How to estimate GARCH model? Maxmimum Likelihood Estimation

(Updated Spring 2021)

CHAOYI CHEN
Institute of MNB, Corvinus University of Budapest

Empirical Financial Econometrics

Outlines

- How to estimate GARCH model? Maxmimum Likelihood Estimation
  Jump [Online Lecture + Self-study]
- Example of estimating GARCH model by Matlab (using MFE toolbox)

Maxmimum Likelihood Estimation of GARCH Models

- Review
- How your software estimate the model? Application to GARCH process
- Model selection criterion: AIC
- Plug in method

Review

- $f(y_t;\theta)$: marginal probability density function (PDF) of y_t.
- $f(y_1, y_2, ..., y_T;\theta)$: joint PDF of $y_1, y_2, ..., y_T$.
- $L(\theta) = f(y_1, y_2, ..., y_T;\theta)$: likelihood function is joint PDF of data expressed as a function of parameters.
- $\widehat{\theta}_{MLE} = \underset{\theta}{argmax} L(\theta)$. The maximum likelihood estimator (MLE) is the choice of θ that maximizes the likelihood.
- If $y_1, y_2, ..., y_T$ is independent, then:

$$L(\theta) = f(y_1, y_2, ..., y_T;\theta) = f(y_1;\theta)f(y_2;\theta)...f(y_T;\theta) = \Pi_{t=1}^{T} f(y_t;\theta).$$

  - Independence: joint PDF = product of marginal PDF.
  - Analogy: $P(A \cap B) = P(A)P(B)$ if A & B are independent.

Review Cont.

- Likelihood: assume $y_t \sim i.i.dN(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$. Hence,
  $f(y_t, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{-(y_t-\mu)^2}{2\sigma^2})$ and
  $L(\mu, sigma^2) = \Pi_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{-(y_t-\mu)^2}{2\sigma^2}) = (2\pi\sigma^2)^{-T/2} \Pi_{t=1}^{T} exp(\frac{-(y_t-\mu)^2}{2\sigma^2}))$.
- Log-likelihood: because $ln(.)$ is monotonic, we get same estimator by selecting $\widehat{\theta}$ to maximize

$$\begin{aligned} \mathcal{L}(\mu, \sigma^2) &= ln\{L(\mu, \sigma^2)\} \qquad (1) \\ &= -\frac{T}{2} ln(2\pi) - \frac{T}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - \mu)^2 \end{aligned}$$

- First order condition:

$$\frac{\partial \mathcal{L}(\mu, \sigma^2)}{\partial \mu} = -\frac{1}{2\widehat{\sigma}^2} \sum_{t=1}^{T} 2(y_t - \widehat{\mu})(-1) = 0$$

$$\Longrightarrow \sum_{t=1}^{T} (y_t - \mu) = 0 \Longrightarrow \widehat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

$$\frac{\partial \mathcal{L}(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{T}{2\widehat{\sigma^2}} - \frac{(-1)}{2\widehat{\sigma^4}} \sum_{t=1}^{T} (y_t - \widehat{\mu})^2 = 0$$

$$\Longrightarrow \frac{T}{2} = \frac{1}{2\widehat{\sigma}^2} \sum_{t=1}^{T} (y_t - \widehat{\mu})^2 \Longrightarrow \widehat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \widehat{\mu})^2.$$

Review : What if $y_1, y_2, ...y_T$ not i.i.d

- Example:
  1. $y_t \sim ARMA$
  2. $y_t \sim GARCH$
- i.i.d case: Analogy: $P(A \cap B) = P(A)P(B)$
  Factor PDF: $f(y_1, ..., y_T; \theta) = \Pi_{t=1}^{T} f(y_t; \theta)$.
- Non - i.i.d case: Analogy: $P(A \cap B) = P(A|B)P(B)$
  Factor PDF:

$$\underbrace{f(y_1, y_2; \theta)}_{\text{joint PDF}} = \underbrace{f(y_2|y_1; \theta)}_{\text{conditional PDF}} \times \underbrace{f(y_1; \theta)}_{\text{marginal PDF}} \tag{2}$$

Now add y_3:

$$\underbrace{f(y_1, y_2, y_3; \theta)}_{\text{joint PDF}} = \underbrace{f(y_3|y_1, y_2; \theta)}_{\text{conditional PDF}} \times \underbrace{\underbrace{f(y_1, y_2; \theta)}_{\text{marginal PDF}}}_{\text{Now substitute here using(2)}}$$

$$\Longrightarrow f(y_1, y_2, y_3; \theta) = f(y_3|y_1, y_2; \theta) f(y_2|y_1; \theta) f(y_1; \theta)$$

Review : What if $y_1, y_2, ...y_T$ not i.i.d Cont.

