Time-varying second moments

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 predict

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

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:

$\varsigma > 0$

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

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

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:

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

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

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

$$w_t = u_t^2 - h_t$$

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

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

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:

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

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

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"

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

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

Usual asymptotics:

$$\sqrt{T}\,(\hat{\phi} - \phi) = \frac{T^{-1/2} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-1} \sum_{t=1}^{T} y_{t-1}^2}$$
$$E(y_{t-1} u_t)^2 = E(y_{t-1}^2) E(u_t^2)$$
$$T^{-1/2} \sum_{t=1}^{T} y_{t-1} u_t \xrightarrow{L} N(0, E(y_{t-1}^2) E(u_t^2))$$
$$T^{-1} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} E(y_{t-1}^2)$$
$$\sqrt{T}\,(\hat{\phi} - \phi) \xrightarrow{L} N(0, E(u_t^2)/E(y_{t-1}^2))$$

$$\sqrt{T}\,(\hat{\phi} - \phi) \xrightarrow{L} N(0, E(u_t^2)/E(y_{t-1}^2))$$
$$\hat{\sigma}_{\hat{\phi}}^2 = s^2 / \sum_{t=1}^{T} y_{t-1}^2$$
$$T \hat{\sigma}_{\hat{\phi}}^2 \xrightarrow{p} E(u_t^2)/E(y_{t-1}^2)$$
$$t \text{ stat} \xrightarrow{L} N(0,1)$$

However, suppose true $\phi = 0$
(so $y_t = u_t$) and $u_t \sim$ GARCH$(1,1)$

$$E(y_{t-1}u_t)^2 = E(u_{t-1}^2 u_t^2)$$
$$= \rho\left\{E(u_t^4) - [E(u_t^2)]^2\right\} + [E(u_t^2)]^2$$
$$\rho = \frac{[1-(\alpha+\delta)\delta]\alpha}{1+\delta^2-2(\alpha+\delta)\delta}$$

If $\alpha = \delta = 0$ (no GARCH), then $\rho = 0$

$$E(u_{t-1}^2 u_t^2) = E(u_{t-1}^2)E(u_t^2)$$

But with GARCH,

$$E(u_{t-1}^2 u_t^2) > E(u_{t-1}^2)E(u_t^2)$$

t stat $\xrightarrow{L} N(0, V_{11})$

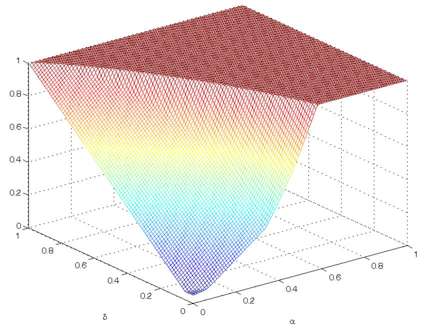
$V_{11} \geq 1$

$V_{11} \xrightarrow{p} \infty$ as

$3\alpha^2 + 2\alpha\delta + \delta^2 \xrightarrow{p} 1$

True size of usual t test > 0.05
As fourth moments become infinite, true size $\rightarrow 1$
All t tests reject the true null hypothesis asymptotically with prob 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

π_t = inflation

y_t = deviation of real GDP from potential

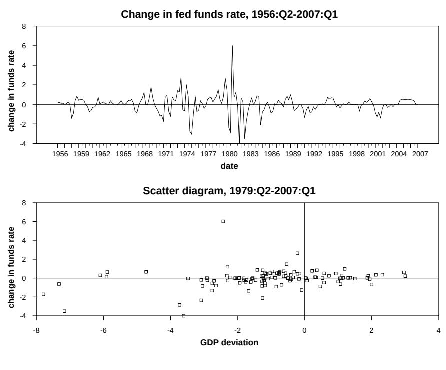
Claim: γ_1 and γ_2 are higher now than in 1970s, which contributes to greater economic stability

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



Time-varying second moments

A. Introduction to ARCH models
B. Extensions

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

GARCH-M (GARCH in mean)

Does uncertainty have effect on level of variable?

