GARCH Residual Processes

Reg Kulperger

June 25, 2003

joint with Hao Yu
and some parts with Janusz Kawczak

Nonlinear time series - important in financial modelling and other areas

ARCH(1)

$$X_{t+1} = \sqrt{h(X_t)}\epsilon_{t+1}$$

where

$$h(x) = \alpha_0 + \alpha_1 x^2$$

ϵ_t are i.i.d. F, $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t^2) = 1$

$\longrightarrow$ last so that $\sigma_{t+1}^2 = h(X_t) =$ conditional variance of X_{t+1} given $\mathcal{F}_t$, the information up to time t.

ARCH-M(1) : Mean term μ allowed : $Y = X + \mu$

$$Y_{t+1} = \mu + \sqrt{h(Y_t - \mu)}\epsilon_{t+1}$$

GARCH(1,1) : different form of conditional variance

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

Conditional variance at time $t + 1$ depends on X_t and σ_t^2, the current conditional variance

Given data X_t, $t = t_0, 1, \ldots, n$ can one estimate distribution F of ϵ?

$t_0 = 0$ for ARCH-M(1)

residuals : $\hat{\epsilon}_t$

Our study : some processes and functionals of residuals

- ARCH-M(p) : EDF process of residuals $\sqrt{n}(\hat{F} - F)$
  - proof is easier than AR residual EDF process due to product form for residuals; p $= 1$

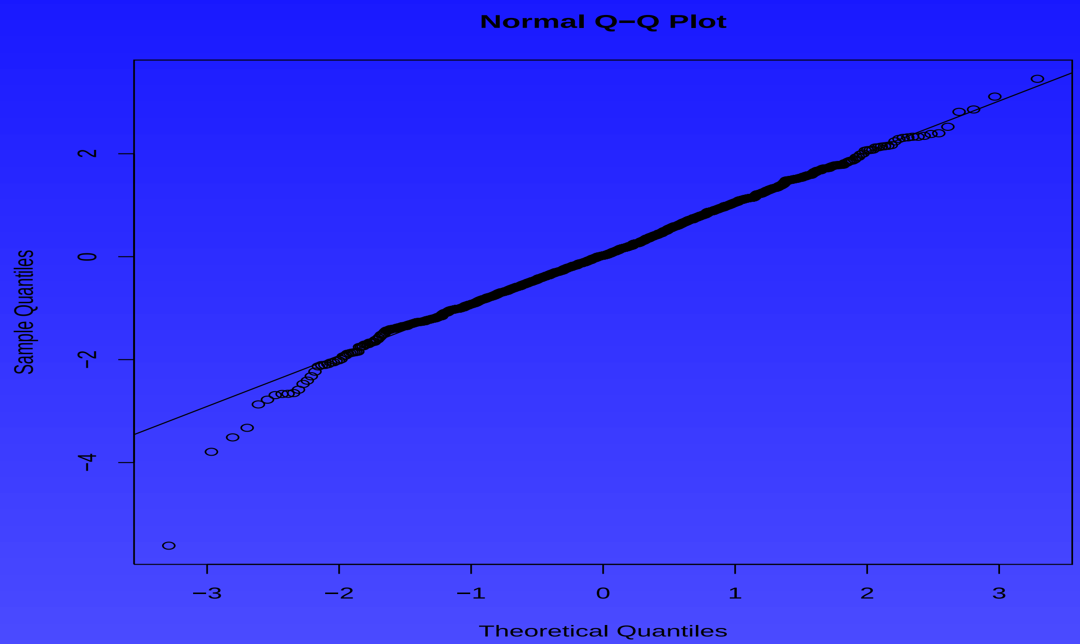
$$
\begin{aligned}
\hat{\epsilon}_t &= \frac{X_t - \hat{\mu}}{\hat{\sigma}_t} = \frac{\mu - \hat{\mu} + \sigma_t \epsilon_t}{\hat{\sigma}_t} \\
&= \frac{\sqrt{n}(\mu - \hat{\mu})}{\sqrt{n\, h(X_{t-1}, \hat{\theta})}} + \epsilon_t \sqrt{\frac{h(X_{t-1}, \theta)}{h(X_{t-1}, \hat{\theta})}}
\end{aligned}
$$

multiplicative structure ; bounded ratio $\frac{h(x,\theta)}{h(x,\hat{\theta})}$ on the set $||\hat{\theta} - \theta|| \leq b$ for any $b > 0$

- does not easily extend to GARCH

converges to Brownian bridge + additional term

Typical residual qqnorm plot from simulated normal ARCH



notice heavy looking tails $\longrightarrow$ makes difficult to interpret

Second nonlinear time series study :

