Confidence Intervals for the Autocorrelations of the Squares of GARCH Sequences

Piotr Kokoszka, Gilles Teyssière and Aonan Zhang

Utah State University, Mathematics and Statistics

NBG Bank

Utah State University, Mathematics and Statistics

piotr@math.usu.edu

stats@gillesteyssiere.net

aonanzhang@cc.usu.edu

ICCS 2004, Kraków: June 8, 2004

Purpose of the paper:

Compare finite sample performance of several methods for finding confidence intervals for autocorrelations of squared returns on speculative assets $X_t^2, t = 1, \ldots, T$, by means of their empirical coverage probability.

Suppose we have a method of constructing, say, a 95% confidence interval $(\hat{l}_n, \hat{u}_n)$ from an observed realization $X_1, X_2, \ldots, X_T$.

We simulate a large number R of realizations from a specific GARCH type model from which we construct R confidence intervals
$(\hat{l}_n^{(r)}, \hat{u}_n^{(r)}),\ r = 1, 2, \ldots, R.$

The percentage of these confidence intervals that contain the population autocorrelation is the ECP, which we want to be as close as possible to the nominal coverage probability of 95%.

Ultimate goal: to recommend a practical procedure for finding confidence intervals for squared autocorrelations which assumes minimal prior knowledge of the stochastic mechanism generating the returns.

Autocorrelations of Squared Returns

$$\hat{\gamma}_{T,X^2}(h) = \frac{1}{T}\sum_{t=1}^{T-h}\left(X_t^2 - \frac{1}{T-h}\sum_{t=1}^{T-h}X_t^2\right)\left(X_{t+h}^2 - \frac{1}{T-h}\sum_{t=h+1}^{T}X_t^2\right)$$

whereas the population autocovariances are

$$\gamma_{X^2}(h) = E\left[(X_0^2 - EX_0^2)(X_h^2 - EX_0^2)\right].$$

The corresponding autocorrelations are

$$\hat{\rho}_{T,X^2}(h) = \frac{\hat{\gamma}_{T,X^2}(h)}{\hat{\gamma}_{T,X^2}(0)}, \qquad \rho_{X^2}(h) = \frac{\gamma_{X^2}(h)}{\gamma_{X^2}(0)}.$$

We focus on the lag 1 autocorrelation, i.e., $h = 1$.

Confidence intervals for autocorrelations of squared returns

Residual Bootstrap

GARCH(1,1) model

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha X_{t-1}^2.$$

1. Estimate $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ and compute $\hat{Z}_t = [\hat{\omega} + \beta\sigma_{t-1}^2 \hat{\alpha} X_{t-1}^2]^{-1/2} X_t$, with $X_0 = \bar{X}_t$.

2. Form B bootstrap realizations
$X_t^2(b) = [\hat{\omega} + \hat{\alpha} X_{t-1}^2(b)]\hat{Z}_t^2(b), \quad t = 1, 2, \ldots, T$, where
$\hat{Z}_1^2(b), \ldots \hat{Z}_T^2(b),\ b = 1, 2, \ldots, B$, are the B bootstrap samples selected with replacement from the squared residuals $\hat{Z}_1^2, \ldots, \hat{Z}_T^2$.

3. Calculate the bootstrap autocorrelations $\rho_{T,X^2}^{(b)}(1),\ b = 1, 2, \ldots, B$ and use their empirical quantiles to find a confidence interval for $\rho_{T,X^2}(1)$.

Confidence intervals for ACF of squared returns (cont 1.)

Denote by $F^*_{\rho(1)}$ the EDF (empirical distribution function) of the $\rho^{(b)}_{T,X^2}(1)$, $b = 1, 2, \ldots, B$.

