大学院計量経済学講義ノートII

劉　慶豊 (Qingfeng Liu)*

平成 21 年 12 月 10 日

小樽商科大学商学部

*qliu.econ@gmail.com

1　標準的な仮定が満たされない場合

A1:$y = X\beta + \varepsilon$ が満たされない場合 nonlinear または nonparametric,semiparametric モデルになる

A3:$E(\varepsilon|X) = 0$ が満たされない場合は推定量の普遍性と一致性がなくなる

A4:$Var(\varepsilon|X) = \sigma^2 I$ が満たされない場合分散不均一、自己相関の問題が発生する

1.1　分散不均一

$$Var(\varepsilon|X) = \sigma^2 \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_{nn} \end{bmatrix} \tag{1}$$

証明を簡単にするため $\sum_{i=1}^{n} w_{ii} = n$ となるようにデータを標準化する。

1.2　OLS 推定量の性質

1.2.1　不偏性

$$\begin{aligned} b &= (X'X)^{-1} X'y = (X'X)^{-1} X' (X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1} X'\varepsilon \end{aligned}$$

$$\begin{aligned} E(b) = E[E(b|X)] &= \beta + E\left\{E\left[(X'X)^{-1} X'\varepsilon|X\right]\right\} \\ &= \beta + E\left\{(X'X)^{-1} X'E(\varepsilon|X)\right\} \\ &= \beta + E\left\{(X'X)^{-1} X' \times 0\right\} = \beta \end{aligned}$$

以上の証明では仮定 $A3, E(\varepsilon|X) = 0$ を利用した。

1.2.2　一致性

$$b = \beta + \left(\frac{X'X}{n}\right)^{-1} \frac{X'\varepsilon}{n}$$

- 仮定 $A5'$　$\frac{X'X}{n} \xrightarrow{p} Q < \infty$,positive definite 正値定符号。

定理 1 (チェビシェフの大数の弱法則) $Z_1, Z_2, \cdots, Z_n$ は互いに独立な確率変数で $E(Z_i) = u_i$、$V(Z_i) = Q_i < \infty$ 正値定符号で、$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} Q_i = 0$ であるとき、$\frac{1}{n}\sum_{i=1}^{n}(Z_i - u_i) \xrightarrow{p} 0$。

ここで、$x_i\varepsilon_i$ を z_i とチェビシェフの大数の弱法則を適用する。

$$
\begin{aligned}
&E(x_i\varepsilon_i) = 0 \\
&V(x_i\varepsilon_i) = \sigma^2 E(w_{ii}x_i x_i')
\end{aligned}
$$

なので、$\sigma^2 E(w_{ii}x_i x_i') < \infty$ が正値定符号で、$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{n^2}\sigma^2 E(w_{ii}x_i x_i')$ の二つの条件の下で、チェビシェフの大数の弱法則より、$\frac{X'\varepsilon}{n} \xrightarrow{p} 0$, そして

$$
b - \beta = \left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n} \xrightarrow{p} Q^{-1} \times 0 = 0
$$

一致性が満たされる。

1.2.3 推定量 b の共分散行列の推定量

- 分散が均一の場合 $s^2(X'X)^{-1}$ が一致推定量となる

$$
\begin{aligned}
V(b|X) &= \sigma^2(X'X)^{-1} \\
s^2 &= \frac{1}{n-k}\sum_{i=1}^{n}(y_i - X_i'b)^2 \\
\hat{V}(b|X) &= s^2(X'X)^{-1} \\
s^2 &\xrightarrow{p} \sigma^2 \\
\hat{V}(b|X) &\xrightarrow{p} V(b|X)
\end{aligned}
$$

- 分散が不均一の場合

$$
\begin{aligned}
V(b|X) &= E\left[(b - E(b|X))(b - E(b|X))'|X\right] \\
&= E\left[(b-\beta)(b-\beta)'|X\right] \\
&= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}|X\right] \\
&= (X'X)^{-1}X'\underbrace{E(\varepsilon\varepsilon')}_{\sigma^2\Omega}X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
s^2 &= \frac{1}{n-k}(y-Xb)'(y-Xb) \\
&= \frac{1}{n-k}\left(y-X(X'X)^{-1}X'y\right)'\left(y-X(X'X)^{-1}X'y\right) \\
&= \frac{1}{n-k}y'\left(I-X(X'X)^{-1}X'\right)y
\end{aligned}
$$

$y = X\beta + \varepsilon$ を代入して上式

$$
\begin{aligned}
&= \frac{1}{n-k}\varepsilon'\varepsilon - \frac{1}{n-k}\varepsilon'X(X'X)^{-1}X'\varepsilon \\
&= \frac{1}{n-k}\varepsilon'\varepsilon - \frac{1}{n-k}\varepsilon'X\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}
\end{aligned}
$$