(1) higher risk $\Rightarrow$ higher expected return

(2) higher uncertainty $\Rightarrow$ macro effects

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

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

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

Time-varying second moments

A. Introduction to ARCH models
B. Extensions
C. Stochastic volatility

GARCH family:

$$
\begin{aligned}
y_t &= \mathbf{x}_t'\boldsymbol{\beta} + u_t \\
u_t &= \sqrt{h_t}\, v_t \\
v_t &\sim \text{ i.i.d. } (0,1) \quad \text{(e.g. } N(0,1)) \\
h_t &= h(u_{t-1}, u_{t-2}, \ldots)
\end{aligned}
$$

Implication:

the difference between the realized value y_t and its conditional expectation $\mathbf{x}_t'\boldsymbol{\beta}$ is the only information useful for forecasting the variance h_t

Stochastic volatility:

Some latent variables in addition to u_{t-j} contribute to h_t

Example:

$$y_t = \exp(h_t/2)v_t$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t$$

$$\begin{bmatrix} v_t \\ \eta_t \end{bmatrix} \sim \text{ i.i.d. } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

argument in favor of stochastic vol:

more natural and flexible

argument in favor of GARCH:

ultimately our forecast

$E(u_t^2|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots)$

will be some function of

$(\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots)$

so why not take this function

as a primitive of the model?

Note sv model above implies

$$y_t^2 = \exp(h_t)v_t^2$$

$$\log y_t^2 = h_t + \log v_t^2$$

$$\log y_t^2 = \mu + (h_t - \mu) + \log v_t^2$$

For $\xi_t = h_t - \mu$ this is a state-space model of the form

$$\xi_t = \phi\xi_{t-1} + \sigma\eta_t$$

$$\log y_t^2 = \mu + \xi_t + \log v_t^2$$

problem: $\log v_t^2$ is not Normally distributed

One solution: approximate with mixture of Normal state-space models, estimate by Monte Carlo Markov Chain (Kim, Shephard, and Chib REStud, 1998; Primiceri, REStud, 2005). Alternative solution: auxiliary particle filter (Chib, Nardaro, Shephard, J. Econometrics, 2002)

$\boldsymbol{\psi} = (\mu, \phi, \sigma)'$

$\Omega_t = \{y_t, y_{t-1}, .., y_1\}$

goal: approximate

$p(\xi_t | \Omega_t, \boldsymbol{\psi})$

$p(y_t | \Omega_{t-1}, \boldsymbol{\psi})$

Input for step $t+1$:

particles $\Lambda_t^{(i)} = \{\xi_t^{(i)}, \xi_{t-1}^{(i)}, \ldots, \xi_1^{(i)}\}$

for $i = 1, \ldots, D$ with weights $1/D$

(1) calculate measure of how useful $\xi_t^{(i)}$ is for predicting y_{t+1}

$\tilde{h}_{t+1}^{(i)} = \mu + \phi(h_t^{(i)} - \mu)$

$\tilde{\tau}_t^{(i)} = \frac{1}{\sqrt{2\pi \exp\left[\tilde{h}_{t+1}^{(i)}/2\right]}} \exp\left(\frac{-y_{t+1}^2}{2\exp\left[\tilde{h}_{t+1}^{(i)}/2\right]}\right)$

(2) Set $\tilde{\omega}_t^{(i)} = \frac{\tilde{\tau}_t^{(i)}}{\sum_{i=1}^{D} \tilde{\tau}_t(i)}$

and resample $\Lambda_t^{(j)}$ with prob $\tilde{\omega}_t^{(j)}$:

$$\Lambda_t^{(j)} = \begin{cases} \Lambda_t^{(1)} & \text{with probability } \tilde{\omega}_t^{(1)} \\ \vdots & \\ \Lambda_t^{(D)} & \text{with probability } \tilde{\omega}_t^{(D)} \end{cases}$$

