Interval Estimation for AR(1) and GARCH(1,1) Models

Liang Peng

School of Mathematics
Georgia Institute of Technology

2010
This talk is based on the following papers:

Outline

Empirical Likelihood

EL for AR(1)

EL for GARCH(1,1)

EL for conditional VaR
Review

- Parametric likelihood ratio test
- Empirical likelihood method
- Estimating equations
- Profile empirical likelihood method
Observations: \(X_1, \cdots, X_n\) iid with pdf \(f(x; g(\mu))\), where \(g\) is a known function, but \(\mu = E(X_1)\) is unknown.

Question: test \(H_0: \mu = \mu_0\) against \(H_a: \mu \neq \mu_0\)

PLRT: Let \(\hat{\mu}\) denote the maximum likelihood estimate for \(\mu\). Then the likelihood ratio is defined as

\[
\lambda = \prod_{i=1}^{n} f(X_i; g(\mu_0))/\prod_{i=1}^{n} f(X_i; g(\hat{\mu})).
\]

The likelihood ratio test is based on the following Wilks Theorem. Under \(H_0\), \(-2 \log \lambda \xrightarrow{d} \chi^2(1)\) as \(n \to \infty\).
Empirical likelihood method

When we do not fit a class of parametric family to $X_i$, but still test $H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$, a similar approach to the parametric likelihood ratio test was introduced by Owen (1988, 1990), which is a nonparametric likelihood ratio test and called empirical likelihood method.
Define the empirical likelihood ratio function for $\mu$ as

$$R(\mu) = \sup \left\{ \prod_{i=1}^{n} (np_i) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_i = \mu \right\}.$$ 

By Lagrange multiplier technique, we have

$$p_i = n^{-1} \left\{ 1 + \lambda^T (X_i - \mu) \right\}^{-1}$$ and

$$-2 \log R(\mu) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \lambda^T (X_i - \mu) \right\},$$

where $\lambda = \lambda(\mu)$ satisfies

$$n^{-1} \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^T (X_i - \mu)} = 0.$$
Wilks Theorem: Under $H_0$,

$$W(\mu_0) := -2 \log R(\mu_0) \xrightarrow{d} \chi^2(d) \quad \text{as} \quad n \to \infty,$$

where $\mu \in \mathbb{R}^d$.

Confidence interval/region: The above theorem can be employed to construct a confidence interval or region for $\mu$ as

$$I_\alpha = \{ \mu : W(\mu) \geq \chi^2_{d,\alpha} \}.$$

Advantages: i) No need to estimate any additional quantities such as asymptotic variance; ii) the shape of confidence interval/region is determined by the sample automatically; iii) Bartlett correctable
A popular way to formulate the empirical likelihood function is via estimating equations.

Observations: $X_1, \cdots, X_n$ iid with common distribution function $F$ and there is a $q$-dimensional parameter $\theta$ associated with $F$.

Conditions: Let $y^T$ denote the transpose of the vector $y$ and

$$G(x; \theta) = (g_1(x; \theta), \cdots, g_s(x; \theta))^T$$

denote $s(\geq q)$ functionally independent functions, which connect $F$ and $\theta$ through the equations $EG(X_1; \theta) = 0$. 
Empirical likelihood function:

\[ R(\theta) = \sup \left\{ \prod_{i=1}^{n} (np_i) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i G(X_i; \theta) = 0 \right\}. \]

Wilks Theorem: \( -2 \log R(\theta_0) \xrightarrow{d} \chi^2(q) \) as \( n \to \infty \).
Profile empirical likelihood method

Suppose we are only interested in a part of $\theta$. Then like the parametric profile likelihood ratio test, we have the profile empirical likelihood method.

**Observations:** $X_1, \cdots, X_n$ iid with common distribution function $F$ and there is a $q$-dimensional parameter $\theta$ associated with $F$. Write $\theta = (\alpha^T, \beta^T)^T$, where $\alpha$ and $\beta$ are $q_1$-dimensional and $q_2$-dimensional parameters, respectively, and $q_1 + q_2 = q$. Now we are interested in $\alpha$.

**Conditions:** Let $y^T$ denote the transpose of the vector $y$ and

$$G(x; \theta) = (g_1(x; \theta), \cdots, g_s(x; \theta))^T$$

denote $s(\geq q)$ functionally independent functions, which connect $F$ and $\theta$ through the equations $EG(X_1; \theta) = 0$. 

Liang Peng

Interval Estimation
Profile empirical likelihood ratio:

\[ l(\alpha) = 2l_E((\alpha^T, \hat{\beta}^T(\alpha))^T) - 2l_E(\tilde{\theta}), \]

where \( l_E(\theta) = \sum_{i=1}^n \log\{1 + \lambda^T G(X_i; \theta)\} \), \( \lambda = \lambda(\theta) \) is the solution of the following equation

\[ 0 = \frac{1}{n} \sum_{i=1}^n \frac{G(X_i; \theta)}{1 + \lambda^T G(X_i; \theta)}, \]

\( \tilde{\theta} = (\tilde{\alpha}^T, \tilde{\beta}^T)^T \) minimizes \( l_E(\theta) \) with respect to \( \theta \), and \( \hat{\beta}(\alpha) \) minimizes \( l_E((\alpha^T, \beta^T)^T) \) with respect to \( \beta \) for fixed \( \alpha \).

Wilks Theorem: \( l(\alpha_0) \xrightarrow{d} \chi^2(q_1) \) as \( n \to \infty \), where \( \alpha_0 \) denotes the true value of \( \alpha \).
Consider the first-order autoregressive process

\[ Y_i = \beta Y_{i-1} + \epsilon_i, \quad i = 1, \ldots, n, \quad Y_0 = 0, \quad (1) \]

where \( \epsilon_1, \ldots, \epsilon_n \) are independent and identically distributed random variables with mean zero and finite variance \( \sigma^2 \).

The process is called stationary, unit root, near-integrated and explosive for the cases \(|\beta| < 1\), \(|\beta| = 1\), \(\beta = \beta(n) = 1 - \gamma/n\) for some constant \(\gamma \in \mathbb{R}\), and \(|\beta| > 1\), respectively.
Although the behaviour of the process depends on $\beta$, the parameter $\beta$ can be estimated by the usual least squares procedure

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^{n} (Y_i - \beta Y_{i-1})^2 = \frac{\sum_{i=1}^{n} Y_i Y_{i-1}}{\sum_{i=1}^{n} Y_{i-1}^2}.$$ 

The asymptotic behaviour of $\hat{\beta}$ has been studied extensively in the literature. Specifically,
\[
\left( \sum_{i=1}^{n} Y_{i-1}^2 \right)^{1/2} \{ \hat{\beta} - \beta \} \xrightarrow{d} \begin{cases} 
N(0, \sigma^2) & \text{if } |\beta| < 1, \\
\sigma \left( \int_0^1 X(t) \, dW(t) \right) \left/ \left( \int_0^1 X^2(t) \, dt \right)^{1/2} \right. & \text{if } \beta = \beta(n) = 1 - \gamma/n, \\
(1 - \beta^{-2})^{-1/2} \text{sgn}(U)V & \text{if } |\beta| > 1,
\end{cases}
\]

where $\gamma \in \mathbb{R}$, $\{X(t) : 0 \leq t \leq 1\}$ is an Ornstein-Uhlenbeck process satisfying the stochastic differential equation $dX(t) = -\gamma X(t) \, dt + \sigma \, dW(t)$ with $X(0) = 0$, $\{W(t) : 0 \leq t \leq 1\}$ is a standard Brownian motion, and $U$ and $V$ are independent random variables having the same distribution as $\sum_{j=1}^{\infty} \beta^{-(j-1)} \epsilon_j$. 