- GARCH(p,q) : k-th moment partial sum processes to construct

I) Joint Distribution of Skewness ($k = 3$) and Kurtosis ($k = 4$) partial sum process
$\longrightarrow$ test of normality : Jarque-Berra (sum of squares of skewness and kurtosis) statistic for GARCH residuals
SAME asymptotics as the i.i.d. innovations !
NOT so for most other statistics
- eg Kolmogorov-Smirnov
II) consistency of residual EDF and density estimate
- justifies first order correct semi-parametric bootstrap

EDF process : ARCH-M(1)

$$\hat{\epsilon}_t = \frac{X_t - \hat{\mu}}{\hat{\sigma}_t} = \frac{\mu - \hat{\mu} + \sigma_t \epsilon_t}{\hat{\sigma}_t}$$

$$= \quad \frac{\sqrt{n}(\mu - \hat{\mu})}{\sqrt{n\, h(X_{t-1}, \hat{\theta})}} + \epsilon_t \sqrt{\frac{h(X_{t-1}, \theta)}{h(X_{t-1}, \hat{\theta})}}$$

Let $s = (s_0, s_1, s_2) \in \mathbf{R}^3$

Define

$$\hat{F}_n(x,s) = \frac{1}{n}\sum_{t=1}^{n} \mathsf{I}\left(\epsilon_t \le \left(x + \frac{s_2}{\sqrt{n\,h(X_{t-1}, \theta + n^{-1/2}s)}}\right) \times \sqrt{1 + \frac{g_n(X_{t-1}, s)}{\sqrt{n}}}\right)$$

where

$$g_n(x,s) = \frac{\sqrt{n}\Big(h(x,\theta+n^{-1/2}s) - h(x,\theta)\Big)}{h(x,\theta)}$$

$$\hat{\epsilon}_t(s) = \frac{\epsilon_t}{\sqrt{1+\frac{g_n(X_{t-1},s)}{\sqrt{n}}}}$$

Then $F_n(x,0)$: EDF of innovations

and

$$\hat{F}_n(x) = \hat{F}_n(x, \sqrt{n}(\hat{\theta}-\theta))$$

$$\sup_{x \in \mathbf{R}} |g_n(x,s)| \leq \sum_{i=1}^{3} \frac{C_i(\theta)||s||_\infty^i}{n^{(i-1)/2}}$$

$$\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{I}\left(\epsilon_t \leq (x + O_p(1/\sqrt{n}))\sqrt{1 + O_p(1/\sqrt{n})} \right)$$

$$\begin{aligned} E_n(x,s) \quad = \quad & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \mathrm{I}\left(\epsilon_t \leq \left(x + \frac{s_2}{\sqrt{n\, h(X_{t-1}, \theta + n^{-1/2}s)}} \right) \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s} \right. \right. \\ & \left. -F\left(\left(x + \frac{s_2}{\sqrt{n\, h(X_{t-1}, \theta + n^{-1/2}s)}} \right) \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)} \right) \right\} \end{aligned}$$

$$
\begin{aligned}
&E_n(x) = \sqrt{n}(F_n(x) - F(x)) = \\
&\quad \sqrt{n}(F_n(x, \sqrt{n}(\hat{\theta} - \theta)) - F(x)) = \\
&\quad E_n\left(x, \sqrt{n}(\hat{\theta} - \theta)\right) + \\
&\quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ F\left(\left\{ x + \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{n\, h(X_{t-1}, \hat{\theta})}} \right\} \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, \sqrt{n}(\hat{\theta} - \theta))} \right) - F(x) \right\}
\end{aligned}
$$

Proposition 1. *The process* $\{X_t, t \geq 0\}$ *: stationary and ergodic*

F *has continuous density* f *that is positive on the open support of* F

$\lim_{x\to\pm\infty} |x|f(x) = 0.$

Then for any $b > 0$

$$\sup_{\|s\|_\infty \leq b} \sup_{x\in\mathbf{R}} |E_n(x,s) - E_n(x,0)| \to 0$$

in probability as $n \to \infty$.