- Now note that the pattern & extrapolate to:

$$L(\theta) = f(y_1, y_2, ..., y_T; \theta) = f(y_T|y_1, ..., y_{T-1}; \theta) f(y_{T-1}|y_1, ..., y_{T-2}; \theta) \tag{3}$$

$$\Longrightarrow \boxed{L(\theta) = \Pi_{t=2}^{T} f(y_t|y_1, ..., y_{t-1}; \theta) f(y_1; \theta)}.$$

- Then, the log-likelihood becomes:

$$\mathcal{L}(\theta) = lnL(\theta) = \sum_{t=2}^{T} ln\Big(f(y_t|y_1, ..., y_{t-1}; \theta)\Big) + ln\Big(f(y_1; \theta)\Big)$$

- Define $\mathcal{I}_t$ as the information set containing the information in all random variables realized at or before time t.
- Then, a simple & more general version of the formulas are

$$L(\theta) = \Pi_{t=2}^{T} f(y_t|\mathcal{I}_{t-1}) f(y_1)$$
$$\mathcal{L}(\theta) = \sum_{t=2}^{T} lnf(y_t|\mathcal{I}_{t-1}) + lnf(y_1).$$

Application to GARCH process

- Define

$$r_t = \mu + \varepsilon_t, \;\; \varepsilon_t = v_t \sqrt{h_{t|t-1}}$$

$v_t \sim i.i.dN(0,1)$, $\mathcal{I}_t$ is the information in all variables realized at or be

$h_{t|t-1}$ follows GARCH specification

- Let $\theta =$ (GARCH parameters, μ). Then, we have

$$r_t = \mu + \varepsilon_t = \mu + \underbrace{v_t}_{N(0,1)} \underbrace{\sqrt{h_{t|t-1}}}_{\text{realized at } t-1}$$

$$\Longrightarrow r_t|\mathcal{I}_{t-1} \sim \mu + N(0,1)\sqrt{h_{t|t-1}} = N(\mu, h_{t|t-1}) \text{ (conditional PDF of } r_t)$$

$$\Longrightarrow f(r_t|\mathcal{I}_{t-1}) = \frac{1}{\sqrt{2\pi h_{t|t-1}}} exp(\frac{-(r_t-\mu)^2}{sh_{t|t-1}})$$

Application to GARCH process Cont.

- Therefore, we can show the likelihood and the log-likelihood function as

$$L(\theta) = \Pi_{t=2}^{T} f(r_t|\mathcal{I}_{t-1}) f(r_1) = \Pi_{t=2}^{T} \frac{1}{2\pi h_{t|t-1}} exp(\frac{-(r_t - \mu)^2}{2h_{t|t-1}}) f(r_1)$$

$$\mathcal{L}(\theta) = \frac{-(T-1)}{2} ln(2\pi) - \frac{1}{2} \sum_{t=2}^{T} \{ ln(h_{t|t-1}) + \frac{(r_t - \mu)^2}{h_{t|t-1}} \} + lnf(r_1)$$

  where $lnf(r_1)$ is hard to deal with but just one term.

- There, we can approximate $\mathcal{L}(\theta)$ as

$$\boxed{approx\ \mathcal{L}(\theta) = \frac{-(T-1)}{2} ln(2\pi) - \frac{1}{2} \sum_{t=2}^{T} \{ ln(h_{t|t-1} + \frac{(r_t - \mu)^2}{h_{t|t-1}}) \}}$$

Model selection criterion: AIC

- AIC criterion:

$$\left.\begin{aligned} AIC &= -2lnL(\widehat{\theta}_{MLE}) + 2k \\ k &= numberofparameters \end{aligned}\right\} \text{general formula}$$

$$\approx (T-1)ln(2\pi) + \sum_{t=2}^{T}\{ln(\widehat{h}_{t|t-1}) + \frac{(r_t - \widehat{\mu})^2}{\widehat{h}_{t|t-1}}\} + 2k \text{ (Applied to GARCH)}$$

- <u>Note</u>:
  1. The term $-(T-1)ln(2\pi)$ is the same constant across all models, so removing it won't change the AIC comparison across models. Therefore, we often redefine the AIC for GARCH process by

     $$AIC \approx \sum_{t=2}^{T}\{ln(\widehat{h}_{t|t-1}) + \frac{(r_t - \mu)^2}{\widehat{h}_{t|t-1}}\} + 2k \text{ (compare to Enders, Ed2, p1}$$

  2. Both $\widehat{\mu}$ and $\widehat{h}_{t|t-1}$ will differ depending on the choice of GARCH model. For example, the formula for $h_{t|t-1}$ will depend on the orders p and q of the GARCH model.

