VII. Time-varying variances

A. Introduction to ARCH models

y_t = return on a stock in period t

μ = population mean return

$y_t = \mu + u_t$

Observation: u_t is almost impossible to predit

$$E(u_t | u_{t-1}, u_{t-2}, \ldots) = 0$$

However: u_t^2 does seem to be quite forecastable

Question 1: how should we forecast u_t^2?

One answer: autoregression on its own lagged values:

$$u_t^2 = \varsigma + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + w_t$$

$E(w_t) = 0$

$E(w_t^2) = \lambda^2$

$E(w_t w_\tau) = 0$ if $t \neq \tau$

Question 2: what kind of data-generating process would imply such a forecast?

$u_t = \sqrt{h_t}\,\varepsilon_t$

$\varepsilon_t \sim$ i.i.d. $(0,1)$ (e.g. $N(0,1)$)

$h_t = \varsigma + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2$

Definition: a regression model with Gaussian $ARCH(m)$ error is characterized by

$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$

$u_t = \sqrt{h_t}\, v_t$

$v_t \sim$ i.i.d. $N(0,1)$

$h_t = \varsigma + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2$

ARCH = autoregressive conditional heteroskedasticity

Note: even though u_t has a distribution that is conditionally Gaussian,

$$u_t | u_{t-1}, u_{t-2} \sim N(0, h_t),$$

its unconditional distribution is non-Gaussian (fatter tails)

parameters of Gaussian $ARCH(m)$
regression: $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \varsigma)'$
estimate by maximum likelihood:

$$\Omega_{t-1} = \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \mathbf{x}_{t-2}, \ldots$$
$$y_t | \Omega_{t-1} \sim N(\mathbf{x}_t'\boldsymbol{\beta}, h_t)$$
$$h_t = \varsigma + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2$$
$$u_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$$
$$f(y_t | \Omega_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{2h_t}\right]$$
$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(y_t | \Omega_{t-1}; \boldsymbol{\theta})$$

choose $\boldsymbol{\theta}$ numerically to maximize
$\mathcal{L}(\boldsymbol{\theta})$ subject to $\varsigma \geq 0,\ \alpha_j \geq 0$
(e.g., set $\alpha_j = \lambda_j^2$)
use first m values of y_t and $\mathbf{x}_t$
for conditioning

Although a Gaussian specification for v_t is natural starting point, stock returns are better modeled using a Student t

$y_t|\Omega_{t-1} \sim$ Student t with $v > 2$ degrees of freedom

conditional mean:

$$E(y_t|\Omega_{t-1}) = \mathbf{x}_t'\boldsymbol{\beta}$$

conditional variance:

$$E[(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2|\Omega_{t-1}] = h_t$$

$$\log f(y_t|\Omega_{t-1};\boldsymbol{\theta}) =$$

$$\log\left\{\frac{\Gamma[(v+1)/2]}{\sqrt{\pi}\,\Gamma(v/2)}(v-2)^{-1/2}\right\} - \frac{1}{2}\log(h_t)$$

$$-\left[\frac{(v+1)}{2}\right]\log\left[1 + \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{h_t(v-2)}\right]$$

$$h_t = \varsigma + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2$$

Issues:

(1) covariance-stationary if

$1 - \alpha_1 z - \cdots - \alpha_m z^m = 0$

implies that $\|z\| > 1$

(2) $E(u_t^2 | u_{t-1}, \ldots, u_{t-m}) > 0$

Sufficient conditions:

$\varsigma > 0$

$\alpha_j \geq 0 \quad j = 1, \ldots, m$

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

Why does the conditional variance matter?

