Finite Sample Performance of the MLE in GARCH(1,1): When the Parameter on the Lagged Squared Residual Is Close to Zero

Suduk Kim*

Department of Economics
Hoseo University
Asan Si, ChungNam,
Korea, 336-795
E-mail: kimsd@dogsuri.hoseo.ac.kr

May 2, 1999

*I am indebted to Hiroki Tsurumi and Choon-Geol Moon for their valuable comments and advice. I thank Robin L. Lumsdaine, Joon Y. Park, Yungsan Kim and two anonymous referees for their comments and helpful discussions. I wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997.
Abstract

We note that the maximum likelihood function misspecified as a GARCH($p,q$) process when $q$ actually is zero, is maximized at all combinations of parameter estimates which equate the sample unconditional variance to true unconditional variance, asymptotically, illustrating a trade-off in parameter estimates. This result leads to our conjecture that the closer the true coefficients for the squared lagged residuals become to zero, the more the estimated conditional variance behaves like a GARCH($p,q$) process with $q = 0$.

To confirm our conjecture and to examine how fast a trade-off disappears with an increasing size of the coefficient for the squared lagged residuals, a Monte Carlo study is done for the case of GARCH(1,1). The results show that this phenomenon continues to prevail even in a correctly specified model with sample size of 1600 and support our conjecture.

We derived a theoretical relationship between the true and empirical persistence of volatility and show that if one parameter estimate for error process is different from the true one, which can happen in finite sample, it can affect other estimates ‘systematically’. It is shown that the empirical estimates for the persistence of volatility reported in the literature might have been over-estimated while the constant term in conditional variance is under-estimated. This result also suggests that confidence interval prediction for one step ahead forecast may be incorrect in finite sample.
1 Introduction

Since its initial introduction by Robert F. Engle (1982), autoregressive conditional heteroscedasticity (ARCH)-type model has been widely applied to modeling in finance (Bollerslev et al., 1992). Depending upon the persistence of past volatility, current volatility in financial time series could be explained using ARCH-type models (Poterba and Summers, 1986). How to model transitory and permanent shocks effectively is a key aspect. Hence the investigation of the problems surrounding parameter estimates in finite sample becomes an important issue. Recent study on the finite sample properties of the maximum likelihood estimator in generalized ARCH (GARCH) is done by Lumsdaine (1995). Her findings are that commonly used statistics such as Lagrange multiplier, likelihood ratio, and Wald do not behave as well as *t* statistics in small sample, and that a pileup effect at the boundary of parameter values is apparent. In addition, estimators, particularly those of the ARCH parameters, are skewed in small samples. Her findings are obtained using the sample size of 500.

Nelson (1990) showed that when coefficients for the lagged squared residuals are zero, conditional variance follows an autoregressive (AR) process which is completely deterministic depending solely on the initial value of the process. That is, parameters in conditional variance are not separately identifiable. Therefore, we can conjecture that identification problem would approximately prevail in finite sample, if the true coefficients of the lagged squared residuals are close to zero.

The purpose of this paper is to answer following three questions: First, suppose we attempt to estimate parameters in conditional variance by maximum likelihood estimation with GARCH(*p,q*) misspecification when the true coefficients of lagged squared residuals are zero. Although we know from above discussion that the parameters in conditional variance are not separately identifiable, it is not immediately clear how the estimates for the parameters would turn out to be, especially those for the coefficients of lagged squared residuals. We try to answer explicitly how these estimates are related to the true value of parameters under this unidentification problem. Second, in addition to that, what would happen to finite sample performance of GARCH(1,1), if we attempt to estimate parameters in conditional variance with MLE when the true coefficient of the lagged squared residuals is close to zero. Third, if the result from the above finite sample performance still contains the characteristics related to the unidentification problem, how can we interpret the empirical findings of the persistence of volatility presented in the articles using GARCH model.

Section II focuses on the model and motivation of this paper. Section III presents a proposition as an answer to the first question. Section IV discusses the Monte Carlo design with data generating process used. The result of Monte Carlo study is also provided to give an answer to the second question. Section V presents a relationship among true, empirical persistence of volatility and the
estimation of constant term in true conditional variance. We give explanations on this relationship and discuss how we can interpret the empirical findings presented in articles using GARCH(1,1). Section VI concludes our investigation. We supply additional materials such as brief proofs used in this paper in Appendix.