We consider two types of confidence intervals:

- Equal-tailed confidence interval: the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of $F^*_{\rho(1)}$ yield an *equal-tailed* $(1-\alpha)$ level confidence interval.
- Symmetric confidence interval: let $F^*_{\rho(1),|\cdot|}$ be the empirical distribution of the B values $|\rho^{(b)}_{T,X^2}(1) - \hat{\rho}_{T,X^2}(1)|$.
  Denote by $q_{|\cdot|}(1-\alpha)$ the $(1-\alpha)$ quantile of $F^*_{\rho(1),|\cdot|}$.
  The *symmetric* confidence interval is

$$\left(\hat{\rho}_{T,X^2}(1) - q_{|\cdot|}(1-\alpha), \quad \hat{\rho}_{T,X^2}(1) + q_{|\cdot|}(1-\alpha)\right).$$

A usual criticism of methods based on a parametric model is that misspecification can lead to large biases.

Confidence intervals for ACF of squared returns (cont 2.)

Block Bootstrap

Method which does not require on a model specification, but relies on the choice of the block size b (a difficult task). We proceed as follows:

1. Having observed the sample $X_1^2, \ldots, X_T^2$, form the $T-1$ vectors $\mathbf{Y}_2 = [X_1^2, X_2^2]', \mathbf{Y}_3 = [X_2^2, X_3^2]', \ldots, \mathbf{Y}_n = [X_{T-1}^2, X_T^2]'$.
2. Choose a block length b and compute the number of blocks $k = [(T-1)/b] + 1$ (if $(T-1)/b$ is an integer we take $k = (T-1)/b$).
3. Choose k blocks with replacement to obtain kb vectors $\mathbf{Y}_{j_1}, \mathbf{Y}_{j_1+1}, \ldots, \mathbf{Y}_{j_1+b-1}, \ldots, \mathbf{Y}_{j_k}, \mathbf{Y}_{j_k+1}, \ldots, \mathbf{Y}_{k_1+b-1}$. This gives us the bootstrap vector process

$$\mathbf{Y}_2^* = [X_1^{*2}, X_2^{*2}]', \mathbf{Y}_3^* = [X_2^{*2}, X_3^{*2}]', \ldots, \mathbf{Y}_T^* = [X_{T-1}^{*2}, X_T^{*2}]'.$$

Confidence intervals for ACF of squared returns (cont 3.) Block Bootstrap.

4. The bootstrap sample autocovariances are computed according to standard formula with the X_t replaced by the X_t^* defined above. The empirical distribution of $\hat{\rho}^*_{T,X^2}(1)$ is then an approximation to the distribution of $\hat{\rho}_{T,X^2}(1)$.

5. The quantiles of the empirical distribution of $|\hat{\rho}^*_{T,X^2}(1) - \hat{\rho}_{T,X^2}(1)|$ can be used to construct symmetric confidence intervals.

Confidence intervals for ACF of squared returns (cont 4.)

Subsampling

Denote

$$U_t = X_t^2 - \frac{1}{T}\sum_{j=1}^{T} X_j^2$$

$$s_T^2(h) = \frac{1}{T}\sum_{j=1}^{T-h} \left(U_{j+h} - \hat{\rho}_T(h) U_j\right)^2, \quad \hat{\sigma}_T^2(h) = \frac{s_T^2(h)}{\sum_{j=h}^{T} U_j^2}$$

and consider the studentized statistic $\hat{\xi}_T = \frac{\hat{\rho}_T(h) - \rho_T(h)}{\hat{\sigma}_T(h)}$.

To construct equal-tailed and symmetric confidence intervals, we would need to know the sampling distribution of $\hat{\xi}_T$ and $|\hat{\xi}_T|$, respectively.

We use subsampling to approximate these distributions.

Confidence intervals for ACF of squared returns (cont 5.)
Subsampling

Consider an integer $b < T$ and the $T - b + 1$ blocks of data $X_t^2, \ldots, X_{t+b-1}^2,\ t = 1, \ldots, T - b + 1.$

From each of these blocks compute $\hat{\rho}_{b,t}(h)$ and $\hat{\sigma}_{b,t}(h)$, but replacing the original data $X_1, \ldots, X_T$ by $X_t, \ldots, X_{t+b-1}$.