1) knowing variance of returns is important for

a) assessing risk

b) portfolio choice

c) options pricing

2) even if you're interested in mean only, correctly modeling the variance could matter for

a) more accurate hypothesis tests

b) more efficient estimates

Hamilton, "Macroeconomics and ARCH"

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$
$$u_t \sim \text{GARCH}(1,1)$$
$$u_t = \sqrt{h_t}\, v_t$$
$$h_t = \kappa + \alpha u_{t-1}^2 + \delta h_{t-1}$$
$$v_t \sim \text{i.i.d. } N(0,1)$$

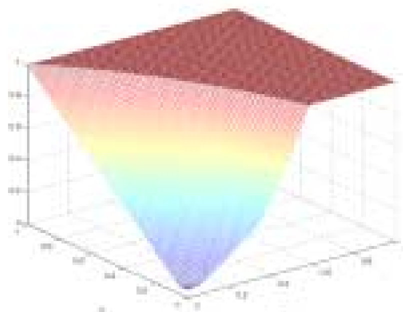
If true $\boldsymbol{\beta} = \mathbf{0}$, then

$$T^{-1/2} \sum_{t=1}^{T} y_{t-1} y_t \xrightarrow{L} N(0, E(u_{t-1}^2 u_t^2))$$
$$E(u_{t-1}^2 u_t^2) = E\{\rho E(u_t^4) + (1-\rho)[E(u_t^2)]^2\}$$
$$> E(u_{t-1}^2) E(u_t^2)$$
$$\rho = \frac{[1-(\alpha+\delta)\delta]\alpha}{1+\delta^2-2(\alpha+\delta)\delta}$$

OLS t statistic:

$$t \xrightarrow{L} N(0, V_{11})$$
$$V_{11} > 1$$
$$V_{11} \xrightarrow{p} \infty \text{ as}$$
$$3\alpha^2 + 2\alpha\delta + \delta^2 \xrightarrow{p} 1$$

Asymptotic rejection probability for OLS *t*-test that autoregressive coefficient is zero as a function of GARCH(1,1) parameters α and δ. Note: null hypothesis is actually true and test has nominal size of 5%.



Taylor rule:

$$\Delta r_t = \gamma_0 + \gamma_1 \pi_t + \gamma_2 y_t + \gamma_3 y_{t-1}$$
$$+ \gamma_r r_{t-1} + \gamma_5 \Delta r_{t-1} + v_t$$

$r_t =$ fed funds rate for quarter t

$\pi_t =$ inflation

$y_t =$ deviation of real GDP from potential

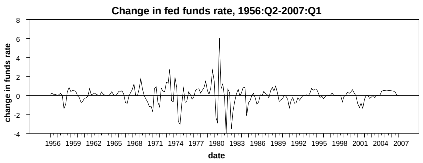
Claim: γ_1 and γ_2 are higher now than in 1970s, which contributes to greater economic stability

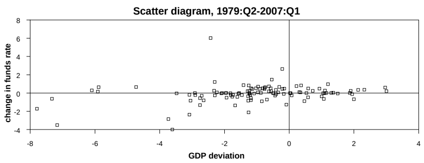
Taylor Rule with separate pre- and post-Volcker parameters as estimated by OLS regression (d_t = 1 for t > 1979:Q2).

Regressor	Coefficient	Std error (OLS)	Std error (White)
constant	0.37	0.19	0.19
π_t	0.17	0.07	0.04
y_t	0.18	0.08	0.07
y_{t-1}	-0.07	0.08	0.07
r_{t-1}	-0.21	0.07	0.06
Δr_{t-1}	0.42	0.11	0.13
d_t	-0.50	0.24	0.30
$d_t\pi_t$	0.26	0.09	0.16
$d_t y_t$	0.64	0.14	0.24
$d_t y_{t-1}$	-0.55	0.14	0.21
$d_t r_{t-1}$	0.05	0.08	0.08
$d_t\Delta r_{t-1}$	-0.53	0.13	0.24

Taylor Rule with separate pre- and post-Volcker parameters as estimated by GARCH-t maximum likelihood (d_t = 1 for t > 1979:Q2).