2 The Model

We present GARCH\((p,q)\) model as follows:

\[
y_t = X'_t\Gamma + \varepsilon_t
\]

\[
\varepsilon_t = \eta_t h_t^{\frac{1}{2}}, \text{ where } \eta_t \sim N(0,1)
\]

\[
h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}
\]  \hspace{1cm} (1)

\[
\eta_t | \Psi_{t-1} \sim N(0,h_t)
\]

where \(y_t\) is the dependent variable, \(X'_t\) and \(\Gamma\) are a \(1 \times k\) vector of explanatory variables and a vector of unknown parameters, respectively. Especially, the GARCH(1,1) process is presented as

\[
h_t = w + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \hspace{1cm} (2)
\]

Following Lumsdaine(1995) with slight modification, a theoretical assumption about the parameter space, \(\Theta\), and the true parameter vector \(\theta_0 = (\Gamma_0, w_0, \alpha_{01}, \alpha_{02}, \cdots, \alpha_{0q}, \beta_{01}, \beta_{02}, \cdots, \beta_{0p})\) is made.

**Assumption 1:** The true parameter vector \(\theta_0 \in \Theta\) is in the interior of \(\Theta\), a compact subspace of a Euclidean space. Specifically, for any vector \(\theta \in \Theta\), assume that \(-\bar{\Gamma}/2 \leq \Gamma \leq \bar{\Gamma}/2, 0 \leq \alpha_i \leq (1 - \delta)\) for \(i = 1, 2, \cdots, q\), and \(0 \leq \beta_i \leq (1 - \delta)\) for \(i = 1, 2, \cdots, p\), for some constant \(\delta > 0, \bar{\Gamma} > 0, \) and \(\sum_{i=1}^{q} \alpha_{0i} + \sum_{i=1}^{p} \beta_{0i} < 1\).

Then, GARCH\((p,q)\) model presented in equations (1) at its true value is presented as

\[
y_t = X'_t \Gamma_0 + \varepsilon_{0t}
\]  \hspace{1cm} (3)

\[
h_{0t} = w_0 + \sum_{i=1}^{q} \alpha_{0i} \varepsilon_{0t-i}^2 + \sum_{i=1}^{p} \beta_{0i} h_{0t-i}
\]  \hspace{1cm} (4)

\[
\varepsilon_{0t} = \eta_{0t} h_{0t}^{\frac{1}{2}}, \text{ where } \eta_t \sim N(0,1)
\]  \hspace{1cm} (5)

As is pointed out by Lumsdaine(1991), the conditional distribution of the estimated \(\varepsilon_t\) is obtained from equations (3) and (4) as

\[
\varepsilon_t = \varepsilon_{0t} + X'_t(\Gamma_0 - \Gamma)
\]  \hspace{1cm} (6)

\[
\varepsilon_t^2 = \varepsilon_{0t}^2 + 2\varepsilon_{0t} X'_t(\Gamma_0 - \Gamma) + (\Gamma_0 - \Gamma)' X_t(\Gamma_0 - \Gamma).
\]  \hspace{1cm} (7)
The conditional distribution of $\varepsilon_t | \Psi_{t-1}$ has mean $X_t'(\Gamma_0 - \Gamma)$ and variance $h_{qt}$. The adjusted error $\varepsilon_t / h_{1/2}^t$ will, in general, have a mean $X_t'(\Gamma_0 - \Gamma)E(h_{1/2}^t) \neq 0$, and therefore, variance not equal to 1. Estimators of the parameters are obtained by maximizing the following with respect to $\theta \in \Theta$.

$$L_T(\theta) = L_T(y_1, \cdots, y_T, x_1, \cdots, x_T, h_0; \theta)$$

$$= \frac{1}{T} \sum_{t=1}^T l_t(\theta)$$

$$= -\frac{1}{2} \ln(2\pi) - \frac{1}{2T} \sum_{t=1}^T \ln(h_t) - \frac{1}{2T} \sum_{t=1}^T \varepsilon_{it}^2 h_{1/2}^t.$$  

(8)

Let’s consider the case where the conditional error follows AR($p$) with $\alpha_0i = 0$ for all $i$ in equation (3). This setup will include the random normal error structure as its special case when $\beta_0i = 0$ for all $i = 1, 2, \cdots p$. In addition, when $\beta_0i \neq 0$, the conditional variance follows an autoregressive process with drift and no error. It is completely deterministic depending solely on the initial values of the process and the estimates of the drift. The true values of parameters, $w$ and $\beta_i$’s, therefore, would not be separately identifiable. Let us define this process with $q = 0$ as ‘AR($p$) conditional variance’ for convenience.

Here, we raise three questions to consider. First, suppose we attempt to estimate those parameters in AR($p$) conditional variance with a MLE. Note that this is a MLE with GARCH($p,q$) misspecification. Although we know $w_0$ and $\beta_0i$’s are not separately identifiable, it is not immediately clear how the estimates for the parameters would turn out to be, especially $\alpha_i$’s, and how these estimates are related to the true value of $w_0$, $\alpha_0i$’s and $\beta_0i$’s under this unidentification problem. Second, in addition to that, what would happen to finite sample performance if we attempt to estimate the parameters in GARCH($p,q$) conditional variance with MLE when the true coefficients of the lagged squared residuals are close to zero. To answer this question, we examined GARCH(1,1) process only. Third, if the result from the above finite sample performance still contains the characteristics related to the unidentification problem, how can we interpret the empirical findings of the persistence of volatility presented in the articles using GARCH model.