Notice : for b large, then with large probability $\sqrt{n}(\hat{\theta} - \theta)$ belongs to $\{s : ||s||_\infty \leq b\}$

Theorem 1. *Same conditions + $\hat{\theta}$ is $\sqrt{n}$-consistent for θ. Then*

$$\sup_{x \in \mathbf{R}} \left| E_n(x) - \left\{ E_n(x,0) + \left\langle \left(\Phi(\theta) + \frac{1}{2} x \Psi(\theta)\right) f(x),\ \sqrt{n}(\hat{\theta} - \theta) \right\rangle \right\} \right|$$

$\longrightarrow 0$ *in probability as* $n \longrightarrow \infty$.

where

$$\begin{aligned} \Psi(\theta) &= \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} g(X_{t-1}) \quad \text{a.s.} \quad \text{and} \\ \Phi(\theta) &= \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} k(X_{t-1}) \quad \text{a.s.} \end{aligned}$$

- constants depending on the parameter; g, k

$$
\begin{aligned}
g(x) &= \frac{(1,\ (x-\mu)^2,\ -2\alpha_1(x-\mu))}{\alpha_0+\alpha_1(x-\mu)^2} \\
k(x) &= \frac{(0,\ 0,\ 1)}{\sqrt{\alpha_0+\alpha_1(x-\mu)^2}}
\end{aligned}
$$

Corollary 1. *Suppose that* $\{E_n(x,0), \sqrt{n}(\hat{\theta}-\theta) : x \in \mathbf{R}\}$ *converges weakly to a Gaussian process* $\{E(F(x)), Z : x \in \mathbf{R}\}$ *on* $D(-\infty,+\infty) \times \mathbf{R}^3$, *where* E *is a standard Brownian bridge. Then* $E_n(x)$ *converges weakly to the Gaussian process*

$$
E(F(x)) + \left\langle (\Phi(\theta) + \frac{1}{2} x \Psi(\theta)) f(x),\ Z \right\rangle
$$

Remark 1 ARCH model with known $\mu = 0$: the term $\Phi(\theta)f(x)$ disappears in the limiting Gaussian process

Remark 2 residuals of a ARCH-M process do not behave as the iid innovations $\{\epsilon_t, t \geq 1\}$.

Kolmogorov-Smirnov does differ from one based on iid sample.

Normal ARCH(1) with $\mu = 0$: extra term $\longrightarrow$ small difference

Normal ARCH(1) with μ estimated : big difference

Normal ARCH with α_1 in the range from .85 to .98 (< 1)

KS 0.95 Critical value for i.i.d. limit : 1.358 (size $=$ 1 - .95)

Table of Monte Carlo estimate of true 0.95 critical value and empirical size when usual critical value (1.358) is used

n	ARCH, μ known		ARCH, μ estimated	
	crit.val	size	MLE	
			crit.val	size
100	1.31	.038	1.07	.005
500	1.32	.038	1.09	.006
1000	1.32	.038	1.09	.007

Subsampling :

deal with distribution and parameter dependent Gaussian limit

Construct the empirical process $\tilde{E}_m(x)$ based on the first m $(< n)$ residuals

$\hat{\theta}$ based on all n observations

$$\tilde{E}_m(x) = \sqrt{m}(\hat{F}_m(x) - F(x)) = \frac{1}{\sqrt{m}} \sum_{t=1}^{m} (\mathsf{I}(\hat{\epsilon}_t \leq x) - F(x))$$

Extra term now becomes $O_p\left(\sqrt{\frac{m}{n}}\right) \to 0$ if $\frac{m}{n} \to 0$
- But looses power

GARCH(p,q) moment sums : joint work with Hao Yu

$X_t = \sigma_t \epsilon_t$ where ϵ_t i.i.d. mean = 0 and $E(\epsilon_0^2) = 1$. For $p = q = 1$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-i}^2 + \beta_1 \sigma_{t-1}^2,$$

GARCH(p,q) residuals representation based on Berkes , Horváth and Kokoszka (2003) $\longrightarrow$ need mean $\mu = 0$ is known.

Hao Yu has a grad student trying to extend this to the mean estimated

k-th moment partial sum of residuals, $0 \leq u \leq 1$:

$$\hat{S}_n^{(k)}(u) = \sum_{t=R}^{[nu]} \hat{\epsilon}_t^k$$

k-th moment partial sum of innovations

$$S_n^{(k)}(u) = \sum_{t=R}^{[nu]} \epsilon_t^k$$

k-th moment : $\mu_k = E(\epsilon_0^k)$

inner product : $\langle \mathbf{x}, \mathbf{y} \rangle$

Theorem 2. *For $\sqrt{n}$ consistent estimator and ergodic process, smooth density and $E|\epsilon_0|^{k+\delta} < \infty$ for some $\delta > 0$ and an integer $k \geq 1$ then*

$$\sup_{0 \leq u \leq 1} \left| \frac{1}{\sqrt{n}} \left(\hat{S}_n^{(k)}(u) - S_n^{(k)}(u) \right) + \frac{ku\mu_k}{2} \langle \psi(\theta), \sqrt{n}(\hat{\theta}_n - \theta) \rangle \right| = o_P(1),$$

where

$$\psi(\theta) = E\left(\partial \log \sigma_0^2(\theta)\right)$$

and $\partial(\cdot) == \partial(\cdot)/\partial\theta$

Remark 1. Asymptotic properties of the high moment partial process depends on the parameter θ unless $\mu_k = 0$