Compute the subsampling counterpart of the studentized statistic $\hat{\xi}_{b,t}(h) = \dfrac{\hat{\rho}_{b,t}(h) - \hat{\rho}_T(h)}{\hat{\sigma}_{b,t}(h)}$ and construct the EDF

$$L_b(x) = \frac{\sum_{t=1}^{T-b+1} \mathbf{1}\left\{\hat{\xi}_{b,t}(h) \leq x\right\}}{\mathcal{N}_b^{-1}}, \quad L_{b,|\cdot|}(x) = \frac{\sum_{t=1}^{T-b+1} \mathbf{1}\left\{|\hat{\xi}_{b,t}(h)| \leq x\right\}}{\mathcal{N}_b^{-1}},$$

with $\mathcal{N}_b = T - b + 1$. The empirical quantiles of L_b and $L_{b,|\cdot|}$ allow us to construct, respectively, equal-tailed and symmetric confidence intervals. For example, denoting by $q_{b,|\cdot|}(1-\alpha)$ the $(1-\alpha)$th quantile of $L_{b,|\cdot|}$, a subsampling symmetric $1-\alpha$ level confidence interval for $\rho_T(h)$ is

$$\left(\hat{\rho}_T(h) - \hat{\sigma}_T(h) q_{b,|\cdot|}(1-\alpha), \quad \hat{\rho}_T(h) + \hat{\sigma}_T(h) q_{b,|\cdot|}(1-\alpha)\right).$$

General Framework: GARCH–type processes

$$X_t = \sigma_t Z_t, \quad E(Z_t) = 0, \quad \mathrm{Var}(Z_t) = 1,$$
$$\sigma_t^2 = g(Z_{t-1}) + c(Z_{t-1})\sigma_{t-1}^2$$

with different specifications for the conditional skedastic function:

1. GARCH$(1, 1)$ process

$$c_{t-1} = \beta + \alpha Z_{t-1}^2, \quad \sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha X_{t-1}^2.$$

2. The GJR-GARCH$(1, 1)$ model, with

$$c_{t-1} = \beta + (\alpha + \phi I(Z_{t-1}))Z_{t-1}^2, \quad \sigma_t^2 = \omega + (\alpha + \phi I(Z_{t-1}))X_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where $I(Z_{t-1}) = 1$ if $Z_{t-1} < 0$, and $I(Z_{t-1}) = 0$ otherwise.

3. The nonlinear GARCH(1,1) model (NL GARCH(1,1,2), with

$$c_{t-1} = \beta + \alpha(1 - 2\eta \ \mathrm{sign}(Z_{t-1}) + \eta^2)Z_{t-1}^2;$$

$$\sigma_t^2 = \omega + \alpha(1 - 2\eta \ \mathrm{sign}(Z_{t-1}) + \eta^2)X_{t-1}^2 + \beta\sigma_{t-1}^2.$$

General Framework: GARCH–type processes (cont. 1)

We denote $\gamma_{ci} = Ec^i(Z_t)$. The fourth unconditional moment of X_t exists if and only if $\gamma_{c2} = Ec_t^2 \in [0, 1]$.

For the three processes considered here, if we assume that $Z_t \sim N(0, 1)$, the values of γ_{c2} and $\rho_{X^2}(1)$ can be computed in a closed form.

If we know the model parameters, we can calculate precisely the population autocorrelation $\rho_{X^2}(1)$ and the value of γ_{c2}.

For each of the three models, we considered five parameter choices, which we labeled as models 1 through 5.

The lag one autocorrelations for these choices are, respectively, approximately equal to .15, .22, .31, .4, .5.

The corresponding values of γ_{c2} are respectively, approximately equal to .1, .3, .5, .7, .9.

Simulation Results

We investigate the performance of the three methods by comparing the empirical coverage properties (ECP) for the 15 data generating processes (3 models $\times$ 5 parameter choices)

To facilitate comparison, models with the same index have similar values of γ_{c2} and $\rho_{X^2}(1)$, e.g. standard GARCH and GJR-GARCH with index 3 both have $\gamma_{c2} \approx .5$ and $\rho_{X^2}(1) \approx .31$.

Consider

- Four sample sizes, $T = 100, 250, 500, 1000$.
- Confidence intervals of 95 %

Simulation Results (cont 1.) Residual Bootstrap

Table 1: ECP of *symmetric* confidence intervals constructed using *residual bootstrap*.