In the following section, we investigate these questions in detail.

### 3 Identification Issue

The first question is how the estimates for the parameters would turn out to be, especially $\alpha_i$’s, when we estimate the parameters in AR($p$) conditional variance

---

2 In general, we can show $h_t = w + \sum_{i=1}^p \beta_i h_{t-i} = w(1- \sum_{i=1}^p \beta_i)^{-1} + \sum_{i=1}^p \beta_i r_i^{-1}$, where $r_i$’s are characteristic roots lying outside the unit circle to have finite conditional variance.

3 We should note that this is different from ARCH($p$) process.

4 However, the same result can be obtained for GARCH($p,q$) models in general. See Kim(1993).
by MLE misspecified as a GARCH($p,q$). To answer the question, we provide a Proposition in this section. But, first, we make following assumptions about the sequence of $\{X_t\}$, and initial value for the true conditional variance. Also, note the definition for convergence in probability and Markov’s Inequality.

**Assumption II:** The sequence $\{X_t: X_t \in \mathbb{R}^k\}$ is assumed to be uniformly bounded such that $|X_t| \leq \sup_{t \in S} |X_t| \sim Q < \infty$, where $\mathbb{R}^k$ is $k$-dimensional Euclidean Space $S=\{1, 2, \cdots\}$, where $t \in S$.

**Assumption III:** Initial value for the true conditional variance, $h_{00}$, is assumed to be finite.

**Definition:** Convergence in Probability to Zero. We say that $\{X_t\}$ converges in probability to zero, written $X_t = o_p(1)$ or $X_t \xrightarrow{P} 0$, if for every $\varepsilon > 0$,

$$P(|X_t| > \varepsilon) \xrightarrow{P} 0 \text{ as } t \to \infty.$$ 

**Markov’s Inequality:** If $E|\Phi|^r < \infty$, $r \geq 0$ and $\varepsilon > 0$, then

$$P(|\Phi| \geq \varepsilon) \leq \frac{E|\Phi|^r}{\varepsilon^r}.$$ 

**Proposition:** For a true process of AR($p$) conditional variance with $\alpha_{0i} = 0$ for all $i$’s in equation (4), $\lim_{T \to \infty} L_T(\theta)$ of log-likelihood function misspecified as GARCH($p,q$) process exists for all $\theta \in \Theta$ and is maximized at $\alpha_i = 0$ for all $i$ and at all combinations of $(w, \beta_i)$ such that $w/(1 - \sum_{i=1}^p \beta_i) = w_0/(1 - \sum_{i=1}^p \beta_{0i})$.

The Proposition is proved in Appendix in brief with related Lemmas. According to Proposition, the misspecified MLE is maximized at $\alpha_i = 0$ for all $i$ and $w_0$ and $\beta_{0i}$’s will not be identified separately. The Proposition leads us to the next question: unidentification problem would prevail even when the true coefficients of the lagged squared residuals are close to zero.

4 A Monte Carlo Study and Simulation Results

To proceed with the second question, we make a conjecture as following:

As the true ARCH-type coefficients ($\alpha_{0i}$’s in our model) approach the boundary of zero, the conditional variance would behave like an AR($p$) case in the finite sample, regardless of their asymptotic properties. This means that in finite sample, the conditions for log-likelihood maximum provided in the Proposition would approximately prevail, and we will see a trade-off among the estimates, especially, between the estimates in numerator and those in denominator.
To confirm our conjecture and to examine how fast the trade-off disappears as the size of the coefficient for the squared lagged residuals changes, a Monte Carlo study is done for the case of GARCH(1,1). Since our study here focuses on GARCH(1,1) process, let us denote $\alpha_1$, $\beta_1$ and $\alpha_0$, $\beta_0$, respectively. The data are generated according to equations from (3) to (5) with $w_0 = 1$, $\alpha_0 \in \{0.45, 0.3, 0.1, 0.05, 0.01\}$, $\beta_0 = 0.5$. We changed sample size $T$ as $T \in \{100, 200, 400, 800, 1600\}$ to observe the effects of varying $\alpha_0$’s on other estimates and of change in sample. We use initial value $h_{00} = 1$, $X_t$ as $T \times 1$ uniform random variable and $\Gamma_0 = 0$. The first 100 observations are discarded to minimize the effect of the initial value. Each case has 500 replications using different standard normal random variates. GAUSS maxlik procedure was used for computation with the BHHH step. We use logit type transformation so that sum of coefficients in conditional variance should lie between 0 and 1.

We summarize simulation results in Figures 1 and Tables 1. The simulation results indicate a persistent trade-off between $w$ and $\alpha + \beta$ estimates. We can easily see that this trade-off causes the pileup effect observed by Lumsdaine(1995). Tables of parameter estimates normalized by standard deviation are not provided here, but they show similar results to those by Lumsdaine.