Liang Peng Interval Estimation
To construct a confidence interval for the autoregressive coefficient $\beta$ or to test $H_0 : \beta = \beta_0$, one usually employs (2) either by means of simulating the limit distribution or using a bootstrap procedure. Since the limit distribution depends on the location of $\beta$, one needs to distinguish the different cases of $\beta$ before simulating from the limit distribution. However, differentiating the near-integrated case ($\beta = 1 - \gamma/n$) from the stationary or the integrated case ($|\beta| \leq 1$) through testing for $\beta$ remains a challenging task as the limit distribution depends on the nuisance parameter $\gamma$ according to (2).
On the other hand, it is known that the full sample bootstrap method (i.e., the resample size is the same as the original sample size) works for both $\hat{\beta} - \beta$ and $(\sum_{i=1}^{n} Y_{i-1}^2)^{1/2}(\hat{\beta} - \beta)$ for the stationary case $|\beta| < 1$; see Bose (1988). For the explosive case $|\beta| > 1$, Basawa et al. (1989) showed that the full sample bootstrap method also works for $(\beta^2 - 1)^{-1}|\beta|^n(\hat{\beta} - \beta)$. But, it follows from the proofs in Basawa et al. (1989) that the full sample bootstrap method is inconsistent for either $\hat{\beta} - \beta$ or $(\sum_{i=1}^{n} Y_{i-1}^2)^{1/2}(\hat{\beta} - \beta)$ when $|\beta| > 1$. Although Ferretti and Romo (1996) showed that the full sample bootstrap method is valid for $(\sum_{i=1}^{n} Y_{i-1}^2)^{1/2}(\hat{\beta} - 1)$ when $\beta = 1$, Datta (1996) proved that it is inconsistent for $(\sum_{i=1}^{n} Y_{i-1}^2)^{1/2}(\hat{\beta} - \beta)$ when $\beta = 1$. 
In summary, neither the limit distribution (2) nor the full sample bootstrap method is applicable to construct a confidence interval for $\beta$ without knowing the true location of $\beta$ a priori.

**Question:** Construct an interval for $\beta$ without knowing the cases of stationary, near unit root or explosive?

**Solution:** First find an estimator which always has a normal limit, but the asymptotic variance may depend on the different cases; Next apply empirical likelihood method.
Weighted estimation: First consider the near-integrated case. Write \( \hat{\beta} - \beta = \sum_{i=1}^{n} \epsilon_i Y_{i-1} / \sum_{i=1}^{n} Y_{i-1}^2 \). When \( \beta = 1 - \gamma/n \), instead of converging to a constant in probability as in the usual case, the quantity \( n^{-2} \sum_{i=1}^{n} E(\epsilon_i^2 Y_{i-1}^2 | F_{i-1}) \) (\( F_i = \sigma(\epsilon_1, \ldots, \epsilon_i) \)) converges in distribution to a random variable. As a result, the limit of \( n^{-1} \sum_{i=1}^{n} \epsilon_i Y_{i-1} \) can no longer be a normal distribution. However, as the proofs of Chan and Wei (1987) indicate, the following equality still holds:

\[
 n^{-1} \sum_{i=1}^{n} E\left( \frac{Y_{i-1}^2}{1 + Y_{i-1}^2} \epsilon_i^2 | F_{i-1} \right) = n^{-1} \sum_{i=1}^{n} E(\epsilon_i^2 | F_{i-1}) + o_p(1) = \sigma^2 + o_p(1).
\]
Motivated by the above equation, a weighted least squares estimate is proposed as follows:

$$\hat{\beta}^w(\delta) = \arg \min_{\beta} \sum_{i=1}^{n} \frac{(Y_i - \beta Y_{i-1})^2}{(1 + \delta Y_{i-1}^2)^{1/2}}$$

$$= \sum_{i=1}^{n} \frac{Y_{i-1} Y_i}{(1 + \delta Y_{i-1}^2)^{1/2}} / \sum_{i=1}^{n} \frac{Y_{i-1}^2}{(1 + \delta Y_{i-1}^2)^{1/2}},$$

where $\delta$ is a positive constant. For such a weighted least squares estimate, asymptotic normality is achieved.
Theorem
Assume model (1) holds with $E|\epsilon_1|^{2+d} < \infty$ for some $d > 0$. Further assume that $\beta$ is either a constant or $\beta = \beta(n) = 1 - \gamma/n$ for some constant $\gamma \in R$. Then for any $\delta > 0$

$$
\left( \sum_{i=1}^{n} \frac{Y_{i-1}^2}{1 + \delta Y_{i-1}^2} \right)^{-1/2} \left( \sum_{i=1}^{n} \frac{Y_{i-1}^2}{(1 + \delta Y_{i-1}^2)^{1/2}} \right) (\hat{\beta}_w(\delta) - \beta) \overset{d}{\rightarrow} N(0, \sigma^2).
$$
Based on this theorem, one constructs a confidence interval for $\beta$ or test $H_0 : \beta = \beta_0$ via estimating $\sigma^2$ by means of
\[
\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \tilde{\beta} Y_{i-1})^2,
\]
where $\tilde{\beta}$ is either the least squares estimate or the weighted least squares estimate $\hat{\beta}^w(\delta)$. Herein, the empirical likelihood method is applied to the weighted score equation of the weighted least squares estimate. Specifically,
define

\[ Z_i(\beta; \delta) = \frac{Y_{i-1}}{(1 + \delta Y_{i-1}^2)^{1/2}} (Y_i - \beta Y_{i-1}) \quad \text{for} \quad i = 2, \ldots, n. \]

Since the weighted least squares estimate \( \hat{\beta}^w(\delta) \) is the solution to the score equation \( \sum_{i=2}^{n} Z_i(\beta; \delta) = 0 \), define the empirical likelihood function as

\[
L(\beta; \delta) = \sup\{ \prod_{i=2}^{n} (np_i) : p_2 \geq 0, \ldots, p_n \geq 0, \sum_{i=2}^{n} p_i = 1, \sum_{i=2}^{n} p_i Z_i(\beta; \delta) = 0 \}.
\]
After applying the Lagrange multiplier technique, one obtains
\[ p_i = n^{-1}(1 + \lambda Z_i(\beta; \delta))^{-1} \]
and the log empirical likelihood ratio becomes
\[ l(\beta; \delta) = -2 \log L(\beta; \delta) = 2 \sum_{i=2}^{n} \log(1 + \lambda Z_i(\beta; \delta)), \]
where \( \lambda = \lambda(\beta, \delta) \) satisfies
\[ n^{-1} \sum_{i=2}^{n} \frac{Z_i(\beta; \delta)}{1 + \lambda Z_i(\beta; \delta)} = 0. \] \( (3) \)

The following theorem shows that the Wilks theorem holds for the proposed empirical likelihood method.
Theorem

Under conditions of Theorem 1, \( l(\beta_0; \delta) \xrightarrow{d} \chi^2(1) \) as \( n \to \infty \) for any given \( \delta > 0 \), where \( \beta_0 \) denotes the true value of \( \beta \).

