24}$$

Since the first terms with p in the logL above artificially increase the logL due to the nature of the conditional/approximate MLE. Also the third term depends only on T and p when we subsitute in for $\hat{\sigma}^2$

The ICs are given by:

$$BIC = -2\log \hat{L}_p + p\log(T) = T\log \hat{\sigma}_p^2 + p\ log(T) \tag{25}$$

$$AIC = -2\log \hat{L}_p + 2p = T\log \hat{\sigma}_p^2 + 2p \tag{26}$$

and $\hat{\phi}_T$ is our usual OLS estimate, regressing y_t on $y_{t-1}, ...y_{t-p}$, with no intercept term.

AIC results

Parameters	p0	p1	p2	p3	p4	p5
A	0.612	0.127	0.102	0.068	0.051	0.04
C	0.885	0.066	0.022	0.015	0.009	0.003
E	0	0.653	0.159	0.08	0.061	0.047
F	0	0.883	0.074	0.024	0.013	0.006
G	0	0.624	0.152	0.103	0.069	0.052
H	0	0.691	0.124	0.067	0.054	0.064

BIC results

Parameters	p0	p1	p2	p3	p4	p5
A	0.931	0.055	0.01	0	0.001	0.003
C	0.978	0.017	0.004	0.001	0	0
E	0	0.921	0.062	0.015	0.002	0
F	0	0.968	0.024	0.005	0.003	0
G	0	0.975	0.025	0	0	0
H	0	0.958	0.026	0.011	0.003	0.002

From theory we would expect that both AIC and BIC would in the limit never pick a p that was too small, and in the limit BIC will also not pick a p that is too big. We also expected AIC to be 'less conservative', i.e., choosing a higher number of lags, and possibly not be consistent towards the true $p = 0$.

The simulation results are well aligned with theory. We see that AIC more often than BIC picks too large p. We also see as $T \to 1000$ that BIC is a lot more consistent towards the true $p = 1$ than AIC for the setups G and H.