5 An Interpretation of the Empirical Persistence of Volatility

Third, if a systematic trade-off among parameter estimates is probable, we must examine the following in detail. Note that, in many papers, $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i$ is used to measure the empirical persistence of volatility. If we can show the systematic trade-off among the parameter estimates, we also can show that there is a certain relationship among the true, the estimated persistence of volatility and the constant term in conditional variance. Note Lemma 1 from Appendix:

**Lemma 1** For $h_t = \varepsilon^2_{0t}$ to hold with probability one, following holds:

$$w_0 \left(1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{i=1}^{p} \beta_{0i}\right)^{-1} = w \left(1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i\right)^{-1} + (\Gamma_0 - \Gamma)'X_t X'_t (\Gamma_0 - \Gamma) \sum_{i=1}^{q} \alpha_i \left(1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i\right)^{-1}.$$

Let us define the true unconditional variance as

$$C \equiv w \left(1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{i=1}^{p} \beta_{0i}\right)^{-1}.$$

5 Normalized tables would be available upon request.
Then, Lemma 1 can be rearranged for the following:

\[
\left( \sum_{i=1}^{q} \alpha_{0i} + \sum_{i=1}^{p} \beta_{0i} \right) - \left( \sum_{i=1}^{q} \alpha_{i} + \sum_{i=1}^{p} \beta_{i} \right)
= \frac{w - w_0}{C} + \frac{(\Gamma_0 - \Gamma)'X_t'X_t(\Gamma_0 - \Gamma)}{C} \sum_{i=1}^{q} \alpha_i.
\]

In a finite sample, true parameter \( \theta_0 \) and its estimates are not automatically equal. From the above equation, therefore, there are two categories we may consider separately.

1. **When \( \Gamma \) is correctly estimated \( (\Gamma_0 = \Gamma) \):**
   
   1.1 If \( w_0 \) is under-estimated, the persistence of volatility will be over-estimated.

   1.2 If \( w_0 \) is over-estimated, the persistence of volatility will be under-estimated.

2. **When \( \Gamma \) is not correctly estimated \( (\Gamma_0 \neq \Gamma) \):**

   2.1 If \( w_0 \) is under-estimated, it is not clear whether persistence of volatility is under- or over-estimated. It depends on the sign of the right hand side in the equation above.

   2.2 If \( w_0 \) is correctly or over-estimated, the persistence of volatility will be under-estimated.

Note that in many empirical studies which employ GARCH(1,1) model, the estimates for \( w_0 \) are very small and those for \( \beta_0 \)'s are usually very large. For example, in Bollerslev (1986), French (1987), Schwert (1990), Day (1992), the estimates for \( w_0 \) are ranging from \( 6.3 \times 10^{-7} \) to 0.007. The empirical persistence of volatility are all larger than 0.9. The sample sizes for the case of Bollerslev, Schwert, and Day are 143, 720, and 319, respectively. Given \( \Gamma_0 = \Gamma \) and \( h_t = \varepsilon_{0t}^2 \) with probability one, let’s assume declining conditional variance for a stable period such as \( h_t < h_{t-1} \). For GARCH(1,1) case, we can easily show

\[
w \leq h_t \leq w/(1 - \alpha - \beta).
\]

Assuming increasing conditional variance for a volatile period such as \( h_t > h_{t-1} \), on the other hand, we can show

\[
h_t > w/(1 - \alpha - \beta).
\]

This result implies that for the period of declining conditional variance, asymptotic confidence interval for the one-step ahead forecast would be narrower than that of the OLS case.

As was illustrated in Bollerslev (1986), with the estimate of 0.007 for \( w_0 \), we may get a very narrow asymptotic confidence interval for one-step ahead forecast.
of $y_t$ for a stable and predictable period, since the conditional heteroskedasticity is bounded to the small value of the estimate, $w$ during this period. If there is a trade-off between the estimates $w$ and $\beta$’s as we conjecture in a finite sample, and if $w_0$ is under-estimated, the empirical result may mistakenly lead to a higher estimate of volatility persistence, as was seen above. And this leads to a narrow interval for one-step ahead asymptotic confidence interval forecast of $y_t$ in a stable period of declining conditional variance.

6 Conclusions

It is shown that the empirical persistence of volatility shown in the literature might have been over-estimated with the under-estimation of the constant term in the conditional variance of GARCH(1,1) process. It, in turn, may lead to a narrower confidence interval prediction for one step ahead forecast. Therefore, it is important to recognize finite sample properties of the maximum likelihood estimators in GARCH(1,1) process. An important implication of our simulation result is that the choice of the true value of parameters cannot be easily identified even in a correctly specified model with sample size as large as 1600, when the coefficients of $\varepsilon^2_{0t-1}$ are very close to zero. And there is a trade-off among the estimates ‘systematically’. GARCH(1,1) maximum likelihood estimation in a finite sample with $\alpha_0$ close to zero is seen to have a difficulty in producing unbiased estimates for the parameters appearing in conditional variance.