$E(X'\varepsilon) = 0$ という仮定から $\frac{1}{n-k}\varepsilon'X \xrightarrow{p} 0$、$\frac{X'\varepsilon}{n} \xrightarrow{p} 0$。$A5'$ から $\frac{X'X}{n} \xrightarrow{p} Q < \infty$ ゆえに $\frac{1}{n-k}\varepsilon'X\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n} \xrightarrow{p} \frac{1}{n-k}\varepsilon'X(Q)^{-1} \times 0 = 0$。

$$
\frac{\varepsilon'\varepsilon}{n-k} = \frac{1}{n-k}\sum_{i=1}^{n}\varepsilon_i^2 = \frac{1}{n-k}\sum_{i=1}^{n}\left(\varepsilon_i^2 - \sigma^2\right) + \frac{n}{n-k}\sigma^2
$$

$E(\varepsilon_i^2) = \sigma_i^2 = \sigma^2 E(w_{ii})$、$E(\varepsilon_i^4) < \infty$ 且つゼロではない、$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}E(\varepsilon_i^4) = 0$ の仮定の下でチェビシェフの大数の法則より $\frac{1}{n-k}\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma_i^2) \xrightarrow{p} 0$。

さらに $\sum_{i=1}^{n} w_{ii} = n$ より $\frac{1}{n-k}\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma^2) = \frac{1}{n-k}\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma_i^2)$

以上から $s^2 \xrightarrow{p} \sigma^2$。

- 分散不均一であるにもかかわらず、$s^2(X'X)^{-1}$ を $V(b|X)$ の推定量として良いのか

$$
Est.V[b|X] - V(b|X) = s^2(X'X)^{-1} - \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}
$$

大標本の場合、近似的に $s^2 = \sigma^2$ として、定式は

$$
\begin{aligned}
&= \sigma^2\left[(X'X)^{-1}X'X(X'X)^{-1} - (X'X)^{-1}X'\Omega X(X'X)^{-1}\right] \\
&= \sigma^2\left[(X'X)^{-1}X'(I-\Omega)X(X'X)^{-1}\right] \\
&= \frac{\sigma^2}{n}\left[\left(\frac{X'X}{n}\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_{ii})x_ix_i'\right]\left(\frac{X'X}{n}\right)^{-1}\right]
\end{aligned}
$$

もし x_i と w_{ii} が独立で x_i が $i.i.d$ であれば、

$$
\begin{aligned}
&E\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_{ii})x_ix_i'\right] \\
&= E(x_ix_i')E\left[\frac{1}{n}\sum_{i=1}^{n}(1-w_{ii})\right] = 0
\end{aligned}
$$

$s^2\left(X'X\right)^{-1}$ は $V\left(b|X\right)$ の不偏な推定量である。しかし、x_i と w_{ii} が独立で x_i が $i.i.d$ という条件が満たされない場合、$s^2\left(X'X\right)^{-1}$ は $V\left(b|X\right)$ の推定量としてバイアスを持つ。

1.2.4 $V\left(b\right)$ の正しい推定量、White の推定量

$$
\begin{aligned}
V\left(b|X\right) &= \left(X'X\right)^{-1}\sigma^2 X'\Omega X\left(X'X\right)^{-1} \\
nV\left(b|X\right) &= \left(\frac{X'X}{n}\right)^{-1}\frac{\sigma^2 X'\Omega X}{n}\left(\frac{X'X}{n}\right)^{-1}
\end{aligned}
$$

$e_i = y_i - x_i'b$ として、ε_i と x_i が独立であれば

$$
p\lim\left(\frac{1}{n}\sum_{i=1}^{n}x_i x_i' e_i^2 - \underbrace{\frac{\sigma^2 X'\Omega X}{n}}_{\frac{1}{n}\sum_{i=1}^{n}x_i x_i'\sigma^2 w_{ii}}\right) = 0
$$

White の heteroscedasticity consistent estimator は

$$
\widehat{V\left(b|X\right)} = \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} S_0\left(\frac{X'X}{n}\right)^{-1}
$$

ただし、$S_0 = \frac{1}{n}\sum_{i=1}^{n}x_i x_i' e_i^2$。

1.2.5 OLS 推定量の性質漸近分布

$b = \left(X'X\right)^{-1}X'y = \beta + \left(X'X\right)^{-1}X'\varepsilon$ なので、ε が正規分布と仮定すれば

$$
b|X \sim N\left[\beta, \sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}\right]
$$

となる。

1.3 撹乱項が互いに相関する場合

例を挙げて説明する

例 2 *Inconsistent OLS*

$y = \mu + \varepsilon$

$$
\Omega = \begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \rho \\
\rho & \cdots & \rho & 1
\end{bmatrix}
$$

μ の最小二乗推定量は y の標本平均 $\bar{y}$ となる。

$$\begin{aligned} Var\left(\bar{y}\right) &= E\left[E\left(\varepsilon\varepsilon'|X=1\right)\right] \\ &= \frac{\sigma^2}{n}\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\Omega X\right)\left(\frac{1}{n}X'X\right)^{-1} \end{aligned}$$

$X=1$ のため

$$Var\left(\bar{y}\right) = \frac{\sigma^2}{n}\left(\frac{1}{n}X'\Omega X\right)$$

さらに $X'\Omega X = (n-1)\,n\rho + n$ であるので

$$Var\left(\bar{y}\right) = \frac{\sigma^2}{n}\left(1-\rho+n\rho\right) \to \sigma^2\rho$$