Regressor	Coefficient	Asymptotic std error
constant	0.13	0.08
π_t	0.06	0.03
y_t	0.14	0.03
y_{t-1}	-0.12	0.03
r_{t-1}	-0.07	0.03
Δr_{t-1}	0.47	0.09
d_t	-0.03	0.12
$d_t\pi_t$	0.09	0.04
$d_t y_t$	0.05	0.07
$d_t y_{t-1}$	0.02	0.07
$d_t r_{t-1}$	-0.01	0.03
$d_t\Delta r_{t-1}$	-0.01	0.11





VII. Time-varying variances

A. Introduction to ARCH models
B. Extensions

generalized autoregressive conditional heteroskedasticity (*GARCH*) Tim Bollerslev dissertation

$$u_t = \sqrt{h_t}\, v_t$$

$$v_t \sim (0,1)$$

$$u_t = \sqrt{h_t}\, v_t$$

$$v_t \sim (0,1)$$

$ARCH(m)$:

$$h_t = \varsigma + \alpha(L) u_t^2$$

$$\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \cdots \alpha_m L^m$$

$ARCH(\infty)$:

$$h_t = \varsigma + \pi(L) u_t^2$$

$$\pi(L) = \sum_{j=0}^{\infty} \pi_j L^j$$

parsimony:

$$\pi(L) = \frac{\alpha_1 L + \alpha_2 L^2 + \cdots \alpha_m L^m}{1 - \delta_1 L - \delta_2 L^2 - \cdots - \delta_r L^r}$$

$$(1 - \delta_1 L - \delta_2 L^2 - \cdots - \delta_r L^r) h_t$$

$$= (1 - \delta_1 - \delta_2 - \cdots - \delta_r)\varsigma$$

$$+ (\alpha_1 L + \alpha_2 L^2 + \cdots \alpha_m L^m) u_t^2$$

$$u_t \sim GARCH(r, m)$$

almost all applications use $GARCH(1,1)$

$$(1-\delta_1 L)h_t = \kappa + \alpha_1 L u_t^2$$

$$h_t = \kappa + \delta_1 h_{t-1} + \alpha_1 u_{t-1}^2$$

$$h_t = \kappa + \delta_1 h_{t-1} + \alpha_1 u_{t-1}^2$$

add u_t^2 to both sides:

$$h_t + u_t^2 = \kappa + \delta_1 u_{t-1}^2 - \delta_1 (u_{t-1}^2 - h_{t-1}) + \alpha_1 u_{t-1}^2 + u_t^2$$

$$u_t^2 = \kappa + (\delta_1 + \alpha_1) u_{t-1}^2 + (u_t^2 - h_t) - \delta_1 (u_{t-1}^2 - h_{t-1})$$

$$E(u_t^2 | u_{t-1}, u_{t-2} \dots) = h_t$$

$$w_t = u_t^2 - h_t$$

$$u_t^2 = \kappa + (\delta_1 + \alpha_1) u_{t-1}^2 + w_t - \delta_1 w_{t-1}$$

$$u_t^2 = \kappa + (\delta_1 + \alpha_1) u_{t-1}^2 + w_t - \delta_1 w_{t-1}$$

conclusion:

$u_t \sim GARCH(1,1)$

$\Rightarrow u_t^2 \sim ARMA(1,1)$

AR coefficient $= \delta_1 + \alpha_1$

MA coefficient $= -\delta_1$

stationarity requires:

$|\alpha_1 + \delta_1| < 1$

more generally:

$u_t \sim GARCH(r,m)$

$\Rightarrow u_t^2 \sim ARMA(\max\{r,m\},r)$

exponential GARCH

(EGARCH, Dan Nelson)

$u_t = \sqrt{h_t}\, v_t$

$\log h_t = \varsigma + \sum_{j=1}^{\infty} \pi_j [|v_{t-j}| - E|v_{t-j}| + \chi v_{t-j}]$