In particular, using this theorem, one can construct a confidence interval for \( \beta_0 \) with level \( \alpha \) as

\[
I_\alpha = \{ \beta : l(\beta; \delta) \leq \chi^2_{1,\alpha} \},
\]

where \( \chi^2_{1,\alpha} \) denotes the \( \alpha \)-quantile of \( \chi^2(1) \).
**Remark 1.** When the sample size $n$ is not large, the rate of convergence of $Z_i(\beta; \delta)$ to infinity is not very fast in the near-integrated case or the explosive case. Hence a small $\delta$ would be preferred. In the simulation study, $\delta = 1$ is used.
Remark 2. Theorem 2 still holds when \( \{\epsilon_i\} \) is a martingale difference sequence with respect to an increasing sequence of \( \sigma \)-fields \( \{\mathcal{F}_t\} \) such that

\[
\begin{align*}
\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(\epsilon^2_t | \mathcal{F}_{t-1}) & \xrightarrow{p} \sigma^2, \\
\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(|\epsilon_t|^{2+d} | \mathcal{F}_{t-1}) & \xrightarrow{p} c < \infty,
\end{align*}
\]

for some \( d > 0 \) as \( n \to \infty \).
Simulation: A total of 10,000 random samples each with size $n = 200$ are drawn from model (1), where the distribution function of $\epsilon_i$ is either $N(0, 1), t_3$ or $t_6$. Consider the near-integrated situations $\beta = 1 - \gamma/n$ with $\gamma = \pm 50, \pm 10, \pm 5, \pm 3, \pm 1$ and 0. Since $n = 200$, the case of $\gamma = \pm 50$ may be perceived as the explosive case and the stationary case, respectively.
For comparison, a bootstrap confidence interval for $\beta$ at level $\alpha$ is constructed as

$$I^*_\alpha = [\hat{\beta} - c^*_2(\sum_{i=1}^{n} Y^2_{i-1})^{-1/2}, \hat{\beta} - c^*_1(\sum_{i=1}^{n} Y^2_{i-1})^{-1/2}],$$

where $c^*_1$ and $c^*_2$ are the $[B(1 - \alpha)/2]$-th largest value and the $[B(1 + \alpha)/2]$-th largest value of the $B$ bootstrapped versions of the statistic $(\sum_{i=1}^{n} Y^2_{i-1})^{1/2}(\hat{\beta} - \beta)$, respectively. Here we took $B = 5,000$. Theoretically, when $\beta$ is the near-integrated case or the explosive case, this bootstrap confidence interval becomes inconsistent. In calculating the coverage probability for the proposed empirical likelihood method, $\delta = 1$ is used in the 'emplik' package in R software.
In Tables 1–3, the empirical coverage probabilities for these two confidence intervals are reported. Note that the cases $\gamma = -1, -3, -50$ clearly indicate the inconsistencies of the bootstrap methods. In most cases, the proposed empirical likelihood method provides more accurate coverage probabilities than the bootstrap method, and the empirical likelihood method performs reasonably well under all circumstances of $\beta$. 
Table 1: Empirical coverage probabilities are calculated for the empirical likelihood confidence intervals $l_\alpha$ and bootstrap confidence intervals $l_{\alpha}^*$, with $\alpha = 0.55, 0.9, 0.95$ and $\epsilon_i$ being standard normal.
<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$I_{0.55}$</th>
<th>$I_{0.9}$</th>
<th>$I_{0.95}$</th>
<th>$I^*_{0.55}$</th>
<th>$I^*_{0.9}$</th>
<th>$I^*_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.5424</td>
<td>0.9016</td>
<td>0.9517</td>
<td>0.5417</td>
<td>0.8988</td>
<td>0.9497</td>
</tr>
<tr>
<td>10</td>
<td>0.5443</td>
<td>0.9005</td>
<td>0.9477</td>
<td>0.5503</td>
<td>0.8944</td>
<td>0.9466</td>
</tr>
<tr>
<td>5</td>
<td>0.5513</td>
<td>0.9020</td>
<td>0.9513</td>
<td>0.5525</td>
<td>0.8981</td>
<td>0.9489</td>
</tr>
<tr>
<td>3</td>
<td>0.5450</td>
<td>0.9012</td>
<td>0.9483</td>
<td>0.5616</td>
<td>0.9018</td>
<td>0.9504</td>
</tr>
<tr>
<td>1</td>
<td>0.5434</td>
<td>0.8992</td>
<td>0.9494</td>
<td>0.5707</td>
<td>0.9011</td>
<td>0.9481</td>
</tr>
<tr>
<td>0</td>
<td>0.5395</td>
<td>0.8974</td>
<td>0.9504</td>
<td>0.5486</td>
<td>0.8942</td>
<td>0.9470</td>
</tr>
<tr>
<td>−1</td>
<td>0.5378</td>
<td>0.8973</td>
<td>0.9511</td>
<td>0.5010</td>
<td>0.8850</td>
<td>0.9420</td>
</tr>
<tr>
<td>−3</td>
<td>0.5474</td>
<td>0.8990</td>
<td>0.9486</td>
<td>0.5632</td>
<td>0.8662</td>
<td>0.9118</td>
</tr>
<tr>
<td>−5</td>
<td>0.5513</td>
<td>0.8987</td>
<td>0.9480</td>
<td>0.5657</td>
<td>0.9216</td>
<td>0.9566</td>
</tr>
<tr>
<td>−10</td>
<td>0.5568</td>
<td>0.8997</td>
<td>0.9481</td>
<td>0.5611</td>
<td>0.9096</td>
<td>0.9555</td>
</tr>
<tr>
<td>−50</td>
<td>0.5484</td>
<td>0.9013</td>
<td>0.9490</td>
<td>0.7502</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 2: Empirical coverage probabilities are calculated for the empirical likelihood confidence intervals $l_\alpha$ and bootstrap confidence intervals $l^*_\alpha$, with $\alpha = 0.55, 0.9, 0.95$ and $\epsilon_i$ being $t_6$. 
<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$I_{0.55}$</th>
<th>$I_{0.9}$</th>
<th>$I_{0.95}$</th>
<th>$I^*_{0.55}$</th>
<th>$I^*_{0.9}$</th>
<th>$I^*_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.5476</td>
<td>0.8968</td>
<td>0.9475</td>
<td>0.5452</td>
<td>0.8913</td>
<td>0.9446</td>
</tr>
<tr>
<td>10</td>
<td>0.5465</td>
<td>0.8952</td>
<td>0.9452</td>
<td>0.5366</td>
<td>0.8893</td>
<td>0.9422</td>
</tr>
<tr>
<td>5</td>
<td>0.5530</td>
<td>0.8939</td>
<td>0.9469</td>
<td>0.5441</td>
<td>0.8955</td>
<td>0.9451</td>
</tr>
<tr>
<td>3</td>
<td>0.5518</td>
<td>0.8958</td>
<td>0.9452</td>
<td>0.5512</td>
<td>0.8974</td>
<td>0.9497</td>
</tr>
<tr>
<td>1</td>
<td>0.5447</td>
<td>0.8943</td>
<td>0.9452</td>
<td>0.5705</td>
<td>0.8939</td>
<td>0.9424</td>
</tr>
<tr>
<td>0</td>
<td>0.5385</td>
<td>0.8924</td>
<td>0.9445</td>
<td>0.5471</td>
<td>0.8879</td>
<td>0.9408</td>
</tr>
<tr>
<td>-1</td>
<td>0.5339</td>
<td>0.8943</td>
<td>0.9439</td>
<td>0.5046</td>
<td>0.8774</td>
<td>0.9367</td>
</tr>
<tr>
<td>-3</td>
<td>0.5365</td>
<td>0.8972</td>
<td>0.9450</td>
<td>0.5555</td>
<td>0.8628</td>
<td>0.9055</td>
</tr>
<tr>
<td>-5</td>
<td>0.5426</td>
<td>0.8971</td>
<td>0.9469</td>
<td>0.5591</td>
<td>0.9191</td>
<td>0.9564</td>
</tr>
<tr>
<td>-10</td>
<td>0.5444</td>
<td>0.8995</td>
<td>0.9493</td>
<td>0.5632</td>
<td>0.9092</td>
<td>0.9557</td>
</tr>
<tr>
<td>-50</td>
<td>0.5436</td>
<td>0.8986</td>
<td>0.9477</td>
<td>0.7401</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 3: Empirical coverage probabilities are calculated for the empirical likelihood confidence intervals $I_{\alpha}$ and bootstrap confidence intervals $I_{\alpha}^*$, with $\alpha = 0.55, 0.9, 0.95$ and $\epsilon_i$ being $t_3$. 
<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$I_{0.55}$</th>
<th>$I_{0.9}$</th>
<th>$I_{0.95}$</th>
<th>$I^*_{0.55}$</th>
<th>$I^*_{0.9}$</th>
<th>$I^*_{0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.5332</td>
<td>0.8860</td>
<td>0.9380</td>
<td>0.5426</td>
<td>0.8915</td>
<td>0.9432</td>
</tr>
<tr>
<td>10</td>
<td>0.5396</td>
<td>0.8885</td>
<td>0.9389</td>
<td>0.5401</td>
<td>0.8955</td>
<td>0.9435</td>
</tr>
<tr>
<td>5</td>
<td>0.5340</td>
<td>0.8866</td>
<td>0.9415</td>
<td>0.5412</td>
<td>0.8970</td>
<td>0.9461</td>
</tr>
<tr>
<td>3</td>
<td>0.5375</td>
<td>0.8860</td>
<td>0.9413</td>
<td>0.5453</td>
<td>0.9014</td>
<td>0.9489</td>
</tr>
<tr>
<td>1</td>
<td>0.5355</td>
<td>0.8859</td>
<td>0.9396</td>
<td>0.5600</td>
<td>0.8996</td>
<td>0.9477</td>
</tr>
<tr>
<td>0</td>
<td>0.5235</td>
<td>0.8836</td>
<td>0.9405</td>
<td>0.5382</td>
<td>0.8959</td>
<td>0.9464</td>
</tr>
<tr>
<td>−1</td>
<td>0.5223</td>
<td>0.8853</td>
<td>0.9410</td>
<td>0.5012</td>
<td>0.8786</td>
<td>0.9406</td>
</tr>
<tr>
<td>−3</td>
<td>0.5319</td>
<td>0.8823</td>
<td>0.9403</td>
<td>0.5549</td>
<td>0.8567</td>
<td>0.9030</td>
</tr>
<tr>
<td>−5</td>
<td>0.5399</td>
<td>0.8910</td>
<td>0.9451</td>
<td>0.5635</td>
<td>0.9182</td>
<td>0.9568</td>
</tr>
<tr>
<td>−10</td>
<td>0.5368</td>
<td>0.8898</td>
<td>0.9427</td>
<td>0.5526</td>
<td>0.9072</td>
<td>0.9513</td>
</tr>
<tr>
<td>−50</td>
<td>0.5365</td>
<td>0.8838</td>
<td>0.9388</td>
<td>0.7494</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Proof of Theorem 1. Define