ゼロにならない。$\bar{y}$ が μ の一致推定量とならないことがわかる。

1.4 分散不均一の検定

1.4.1 White(1980) の検定

$$\begin{aligned} &H_0 : \sigma_i^2 = \sigma^2 \text{ for all } i \\ &H_1 : \text{Not } H_0 \end{aligned}$$

White の発想はもし分散均一であれば $S_0 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' e_i^2 \xrightarrow{p} \sigma^2 Q \approx S^2\left(X'X\right)^{-1}$ のはずである。もし分散が不均一であるならそうならない。$\psi_i \equiv x_i x_i', \hat{\psi} = 1/n\sum_{i-1}^{n}\psi_i$ と定義する。

$$\begin{aligned} S_0 - S^2\frac{X'X}{n} &= \frac{1}{n}\sum_{i=1}^{n} x_i x_i' e_i^2 - \frac{1}{n}S^2\sum_{i=1}^{n} x_i x_i' \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i x_i'\left(e_i^2 - S^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\psi_i\left(e_i^2 - S^2\right) \end{aligned}$$

White の統計量

$$\begin{aligned} T \equiv n &\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_i\left(e_i^2-S^2\right)\right\} \times \left\{\frac{1}{n}\sum_{i=1}^{n}\left(e_i^2-S^2\right)\left(\psi_i-\hat{\psi}\right)\left(e_i^2-S^2\right)\right\} \\ &\times \left\{\frac{1}{n}\sum_{i=1}^{n}\psi_i\left(e_i^2-S^2\right)\right\} \xrightarrow{d} \chi^2_{k(k+1)/2} \end{aligned}$$

と定義される。弱点として、White(1980) の検定を行ってから、分散不均一が検出されても、分散共分散行列の構造がわからないため、次はどう対処すべきかわからない。

1.4.2 Goldfeld-Quandt Test

$$F=\frac{S_1^2}{S_2^2}=\frac{e_1'e_1/n_1}{e_2'e_2/n_2}\xrightarrow{d}F\left(n_1,n_2\right)$$

1.5 効率的な推定 GLS,MLE

1.5.1 GLS 一般化最小二乗法（Ω が既知の場合）

モデル

$$\begin{aligned}y_i&=x_i'\beta+\varepsilon_i\\E\left(\varepsilon_i|x_i\right)&=0\\V\left(\varepsilon_i|x_i\right)&=\sigma^2w_i\end{aligned}$$

変形したモデル

$$\frac{y_i}{\sqrt{w_i}}=\frac{x_i'}{\sqrt{w_i}}\beta+\frac{\varepsilon_i}{\sqrt{w_i}}$$

分散均一になる。普通の OLS が効率的な推定量となる。

一般的な表現 $Var\left(\varepsilon|X\right)=\Omega$ として、分散も共分散も不均一（共分散が一定の条件を満たす）とする。モデルの行列表現が $y=X\beta+\varepsilon$ とする。分散共分散行列を $\Omega=\Gamma\Gamma'$[1] と三角分解し、モデルを変形する

$$\Gamma^{-1}y=\Gamma^{-1}X\beta+\Gamma^{-1}\varepsilon$$

さらに書き換えて

$$y^*=X^*\beta+\varepsilon^*$$

となる。

$$\begin{aligned}Var\left(\varepsilon^*|X\right)&=E\left(\varepsilon^*\varepsilon^{*\prime}|X\right)\\&=E\left(\Gamma^{-1}\varepsilon\varepsilon'\left(\Gamma^{-1}\right)'|X\right)\\&=\Gamma^{-1}E\left(\varepsilon\varepsilon'|X\right)\left(\Gamma^{-1}\right)'\\&=\Gamma^{-1}\Omega\left(\Gamma^{-1}\right)'=I\end{aligned}$$

なので、ε^* が分散均一となる。モデル $y^*=X^*\beta+\varepsilon^*$ に関して OLS 推定を行う、その推定量はもとのモデル $y=X\beta+\varepsilon$ の GLS 推定量となる

$$\begin{aligned}\beta_{GLS}&=\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}y^*\\&=\left[\left(\Gamma^{-1}X\right)'\left(\Gamma^{-1}X\right)\right]^{-1}\left(\Gamma^{-1}X\right)'\Gamma^{-1}y\\&=\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y\end{aligned}$$

[1] 任意の正値定符号行列は下側三角行列 Γ によって $A=\Gamma\Gamma'$ と分解できる。さらに、$\Gamma^{-1}A\left(\Gamma^{-1}\right)'=I$ となる。

