ECMT3150: The Econometrics of Financial Markets

2b. Conditional Heteroskedastic Models

Simon Kwok
University of Sydney

Semester 2, 2020

Outline

Log-likelihood Function

Data: $\{r_t : t = 1, 2, \ldots, T\}$. Let $\mathbf{r} = (r_1, \ldots, r_T)'$.
Let $\mathcal{F}_{t-1} = \{r_s : s \leq t-1\}$ and let $f_{t|t-1}$ be the conditional density function of r_t given $\mathcal{F}_{t-1}$.
The joint density function of $\mathbf{r}$ is:

$$f(\mathbf{r}) = \prod_{t=1}^{T} f_{t|t-1}(r_t | \mathcal{F}_{t-1}).$$

E.g.: Assume that ε_t are *iid* $N(0,1)$.
This implies that $r_t | \mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2)$.
Define $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')' = (\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_s)'$.
The joint likelihood and log-likelihood functions are:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{r}) &= \prod_{t=1}^{T} f_{t|t-1}(r_t | \mathcal{F}_{t-1}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{ -\frac{(r_t - \mu_t)^2}{2\sigma_t^2} \right\}, \\ \ell(\boldsymbol{\theta}; \mathbf{r}) &= \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}; r_t) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{r_t - \mu_t}{\sigma_t}\right)^2. \end{aligned}$$

Maximum Likelihood Estimation

Q: How to estimate a GARCH model?

1. Assume a distribution on the error: e.g., ε_t are i.i.d. $N(0, 1)$, so that $a_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$. Other error distributions may be assumed: e.g., Student-t, generalized error distribution.
2. Obtain the log-likelihood $\ell(\boldsymbol{\theta}; \mathbf{r})$ as a function of parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$.
3. Define the pre-sample values of a_t^2 and σ_t^2 (when $t \leq 0$). Some options:
    - 3.1 set to zero;
    - 3.2 set to the unconditional variance in (7) of slide 2a;
    - 3.3 set to the mean squared residuals $\frac{1}{T-p-q} \sum_{t=p+q+1}^{T} \hat{a}_t^2$; or
    - 3.4 treat them as additional parameters.
4. Maximize $\ell(\boldsymbol{\theta}; \mathbf{r})$ w.r.t. $\boldsymbol{\theta}$ and obtain the MLE $\hat{\boldsymbol{\theta}}$.

Maximum Likelihood Estimation

Let $\boldsymbol{\theta}_0$ be the true parameter vector. Under some regularity assumptions, the MLE $\hat{\boldsymbol{\theta}}$ is:

- consistent for $\boldsymbol{\theta}_0$, i.e., $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $T \to \infty$.
- asymptotically normal:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} MN\left(\mathbf{0}, \mathcal{I}\left(\boldsymbol{\theta}_0\right)^{-1}\right)$$

  as $T \to \infty$. The limit $\mathcal{I}\left(\boldsymbol{\theta}_0\right)$ is known as the *asymptotic information matrix*: $\mathcal{I}\left(\boldsymbol{\theta}_0\right) := -\lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \ell_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]$.
- efficient (Cramér-Rao): Among all asymptotically unbiased estimators $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}$ has the smallest variance asymptotically, i.e., $Var\left(\sqrt{T}\tilde{\boldsymbol{\theta}}\right) - Var\left(\sqrt{T}\hat{\boldsymbol{\theta}}\right)$ is positive definite in the limit as $T \to \infty$.

GARCH Model - Estimation

Q: How to estimate the s.e. of $\hat{\theta}_j$ for $j = 1, \ldots, m + s$?
A: Get the inverse of the empirical Hessian matrix evaluated at $\hat{\boldsymbol{\theta}}$:

$$H^{-1}(\hat{\boldsymbol{\theta}}) = \left(- \frac{\partial^2 \ell \left(\hat{\boldsymbol{\theta}} \right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} = \left(- \sum_{t=1}^{T} \frac{\partial^2 \ell_t \left(\hat{\boldsymbol{\theta}} \right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} .$$

Then the s.e. of $\hat{\theta}_j$ is the square root of the jth diagonal element of $H^{-1}(\hat{\boldsymbol{\theta}})$.
Under correct model specification and some regularity assumptions, $T \cdot H^{-1}(\hat{\boldsymbol{\theta}}) \xrightarrow{a.s.} \mathcal{I}\left(\boldsymbol{\theta}_0\right)^{-1}$ as $T \to \infty$.

Q: What if the $N(0, 1)$ error distribution is wrong?
A: The set of MLE procedures becomes *quasi-maximum likelihood estimation* (QMLE). It gives consistent but inefficient parameter estimate.

Estimators of Asymptotic Variance

Different estimators for $AVar(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\mathcal{I}\left(\boldsymbol{\theta}_0\right)^{-1}$ (i.e., asymptotic variance of $\hat{\boldsymbol{\theta}}$):

1. Empirical Hessian estimator: $\left(-\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)^{-1}$.
2. Information matrix estimator: $\left(E\left[-\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right)^{-1}$.
3. Outer-product-of-the-gradient estimator:

$$\left(\sum_{t=1}^{T} \frac{\partial \ell_t(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} \frac{\partial \ell_t(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right)^{-1}.$$

4. Sandwich estimator:

$$\left(-\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)^{-1} \left(\sum_{t=1}^{T} \frac{\partial \ell_t(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} \frac{\partial \ell_t(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right) \left(-\frac{\partial^2 \ell(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)^{-1}.$$

Model Diagnostics

Check whether the standardized residuals $\tilde{a}_t = \frac{\hat{a}_t}{\sigma_t}$ form an iid sequence.