An explanation for the failure of normality test for the adjusted error may be possible using the results obtained above.\footnote{Even if $\Gamma = \Gamma_0$ in equations (6) and (10), the failure of normality test is still possible from our result.} The issue here seems to be a trade-off between $w$ and $\beta$ or $w$ and $\alpha + \beta$, rather than the satisfaction of the higher moment conditions. Note that our cases of $\alpha_0 \in \{0.01, 0.05, 0.1\}$ with $\beta_0 = 0.5$ satisfy the moment conditions up to 12-th. But the trade-off in these cases are much severer. The case of $\alpha_0 = 0.45$ with $\beta_0 = 0.5$, on the other hand, satisfies only 2nd moment condition while its finite sample performance in large sample is the best.
Figure 1: Graphical Presentation of a Trade-off Between $\alpha + \beta$ and $w$ with $T \in \{100, 200, 400, 800, 1600\}$ and $\alpha_0 \in \{0.45, 0.3, 0.1, 0.05, 0.01\}$

**Vertical-Axis:** $\alpha + \beta$, **Horizontal-Axis:** $w$

From the top to the bottom rows, $\alpha$ values are assigned as $\{0.45, 0.3, 0.1, 0.05, 0.01\}$. From the left to the right rows, $T$ values are assigned as $\{100, 200, 400, 800, 1600\}$. 

8
Table 1:
Sample Standard Deviation and Correlation Coefficient
Summary Statistics for GARCH(1,1)

<table>
<thead>
<tr>
<th>α₀</th>
<th>sample size</th>
<th>s_γ</th>
<th>s_ω</th>
<th>s_α</th>
<th>s_β</th>
<th>s_α+β</th>
<th>r_ω,α</th>
<th>r_ω,β</th>
<th>r_α,β</th>
<th>r_ω,α+β</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.45</td>
<td>T=100</td>
<td>0.413</td>
<td>1.161</td>
<td>0.179</td>
<td>0.208</td>
<td>0.154</td>
<td>0.292</td>
<td>-0.729</td>
<td>-0.692</td>
<td>-0.642</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.278</td>
<td>0.655</td>
<td>0.124</td>
<td>0.129</td>
<td>0.100</td>
<td>0.293</td>
<td>-0.757</td>
<td>-0.688</td>
<td>-0.610</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.194</td>
<td>0.394</td>
<td>0.088</td>
<td>0.086</td>
<td>0.067</td>
<td>0.291</td>
<td>-0.790</td>
<td>-0.710</td>
<td>-0.637</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.134</td>
<td>0.220</td>
<td>0.061</td>
<td>0.052</td>
<td>0.041</td>
<td>0.240</td>
<td>-0.700</td>
<td>-0.747</td>
<td>-0.524</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>0.095</td>
<td>0.148</td>
<td>0.044</td>
<td>0.035</td>
<td>0.029</td>
<td>0.222</td>
<td>-0.682</td>
<td>-0.751</td>
<td>-0.477</td>
</tr>
<tr>
<td>0.3</td>
<td>T=100</td>
<td>0.330</td>
<td>0.890</td>
<td>0.167</td>
<td>0.270</td>
<td>0.214</td>
<td>0.299</td>
<td>-0.862</td>
<td>-0.608</td>
<td>-0.852</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.229</td>
<td>0.667</td>
<td>0.120</td>
<td>0.197</td>
<td>0.161</td>
<td>0.278</td>
<td>-0.877</td>
<td>-0.576</td>
<td>-0.865</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.162</td>
<td>0.440</td>
<td>0.083</td>
<td>0.137</td>
<td>0.110</td>
<td>0.298</td>
<td>-0.904</td>
<td>-0.591</td>
<td>-0.896</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.112</td>
<td>0.276</td>
<td>0.056</td>
<td>0.086</td>
<td>0.067</td>
<td>0.308</td>
<td>-0.886</td>
<td>-0.631</td>
<td>-0.884</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>0.080</td>
<td>0.172</td>
<td>0.039</td>
<td>0.054</td>
<td>0.042</td>
<td>0.279</td>
<td>-0.863</td>
<td>-0.641</td>
<td>-0.856</td>
</tr>
<tr>
<td>0.1</td>
<td>T=100</td>
<td>0.273</td>
<td>0.772</td>
<td>0.116</td>
<td>0.364</td>
<td>0.301</td>
<td>0.467</td>
<td>-0.944</td>
<td>-0.652</td>
<td>-0.960</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.189</td>
<td>0.708</td>
<td>0.085</td>
<td>0.323</td>
<td>0.279</td>
<td>0.450</td>
<td>-0.960</td>
<td>-0.610</td>
<td>-0.972</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.136</td>
<td>0.635</td>
<td>0.062</td>
<td>0.284</td>
<td>0.254</td>
<td>0.434</td>
<td>-0.975</td>
<td>-0.571</td>
<td>-0.985</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.095</td>
<td>0.545</td>
<td>0.043</td>
<td>0.238</td>
<td>0.218</td>
<td>0.420</td>
<td>-0.983</td>
<td>-0.540</td>
<td>-0.991</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>0.066</td>
<td>0.433</td>
<td>0.031</td>
<td>0.186</td>
<td>0.175</td>
<td>0.341</td>
<td>-0.987</td>
<td>-0.451</td>
<td>-0.993</td>
</tr>
<tr>
<td>0.05</td>
<td>T=100</td>
<td>0.260</td>
<td>0.747</td>
<td>0.096</td>
<td>0.380</td>
<td>0.329</td>
<td>0.449</td>
<td>-0.955</td>
<td>-0.616</td>
<td>-0.970</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.181</td>
<td>0.704</td>
<td>0.070</td>
<td>0.354</td>
<td>0.312</td>
<td>0.531</td>
<td>-0.972</td>
<td>-0.653</td>
<td>-0.981</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.131</td>
<td>0.649</td>
<td>0.052</td>
<td>0.319</td>
<td>0.289</td>
<td>0.544</td>
<td>-0.984</td>
<td>-0.642</td>
<td>-0.991</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.091</td>
<td>0.607</td>
<td>0.036</td>
<td>0.293</td>
<td>0.272</td>
<td>0.545</td>
<td>-0.990</td>
<td>-0.623</td>
<td>-0.995</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>0.063</td>
<td>0.545</td>
<td>0.027</td>
<td>0.257</td>
<td>0.245</td>
<td>0.430</td>
<td>-0.993</td>
<td>-0.503</td>
<td>-0.996</td>
</tr>
<tr>
<td>0.01</td>
<td>T=100</td>
<td>0.250</td>
<td>0.725</td>
<td>0.074</td>
<td>0.391</td>
<td>0.352</td>
<td>0.470</td>
<td>-0.988</td>
<td>-0.599</td>
<td>-0.977</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.174</td>
<td>0.707</td>
<td>0.055</td>
<td>0.373</td>
<td>0.342</td>
<td>0.501</td>
<td>-0.977</td>
<td>-0.611</td>
<td>-0.985</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.126</td>
<td>0.669</td>
<td>0.039</td>
<td>0.348</td>
<td>0.326</td>
<td>0.539</td>
<td>-0.990</td>
<td>-0.614</td>
<td>-0.994</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.087</td>
<td>0.670</td>
<td>0.025</td>
<td>0.341</td>
<td>0.327</td>
<td>0.533</td>
<td>-0.995</td>
<td>-0.587</td>
<td>-0.997</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>0.061</td>
<td>0.624</td>
<td>0.019</td>
<td>0.315</td>
<td>0.305</td>
<td>0.498</td>
<td>-0.997</td>
<td>-0.540</td>
<td>-0.998</td>
</tr>
</tbody>
</table>