β_{GLS} は分散均一のモデル

$$y^* = X^*\beta + \varepsilon^*$$

の OLS 推定量なので、以前学んだ OLS 推定量の性質（Gauss-Markov Theorem(BLUE, Best Linear Unbiased Estimator)）からこの推定量は効率性を達成することがわかる。
標本分散は

$$V\left(\beta_{GLS}\right) = \sigma^2\left(X'\Omega^{-1}X\right)^{-1}$$

となる。そして、

分散不均一のとき最小二乗推定量 b が効率性を持たない

$$Var\left(b|X\right) \geq Var\left(\beta_{GLS}|X\right)$$

となり、b が効率的ではないこともわかる。

1.5.2 GLS 一般化最小二乗法（Ω が未知の場合）

Ω を推定してから β_{GLS} に代入することが基本的な考えだが、場合によって Ω の推定は計算上困難な時がある。簡単に推定できる例をあげる。

$$Var\left(\varepsilon|X\right) = \Omega = \begin{bmatrix} \alpha z_1 & 0 & \cdots & 0 \\ 0 & \alpha z_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha z_n \end{bmatrix} \tag{2}$$

z が観測できる変数、$z = x$ でもいい。
2 段階推定を行う

1. OLS を行って残差 e_i を求めて、e_i を z_i に回帰する α そして Ω を推定する
2. Ω の推定量を β_{GLS} に代入して feasible な β_{GLS} を計算する

$$\beta_{FGLS} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}y$$

1.5.3 最尤法 (MLE)

尤度関数：y_i が $i.i.d.$ とする。

$$L\left(\theta|\boldsymbol{y}\right) \equiv f\left(y_1, \cdots, y_n|\theta\right) = \prod_{i=1}^{n} f\left(y_i|\theta\right)$$

大数尤度：

$$\ln L\left(\theta|\boldsymbol{y}\right) = \sum_{i=1}^{n} \ln f\left(y_i|\theta\right)$$

真のモデルが $f(\theta_0)$ とする。最尤推定量

$$\theta_{MLE} = \arg\max \ln L(\theta|\boldsymbol{y})$$

と定義される。

最尤法の直感的な理解

データ $\{y_1, y_2, \cdots, y_n\}$ が観測されて、真のモデルの構造はが $f(\theta_0)$ によって規定されていることがわかるが、正確な θ_0 の値がわからない、できるだけ θ_0 に近い値を割り出したい。そのとき、様々な θ の値 $\theta_1, \theta_2, \cdots$ そして様々な $f(\theta)$、$f(\theta_1), f(\theta_2), \cdots$ が考えられるが、その中から、この θ でこの $f(\theta)$ であれば、今観測されたデータ $\{y_1, y_2, \cdots, y_n\}$ が一番現れやすいというような θ を選ぶ。これは MLE が行っていることである。

- 分散不均一のときの MLE、ε_i が $i.i.d.$ で正規分布に従うと仮定する[2]

$$Var(\varepsilon|X) = \Omega = \begin{bmatrix} \omega_1(\alpha) & 0 & \cdots & 0 \\ 0 & \omega_2(\alpha) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \omega_n(\alpha) \end{bmatrix} \tag{3}$$

とする

$$\ln L(\beta, \alpha|\mathbf{y}) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n} \ln w_i(\alpha) - \frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - x_i'\beta)^2}{w_i(\alpha)}$$

$\ln L(\beta, \alpha|\mathbf{y})$ を最大にするような β と α が β_{MLE} と α_{MLE} となる。最大化の計算は普通統計ソフトで数値的に行う。

2 ARCH 型モデル (Autoregressive Conditional Heteroscedasticity Type Models)

2.1 ARCH(q) モデルと GARCH(p,q) モデル

ARCH(q) モデル (Engle (1982)) と GARCH(p,q) モデル (Bollerslev(1986)) は ARCH 型モデルの基礎をなす二つのモデルである。この二つのモデルを理解しておくことは, ARCH 型モデル全体の発想や考え方を理解するために不可欠である。

---

[2] 正規分布の密度関数

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

2.1.1 ARCH 型モデルの起源 ARCH(q) モデル

Engle(1982) は時系列データの分散不均一性を捉えるためのモデルとして，ARCH(q) モデルを提唱した。ARCH(q) モデルにおいては，条件付分散が不均一であることを仮定されている。$t-1$ 期までの情報集合 ψ_{t-1} に含まれる内生変数のラグと外生変数の線形結合 $X_t\beta$ によって確率変数 y_t の条件付平均があらわされる。またその条件付分散を h_t とする。そのモデル式は次のようにあらわされる。

$$y_t = X_t\beta + \varepsilon_t \quad or \quad y_t|\psi_{t-1} \sim N[X_t\beta, h_t] \tag{4}$$

$$\varepsilon_t = \sqrt{h_t}z_t \quad z_t \sim N[0,1]\ i.i.d.. \tag{5}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{6}$$

ここの $\sqrt{h_t}$ はファイナンスモデルにおけるボラティリティに該当する。モデル式を見れば分かるように，ARCH(q) モデルはボラティリティが過去の予期せぬリターンのショックの二乗によって影響されることを表現している。

h_t は条件付分散なので,非負でなければならない。従って,$\alpha_0 > 0, \alpha_i \geq 0$ という制約が必要となる。この制約により，一旦リターンに予期せぬショックが発生すると，次期のボラティリティが大きくなり、その影響が長期的な効果を及ぼす。これは金融市場における収益率が激しく変動すると，しばらくの間，激しい変動が続くという事実と合致する。このように ARCH(q) モデルはいわゆる Volatility Clustering という現象を表現できる。これが ARCH 型モデルがファイナンスの分野で広く利用されるようになった理由である。