T	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)
STD GARCH	1	2	3	4	5
100	99.6	85.3	86.0	80.4	77.4
250	92.9	91.3	92.1	89.4	84.4
500	93.4	93.4	94.1	93.7	92.7
1000	95.1	96.8	97.6	97.6	94.4
GJR GARCH	1	2	3	4	5
100	97.7	94.8	92.0	89.5	81.5
250	96.2	96.6	97.0	96.4	92.3
500	98.3	99.2	98.9	99.1	96.5
1000	99.0	99.4	99.6	99.8	98.8

Simulation Results (cont 2.)

Table 2: ECP of *symmetric* confidence intervals constructed using *residual bootstrap*.

T	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)
NL GARCH	1	2	3	4	5
100	95.5	83.8	79.8	74.7	66.0
250	91.7	87.3	84.3	81.0	73.6
500	91.7	93.1	88.5	82.1	77.3
1000	96.4	93.3	92.9	87.0	81.0

Simulation Results (cont 3.)

- Equal tailed and symmetric confidence intervals perform equally well for standard GARCH and GJR–GARCH,
- For NL–GARCH, symmetric confidence is better than equal tailed,
- The ECP decreases as γ_{c2} approaches 1. ($\gamma_{c2} < 1$ is required for the population autocovariances to exist)
- For the NL–GARCH, results are unsatisfactory except when $\gamma_{c2} < .3$
- Bad results for the NL–GARCH model can be caused by parameter identification problems: when η is large, parameter biases are very large. (Furthermore, large η corresponds to large γ_{c2}).
- These identification problems are less severe for the GJR–GARCH.

Simulation Results (cont 4.)