\[ X_i = \frac{Y_{i-1}}{n^{1/2}(1 + \delta Y_{i-1}^2)^{1/2}} \epsilon_i \quad \text{and} \quad F_i = \sigma(\epsilon_1, \ldots, \epsilon_i) \]

for \( i = 1, \ldots, n \). Then

\[ \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} \frac{Y_{i-1}^2}{(1 + \delta Y_{i-1}^2)^{1/2}} \right) (\hat{\beta}_w^*(\delta) - \beta) = \sum_{i=1}^{n} X_i. \]  

(4)

When \( |\beta| < 1 \), it is easy to check that

\[ \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{i-1}^2}{1 + \delta Y_{i-1}^2} \xrightarrow{p} \mathbb{E} \frac{Y_1^2}{1 + \delta Y_1^2}. \]  

(5)
When \( \beta = 1 - \gamma/n \) for some constant \( \gamma \), Lemma 2.1 in Chan and Wei (1987) shows that there is a standard Brownian motion \( W_1(t) \) such that

\[
n^{-1/2} \beta[nt]^{-n} Y_{[nt]} \overset{D}{\to} W_1(e^{-2\gamma(e^{2t\gamma} - 1)/(2\gamma)})
\]

in \( D[0, 1] \) as \( n \to \infty \), which implies that

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{Y_{i-1}^2}{1 + \delta Y_{i-1}^2} \overset{p}{\to} \delta^{-1}
\]

as \( n \to \infty \).
When $|\beta| > 1$, it follows that

$$\beta^{2-T} Y_{T-1} \xrightarrow{d} \sum_{i=1}^{\infty} \beta^{1-i} \epsilon_i$$

as $T \to \infty$, which implies that as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{Y_{i-1}^2}{1 + \delta Y_{i-1}^2} \xrightarrow{p} \delta^{-1}. \quad (7)$$
Hence, by (5)–(7),

\[
\sum_{i=1}^{n} \mathbb{E}(X_i^2 | F_i) = \sigma^2 \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{i-1}^2}{1 + \delta Y_{i-1}^2} \xrightarrow{p} \begin{cases}
\sigma^2 \mathbb{E} \frac{Y_1^2}{1 + \delta Y_1^2} & \text{if } |\beta| < 1, \\
\delta^{-1} & \text{if } |\beta| > 1, \\
\delta^{-1} & \text{if } \beta = 1 - \gamma/n.
\end{cases}
\] (8)

Similar to the proofs of (5)–(7), it can be shown that
\[
\sum_{i=1}^{n} \mathbb{E}(X_i^2 I(|X_i| > c) | \mathcal{F}_i) \\
\leq c^{-2-d} \sum_{i=1}^{n} \mathbb{E}(|X_i|^{2+d} | \mathcal{F}_i) \\
= c^{-2-d} \mathbb{E}|\epsilon_1|^{2+d} \sum_{i=1}^{n} \frac{|Y_{i-1}|^{2+d}}{n^{(2+d)/2}} \{1 + \delta Y_{i-1}^2\}^{(2+d)/2} \\
\xrightarrow{p} 0
\]  

for any \( c > 0 \) as \( n \to \infty \). Hence, Theorem 1 follows from (4)–(9) and Corollary 3.1 of Hall and Heyde (1980).
Proof of Theorem 2. It follows from (5)–(7) that

\[
\frac{1}{n} \sum_{i=1}^{n} Z_i^2(\beta_0; \delta) \xrightarrow{p} \begin{cases} 
\sigma^2 E \frac{Y_1^2}{1+\delta Y_1^2} & \text{if } |\beta| < 1, \\
\delta^{-1} & \text{if } \beta = 1 - \gamma/n, \\
\delta^{-1} & \text{if } |\beta| > 1.
\end{cases}
\] (10)

By Corollary 3.1 of Hall and Heyde (1980),

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(\beta_0; \delta) \xrightarrow{d} \begin{cases} 
N(0, \sigma^2 E \frac{Y_1^2}{1+\delta Y_1^2}) & \text{if } |\beta| < 1, \\
N(0, \sigma^2 \delta^{-1}) & \text{if } \beta = 1 - \gamma/n, \\
N(0, \sigma^2 \delta^{-1}) & \text{if } |\beta| > 1.
\end{cases}
\] (11)
It is easy to check that

$$\max_{1 \leq i \leq n} |Z_i(\beta_0; \delta)| = o(n^{-1/2}) \quad \text{a.s.} \quad (12)$$

Hence, Theorem 2 can be proved by using (10)–(12) and standard arguments in Chapter 11 of Owen (2001).
Consider a generalized autoregressive conditional heteroskedastic (GARCH) model given by

\[ Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + a\sigma_{t-1}^2 + bY_{t-1}^2, \] (13)

for \( t = 1, \ldots, n \), where \( \{\varepsilon_t\} \) are independent and identically distributed random variables with mean zero and variance one, \( \omega > 0, a \geq 0 \) and \( b \geq 0 \).