2.1.2 GARCH(p,q) モデル

実務上，ARCH(q) モデルを金融データに適用して，ボラティリティに対するショックの持続性を表現しようとすると，ARCH(q) モデルの次数が長くなるという問題が生じる。この問題を回避するために考察されたのが GARCH(p,q) モデルである。GARCH(p,q) モデルは ARCH(q) モデルにボラティリティのラグの項を追加したものである。GARCH(p,q) モデルのボラティリティに関する式は次のようになる。

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{7}$$

AR モデルに MA 項を追加することで, AR モデルの次数を減らせることと同様に,GARCH モデルにおいてはボラティリティのラグ項を導入することで,短い次数，普通は GARCH(1,1) モデルで，データの特性を捉えることができるようになる。

GARCH モデルにおいても，ボラティリティの非負性を保つため，各パラメータに制約をおく必要がある。ARCH モデルにおける制約

$$\alpha_0 > 0, \alpha_i \geq 0 \quad i = 1, 2, \cdots, q$$

に加えて，$\beta_j \geq 0, j = 1, 2, \cdots p$ の制約条件が必要となる。

2.1.3 ARCH(q),GARCH(p,q) モデルの定常条件および推定法

Franses and Dijk (2000) の方法に従って，(2) 式の両辺に ε_t^2 を足し，h_t を引くことによって，ARCH(q) モデルの新しい表現が得られる。

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-q}^2 + v_t.$$

ただし $v_t \equiv \varepsilon_t^2 - h_t$ である。このとき明らかに，$E[v_t] = E[E[v_t|\psi_{t-1}]] = 0$ が成立する。見方を変えると，(6) 式は ε_t^2 に関する AR(q) モデルとなる。この AR(q) モデルすなわち ARCH(q) モデルが共分散定常になるためには，z に関する方程式

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_q z^q = 0$$

の解が単位円の外にある必要がある。このとき，α_i の非負制約によって，結局共分散定常または弱定常条件は

$$\alpha_1 + \alpha_2 + \cdots + \alpha_q < 1$$

となる。結果として，ε_t の無条件分散 σ^2 は

$$\sigma^2 = E(\varepsilon_t^2) = \alpha_0/(1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_q))$$

となる。GARCH(p,q) モデルの場合，ARMA モデルに書き換えることで，その弱定常条件を求めることができる (Hamilton (1994))。それは

$$\alpha_1 + \alpha_2 + \cdots + \alpha_q + \beta_1 + \beta_2 + \cdots + \beta_p < 1$$

である。

ARCH(q) モデルと GARCH(p,q) モデルを含む ARCH 型モデルの推定には最尤法が用いられることが多い。その対数尤度関数は基本的に以下のような関数である。

$$L_t(\theta) = -T\frac{1}{2}\log 2\pi - \sum_{t=1}^{T}\frac{1}{2}\log h_t - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2h_t}. \tag{8}$$

ただし，T はサンプル期間である。

2.2 GARCH(p,q) モデル以後の発展

GARCH(p,q) モデルが提起されてから，広く実証研究に用いられてきたが，実際のデータの更なる特徴を捉えられるモデルが求められるようになってきた。一連の流れの中で IGARCH モデル, EGARCH モデル, TGARCH モデルなど様々なモデルが提唱されてきた。

2.2.1 弱定常性に関する問題

GARCH(1,1) モデルを金融データに適用するとき，$\alpha_1 + \beta_1$ の推定値が 1 に近い値をとることが多い。特に為替レートのデータの場合では顕著である (Bollerslev, Chou and Kroner (1992))。このことは，GARCH(1,1) モデルの弱定常条件 $\alpha_1 + \beta_1 < 1$ が満たされず，実際には $\alpha + \beta_1 = 1$ となっている可能性を示唆している。Engle and Bollerslev (1986) は，このような特殊なケースをうまく説明するためのモデルとして IGARCH モデルを提唱した。

2.2.2 非対称性に関する問題

実証研究を積み重ねるうちに，金融データのボラティリティに関する様々な特徴が明らかになってきた。その中の一つとして，ボラティリティ変動の非対称性があげられる。非対称性とは，前日の株価変動の方向の違いによって，ボラティリティの増減の傾向が変わるという特性である。経験的には前日の株価が下落した場合には，前日の株価が上昇した場合よりも，今日のボラティリティは大きくなる傾向が強い。しかし，GARCH モデルでは，この非対称性を捉えることができない。そこで，GJR-GARCH モデル，EGARCH モデル，TGARCH モデルなどが提案された。. Glosten, Jagannathan and Runkle (1993) によって提案された GJR-GARCH モデルでは非対称性を表すために，ε_{t-i}^2 の係数がショックすなわち ε_{t-i} の符号に依存すると仮定されている。GJR-GARCH(p,q) モデルのモデル式は

$$h_t = \alpha_0 + \sum_{i=1}^{q} (\alpha_i \varepsilon_{t-i}^2 + \gamma_i \varepsilon_{t-i}^2 I[\varepsilon_{t-i} < 0]) + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j}^2$$

$$\alpha_0 > 0, \ \ \alpha_i, \beta_j, \gamma_i \geq 0.$$

と定式化される。このように定式化することによって，ε_{t-i} がマイナスのとき，ほかの期のショックを 0 として，$t-i$ 期にのみショックを与えて，同じサイズのショックがマイナスであるほうがプラスであるときよりもボラティリティの変動に与える影響を大きくすることが可能となる。つまり，ボラティリティ変動の非対称性を表現できるようになった。