Model selection criterion: SBC

- SBC criterion:

$$SBC = -2lnL(\widehat{\theta}_{MLE}) + kln(T) \text{ (general formula)}$$

$$\approx (T-1)ln(2\pi) + \sum_{t=2}^{T}\{ln(\widehat{h}_{t|t-1} + \frac{(r_t-\mu)^2}{\widehat{h}_{t|t-1}})\} + kln(T)$$

<u>or</u>, igoring the constant term

$$SBC = \sum_{t=2}^{T}\{ln(\widehat{h}_{t|t-1} + \frac{(r_t-\mu)^2}{\widehat{h}_{t|t-1}})\} + kln(T)$$

Comparison to model selection without GARCH erros

- e.g. $y_t = \mu + \varepsilon_t$,
  $\varepsilon_t \sim iidN(0, \sigma^2)$. - Replace GARCH errors by iid errors
  Therefore, $r_t \sim iid(\mu, \sigma^2)$.
- From equation (1), we have

$$\mathcal{L}(\mu, \sigma^2) = lnL(\mu, \sigma^2) = -\frac{T}{2} ln(2\pi) - \frac{T}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - \mu)^2$$

$$AIC = -2ln(\widehat{\mu}_{MLE}, \widehat{\sigma}^2_{MLE}) + 2k \text{ (general formular)}$$

$$= Tln(2\pi) + Tln(\widehat{\sigma}^2) + \frac{1}{\widehat{\sigma}^2} \sum_{t=1}^{T} (y_t - \widehat{\mu}) + 2k.$$

And if we ignore the constant $Tln(2\pi)$ then

$$AIC = Tln(\widehat{\sigma}^2) + \frac{1}{\widehat{\sigma}^2} \sum_{t=1}^{T} (y_t - \widehat{\mu})^2 + 2k.$$

Comparison to model selection without GARCH erros Cont.

- Also common to re-express in terms of the sum of squared residuals (SSR) as follows:

$$\widehat{\sigma}^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T}(y_t - \widehat{\mu})^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T}\widehat{\varepsilon}_t^2 = \frac{SSR}{T}$$
$$\Longrightarrow AIC = Tln(\widehat{\sigma}^2) + \frac{1}{\widehat{\sigma}^2}T\widehat{\sigma}^2 + 2k$$
$$= Tln(\widehat{\sigma}^2) + T + 2k.$$

- Again, the term T does not depend on the model parameters so it can be dropped with AIC redefined as:

$$AIC = Tln(\widehat{\sigma}^2) + 2k = Tln(\frac{SSR}{T}) + 2k,$$

(compare to Enders, p69, which may have a typo - it does not divide SSR by T).

Similarly, we can express $SBC = Tln(\frac{SSR}{T}) + kln(T)$.

Comparison to model selection without GARCH erros: Remark

- Expressing of AIC & SBC in terms of SSR also holds for regression models and autoregressive models.
- However, there is no simple way to express AIC & SBC for GARCH models in terms of the SSR - the likelihood based on variant must be used if GARCH models are included among the comparison.
- Notice that there are different definitions depending on whether constants dropped or not. The important thing is to be consistent - make sure the same definition of AIC or SBC is used across all models being compared.