$v_t \sim$ i.i.d. $(0,1)$

$\pi_j > 0 \Rightarrow$ if $|v_{t-j}| \uparrow$, then $h_t \uparrow$

$\chi = 0 \Rightarrow$ positive v_{t-j} and negative v_{t-j} has identical effects on variance

$\log h_t = \varsigma + \sum_{j=1}^{\infty} \pi_j [|v_{t-j}| - E|v_{t-j}| + \chi v_{t-j}]$

$\chi < 0 \Rightarrow$ a decrease in stock price increases variance more than an increase in stock prices (called “leverage effect”) parsimony:

$$\pi(L) = \frac{\alpha(L)}{1 - \delta(L)}$$

EGARCH(1,1):

$$\log h_t = \kappa + \delta_1 \log h_{t-1} + \alpha_1 \{|v_{t-1}| - E|v_{t-1}| + \chi v_{t-1}\}$$

Nelson proposed generalized error distribution (GED) for v_t

$$f(v_t; \eta) = c_\eta \exp\{-(1/2)|v_t/\lambda_\eta|^\eta\}$$

where c_η and λ_η are constants to make the density integrate to 1 and have unit variance

$f(v_t;\eta) = c_\eta \exp\{-(1/2)|v_t/\lambda_\eta|^\eta\}$

$\eta = 2 \Rightarrow$

$f(v_t;\eta = 2) = c_2 \exp\{-(1/2)v_t^2/\lambda_2\}$

$\sim N(0,1)$

$\eta = 1 \Rightarrow$ double exponential

$\eta < 2 \Rightarrow$ fatter tails than Normal

$\eta > 2 \Rightarrow$ thinner tails than Normal

Application (Hamilton and Susmel):
weekly stock returns, 1962-1987

Gaussian GARCH(1,1):

$y_t = \alpha + \phi y_{t-1} + u_t$

$u_t = \sqrt{h_t}\, v_t$

$v_t \sim N(0,1)$

$h_t = a_0 + a_1 u_{t-1}^2 + b_1 h_{t-1}$

$$y_t = \underset{(0.056)}{0.427} + \underset{(0.032)}{0.281}\, y_{t-1} + u_t$$

$$h_t = \underset{(0.258)}{0.585} + \underset{(0.071)}{0.367}\, u_{t-1}^2 + \underset{(0.087)}{0.619}\, h_{t-1}$$

log likelihoods:

constant variance: -3097.2

GARCH(1,1): -2949.7

forecast comparisons:

MSE = $T^{-1}\sum_{t=1}^{T}(\hat{u}_t^2 - h_t)^2$

MAE = $T^{-1}\sum_{t=1}^{T}|\hat{u}_t^2 - h_t|$

Percent improvement of GARCH(1,1) over constant variance:

MSE = -0.14

MAE = -0.08

Possible improvements:

(1) Use Student t instead of Normal innovations

(2) Allow for asymmetry (= leverage effect)

$$y_t = \alpha + \phi y_{t-1} + u_t$$

$$u_t = \sqrt{h_t}\, v_t$$

v_t ~ Student t with v degrees of freedom and unit variance

$$h_t = a_0 + a_1 u_{t-1}^2 + \xi u_{t-1}^2 (\delta_{u_{t-1} \leq 0}) + b_1 h_{t-1}$$

v_t ~ Student t with v degrees of freedom and unit variance

$$h_t = a_0 + a_1 u_{t-1}^2 + \xi u_{t-1}^2 (\delta_{u_{t-1} \leq 0}) + b_1 h_{t-1}$$

degrees of freedom parameter:

$$\hat{v} = \underset{(0.8)}{5.6}$$

leverage parameter

$$\hat{\xi} = \underset{(0.06)}{0.23}$$

log likelihoods:

constant variance: -3097.2

Gaussian GARCH(1,1): -2949.7

Student t GARCH-L(1,1): -2890.0

Percent improvement of Student t GARCH-L(1,1) in forecasting over constant variance specification:

MSE $= -0.08$

MAE $= 0.03$

What seems to go wrong?