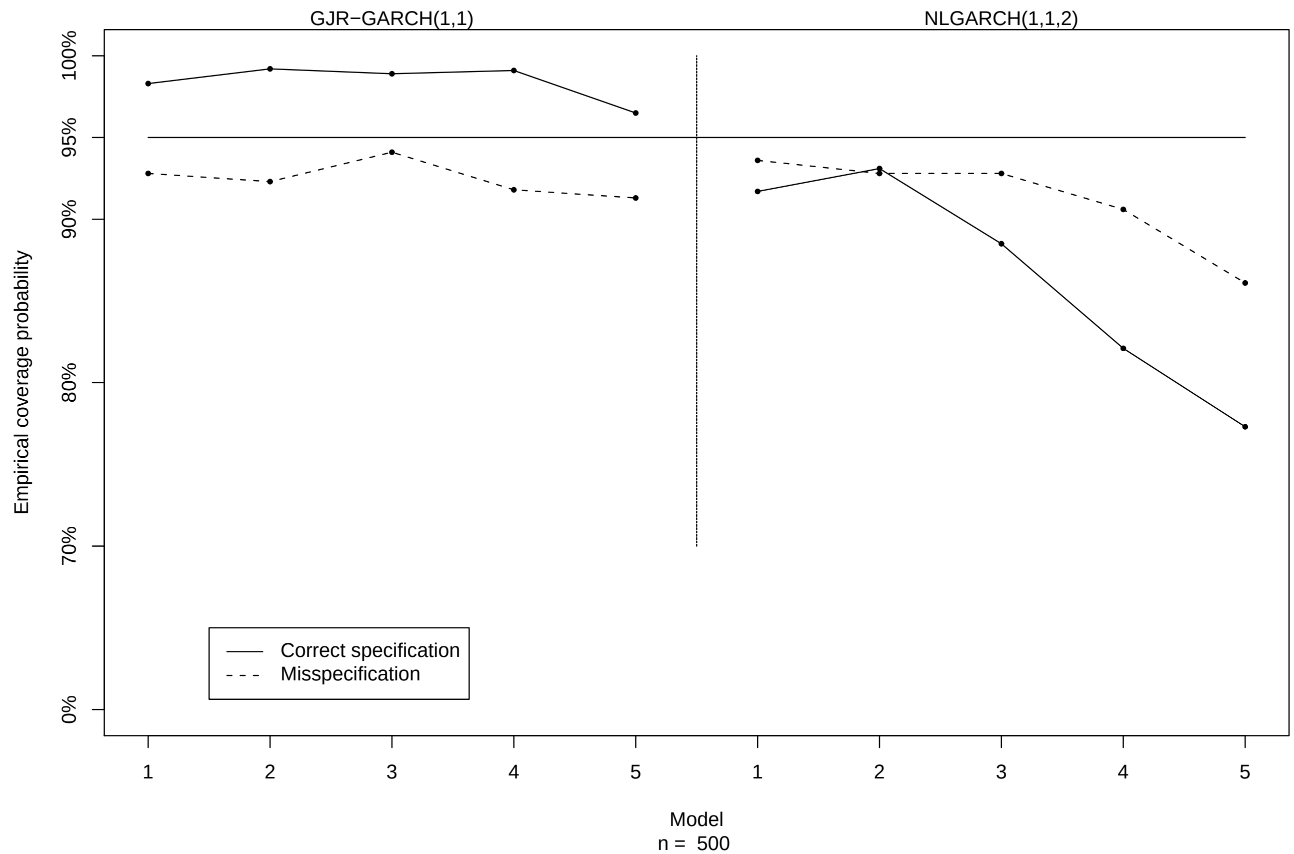

Figure 1: Comparison of ECP's for symmetric residual bootstrap confidence intervals based on standard GARCH and a correct specification. The nominal coverage of 95% is marked by the solid horizontal line. The sample size is $T = 500$.

Simulation Results (cont 5.)

- Figure 1 shows that estimating the standard GARCH model on all three DGP's might lead to improvements in ECP's, for symmetric confidence intervals and series of length 500.
- The results for other series lengths look very much the same and are therefore not presented.
- The residual bootstrap method works best if symmetric confidence intervals are used and the standard GARCH model is estimated.
- Thus, in our context, misspecifying a model improves the performance of the procedure.

Simulation Results (cont 6.)

Table 3: ECP of *symmetric* confidence intervals based on the *block bootstrap* method for the five parameter choices in the *GJR-GARCH* model.

Model		1	2	3	4	5
T	b	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)
500	3	87.0	82.0	78.4	65.5	61.4
	5	89.1	83.8	73.4	63.0	58.5
	10	87.9	81.8	71.4	60.6	51.9
	15	84.5	78.7	71.8	63.8	52.7
	30	85.6	79.0	69.6	61.3	50.0
1000	5	87.7	84.4	75.2	67.9	59.6
	10	88.6	85.1	70.8	61.0	52.6
	15	89.7	83.0	72.7	63.6	53.3
	30	87.8	80.9	72.7	59.7	51.2

Simulation Results (cont 7.)

- Empirical coverage probabilities are too low for all the choices of T and b,
- ECP are in the range $[0.80, 0.90]$ only for $\gamma_{c2} < 0.3$,
- ECP are slightly above 50% when $\gamma_{c2} = 0.9$,
- We recommend using $b = 3, 5$, although results do not depend too much on the choice of b,
- QML estimator underestimate the true value of the autocorrelation, which causes under–coverage.

Simulation Results (cont 8.)

Table 4: Empirical coverage probabilities of *symmetric* confidence interval based on the *subsampling* method for the five parameter choices in the *NLGARCH* model. Sample size $T = 500$.

Model	1	2	3	4	5
b	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)	*e.c.p.* (%)
3	97.2	95.3	91.6	82.3	70.4
6	94.1	95.5	79.9	67.9	51.5
8	90.1	83.0	75.1	63.3	50.2
10	85.4	80.9	71.4	57.5	44.5
50	80.2	76.1	63.9	54.1	41.2

Simulation Results (cont 9.) Subsampling.

- Symmetric CI have a much better ECP than equal tailed CI,
- Subsampling method is very sensitive to the choice of b,
- Choosing small b, e.g., $b = 3, 6$, we get ECP close to 95% for $\gamma_{c2} < 0.6$, and fair coverage for higher values of γ_{c2}.
- Such low value for b is surprising, as autocovariances are computed from very short sub–series,
- ECP are too low for equal tailed CI, and as γ_{c2} approaches one, ECP tends to 10%.

Simulation Results (cont 10.)