The values in the parenthesis represent Bootstrap standard deviation.
References


Lemma 1 For \( h_t = \varepsilon_{0t}^2 \) to hold with probability one,
\[
0 \left( 1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{i=1}^{p} \beta_{0i} \right)^{-1} = \frac{1}{w_0} \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1} \\
+ (\Gamma_0 - \Gamma)'X_tX_t' (\Gamma_0 - \Gamma) \sum_{i=1}^{q} \alpha_i \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1}.
\]

Proof of Lemma 1: Applying Markov’s Inequality,
\[
P(\|h_t - \varepsilon_{0t}^2\| \geq \varepsilon) \leq \frac{1}{\varepsilon} E|h_t - \varepsilon_{0t}^2| = 0. 
\]
Therefore, for \( h_t = \varepsilon_{0t}^2 \) to hold with probability one,
\[
E(h_t) = E(\varepsilon_{0t}^2).
\]
Since \( E(\varepsilon_{0t}^2 | \Psi_{t-1}) = h_{0t} \), its unconditional expectation is
\[
E(\varepsilon_{0t}^2) = \lim_{t \to \infty} E(h_{0t} | \Psi_{t-1}) = E(h_{0t}) \\
= w_0 \left( 1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{i=1}^{p} \beta_{0i} \right)^{-1}.
\]
For \( E(h_t) \),
\[
E(h_t) = w + \sum_{i=1}^{q} \alpha_i E(\varepsilon_{t-1}^2) + \sum_{i=1}^{p} \beta_i E(h_{t-1})
\]
\[
(1 - B(L))E(h_t) = w + \sum_{i=1}^{q} \alpha_i \{ E(\varepsilon_{0t-1}^2) + (\Gamma_0 - \Gamma)'X_tX_t' (\Gamma_0 - \Gamma) \}
\]
\[
\quad = w + \sum_{i=1}^{q} \alpha_i \{ E(h_{0t}) + (\Gamma_0 - \Gamma)'X_tX_t' (\Gamma_0 - \Gamma) \}
\]
\[
E(h_t) = w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} + \sum_{i=1}^{q} \alpha_i E(h_{0t}) \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}
\]
\[
\quad + \sum_{i=1}^{q} \alpha_i (\Gamma_0 - \Gamma)'X_tX_t' (\Gamma_0 - \Gamma) \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}.
\]
From the last equation above, the condition for $E(h_t) = E(\varepsilon^2_{0t})$ to hold is,