(3) Generate $h_{t+1}^{(j)}$ from $N(\mu + \phi(h_t^{(j)} - \mu), \sigma^2)$ for $j = 1, \ldots, D$

(4) Calculate weights

$$\omega_{t+1}^{(j)} = \frac{1}{\tilde{\omega}_t^{(j)}} \frac{1}{\sqrt{2\pi \exp\left[h_{t+1}^{(j)}/2\right]}} \exp\left(\frac{-y_{t+1}^2}{2 \exp\left[h_{t+1}^{(j)}/2\right]} \right)$$

$$\hat{p}(y_{t+1}|\Omega_t; \psi) = D^{-1} \sum_{j=1}^{D} \omega_{t+1}^{(j)}$$

$$\hat{\omega}_{t+1}^{(j)} = \frac{\omega_{t+1}^{(j)}}{D^{-1} \sum_{j=1}^{D} \omega_{t+1}^{(j)}}$$

$$\hat{E}(h_{t+1}|\Omega_{t+1}; \psi) = \sum_{j=1}^{D} \hat{\omega}_{t+1}^{(j)} h_{t+1}^{(j)}$$

(5) Resample

$$\Lambda_{t+1}^{(i)} = \begin{cases} \Lambda_{t+1}^{(1)} & \text{with probability } \hat{\omega}_{t+1}^{(1)} \\ \vdots & \\ \Lambda_{t+1}^{(D)} & \text{with probability } \hat{\omega}_{t+1}^{(D)} \end{cases}$$

$$\mathcal{L}(\boldsymbol{\psi}) = \sum_{t=0}^{T-1} \log \hat{p}(y_{t+1}|\Omega_t;\boldsymbol{\psi})$$

Note structure is no more difficult for generalizations, e.g.,

$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \exp(h_t/2)v_t$

$v_t \sim$ Student t $(0, 1, \eta)$

Just replace $N(0, \exp(h_t/2))$ densities above with Student t $(\mathbf{x}_t'\boldsymbol{\beta}, \exp(h_t/2), \eta)$

Time-varying second moments

A. Introduction to ARCH models
B. Extensions
C. Stochastic volatility
D. Realized volatility

Consider continuous-time process:

$p(t) = \mu t + \sigma W(t)$

$W(t) \sim$ standard Brownian motion

e.g., $p(t) =$ log of asset price at t

$p(t) - p(t-h) \sim N(\mu h, \sigma^2 h)$

Divide interval $[t-h, t]$ into n segments each of length $\Delta = h/n$

segment i starts at $t - h + (i-1)\Delta$ and ends at $t - h + i\Delta$

segment $i = 1$: $[t-h, t-h+\Delta]$

segment $i = n$: $[t-\Delta, t]$

r_i = return over segment i

$= p(t - h + i\Delta) - p(t - h + (i-1)\Delta)$

$\sim N(\mu\Delta, \sigma^2\Delta)$

Question 1: Can we get better inference about μ by dividing fixed interval $[t-h, t]$ into smaller segments, that is, by making n bigger?

Answer: no

$\hat{\mu}_n = n^{-1}\sum_{i=1}^{n} r_i\Delta^{-1}$

Recall $r_i \sim N(\mu\Delta, \sigma^2\Delta)$ and $\Delta = h/n$

$\hat{\mu}_n = h^{-1}\sum_{i=1}^{n}[p(t-h+i\Delta) -$
$p(t-h+(i-1)\Delta)]$

$= h^{-1}[p(t) - p(t-h)]$

same estimate regardless of n

$\hat{\mu}_n \sim N(\mu, \sigma^2/h)$
unbiased but not consistent as
$n \to \infty$
To get better estimate, need longer time period (bigger h) not more observations for fixed period (bigger n)