Nelson (1991) によって提案された EGARCH(Exponential GARCH) モデルは非説明変数の h_t を $\ln(h_t)$ に置き換えることによって，ボラティリティ変動の非対称性を捉えることを可能にし，また GARCH モデルにおけるパラメータの非負制約を除去することにも成功した。EGARCH(p,q) モデルは以下の式で表される。

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i [\theta z_{t-i} + \gamma(|z_{t-i}| - E|z_{t-i}|)] + \sum_{j=1}^{p} \beta_j \ln(h_{t-j})$$

$$\varepsilon_t = \sqrt{h_t} z_t \ \ \ z_t \sim N[0,1] \ i.i.d.$$

ここで，$\theta < 0$ であれば，前述した経験的事実のボラティリティ変動の非対称性が表現されることになる。$\ln(h_t)$ がマイナスの値をとることもできるので，パラメータに関す

る非負制約が不要となる。ただし,EGARCH モデルには欠点もある。GARCH モデルと異なり，ε_t の代わりに z_t でモデルを定式化したため，特定のショック ε_t のボラティリティの変動に与える影響がわかりにくくなるのである。

ボラティリティ変動の非対称性を捉えると同時に,EGARCH モデルの欠点を克服するモデルとして，Zakoian (1994) は TGARCH(Threshold GARCH) モデルを提案した。TGARCH モデルは次のように定式化される。

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^{q}(\alpha_i^+ \varepsilon_{t-i}^+ + \alpha_i^- \varepsilon_{t-i}^-) + \sum_{j=1}^{p}\beta_j h_{t-j} \\ \varepsilon_t &= \sqrt{h_t} z_t \quad z_t \sim N[0,1]\ i.i.d. \\ \varepsilon_{t-i}^+ &= \max(\varepsilon_t, 0) \quad \varepsilon_{t-i}^- = \min(\varepsilon_t, 0) \\ \alpha_0 &> 0, \quad \alpha_i^+ \geq 0, \quad \alpha_i^- \geq 0, \quad \beta_j \geq 0 \end{aligned}$$

明らかに，$\alpha_i^- \geq \alpha_i^+$ であれば，i 期前に正のショックがあった場合より，負のショックがあった場合，t 期のボラティリティがより上昇する。TGARCH モデルにはパラメータの非負制約が存在するが，モデル式がよりシンプルになって，ショック ε_t のラグと h_t の関係が明瞭となり，取り扱いが容易となっている。

2.2.3 Regime Switching に関する問題

GARCH(p,q) モデルでは，全サンプル期間においてモデルの構造が不変である，すなわちモデルの定式化やパラメータの不変性が仮定されている。しかし，現実の金融データにおいては，構造の変化，いわゆるレジームの切り替えがないとはとても考えにくい。ボラティリティのダイナミックな動きを捉えるためには，Regime Switching をモデルに取り込む必要がある。

ボラティリティの Regime Switching モデルが多数提案された。最近，注目されているモデルの一つとして MS-GARCH(Markov-Switching GARCH) モデルがある。2 レジーム MS-GARCH(1,1) モデルの一般的な定式化は Klaassen(1999) によれば，次のようになる。

$$\begin{aligned} h_t = &[\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}]I[s_t = 1] \\ &+ [\gamma_0 + \gamma_1\varepsilon_{t-1}^2 + \delta_1 h_{t-1}]I[s_t = 2] \end{aligned}$$

ここで，s_t は 1 次のマルコフ過程である。すなわち，今のレジームの状態 s_t は 1 期前のレジームの状態 s_{t-1} にしか依存しない。レジーム s_t の推移確率は次のように定義される。

$$\begin{aligned} &P(s_t = 1|s_{t-1} = 1) = p_{11} \quad P(s_t = 2|s_{t-1} = 1) = p_{12} \\ &P(s_t = 1|s_{t-1} = 2) = p_{21} \quad P(s_t = 2|s_{t-1} = 2) = p_{22} \end{aligned}$$

この MS-GARCH(1,1) モデルのもとでは，異なる s_t の値に対して，モデルは 2 つのレジームの間で往来するようになり，実際のデータの動きをよりうまく捉えるようになる。

ただし，MS-GARCH(1,1) モデルは一つの欠点を持っている。それはレジームの切り替えをもたらす要因がモデルだけで特定できないというところにある。何らかの原因でレジームが切り替えられたという事実のみしか説明できない。従って，ボラティリティの予測などに用いることができるが，ボラティリティ変動の構造変化に関する経済学的な分析には役に立たない。

前述のように，ARCH 型モデルに関する研究は，実際のデータの動きをうまく捉えるようにモデルを改良し，ないしは多様化してきた。本稿でも，今までの多数の研究を参考にし，視点を変えながら，可能な限り実際のデータの動きをうまく捉えるように新しいモデルを構築していく。

参考文献

[1] **Barberis, N., A. Shleifer, and R. Vishny**, “A model of investor sentiment,” *Journal of Financial Economics*, 1998, *49*, 307–343.