$$E(h_{0t}) = w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} + \sum_{i=1}^{q} \alpha_i E(h_{0t}) \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}$$

$$+ \sum_{i=1}^{q} \alpha_i (\Gamma_0 - \Gamma) X_t' X'_t (\Gamma_0 - \Gamma) \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}.$$

After some arrangement,

$$w_0 \left( 1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{i=1}^{p} \beta_{0i} \right)^{-1} = w \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1}$$

$$+ (\Gamma_0 - \Gamma) X_t' X'_t (\Gamma_0 - \Gamma) \sum_{i=1}^{q} \alpha_i \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1}.$$

**Lemma 2** For $h_t$ defined in Equation (9),

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \ln(h_t) \geq \ln w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}.$$

**Proof of Lemma 2:**

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

$$\geq w + \sum_{i=1}^{p} \beta_i h_{t-i}$$

$$= w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} + \sum_{i=1}^{p} A_i r_i^{t-i} \tag{9}$$

where the first inequality follows from $\alpha_i \geq 0$. Equality holds only when $\alpha_i = 0$ for all $i \in \{1, 2, \ldots, q\}$. Second equality follows from the common solution for difference equation. Characteristic roots $r_i$'s lie outside the unit circle by assumption.

$$\ln(h_t) \geq \ln \left( \frac{w}{1 - \sum_{i=1}^{p} \beta_i} + \sum_{i=1}^{p} A_i r_i^{t-i} \right)$$

$$= \ln w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}.$$
\[ + \ln \left( 1 + \left( 1 - \sum_{i=1}^{p} \beta_i \right) w^{-1} \sum_{i=1}^{p} A_i r_{t-i}^i \right) \]  

(10)

\[ \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \ln(h_t) \geq \ln w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} \]

\[ + \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \ln \left( 1 + \left( 1 - \sum_{i=1}^{p} \beta_i \right) w^{-1} \sum_{i=1}^{p} A_i r_{t-i}^i \right) \]

After some arrangement, we can show that the second term vanishes such as

\[ \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \ln \left( 1 + \left( 1 - \sum_{i=1}^{p} \beta_i \right) w^{-1} \sum_{i=1}^{p} A_i r_{t-i}^i \right) = 0. \]

**Proposition:** For a true process of AR\((p)\) conditional variance with \(\alpha_{0i} = 0\) for all \(i\)'s in equation (8), \(\lim_{T \to \infty} L_T(\theta)\) of log-likelihood function misspecified as GARCH\((p,q)\) process exists for all \(\theta \in \Theta\) and is maximized at \(\alpha_{i} = 0\) for all \(i\) and at all combinations of \((w, \beta_i)\) such that \(w/(1 - \sum_{i=1}^{p} \beta_i) = w_0/(1 - \sum_{i=1}^{p} \beta_0i)\).

**Proof of Proposition:** It is noted that the only assumptions used here are the boundedness of the sequence of exogenous variable \(X_t\). We do not apply ergodic theorem to calculate \(\lim_{T \to \infty} L_T(\theta)\). From equation (8),

From Equation (8),

\[ \lim_{T \to \infty} L_T(\theta) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \ln(h_t) - \frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \varepsilon_t^2 h_t^{-1} \]

\[ = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \ln(h_t) - \frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \varepsilon_0^2 h_t^{-1} \]

\[ - \frac{1}{2} \lim_{T \to \infty} (\Gamma_0 - \Gamma)' \frac{1}{T} T \sum_{t=1}^{T} \frac{X_t X_t'}{h_t} (\Gamma_0 - \Gamma) \]

\[ - \lim_{T \to \infty} (\Gamma_0 - \Gamma)' \frac{1}{T} T \sum_{t=1}^{T} \frac{\varepsilon_0 X_t}{h_t}. \]  

(12)

The second equality above follows from Equation (9). Define \(F\) such that,

\[ F = -\frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \ln(h_t) - \frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} \varepsilon_0^2 h_t^{-1} \]

\[ = \frac{1}{2} \lim_{T \to \infty} \frac{1}{T} T \sum_{t=1}^{T} (\ln h_t^{-1} - \varepsilon_0^2 h_t^{-1}) \]
\begin{align*}
\leq & -\frac{1}{2} \left( 1 + \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \ln h_t \right) \\
\leq & -\frac{1}{2} \frac{1}{2} \ln w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} 
\end{align*}