Plugging in for $h_{t|t-1}$

- So far expressed likelihood generically in terms of $h_{t|t-1}$.
- For a particular GARCH specification we can provide a more detailed description of the likelihood by substituting in for $h_{t|t-1}$.
- This is more easily done for ARCH process

Plugging in for $h_{t|t-1}$: An example for ARCH(q)

- e.g. Log-likelihood of ARCH(q)

$$r_t = \mu + \varepsilon_t,$$

$$h_{t|t-1} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t=i}^2,$$

$$\alpha = (\alpha_0, \alpha_1, ..., \alpha_q)', \ \theta = (\mu, \alpha^{'}).$$

$$\mathcal{L}(\theta) \approx -\frac{(T-1)}{2} ln(2\pi) - \frac{1}{2}\sum\nolimits_{t=2}^{T}\left\{ln(h_{t|t-1}) + \frac{(r_t-\mu)^2}{h_{t|t-1}}\right\}$$

$$= -\frac{(T-1)}{2} ln(2\pi) - \frac{1}{2}\sum\nolimits_{t=2}^{T}\left\{ln\left(\alpha_t + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2\right) + \frac{\varepsilon_t^2}{(\alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2)}\right\}$$

$$= -\frac{(T-1)}{2} ln(2\pi) - \frac{1}{2}\sum\nolimits_{t=2}^{T}\left\{ln\left(\alpha_t + \sum_{i=1}^{q}\alpha_i(r_{t-i}-\mu)^2\right)\right.$$

$$\left. + \frac{(r_t-\mu)^2}{(\alpha_0 + \sum_{i=1}^{q}\alpha_i(r_{t-i}-\mu)^2)}\right\},$$

which is now expressed in terms of data and parameters.

Plugging in for $h_{t|t-1}$: An example for GARCH(1,1)

- e.g. Log-likelihood of GARCH(1,1)

$$
\begin{aligned}
& h_{t|t-1} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1|t-2} \\
& = \alpha_0 + \alpha_1 L\varepsilon_t^2 + \beta_1 L h_{t|t-1} \\
& \Longrightarrow (1 - \beta_1 L) h_{t|t-1} = \alpha_0 + \alpha_1 L \varepsilon_t^2 \\
& \Longrightarrow h_{t|t-1} = \frac{\alpha_0}{(1 - \beta_1 L)} + \frac{\alpha_1}{(1 - \beta_1 L)} L\varepsilon_t^2 \\
& = \frac{\alpha_0}{(1 - \beta_1)} + \alpha_1 \sum_{j=0}^{\infty} \beta_i^j L^j (L\varepsilon_t^2) \\
& \Longrightarrow h_{t|t-1} = \frac{\alpha_0}{(1 - \beta_1)} + \alpha_1 \sum_{j=0}^{\infty} \beta_i^j L^j \varepsilon_{t-j-1}^2,
\end{aligned}
$$

An ARCH(∞) representation.

Plugging in for $h_{t|t-1}$: An example for GARCH(1,1) Cont.

- Because we only have data starting at $t=1$, to make this practically feasible, we will need to initialize by setting $\varepsilon_t \equiv 0$ and $t \leq 0$.
- This gives us

$$h_{t|t-1} = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{j=0}^{t-2} \beta_1^j \varepsilon_{t-j-1}^2.$$

- Now we can plug this into $\mathcal{L}(\theta)$,

$$\mathcal{L}(\theta) \approx -\frac{(T-1)}{2} ln(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\left\{ ln(h_{t|t-1}) + \frac{(r_t-\mu)^2}{h_{t|t-1}} \right\}$$

$$= -\frac{(T-1)}{2} ln(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\left\{ ln\left(\frac{\alpha_0}{(1-\beta_1)} + \alpha_1 \sum_{j=0}^{t-2} \beta_1^j \varepsilon_{t-j-1}^2\right) + \frac{\varepsilon_t^2}{\frac{\alpha_0}{(1-\beta_1)} + \alpha_1 \sum_{j=0}^{t-2} \beta_1^j \varepsilon_{t-j-1}^2} \right\}.$$

Plugging in for $h_{t|t-1}$: An example for GARCH(1,1)
Remark

- Notice that we had to "solve" for $h_{t|t-1}$ before we could substitute it into the log-likelihood.
- When we "solved" for $h_{t|t-1}$, we obtained the ARCH(∞) representation.
- That involved squared errors infinitely far back.
- But, in practice, we cannot fit squared errors before the start of the data series.
- Therefore, the ARCH(∞) had to be truncated by imposing $\varepsilon_t = 0$ for $t \leq 0$ before we could substitute $h_{t|t-1}$ into the likelihood function.

Obtaining parameter estimates

- In simpler models, such as regression and AR(p) models, we can solve the first order conditions to get explicit formulas for the ML estimators.
- In ARCH & GARCH models, this is not possible.
- Instead numerical methods are used on computers to find the parameters values that maximize the likelihood.