Question 2: Can we get better inference about σ^2 by dividing fixed interval $[t-h, t]$ into smaller segments, that is, by making n bigger?
Answer: yes

Recall $r_i \sim N(\mu\Delta, \sigma^2\Delta)$ and $\Delta = h/n$

$\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n r_i^2\Delta^{-1} = h^{-1}\sum_{i=1}^n r_i^2$

$\hat{\sigma}_n^2 = h^{-1}\sum_{i=1}^n [p(t-h+i\Delta) - p(t-h+(i-1)\Delta)]^2$

$= h^{-1}\sum_{i=1}^n (\mu\Delta + \sigma\sqrt{\Delta}\,x_i)^2$

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

$\hat{\sigma}_n^2 = h^{-1} \sum_{i=1}^{n} (\mu^2 \Delta^2 + 2\mu\sigma\Delta^{3/2} x_i + \sigma^2 \Delta x_i^2)$

As $n \to \infty$,

$h^{-1} \sum_{i=1}^{n} \mu^2 \Delta^2 = h^{-1} n \Delta^2 \mu^2$

$= (h/n)\mu^2 \to 0$

$h^{-1} \sum_{i=1}^{n} 2\mu\sigma\Delta^{3/2} x_i = 2\mu\sigma(h/n)^{1/2} n^{-1} \sum_{i=1}^{n} x_i$

$\xrightarrow{p} 0$

$h^{-1} \sum_{i=1}^{n} \sigma^2 \Delta x_i^2 = \sigma^2 n^{-1} \sum_{i=1}^{n} \Delta x_i^2 \xrightarrow{p} \sigma^2$

Conclusion:

$\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$ for any h

More generally, if

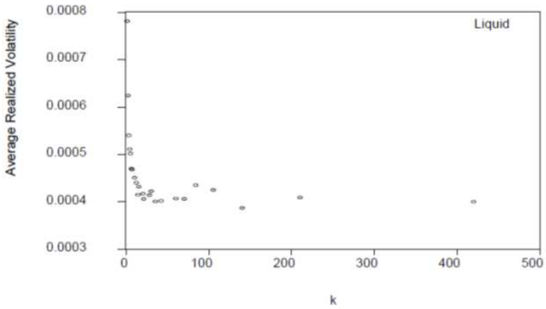
$dp(t) = \mu(t)dt + \sigma(t)dW(t)$

$\forall \xi > 0 \; \exists h > 0 :$

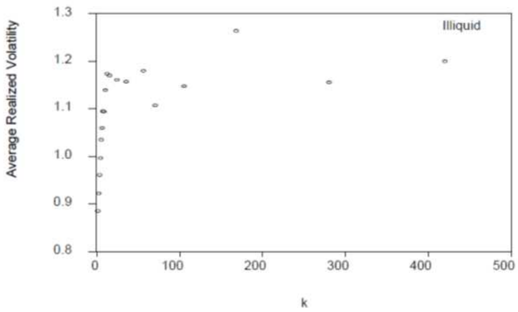
$\sup_{t-h \le \tau \le t} |\sigma^2(\tau) - \sigma^2(t)| < \xi$ (a.s.)

then $\operatorname{plim}_{n\to\infty, h\to 0} \hat{\sigma}_{n,h,t}^2 = \sigma^2(t)$

For liquid security, realized vol shoots up as Δ (measured in minutes k) gets small due to bid-ask bounce



For illiquid security, realized vol shoots up as Δ (measured in minutes k) gets big due to nontrading



Time-varying second moments

A. Introduction to ARCH models
B. Extensions
C. Stochastic volatility
D. Realized volatility
E. Dynamic conditional correlation

Question: how does GARCH generalize to multivariate setting?