It follows from Kesten (1973) or Goldie (1991) that under some regularity conditions, there exists a constant \( \alpha \) such that

\[ E(a + b\varepsilon_t^2)^\alpha = 1, \] (14)

where \( \alpha \) is the tail index of the time series \( \{Y_t\} \).
Specifically, Goldie (1991) showed that there exists a positive constant $c$ such that

$$P(|Y_1| > x) = cx^{-\alpha} \{1 + o(1)\}, \quad \text{as } x \to \infty. \quad (15)$$

Consequently, knowing the tail index $\alpha$ would offer important information to understand many properties of $\{Y_t\}$. 
For example, knowing whether $\alpha$ exceeds $1/2$ or 1 is instrumental in testing for structural changes in stock price, see Quinton, Fan and Phillips (2001). Similarly, the limit distribution of the sample covariances of $\{Y_t\}$ depends on whether $\alpha \in (0, 4]$, $\alpha \in (4, 8]$ or $\alpha > 8$, see Mikosch and Stărică (2000). Also, the tail empirical process of $\{Y_t\}$ is determined by the value of $\alpha$ (see Drees (2000)) and for a fixed integer $m$, the tail index of the joint distribution of $(Y_1, \ldots, Y_m)$ is simply $\alpha$ (see Basrak, Davis and Mikosch (2002)).
In view of (14), Mikosch and Stărică (2000) and Berkes, Horváth and Kokoszka (2003) proposed estimating $\alpha$ by solving the equation

$$\frac{1}{n} \sum_{i=1}^{n} (\hat{a} + b\hat{\varepsilon}_t^2)^\alpha = 1, \quad (16)$$

where $\hat{a}$ and $\hat{b}$ are the quasi maximum likelihood estimate (QMLE) of $a$ and $b$, and $\hat{\varepsilon}_t = Y_t/\hat{\sigma}_t$ is an estimator of $\varepsilon_t$ with $(\omega, a, b)^T$ with being replaced by the QMLE. Denote such an estimate by $\hat{\alpha}$. For a detailed study of the asymptotic limit of the quasi maximum likelihood estimation of a GARCH($p$, $q$) model, see Hall and Yao (2003).
Based on the asymptotic distribution of $\hat{\alpha}$ given in Berkes, Horváth and Kokoszka (2003), one can construct an interval estimate of $\alpha$ by means of estimating the asymptotic variance of $\hat{\alpha}$. Since the plug-in estimators $\hat{a}$ and $\hat{b}$ are employed, the proposed estimator $\hat{\alpha}$ has a non-trivial asymptotic variance that depends on the asymptotic limit of the plug-in estimators. To circumvent the difficulty of estimating the asymptotic variance, we propose to apply the empirical likelihood method to construct a confidence interval for $\alpha$. 
Methodology: Let $\theta = (\omega, a, b)^T$ and assume that in model (13) $E \log(a + b\epsilon_t^2) < 0$. This implies that

$$
\sigma_t^2 := \sigma_t^2(\theta) = \frac{\omega(1 - a^t)}{1 - a} + \sum_{k=0}^{t-1} ba^k Y_{t-1-k} + a^t \sigma_0^2(\theta). \quad (17)
$$

For $t = 1, \ldots, n$, define
\[
\bar{\sigma}_t^2(\theta) = \frac{1 - a^t}{1 - a} + b \sum_{0 \leq k \leq t-1} a^k Y_{t-k-1}^2,
\]
\[
Z_{t,1}(\theta, \alpha) = \left\{ a + b Y_t^2 / \sigma_t^2(\theta) \right\}^\alpha - 1,
\]
\[
\tilde{Z}_{t,1}(\theta, \alpha) = \left\{ a + b Y_t^2 / \bar{\sigma}_t^2(\theta) \right\}^\alpha - 1,
\]
\[
Z_{t,2}(\theta) = \frac{\partial}{\partial \omega} \left\{ Y_t^2 / \sigma_t^2(\theta) + \log(\sigma_t^2(\theta)) \right\},
\]
\[
\tilde{Z}_{t,2}(\theta) = \frac{\partial}{\partial \omega} \left\{ Y_t^2 / \bar{\sigma}_t^2(\theta) + \log(\bar{\sigma}_t^2(\theta)) \right\},
\]
\[
Z_{t,3}(\theta) = \frac{\partial}{\partial a} \left\{ Y_t^2 / \sigma_t^2(\theta) + \log(\sigma_t^2(\theta)) \right\},
\]
\[
\tilde{Z}_{t,3}(\theta) = \frac{\partial}{\partial a} \left\{ Y_t^2 / \bar{\sigma}_t^2(\theta) + \log(\bar{\sigma}_t^2(\theta)) \right\},
\]
\[
Z_{t,4}(\theta) = \frac{\partial}{\partial b} \left\{ Y_t^2 / \sigma_t^2(\theta) + \log(\sigma_t^2(\theta)) \right\},
\]
\[
\tilde{Z}_{t,4}(\theta) = \frac{\partial}{\partial b} \left\{ Y_t^2 / \bar{\sigma}_t^2(\theta) + \log(\bar{\sigma}_t^2(\theta)) \right\}.
\]
Then the quasi maximum likelihood estimate for $\theta$ is the solution to the score equations \( \sum_{t=1}^{n} \tilde{Z}_{t,i}(\theta) = 0 \) for \( i = 2, 3, 4 \). We propose to apply the profile empirical likelihood method based on estimating equations in Qin and Lawless (1994) to the equations \( EZ_t(\theta, \alpha) = 0 \), where \( Z_t(\theta, \alpha) = (Z_{t,1}(\theta, \alpha), Z_{t,2}(\theta), Z_{t,3}(\theta), Z_{t,4}(\theta))^T \) for \( t = 1, \ldots, n \).
In other words, define the empirical likelihood function of $(\theta, \alpha)$ as

$$L(\theta, \alpha) = \sup \{ \prod_{t=1}^{n} (np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \bar{Z}_t(\theta, \alpha) = 0 \},$$

where $\bar{Z}_t(\theta, \alpha) = (\bar{Z}_{t,1}(\theta, \alpha), \bar{Z}_{t,2}(\theta), \bar{Z}_{t,3}(\theta), \bar{Z}_{t,4}(\theta))^T$ for $t = 1, \ldots, n$. 
By virtue of the Lagrange multipliers, it is clear that
\[ p_t = n^{-1} \{ 1 + \lambda^T \tilde{Z}_t(\theta, \alpha) \}^{-1} \text{ for } t = 1, \ldots, n \] and

\[ l(\theta, \alpha) := -2 \log L(\theta, \alpha) = 2 \sum_{t=1}^{n} \log \{ 1 + \lambda^T \tilde{Z}_t(\theta, \alpha) \}, \]

where \( \lambda = \lambda(\theta, \alpha) \) satisfies

\[ \sum_{t=1}^{n} \frac{\tilde{Z}_t(\theta, \alpha)}{1 + \lambda^T \tilde{Z}_t(\theta, \alpha)} = 0. \quad (18) \]

Since we are only interested in the tail index \( \alpha \), consider the profile empirical likelihood function \( l_p(\alpha) = l(\tilde{\theta}(\alpha), \alpha) \), where

\[ \tilde{\theta} = \tilde{\theta}(\alpha) := \arg \min_{\theta} l(\theta, \alpha). \]
Throughout we use $\theta_0$ and $\alpha_0$ to denote the true values of $\theta$ and $\alpha$ respectively. First we show the existence and consistency of $\tilde{\theta}(\alpha_0)$. 
Proposition