[2] **Berndt, E., B. Hall, R. Hall, and J. Hausman**, “Estimation and Inference in Nonlinear Structural Models,” *Annals of Economic and Social Measurement*, 1974, *3*, 653–665.

[3] **Bollerslev, T.**, “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 1986, *31*, 307–327.

[4] **___ , R. Y. Chou, and K. F. Kroner**, “ARCH modelling in finance,” *Journal of Econometrics*, 1992, *52*, 5–59.

[5] **Chen, Chien-Hsun and Hui-Tzu Shih**, *The evolution of the stock market in China’s transitional economy* Advances in Chinese economic studies, Edward Elgar Publishing Limited, 2002. 37- 54.

[6] **Engle, R. F.**, “Autoregressive Conditional Heteroskedasticity with estimates of the variance of United Kingdom inflation,” *Economica*, 1982, *50*, 987–1008.

[7] **___ and T. Bollerslev**, “Modelling the persistence of conditional variances,” *Econometric Reviews*, 1986, *5*, 1–50.

[8] **___ and V. K. Ng**, “Measuring and Testing the Impact of News on Volatility,” *Journal of Finance*, 1993, *48*, 1749–1778.

[9] **Franses, Philip Hans and Dick Van Dijk**, *Non-linear time series models in empirical finance*, Cambridge University Press, 2000. 135-205.

[10] **Glosten, L. R., R. Jagannathan, and D. E. Runkle**, “On the relation between the expected value and the volatility of the nominal excess return on stocks,” *Journal of Finance*, 1993, *48*, 1779–1801.

[11] **Greene, W. H.**, *Econometric Analysis*, 4 ed., Prentice Hall Intetnational, 2000.

[12] **Hamilton, James D.**, *Time Series Analysis*, Princeton University Press, 1994. 665-667.

[13] **Kang, J., M. H. Liu, and S. X. Ni**, "Contrarian and momentum strategies in the China stock market: 1993-2000," *Pacific-Basin Finance Journal*, 2002, *10*, 243–265.

[14] **Nelson, D.B.**, "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 1990, *6*, 318–334.

[15] ___ , "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, 1991, *59*, 347–370.

[16] **Weiss, A. A.**, "Asymptotic theory for ARCH models: Estimation and testing," *Econometric Theory*, 1986, *2*, 107–131.

[17] **Zakoian, J. M.**, "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 1994, *18*, 931–955.

3 系列相関

例 3 *P205 Example 12.2*

モデル

$$y = X\beta + \varepsilon$$

強定常（Stong Statinarity） 任意の整数 $s1, s2, \cdots, sm$ に関して、確率変数 X_t の有限次元の同時分布

$$f\left(x_t, x_{t+s1}, x_{t+s2}, \cdots, x_{t+sm}\right) = f\left(x_{t'}, x_{t'+s1}, x_{t'+s2}, \cdots, x_{t'+sm}\right)$$

が成り立つ時、X_t は強定常である。

自己共分散（Autocovariance） $Cov\left(X_t, X_{t-s}\right)$

弱定常（分散共分散定常、Weak Statinarity） 任意の整数 t, t' と s に関して $Cov\left(X_t, X_{t-s}\right) = Cov\left(X_{t'}, X_{t'-s}\right)$ が成り立つ時 X_t は弱定常である。γ_s で弱定常の確率変数の自己共分散を表す。

自己相関係数（Autocorrelation） 弱定常の確率変数の自己相関係数

$$Corr\left(\varepsilon_t, \varepsilon_{t-s} | X\right) = \rho_s = \frac{\gamma_s}{\gamma_0}$$

自己相関行列 $E\left(\varepsilon\varepsilon'|X\right)$

エルゴード性（Ergodicity） 定常な確率変数の列 $\{X_t\}, t=-\infty,\cdots-2,-1,0,1,2,\cdots,\infty$, と有界な関数 $f:R^l\to R,\quad g:R^m\to R$ に関して

$$\lim_{k\to\infty}\left|E\left[f\left(x_t,x_{t+1},\cdots,x_{t+l}\right)\right]g\left(x_{t+k},x_{t+k+1},\cdots,x_{t+k+l}\right)\right|$$
$$=\left|E\left[f\left(x_t,x_{t+1},\cdots,x_{t+l}\right)\right]\right|\left|E\left[g\left(x_{t+k},x_{t+k+1},\cdots,x_{t+k+l}\right)\right]\right|$$

が成り立つとき、$\{X_t\}$ がエルゴード性を持つという。

定常性の確認は行うときがあるが、殆どの場合エルゴード性が所与であるとする。定常性やエルゴード性が様々な推定法がうまく機能することを保障するための基本的な条件である。