- Ljung-Box test on $\tilde{a}_t$ detects autocorrelations, which may hint to misspecification of the conditional mean model μ_t.
- QQ plot and tests of skewness and kurtosis of $\tilde{a}_t$ check the validity of the distributional assumption on ε_t.

IGARCH

- Motivation: During turbulent periods, the volatility process may display persistence (unit-root behaviour), which the stationary GARCH model cannot accommodate.
- The *Integrated GARCH* (IGARCH) model relaxes the stationarity restriction by allowing for the presence of unit roots in the AR polynomial associated with (9) of slide 2a, i.e., some roots to the polynomial equation

$$1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) x^i = 0$$

  lie *on* the unit circle.
- E.g.: For IGARCH(1,1), the root $x = \frac{1}{\alpha_1 + \beta_1}$ lying on the unit circle means that $\alpha_1 + \beta_1 = 1$, so that the necessary condition for stationarity ((8) of slide 2a) is violated.

GARCH-M

- Motivation: asset returns may depend on its volatility. Usually, investors seek a higher rate of returns on assets that display higher volatility.
- The *GARCH-in-mean* (GARCH-M) model includes the conditional variance σ_t^2 as additional regressor of r_t. E.g., a GARCH(1,1)-M model is:

$$\begin{aligned} r_t &= \mu + c\sigma_t^2 + a_t, \\ a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

- The coefficient c is interpreted as the volatility risk premium - the average increase in log-return for a unit increase in conditional variance.

EGARCH

- Motivation: asset volatility may exhibit asymmetric effect.
- The *exponential GARCH* (EGARCH) model:
  - specifies the dynamics of the log volatility process $\ln(\sigma_t^2)$ to ensure positive σ_t^2;
  - allows for asymmetric effect of a past standardized shock ε_{t-1} on log volatility $\ln(\sigma_t^2)$, captured by the asymmetry function:

$$\begin{aligned} g(\varepsilon_t) &= \theta\varepsilon_t + \gamma\left[|\varepsilon_t| - E(|\varepsilon_t|)\right] \\ &= \begin{cases} (\theta+\gamma)\varepsilon_t - \gamma E(|\varepsilon_t|) & \text{if } \varepsilon_t \geq 0, \\ (\theta-\gamma)\varepsilon_t - \gamma E(|\varepsilon_t|) & \text{if } \varepsilon_t < 0. \end{cases} \end{aligned}$$

- E.g., EGARCH(1,1):

$$\begin{aligned} r_t &= \mu_t + a_t, \qquad a_t = \sigma_t\varepsilon_t, \\ \ln(\sigma_t^2) &= \omega + \alpha\ln(\sigma_{t-1}^2) + g(\varepsilon_{t-1}) \\ &= \omega + \alpha\ln(\sigma_{t-1}^2) + \theta\varepsilon_{t-1} + \gamma\left[|\varepsilon_{t-1}| - E(|\varepsilon_{t-1}|)\right]. \end{aligned}$$

TGARCH

- Motivation: asset volatility may be governed by different dynamics depending on the value of the past return shocks.
- Define the indicator $1_A = 1$ if A is true, and $1_A = 0$ otherwise.
- The *threshold GARCH* (TGARCH) model assumes different marginal impact of past squared shocks a_{t-i}^2 on σ_t^2, depending on the values of a_{t-i}^2.
  E.g., TGARCH(1,1):

$$
\begin{aligned}
r_t &= \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + (\alpha_1 + \gamma_1 1_{\{a_{t-1}<0\}}) a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \begin{cases} \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & \text{if } a_{t-1} \geq 0, \\ \alpha_0 + (\alpha_1 + \gamma_1) a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & \text{if } a_{t-1} < 0. \end{cases}
\end{aligned}
$$

  TGARCH(m, s):

$$
\begin{aligned}
r_t &= \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i 1_{\{a_{t-i}<0\}}) a_{t-i}^2 + \sum_{i=1}^{m} \beta_i \sigma_{t-i}^2 .
\end{aligned}
$$

Stochastic Volatility Model

- Motivation: conditional variance may be driven by an independent random source different from the return innovations.
- Define the *stochastic volatility* (SV) model:

$$\begin{aligned} r_t &= \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \\ \ln(\sigma_t^2) &= \alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + \cdots + \alpha_m \ln(\sigma_{t-m}^2) + v_t, \end{aligned}$$

  where ε_t are $iid(0,1)$, v_t are $iid(0,\sigma_v^2)$, and $\{\varepsilon_t\}$ and $\{v_t\}$ are independent.
- Stationarity condition: the roots of the AR polynomial equation $1 - \sum_{i=1}^m \alpha_i x^i = 0$ lie outside the unit circle.
- Compared to GARCH-type models, the separate random source in SV model adds more flexibility to the volatility dynamics.

Long Memory Stochastic Volatility Model

We may introduce long memory to an SV model:

$$\begin{aligned} r_t &= \mu_t + a_t, \qquad a_t = \sigma_t \varepsilon_t, \\ \ln(\sigma_t^2) &= \alpha_0 + u_t, \\ (1-L)^d u_t &= \eta_t, \end{aligned}$$

where ε_t are $iid(0,1)$, η_t are $iid(0,\sigma_\eta^2)$, $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent, and $d \in (0, 0.5)$.
The fractional differencing leads to long-range positive dependence of u_t, with its ACF decaying hyperbolically with lag order ($\rho_j \sim j^{2d-1}$ as $j \to \infty$).