$\mu_1 = 0$: Ordinary residual partial sum process $\hat{S}_n^{(1)}(u)$ has same asymptotics as the unobservable innovations partial sum process

does not depend on F or f except through $\hat{\theta}$

Theorem 3. *Same conditions :*

$$\left| \frac{1}{\sqrt{n}} \sum_{t=R}^{n} |\hat{\epsilon}_t^k - \epsilon_t^k| - \frac{k}{2} \psi_k \left(\sqrt{n} (\hat{\theta}_n - \theta) \right) \right| = o_P(1),$$

where

$$\psi_k(\mathbf{u}) = E \left| \langle \epsilon_0^k \partial \log \sigma_0^2(\theta), \ \mathbf{u} \rangle \right|, \ \mathbf{u} \in \mathbf{R}^{p+q+1}.$$

Remark 2. For $h_n > 0;\ h_n \to 0;\ \sqrt{n} h_n^2 \to \infty$ then

$$\frac{1}{n h_n^2} \sum_{t=R}^{n} |\hat{\epsilon}_t^k - \epsilon_t^k| = o_P(1)$$

Case $k = 1$: consistency of kernel density estimation based on the residuals

To study residual sample skewness and kurtosis need

The kth order centered moment partial sum process residuals

$$\hat{T}_n^{(k)}(u) = \sum_{t=R}^{[nu]} \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}}\right)^k$$

counterpart based on iid innovations

$$T_n^{(k)}(u) = \sum_{t=R}^{[nu]} \left(\epsilon_t - \bar{\epsilon}\right)^k,$$

Residuals sample variance $\hat{\sigma}^2_{(n)} = \hat{T}_n^{(2)}(1)/n$

$\hat{\sigma}^2_{(n)} \to \mu_2$ in probability under the minimum condition $\mu_2 < \infty$

Since $\mu_2 = 1$ for GARCH models, then $\hat{\sigma}^2_{(n)}$ seems a useless estimator.

- Plays important role to self normalize $\hat{T}_n^{(k)}(u)$

Innovations sample variance : $\sigma^2_{(n)} = T_n^{(2)}(1)/n$.

Theorem 4. *$\sqrt{n}$ consistent estimator, density condition $E|\epsilon_0|^{k+\delta} < \infty$ for some $\delta > 0$ and an integer $k \geq 2$ or if $k = 1$ and $E\epsilon_0^2 < \infty$ then*

$$\sup_{0 \leq u \leq 1} \frac{1}{\sqrt{n}} \left| \frac{\hat{T}_n^{(k)}(u)}{\hat{\sigma}_{(n)}^k} - \frac{T_n^{(k)}(u)}{\sigma_{(n)}^k} \right| = o_P(1).$$

Remark 3. The self normalized moment sums have same asymptotics for residuals and innovations scaled moment : $\lambda_k = \mu_k / \mu_2^{k/2}$

Corollary 2. *Under these same conditions, for* $k \geq 2$

$$\frac{1}{\sqrt{n}} \left(\frac{\hat{T}_n^{(k)}(u)}{\hat{\sigma}_{(n)}^k} - nu\lambda_k \right)$$

converges weakly to a Gaussian process $\{B^{(k)}(u),\ 0 \leq u \leq 1\}$

Also joint limit process for various k is obtained
$k = 3, 4$ needed for skewness and kurtosis

Remark 4. partial sum for $k = 1$: $\lambda_0 = 1$, $\lambda_1 = 0$ and $\lambda_2 = 1$

The covariance : $E\left[B^{(1)}(u)B^{(1)}(v)\right] = u \wedge v - uv$, $0 \le u, v \le 1$

$\{B^{(1)}(u),\ 0 \le u \le 1\}$ is a Brownian bridge

For $k = 2$, Covariance

$E\left[B^{(2)}(u)B^{(2)}(v)\right] = (\lambda_4 - 1)(u \wedge v - uv)$ for any $0 \le u, v \le 1$

- scaled Brownian bridge

In general, the Gaussian process $\{B^{(k)}(u),\ 0 \le u \le 1\}$ depends on the moments of the innovation distribution

- Neither a Brownian motion nor a Brownian bridge.