Suppose (13) holds with $E\log(a + b\epsilon_t^2) < 0$, $a < 1$ and $E|\epsilon_1|^{4\delta} < \infty$ for some $\delta > \max(1, \alpha_0)$. Then, with probability tending to 1, $l(\theta, \alpha_0)$ attains its minimum value at some point $\tilde{\theta}$ in the interior of the ball $\Theta = \{\theta : \|\theta - \theta_0\| \leq Cn^{-1/(2\gamma)}\}$ for any given $\gamma \in (1, \min\{\delta/\alpha_0, \delta\})$ and $C > 0$. Moreover $\tilde{\theta}$ and $\tilde{\lambda} = \tilde{\lambda}(\tilde{\theta}, \alpha_0)$ satisfy

$$Q_{1n}(\tilde{\theta}, \tilde{\lambda}) = 0 \quad \text{and} \quad Q_{2n}(\tilde{\theta}, \tilde{\lambda}) = 0,$$

(19)

where $Q_{1n}(\theta, \lambda) = \frac{1}{n} \sum_{t=1}^{n} \frac{\bar{Z}_t(\theta, \alpha_0)}{1 + \lambda^T \bar{Z}_t(\theta, \alpha_0)}$, $Q_{2n}(\theta, \lambda) = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{1 + \lambda^T \bar{Z}_t(\theta, \alpha_0)} \left\{ \frac{\partial}{\partial \theta} \bar{Z}_t(\theta, \alpha_0) \right\}^T \lambda$. 
Theorem
Under conditions of Proposition 2.1, the random variable \( l_p(\alpha_0) \), with \( \tilde{\theta} \) given in Proposition 2.1, converges in distribution to \( \chi^2(1) \) as \( n \to \infty \).

Corollary
For any \( 0 < \varsigma < 1 \), let \( \mu_\varsigma \) denote the \( \varsigma \)-th quantile of a \( \chi^2(1) \) random variable and define the empirical likelihood confidence interval with level \( \varsigma \) as \( I_\varsigma = \{ \alpha \mid l_p(\alpha) \leq \mu_\varsigma \} \). Then
\[
P(\alpha_0 \in I_\varsigma) \longrightarrow \varsigma \quad \text{as} \quad n \to \infty.
\]
Remark: Note that instead of assuming $\delta > \alpha_0/2$ as stipulated in Berkes, Horváth and Kokoszka (2003), we assume that $\delta > \alpha_0$. The reason is that to establish the asymptotic normality of \[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{(a_0 + b_0 \varepsilon_t^2)^{\alpha_0} - 1\}, \]
the condition \[
E\{(a_0 + b_0 \varepsilon_t^2)^{\alpha_0} - 1\}^2 < \infty \] is required.
Simulation:
Consider model (13) with \((\omega, a, b) = (1, 0.1, 0.8), (1, 0.8, 0.1)\) and \(\varepsilon_t \sim N(0, 1)\) or \(t(\nu)/\sqrt{\nu}/(\nu - 2)\). First we draw 100,000 random samples of size \(n = 10,000\) for the errors in (13). For each random sample, we solve the equation (14) with the right-hand side replaced by the sample average to obtain \(\alpha\), say \(\tilde{\alpha}\). Then the true value \(\alpha_0\) is calculated as the average of these 100,000 \(\tilde{\alpha}\)'s.
Next we draw 1,000 random samples each of size $n = 500$ and 2,000 from (13). Since the asymptotic limit of $\hat{\alpha}$ derived in Berkes, Horváth and Kokoszka (2003) is quite complicated, we consider the percentile bootstrap confidence interval for $\alpha$ based on $\hat{\alpha}$. Specifically, draw $B = 200$ bootstrap samples each of size $n$ from $\hat{\varepsilon}_1, \cdot \cdot \cdot, \hat{\varepsilon}_n$ with replacement, say $\varepsilon^*_1, \cdot \cdot \cdot, \varepsilon^*_n$, and then generate observations from model (13) with $(\omega, a, b)$ and $\varepsilon'_t$'s being replaced by the QMLE and $\varepsilon^*_t$'s. Based on these generated observations, the bootstrap version of $\hat{\alpha}$ is calculated. Hence the percentile bootstrap confidence interval with level $\varsigma$ is obtained via the $B$ bootstrapped estimators of $\alpha$. Denote it by $I^b_\varsigma$. 
The empirical coverage probabilities for the percentile bootstrap confidence interval $I_{\varsigma}^b$ and the proposed empirical likelihood confidence interval $I_{\varsigma}$ at levels $\varsigma = 0.9$ and $0.95$ are reported in Tables 4 and 5. Since computing the coverage probability of the percentile bootstrap interval $I_{\varsigma}^b$ is very time-consuming, which is almost $B$ times of the time required for the empirical likelihood method, we only calculate $I_{\varsigma}^b$ for sample size $n = 500$. 
We observe from Tables 4 and 5 that: (i) the proposed empirical likelihood method performs much better than the percentile bootstrap method when \((\omega, a, b) = (1, 0.1, 0.8)\), (ii) the performance of \(\hat{\alpha}\) is very poor when \((\omega, a, b) = (1, 0.8, 0.1)\) and \(n = 500\), and (iii) both the accuracy of the proposed empirical likelihood method and the tail index estimator improve when the sample size increases. In conclusion, the simulation results show that the proposed empirical likelihood method offers a more accurate coverage probability than the percentile bootstrap method and the computational challenge of the proposed method is much less severe than the bootstrap method.
Table 4: Empirical coverage probabilities of the proposed empirical likelihood interval $I_\varsigma$ and the percentile bootstrap interval $I^b_\varsigma$ at levels $\varsigma = 0.9$ and 0.95. The true value of the tail index $\alpha$, the sample mean and sample standard deviation of $\hat{\alpha}$ are respectively given in the last three columns of the table. Here $n = 500$. 
<table>
<thead>
<tr>
<th>$\omega, a, b, \epsilon_t$</th>
<th>$l_{0.9}$</th>
<th>$l_{0.9}^b$</th>
<th>$l_{0.95}$</th>
<th>$l_{0.95}^b$</th>
<th>$\alpha_0$</th>
<th>$E(\hat{\alpha})$</th>
<th>$SD(\hat{\alpha})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1, 0.1, 0.8, N)$</td>
<td>0.893</td>
<td>0.757</td>
<td>0.934</td>
<td>0.767</td>
<td>1.194</td>
<td>1.253</td>
<td>0.217</td>
</tr>
<tr>
<td>$(1, 0.8, 0.1, N)$</td>
<td>0.846</td>
<td>0.702</td>
<td>0.906</td>
<td>0.768</td>
<td>6.282</td>
<td>7.824</td>
<td>7.353</td>
</tr>
<tr>
<td>$(1, 0.1, 0.8, t(5))$</td>
<td>0.810</td>
<td>0.728</td>
<td>0.877</td>
<td>0.764</td>
<td>1.123</td>
<td>1.238</td>
<td>0.268</td>
</tr>
<tr>
<td>$(1, 0.8, 0.1, t(5))$</td>
<td>0.718</td>
<td>0.849</td>
<td>0.799</td>
<td>0.880</td>
<td>2.806</td>
<td>4.969</td>
<td>5.994</td>
</tr>
<tr>
<td>$(1, 0.1, 0.8, t(10))$</td>
<td>0.871</td>
<td>0.709</td>
<td>0.933</td>
<td>0.724</td>
<td>1.161</td>
<td>1.232</td>
<td>0.240</td>
</tr>
<tr>
<td>$(1, 0.8, 0.1, t(10))$</td>
<td>0.811</td>
<td>0.807</td>
<td>0.872</td>
<td>0.836</td>
<td>4.249</td>
<td>6.281</td>
<td>7.022</td>
</tr>
</tbody>
</table>
Table 5: Empirical coverage probabilities of the proposed empirical likelihood interval $I_{\varsigma}$ and the percentile bootstrap interval $I_{\varsigma}^b$ at levels $\varsigma = 0.9$ and 0.95. The true value of the tail index $\alpha$, the sample mean and sample standard deviation of $\hat{\alpha}$ are respectively given in the last three columns of the table. Here $n = 2000$. 
<table>
<thead>
<tr>
<th>$(\omega, a, b, \epsilon_t)$</th>
<th>$l_{0.9}$</th>
<th>$l_{0.9}^b$</th>
<th>$l_{0.95}$</th>
<th>$l_{0.95}^b$</th>
<th>$\alpha_0$</th>
<th>$E(\hat{\alpha})$</th>
<th>$SD(\hat{\alpha})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1, 0.1, 0.8, N)$</td>
<td>0.911</td>
<td></td>
<td>0.952</td>
<td></td>
<td>1.194</td>
<td>1.208</td>
<td>0.107</td>
</tr>
<tr>
<td>$(1, 0.8, 0.1, N)$</td>
<td>0.885</td>
<td></td>
<td>0.932</td>
<td></td>
<td>6.282</td>
<td>6.616</td>
<td>1.711</td>
</tr>
<tr>
<td>$(1, 0.1, 0.8, t(5))$</td>
<td>0.865</td>
<td></td>
<td>0.924</td>
<td></td>
<td>1.123</td>
<td>1.149</td>
<td>0.129</td>
</tr>
<tr>
<td>$(1, 0.8, 0.1, t(5))$</td>
<td>0.803</td>
<td></td>
<td>0.893</td>
<td></td>
<td>2.806</td>
<td>3.192</td>
<td>1.032</td>
</tr>
<tr>
<td>$(1, 0.1, 0.8, t(10))$</td>
<td>0.883</td>
<td></td>
<td>0.935</td>
<td></td>
<td>1.161</td>
<td>1.180</td>
<td>0.121</td>
</tr>
<tr>
<td>$(1, 0.8, 0.1, t(10))$</td>
<td>0.841</td>
<td></td>
<td>0.898</td>
<td></td>
<td>4.249</td>
<td>4.640</td>
<td>1.268</td>
</tr>
</tbody>
</table>
EL for conditional VaR