(13)

Since $\ln x_t - cx_t$ has a maximum at $1/c$, $F$ will have a maximum when $h_t = \varepsilon_0^2$ with probability one and the first inequality above follows. The second inequality follows from Lemma 2. From Assumption II and Definition, we can show that

\begin{align*}
\lim_{T \to \infty} (\Gamma_0 - \Gamma)' \frac{1}{T} \sum_{t=1}^{T} \frac{X_t X_t'}{h_t} (\Gamma_0 - \Gamma) & = (\Gamma_0 - \Gamma)' M_{\theta} (\Gamma_0 - \Gamma) \\
\lim_{T \to \infty} (\Gamma_0 - \Gamma)' \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{0t} X_t & \xrightarrow{P} 0
\end{align*}

(14)

(15)

where $M_{\theta} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{X_t X_t'}{h_t}$. Therefore, from equations (13)–(15),

\begin{align*}
L(\theta) & = \lim_{T \to \infty} L_T(\theta) \\
& \leq -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \left\{ 1 + \ln w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} \right\} \\
& \qquad -\frac{1}{2} (\Gamma_0 - \Gamma)' M_{\theta} (\Gamma_0 - \Gamma)
\end{align*}

(17)

where the equality follows when all $\alpha_i = 0$ from Lemma 2. This proves the first half of Proposition.

For the proof of the second half of Proposition, we proceed by contradiction. Suppose $L(\theta)$ has a unique maximum at $\theta_0$, from Equation (17),

\begin{align*}
L(\theta) & = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \left\{ \ln w_0 \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} + 1 \right\}.
\end{align*}

Then,

\begin{align*}
L(\theta) - L(\theta_0) \\
\leq & -\frac{1}{2} \left\{ \ln \frac{w}{w_0} \left( 1 - \sum_{i=1}^{p} \beta_i \right) \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} \right\}
\end{align*}

16
\[- \frac{1}{2} (\Gamma_0 - \Gamma)' M \theta (\Gamma_0 - \Gamma) \leq \frac{1}{2} \left\{ \ln \frac{w}{w_0} \left( 1 - \sum_{i=1}^{p} \beta_{0i} \right) \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} \right\} \leq 0 \]  

(18)

The second inequality follows because \((\Gamma_0 - \Gamma)' M \theta (\Gamma_0 - \Gamma) \geq 0\), the equality holds only when \(\Gamma_0 = \Gamma\) (since \(M \theta\) a symmetric positive definite). The last inequality follows from our assumption of a unique maximum at \(\theta_0\). Then the condition for \(L(\theta_0) \geq L(\theta)\) is

\[ w_0 \left( 1 - \sum_{i=1}^{p} \beta_{0i} \right)^{-1} \leq w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} \]  

(20)

But from Lemma 1, for \(h_t = \epsilon_{0t}^2\) to hold with probability one,

\[ w_0 \left( 1 - \sum_{i=1}^{q} \alpha_{0i} - \sum_{i=1}^{p} \beta_{0i} \right)^{-1} = w \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1} + (\Gamma_0 - \Gamma)' X_t X_t' (\Gamma_0 - \Gamma) \sum_{i=1}^{q} \alpha_i \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1} \]  

(21)

By the assumption of the AR\((p)\) conditional variance, \(\alpha_{0i} = 0\) for all \(i\). Then, Equation (21) becomes

\[ w_0 \left( 1 - \sum_{i=1}^{p} \beta_{0i} \right)^{-1} = w \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1} \]

\[ + (\Gamma_0 - \Gamma)' X_t X_t' (\Gamma_0 - \Gamma) \sum_{i=1}^{q} \alpha_i \left( 1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i \right)^{-1} \]  

\[ \geq w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1} \]  

(22)

In the above inequality, the equality holds if \(\sum_{i=1}^{q} \alpha_i = 0\). This is equivalent to the condition of \(\alpha_i = 0\) for all \(i\), since \(\alpha_i\)'s are restricted to be non-negative.

From Equations (20)–(22), the only condition for maximum is

\[ w_0 \left( 1 - \sum_{i=1}^{p} \beta_{0i} \right)^{-1} \equiv C = w \left( 1 - \sum_{i=1}^{p} \beta_i \right)^{-1}, \]

and this contradicts our assumption that \(L(\theta)\) has a unique maximum at \(\theta_0\).

This proves the second half of the Proposition that \(L(\theta)\) is maximized at all combinations \((w, \beta_i)\) such that \(w/(1 - \sum_{i=1}^{p} \beta_{0i}) = w_0/(1 - \sum_{i=1}^{p} \beta_{0i})\) and \(\alpha_i = 0\) for all \(i\). Note that \(w_0/(1 - \sum_{i=1}^{p} \beta_{0i})\) is a constant.