APPLICATIONS

Test of normality : Jarque-Berra test statistic

sample skewness and kurtosis partial sum process as

$$\hat{\gamma}_n(u) = \frac{\hat{T}_n^{(3)}(u)/n}{\hat{\sigma}_{(n)}^3}$$

$$\hat{\kappa}_n(u) = \frac{\hat{T}_n^{(4)}(u)/n}{\hat{\sigma}_{(n)}^4}$$

for $0 \leq u \leq 1$

residual sample skewness $\hat{\gamma}_n(1)$ and sample kurtosis $\hat{\kappa}_n(1)$

Omnibus statistics based on sample skewness and kurtosis have been used to test normality : Bowman and Shenton (1975) and Gasser (1975) basic idea

$$\frac{n}{\sigma_\gamma^2}\left(\hat{\gamma}_n(1) - \lambda_3\right)^2 + \frac{n}{\sigma_\kappa^2}\left(\hat{\kappa}_n(1) - \lambda_4\right)^2,$$

$$\sigma_\gamma^2 = E(B^{(3)}(1))^2$$
$$= \ (\lambda_6 - \lambda_3^2) + 3(3 + 3\lambda_3^2 - 2\lambda_4) + 3\lambda_3(\lambda_3/4 + 3\lambda_3\lambda_4/4 - \lambda_5)$$

$$\sigma_\kappa^2 = E(B^{(4)}(1))^2 = (\lambda_8 - \lambda_4^2) + 4\lambda_3(4\lambda_3 + 4\lambda_3\lambda_4 - 2\lambda_5) + 4\lambda_4(\lambda_4^2 - \lambda_6).$$

Our results yield : If the innovation distribution is symmetric about 0

$$\frac{n}{\sigma_\gamma^2}\left(\hat{\gamma}_n(1) - \lambda_3\right)^2 + \frac{n}{\sigma_\kappa^2}\left(\hat{\kappa}_n(1) - \lambda_4\right)^2 \xrightarrow{\mathcal{D}} \chi^2(2).$$

Standard normal innovation distribution :

$$\lambda_3 = 0,\ \lambda_4 = 3,\ \sigma_\gamma^2 = 6\ , \sigma_\kappa^2 = 24$$

then

$$JB = \frac{n}{6}\hat{\gamma}_n^2(1) + \frac{n}{24}\left(\hat{\kappa}_n(1) - 3\right)^2 \xrightarrow{\mathcal{D}} \chi^2(2) \tag{1}$$

The statistic JB in (1) is the exact Jarque-Berra normality test widely used in econometrics and implemented in standard statistical packages such as Splus

The previously unjustified asymptotic $\chi^2(2)$ distribution is now justified for GARCH residuals

Jarque and Berra (1987) : JB is a Lagrange multiplier test statistic

asymptotically equivalent to the likelihood ratio test
- implying it has the same asymptotic power characteristics including maximum local asymptotic power (Cox and Hinkley, 1974)
- JB statistic is asymptotically locally most powerful

Instead of asymptotic $\chi^2(2)$ critical value can use Monte Carlo to find finite sample correction coefficients based on polynomial approximation in $n^{-1/2}$
Lu (2001) : 5 per cent critical values of JB for a given sample size n

$$JB_{0.05} = 4.60517 - 9.78n^{-1/2} + 132.25n^{-1} - 1696n^{-3/2},\ n \geq 100.$$

Tests of structural change : change point

Test in the literature : cusums usually based on original data

Tests based on cusum of residuals : much better power

Kernel density estimation of the innovation distribution

Innovation distribution : uniformly continuous density $f(x)$

$h_n > 0$ and $K(x)$ be a probability density function (kernel)

kernel density estimation of $f(x)$ based on the residuals

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{t=R}^{n} K\left(\frac{x - \hat{\epsilon}_t}{h_n}\right), \; x \in \mathbf{R}.$$

counterpart based on iid innovations

$$f_n(x) = \frac{1}{nh_n} \sum_{t=R}^{n} K\left(\frac{x - \epsilon_t}{h_n}\right), \; x \in \mathbf{R}.$$

Theorem 5. *$\sqrt{n}$ consistent estimator, density condition and $E|\epsilon_0| < \infty$*

(i) $h_n > 0$; $h_n \to 0$; $\sqrt{n}h_n^2 \to \infty$,

(ii) $\sup_{|x|>b} |x|K(x) \to 0$ *as* $b \to \infty$,

(iii) K *is Lipschitz, i.e., there exists a constant* C *such that*

$$|K(x) - K(y)| \leq C|x - y|, \ \forall\ x, y \in \mathbf{R}.$$

Then

$$\sup_{x \in \mathbf{R}} |\hat{f}_n(x) - f_n(x)| = o_P(1).$$

semi-parametric bootstrap based on resampling from f_n will be first order correct.