定理 4 (The Ergodic Theorem) $\{X_t\}_{t=-\infty}^{t=\infty}$ が定常なエルゴード性を持つであるとき、

$$\bar{x}_t\overset{a.s.}{\to}\mu$$

ただし、$\bar{x}_t=\frac{1}{n}\sum_{t=1}^{n}x_t$、$\mu=E\left(x_t\right)$。

3.1 自己相関のある撹乱項を持つモデルの OLS 推定

$$b=\left(X'X\right)^{-1}X'y$$

1. $E\left(\varepsilon|X\right)=0$ であれば不偏性を持つ。
2. $\frac{1}{n^2}X'\Omega X\overset{p}{\to}0$ であれば一致性を持つ。b の条件付分散

$$\begin{aligned}V\left(b|X\right)&=\sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}\\&=\sigma^2\left(\frac{X'X}{n}\right)^{-1}\frac{X'\Omega X}{n^2}\left(\frac{X'X}{n}\right)^{-1}\end{aligned}\tag{9}$$

   が 0 に確率収束するため、$b\overset{p}{\to}\beta$[3]。

3. 分散共分散行列の推定

$$V\left(b|X\right)=\frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\sigma^2\frac{X'\Omega X}{n}\left(\frac{X'X}{n}\right)^{-1}\tag{10}$$

[3]Chebychev's Inequality: 任意の正の数 ε に関して $P\left(|X_t-\mu|>\varepsilon\right)\leq\sigma^2/\varepsilon^2$ が成り立つ。

$\sigma^2 \frac{X'\Omega X}{n}$ を Newey-West の推定量[4]

$$S_* = S_0 + \frac{1}{n} \sum_{j=1}^{L} \sum_{t=j+1}^{n} \left(1 - \frac{j}{L+1}\right) e_t e_{t-j} \left[x_t x'_{t-j} + x_{t-j} x'_t\right]$$

$$S_* - \sigma^2 \frac{X'\Omega X}{n} \xrightarrow{p} 0 \tag{11}$$

で推定。ここでは $S_0 = \frac{1}{n}\sum_{i=1}^{n} x_i x'_i e_i^2$。$V(b|X)$ の推定量は

$$\frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} S_* \left(\frac{X'X}{n}\right)^{-1} \tag{12}$$

となる。Newey-West の推定量は正値定符号を保障する。

3.2 効率的な推定量

AR(1) 撹乱項モデルの例

$$y_t = \mu + x_t\beta + \varepsilon_t \tag{13}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

u_t が定常な white noise（自己相関がない）。$|\rho| < 1$。$Var(u_t) = \sigma_u^2$。$E(u_t) = 0$。

練習 5 上のモデルの ε_t の分散と自己共分散、自己相関係数を求めなさい。

3.2.1 GLS

$$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}y\right)$$

AR(1) の場合

$$\Omega = \frac{\sigma_u^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & & \rho^{T-2} \\ \rho^2 & \rho & 1 & & \rho^{T-3} \\ \vdots & & \vdots & \cdots & \\ \rho^{T-1} & & & & 1 \end{bmatrix}$$

Cochrane-Orcutt 法 $\hat{\rho} = \frac{n-K}{n-1}\frac{\sum_{t=2}^{n} e_t e_{t+1}}{\sum_{i=1}^{n} e_t^2}$ を利用して Ω の推定値 $\hat{\Omega}$ を求める、それを $\hat{\beta}_{GLS}$ に代入して、β, ρ, σ_u^2 を推定する、その ρ の推定値を利用して新しい $\hat{\Omega}$ を求める、それを $\hat{\beta}_{GLS}$ に代入して、β, ρ, σ_u^2 を再度推定する。この手続きを繰り返して行う。

[4] 前節紹介した White の推定量は分散不均一で自己相関がないモデルに適用できるが、自己相関がある場合は利用できない。

3.2.2 MLE

$$\ln L = -\frac{T}{2}\left(\ln 2\pi + \ln \sigma_u^2\right) - \frac{\sum_{t=1}^{n} u_t^2}{2\sigma_u^2} + \frac{1}{2}\ln\left(1-\rho^2\right)$$

ただし、$u_1 = \varepsilon_1$ とする。

3.2.3 自己相関の検定

3.2.4 ダービンーワトソン検定（Dubbin-Watson test）

$$DW = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}(\hat{u}_t)^2} \tag{14}$$

$H_0 : \rho = 0, H_1 : \rho > 0$。$DW < d_L\left(n, K\right)$ なら H_0 を棄却、$DW > d_U\left(n, K\right)$ なら H_0 を棄却しない、$d_L\left(n, K\right) < DW < d_U\left(n, K\right)$ なら決定不可能となる。$d_L\left(n, K\right), d_U\left(n, K\right)$ はサンプル数が n でパラメータの数が K のときの DW の下限と上限である。

3.2.5 ラグランジュ乗数法検定（LM test）

$$LM = n\left(\frac{e'X_0\left(X_0'X_0\right)^{-1}X_0'e}{e'e}\right) = nR_0^2 \sim \chi^2\left(1\right)$$

自由度 1 の χ^2 分布に従う。$X_0 = [X; e_{t-1}]$ で、説明変数 X の行列に残差のラグ e_{t-1} のベクトルを追加した行列となる。R_0^2 は回帰式

$$e_t = X\beta^* + e_{t-1}\tilde{\beta} + v_t$$

の決定係数である。