(1) Parameter instability– GARCH models do better fit to shorter samples

(2) When fit to longer samples, this instability shows up as too much persistence

VII. Time-varying variances

A. Introduction to ARCH models
B. Extensions
C. Markov-switching GARCH

$$
\begin{aligned}
E(u_t^2 | u_{t-1}^2, u_{t-2}^2, \ldots) &= h_t \\
&= \omega_{s_t} \\
&= \omega_0 + \omega_1 s_t \\
s_t &\in \{0, 1\}
\end{aligned}
$$

$$
\begin{aligned}
P(s_t = 0 | s_{t-1} = 0) &= q \\
P(s_t = 1 | s_{t-1} = 1) &= p \\
E(s_t | s_{t-1}, s_{t-2}, \ldots) &= (1 - q) + \lambda s_{t-1} \\
\lambda &= -1 + p + q
\end{aligned}
$$

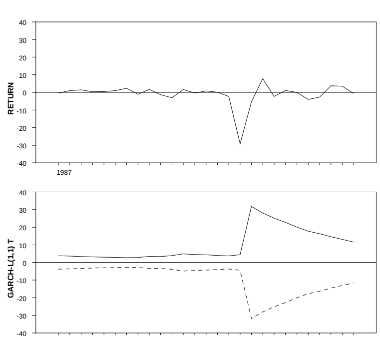
$$s_t = (1-q) + \lambda s_{t-1} + \varepsilon_t$$
$$u_t^2 = \omega_0 + \omega_1 s_t + w_t$$
$$(1-\lambda L)u_t^2 = (1-\lambda)\omega_0 + \omega_1[(1-q)+\varepsilon_t] + (1-\lambda L)w_t$$
$$u_t^2 \sim ARMA(1,1)$$

AR coefficient:

$$\lambda = -1 + p + q \approx 1$$

Estimated GARCH(1,1) parameters:

$$\hat{b}_1 = 0.619$$
$$\hat{a}_1 = 0.367$$
$$\hat{a}_1 + \hat{b}_1 = 0.986$$



Goal: parsimonious model of regime changes

$$y_t = \alpha + \phi y_{t-1} + g_{s_t} u_t$$
$$u_t = \sqrt{h_t}\, v_t$$

$v_t \sim$ Student t with ν degrees of freedom and unit variance

$$h_t = a_0 + \sum_{j=1}^{q} a_j u_{t-j}^2 + \xi u_{t-1}^2 (\delta_{u_{t-1} \leq 0})$$
$$P(s_t = j | s_{t-1} = i) = p_{ij}$$

for each $j \in \{1,2,\ldots,N\}$,

$$u_t(j) = (y_t - \alpha - \phi y_{t-1})/g_j$$
$$g_1 \equiv 1$$

for each $j_1, j_2, \ldots, j_q \in \{1,2,\ldots,N\}$,

$$h_t(j_1,j_2,\ldots,j_q) = a_0 + a_1[u_{t-1}(j_1)]^2 + a_2[u_{t-2}(j_2)]^2 + \cdots + a_q[u_{t-q}(j_q)]^2 + \xi[u_{t-1}(j_1)]^2(\delta_{u_{t-1}(j_1) \leq 0})$$

$$f(y_t | y_{t-1}, y_{t-2}, \ldots, s_t = j_0, s_{t-1} = j_1, \ldots, s_{t-q} = j_q) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi(\nu-2) g_{j_0} h_t(j_1,j_2,\ldots,j_q)}} \times \left\{1 + \frac{(y_t - \alpha - \phi y_{t-1})^2}{(\nu-2) g_{j_0} h_t(j_1,j_2,\ldots,j_q)}\right\}$$