GARCH\((p,q)\) model

\[ X_t = \epsilon_t \sqrt{h_t} \quad \text{and} \quad h_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \quad (20) \]

where \( \omega > 0, \alpha_i, \beta_j \geq 0 \) are constants and \( \epsilon_t \) (error): i.i.d. random variables with mean 0 and variance 1. Denote \( \lambda = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T \in \Theta \) and write \( h_t \) as \( h_t(\lambda) \).

- \( \omega \): the baseline variance
- \( \alpha_i \): quantifies the impact of shocks
- \( \beta_j \): quantifies the persistence of shocks.
Remark

- Put $\beta_j = 0$ in model (20), it becomes ARCH($p$) model.

- GARCH models are applied in finance and economics used for option pricing estimating volatility for risk management purposes.
For \( r \in (0, 1) \), the one-step ahead 100\( r \)% CVaR, given \( X_1, \cdots , X_n \), is defined as

\[
q_r = \inf \{ x : P(X_{n+1} \leq x \mid X_{n+1-k}, k \geq 1) \geq r \}.
\]

An obvious estimator for CVaR \( q_r \) is

\[
\hat{q}_r = \sqrt{h_{n+1}(\hat{\lambda})} \quad \hat{\theta}_{\epsilon,r},
\]

where \( \hat{\lambda} \) is an estimator for \( \lambda \) and \( \hat{\theta}_{\epsilon,r} \) is an estimator of the 100\( r \)% quantile of \( \epsilon_t \).
\hat{\lambda} \text{ can be the quasi-maximum likelihood estimator.}

\hat{\theta}_{\epsilon,r} \text{ can be estimated as the } r\text{th sample quantile based on the estimated errors } \hat{\epsilon}_1, \ldots, \hat{\epsilon}_n.

However, the asymptotic variance is complicated since the variability of parameter estimation involves.

Here we propose the estimating equation approach of EL to take the variability of parameter estimation into account.
Define the quasi maximum likelihood as

\[ L(\lambda) = \sum_{t=1}^{n} l_t(\lambda) \quad \text{and} \quad l_t(\lambda) = -\frac{1}{2} \log h_t(\lambda) - \frac{X_t^2}{2h_t(\lambda)}, \]

Score equation: \( \sum_{t=1}^{n} D_t(\lambda) = 0, \) where

\[ D_t(\lambda) = \frac{\partial l_t(\lambda)}{\partial \lambda} = \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \lambda} \left\{ \frac{X_t^2}{h_t(\lambda)} - 1 \right\}. \]

Let \( K \) be a symmetric density function with support in \([-1, 1]\). Put \( G(x) = \int_{-\infty}^{x} K(y)dy. \)
Then we can construct the EL ratio for the $r$th CVaR $\theta$ as

$$L_n(r; \theta, \lambda) = \sup \left\{ \prod_{t=1}^{n} (np_t) : \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t G\left(\frac{\theta}{\sqrt{h_{n+1}(\lambda)}} - \epsilon_t(\lambda)\right) = r, \right\},$$

$$\sum_{t=1}^{n} p_t D_t(\lambda) = 0, \quad p_i > 0, \ i = 1, \ldots, n,$$

where $\epsilon_t(\lambda) = \frac{X_t}{\sqrt{h_t(\lambda)}}$, $r \in (0, 1)$ and $h = h(n) > 0$ is a bandwidth.
Denote

\[ g_t(\theta, \lambda) = \left( G\left( \frac{\theta / \sqrt{h_n + 1(\lambda)} - \epsilon_t(\lambda)}{h} \right) - r, D_t^T(\lambda) \right)^T = (\omega_t(\lambda), D_t^T(\lambda))^T. \]

Then the EL ratio can be rewritten as

\[ L_n(r; \theta, \lambda) = \sup \left\{ \prod_{t=1}^{n} (np_t) : \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t g_t(\theta, \lambda) = 0, \right. \]

\[ \left. p_i > 0, i = 1, \ldots, n \right\}. \]
By the argument of Lagrange multipliers, we get the log EL ratio

$$l_n(r; \theta, \lambda) = 2 \sum_{t=1}^{n} \log \left\{ 1 + b^T(\theta, \lambda)g_t(\theta, \lambda) \right\},$$

where $b(\theta, \lambda)$ satisfies

$$\frac{1}{n} \sum_{t=1}^{n} \frac{g_t(\theta, \lambda)}{1 + b^T(\theta, \lambda)g_t(\theta, \lambda)} = 0.$$
Our interest is $\theta$, we consider the profiled likelihood ratio $l_n(r; \theta, \hat{\lambda}(\theta))$, where $\hat{\lambda}(\theta) = \arg \min_\lambda l_n(r; \theta, \lambda)$.

**Proposition.** Suppose $E \epsilon_{t}^{4+\delta} < \infty$ for some $\delta > 0$, and $n^{1-\sigma} h^2 \to \infty$ and $nh^4 \to 0$ for some $\sigma \in (0, 1/2)$ as $n \to \infty$. Then, with probability tending to one, $l_n(r; \theta_0, \lambda)$ attains its minimum value at some point $\hat{\lambda}_n(\theta_0)$ in the interior of $V_n = \{ \lambda : |\lambda - \lambda_0| \leq n^{-0.5+\sigma/4} \}$. 
Theorem. Under conditions of the above proposition, we have

\[ l_n(r; \theta_0, \hat{\lambda}_n(\theta_0)) \xrightarrow{d} \chi^2_1 \quad \text{as} \quad n \to \infty, \]

where \( \hat{\lambda}_n(\theta_0) \) is given in Proposition 1.