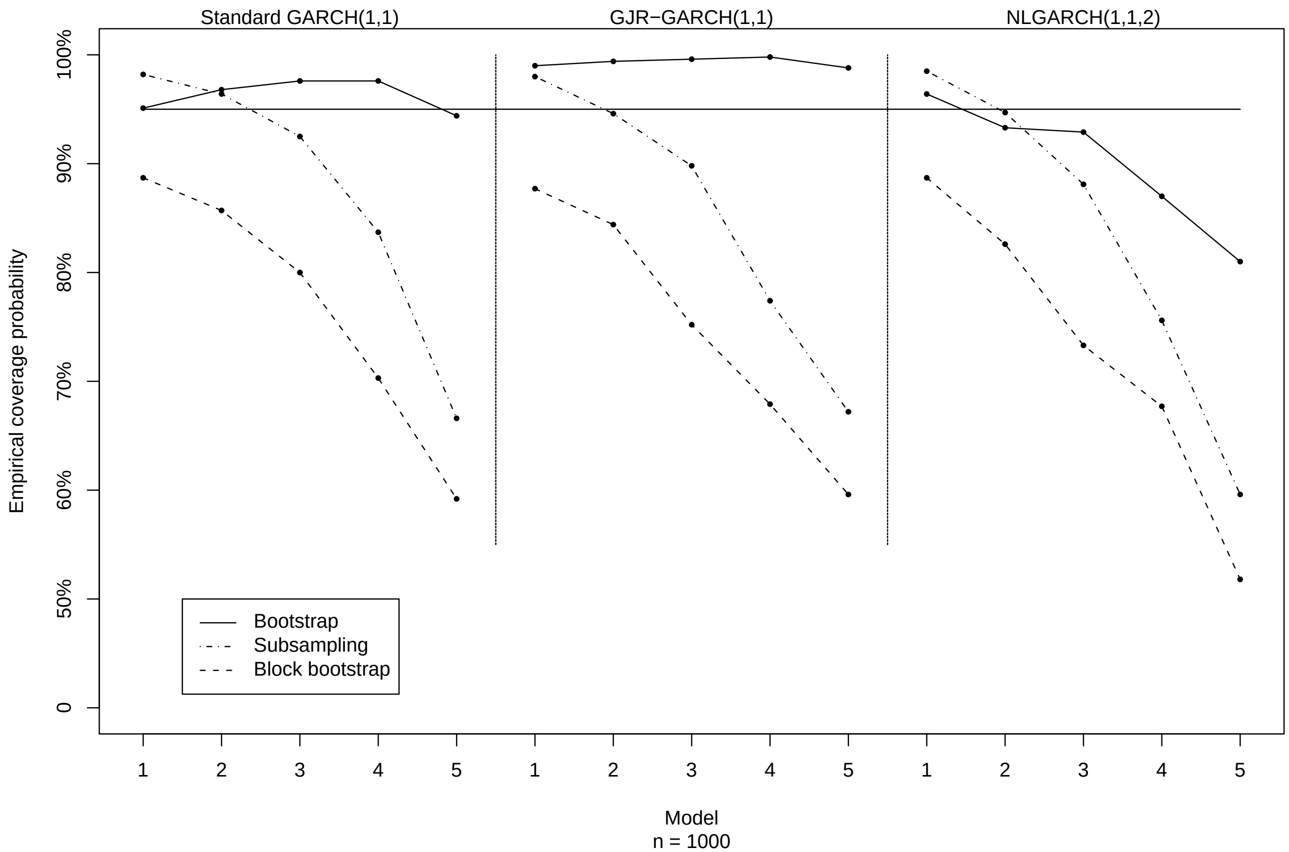

Figure 2: Comparison of ECP's for symmetric confidence intervals. The nominal coverage 95% is marked by solid horizontal line. The series length is $T = 1000$. For block bootstrap, $b = 5$, for subsampling $b = 3$.

Conclusion and practical recommendations

- The best method is residual bootstrap with the assumption that the model is a standard GARCH(1,1),
- The residual bootstrap confidence intervals based on a misspecified model can produce good coverage probabilities.