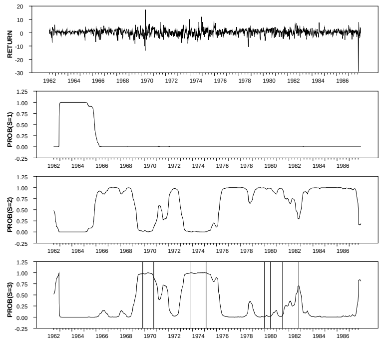
Define aggregate state s_t^* as

$$s_t^* = 1 \text{ if } s_t = 1, s_{t-1} = 1, \ldots, s_{t-q} = 1$$
$$s_t^* = 2 \text{ if } s_t = 2, s_{t-1} = 1, \ldots, s_{t-q} = 1$$
$$\vdots$$
$$s_t^* = N^* \text{ if } s_t = N, s_{t-1} = N, \ldots, s_{t-q} = N$$

Collect the densities that might be associated with each of the N^* states in an $(N^* \times 1)$ vector

$$\boldsymbol{\eta}_t = \begin{bmatrix} p(y_t|s_t^* = 1, \Omega_{t-1}) \\ p(y_t|s_t^* = 2, \Omega_{t-1}) \\ \vdots \\ p(y_t|s_t^* = N^*, \Omega_{t-1}) \end{bmatrix}$$

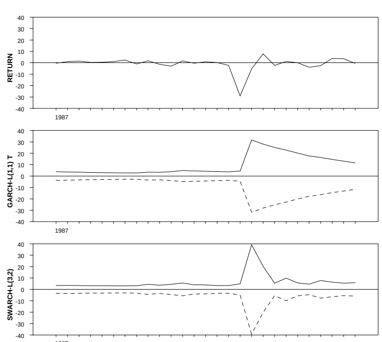
$$\frac{(\boldsymbol{\eta}_t \odot \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1})}{\mathbf{1}'(\boldsymbol{\eta}_t \odot \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1})} = \hat{\boldsymbol{\xi}}_{t|t}$$

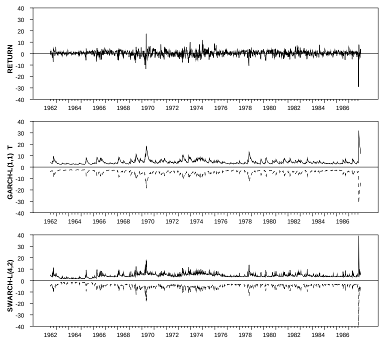


Percent improvement of Student t SWARCH-L(4,4) in forecasting over constant variance specification:

MSE = 0.06

MAE = 0.13





Why ARCH and not GARCH? Want $h_t(j_1, j_2, \ldots, j_q)$ rather than $h_t(j_1, j_2, \ldots, j_t)$

Options for Markov-switching GARCH:
(1) Gray (1996)
Replace
$$h_t = \gamma_{s_t} + \alpha_{s_t} u_{t-1}^2 + \beta_{s_t} h_{t-1}$$
with
$$h_t = \gamma_{s_t} + \alpha_{s_t} u_{t-1}^2 + \beta_{s_t} \tilde{h}_{t-1}$$
$$\tilde{h}_{t-1} = \sum_{i=1}^{N} \hat{\xi}_{i,t-1|t-2} (\gamma_i + \alpha_i y_{t-2}^2 + \beta_i \tilde{h}_{t-2})$$

Options for Markov-switching GARCH:
(2) Haas, Mittnik, and Paolella (2004)

$$h_{jt} = \gamma_j + \alpha_j u_{t-1}^2 + \beta_j h_{j,t-1}$$
$$y_t = \sqrt{h_{s_t,t}}\, u_t$$

(3) your name here (2009)

$h_t(s_t, s_{t-1}, \ldots, s_1)$

$(s_T, s_{T-1}, \ldots, s_1)$ generated as block of Gibbs sampler

problem: Even with Gaussian v_t, density of $\mathbf{s}|\mathbf{y}, \theta$ not known analytically

solution: (?) Kim, Shephard, Chib methods might work for $\mathbf{s}|\mathbf{y}, \theta$