Based on the above theorem, a confidence interval of \( \theta_0 \) with level \( \gamma \) can be obtained as

\[ I_\gamma(r) = \{ \theta : l_n(r; \theta, \hat{\lambda}_n(\theta)) \leq \chi^2_{1,\gamma} \}, \]

where \( \chi^2_{1,\gamma} \) is the \( \gamma \)th quantile of \( \chi^2_1 \).
Simulation
Consider ARCH(1) models and the conditional VaR at \( X_n = 0 \) and 1 (i.e., \( z_1 = 0, 1 \)) with level \( r = 0.90, 0.95, \) and 0.99. In each case, data are generated from an ARCH(1) model with parameters \((\omega, \alpha_1) = (0.5, 0.7) \) or \((0.5, 0.9)\), and the error \( \epsilon_t \) has a \( N(0, 1) \) distribution and a standardized \( t \)-distribution with the degree of freedom 5, i.e., \( \sqrt{\frac{3}{5}} t(5) \). From each model, we draw random samples with the sample sizes \( n = 1000 \) and 3000. The number of replications is set to 5000. We employ the kernel \( K(x) = \frac{15}{16} (1 - x^2)^2 I(|x| \leq 1) \) and choose the bandwidth \( h = n^{-1/3} \).
Based on the 5000 replications, we report the empirical coverage probabilities with confidence levels \( \gamma = 0.90, 0.95 \).
Table 1: Coverage probabilities for $\epsilon_t \sim N(0, 1)$.

<table>
<thead>
<tr>
<th>$(n, \alpha_1, z_1, \gamma)$</th>
<th>$\epsilon_t \sim N(0, 1)$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r=0.90$</td>
<td>$r=0.95$</td>
</tr>
<tr>
<td>(1000, 0.7, 0, 0.90)</td>
<td>0.9070</td>
<td>0.9120</td>
</tr>
<tr>
<td>(3000, 0.7, 0, 0.90)</td>
<td>0.9088</td>
<td>0.8980</td>
</tr>
<tr>
<td>(1000, 0.9, 0, 0.90)</td>
<td>0.9040</td>
<td>0.9106</td>
</tr>
<tr>
<td>(3000, 0.9, 0, 0.90)</td>
<td>0.9010</td>
<td>0.9042</td>
</tr>
<tr>
<td>(1000, 0.7, 1, 0.90)</td>
<td>0.9000</td>
<td>0.9018</td>
</tr>
<tr>
<td>(3000, 0.7, 1, 0.90)</td>
<td>0.9048</td>
<td>0.9052</td>
</tr>
<tr>
<td>(1000, 0.9, 1, 0.90)</td>
<td>0.9030</td>
<td>0.9088</td>
</tr>
<tr>
<td>(3000, 0.9, 1, 0.90)</td>
<td>0.9038</td>
<td>0.8988</td>
</tr>
<tr>
<td>((n, \alpha_1, z_1, \gamma))</td>
<td>(\epsilon_t \sim N(0, 1))</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(r=0.90)</td>
<td>0.9526</td>
</tr>
<tr>
<td>(1000, 0.7, 0, 0.95)</td>
<td>0.9536</td>
<td>0.9504</td>
</tr>
<tr>
<td>(3000, 0.7, 0, 0.95)</td>
<td>0.9546</td>
<td>0.9576</td>
</tr>
<tr>
<td>(1000, 0.9, 0, 0.95)</td>
<td>0.9488</td>
<td>0.9514</td>
</tr>
<tr>
<td>(3000, 0.9, 0, 0.95)</td>
<td>0.949</td>
<td>0.9564</td>
</tr>
<tr>
<td>(1000, 0.7, 1, 0.95)</td>
<td>0.9504</td>
<td>0.9496</td>
</tr>
<tr>
<td>(3000, 0.7, 1, 0.95)</td>
<td>0.9498</td>
<td>0.9542</td>
</tr>
<tr>
<td>(1000, 0.9, 1, 0.95)</td>
<td>0.9520</td>
<td>0.9474</td>
</tr>
<tr>
<td>(3000, 0.9, 1, 0.95)</td>
<td>0.9526</td>
<td>0.9584</td>
</tr>
</tbody>
</table>
### Table 2: Coverage probabilities for $\epsilon_t \sim \sqrt{\frac{3}{5}} t(5)$

<table>
<thead>
<tr>
<th>$(n, \alpha_1, z_1, \gamma)$</th>
<th>$r=0.90$</th>
<th>$r=0.95$</th>
<th>$r=0.99$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1000, 0.7, 0, 0.90)</td>
<td>0.8816</td>
<td>0.8880</td>
<td>0.9252</td>
</tr>
<tr>
<td>(3000, 0.7, 0, 0.90)</td>
<td>0.9030</td>
<td>0.8932</td>
<td>0.9060</td>
</tr>
<tr>
<td>(1000, 0.9, 0, 0.90)</td>
<td>0.8694</td>
<td>0.8798</td>
<td>0.9396</td>
</tr>
<tr>
<td>(3000, 0.9, 0, 0.90)</td>
<td>0.8870</td>
<td>0.8862</td>
<td>0.9110</td>
</tr>
<tr>
<td>(1000, 0.7, 1, 0.90)</td>
<td>0.8928</td>
<td>0.8904</td>
<td>0.9408</td>
</tr>
<tr>
<td>(3000, 0.7, 1, 0.90)</td>
<td>0.8980</td>
<td>0.9044</td>
<td>0.9066</td>
</tr>
<tr>
<td>(1000, 0.9, 1, 0.90)</td>
<td>0.8772</td>
<td>0.8966</td>
<td>0.9468</td>
</tr>
<tr>
<td>(3000, 0.9, 1, 0.90)</td>
<td>0.8918</td>
<td>0.8952</td>
<td>0.9276</td>
</tr>
<tr>
<td>((n, \alpha_1, z_1, \gamma))</td>
<td>(\epsilon_t \sim \sqrt{\frac{3}{5}} t(5))</td>
<td>(r=0.90)</td>
<td>(r=0.95)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>(1000, 0.7, 0, 0.95)</td>
<td>0.9358</td>
<td>0.9406</td>
<td>0.9766</td>
</tr>
<tr>
<td>(3000, 0.7, 0, 0.95)</td>
<td>0.9514</td>
<td>0.9468</td>
<td>0.9542</td>
</tr>
<tr>
<td>(1000, 0.9, 0, 0.95)</td>
<td>0.9364</td>
<td>0.9356</td>
<td>0.9788</td>
</tr>
<tr>
<td>(3000, 0.9, 0, 0.95)</td>
<td>0.9388</td>
<td>0.9386</td>
<td>0.9598</td>
</tr>
<tr>
<td>(1000, 0.7, 1, 0.95)</td>
<td>0.9420</td>
<td>0.9420</td>
<td>0.9876</td>
</tr>
<tr>
<td>(3000, 0.7, 1, 0.95)</td>
<td>0.9494</td>
<td>0.9524</td>
<td>0.9554</td>
</tr>
<tr>
<td>(1000, 0.9, 1, 0.95)</td>
<td>0.9380</td>
<td>0.9470</td>
<td>0.9856</td>
</tr>
<tr>
<td>(3000, 0.9, 1, 0.95)</td>
<td>0.9458</td>
<td>0.9460</td>
<td>0.9656</td>
</tr>
</tbody>
</table>
THANKS!