Consider a collection of zero-mean GARCH(1,1) processes:

$r_{it} = \sqrt{h_{it}}\,\varepsilon_{it}$

$\varepsilon_{it} \sim$ i.i.d. $(0,1)$

$h_{it} = \omega_i + \kappa_i r^2_{i,t-1} + \lambda_i h_{i,t-1}$

$i = 1,\ldots,n$

$q_{ijt} = s_{ij} + \alpha(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - s_{ij}) + \beta(q_{ij,t-1} - s_{ij})$

$\alpha + \beta \leq 1$

If $\alpha + \beta = 1,$ amounts to forecast $\varepsilon_{it}\varepsilon_{jt}$ by exponential smoothing.

$s_{ij} = E(\varepsilon_{it}\varepsilon_{jt})$ (unconditional correlation)

$\mathbf{Q}_t = (1-\alpha-\beta)\mathbf{S} + \alpha\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} + \beta\mathbf{Q}_{t-1}$

If $\mathbf{Q}_0$ is positive definite then so is $\{\mathbf{Q}_t\}_{t=1}^{T}$

Define $\rho_{ijt} = \frac{q_{ijt}}{\sqrt{q_{iit}}\sqrt{q_{jjt}}}$

$$\mathbf{R}_t = \begin{bmatrix} \rho_{11t} & \cdots & \rho_{1nt} \\ \vdots & \cdots & \vdots \\ \rho_{n1t} & \cdots & \rho_{nnt} \end{bmatrix}$$

positive definite with ones along diagonal (a correlation matrix)

More generally, could consider

$$\mathbf{Q}_t = \mathbf{S} \circ (\mathbf{11}' - \mathbf{A} - \mathbf{B}) + \mathbf{A} \circ \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \mathbf{B} \circ \mathbf{Q}_{t-1}$$

so each correlation gets its own α_{ij}, β_{ij} instead of $\alpha_{ij} = \alpha, \beta_{ij} = \beta$.

Need $\mathbf{A}$, $\mathbf{B}$, $(\mathbf{11}' - \mathbf{A} - \mathbf{B})$ p.d.

Likelihood function for $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I}_n)$

$\Omega_t = \{\mathbf{r}_t, \mathbf{r}_{t-1}, \ldots, \mathbf{r}_1\}$

$\mathbf{r}_t|\Omega_{t-1} \sim N(\mathbf{0}, \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t)$

$\mathbf{D}_t = \text{diag}\{\sqrt{h_{it}}\}$

$h_{it} = \omega_i + \kappa_i r_{i,t-1}^2 + \lambda_i h_{i,t-1}$

$\boldsymbol{\varepsilon}_t = \mathbf{D}_t^{-1}\mathbf{r}_t$

$\mathbf{Q}_t = \mathbf{S} \circ (\mathbf{11}' - \mathbf{A} - \mathbf{B}) + \mathbf{A} \circ \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \mathbf{B} \circ \mathbf{Q}_{t-1}$

$\mathbf{Q}_t^* = \text{diag}\{\sqrt{q_{iit}}\}$

$\mathbf{R}_t = \mathbf{Q}_t^{*-1}\mathbf{Q}_t\mathbf{Q}_t^{*-1}$

$$\begin{aligned}
\mathcal{L} &= -(1/2)\sum_{t=1}^T \{n\log(2\pi) \\
&\quad + \log|\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t| + \mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}\mathbf{r}_t\} \\
&= -(1/2)\sum_{t=1}^T \{n\log(2\pi) + 2\log|\mathbf{D}_t| \\
&\quad + \log|\mathbf{R}_t| + \boldsymbol{\varepsilon}_t'\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t\} \\
&= -(1/2)\sum_{t=1}^T \{n\log(2\pi) + 2\log|\mathbf{D}_t| + \mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{D}_t^{-1}\mathbf{r}_t \\
&\qquad - \boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t + \log|\mathbf{R}_t| + \boldsymbol{\varepsilon}_t'\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t\}
\end{aligned}$$

First component:

$-(1/2)\sum_{t=1}^{T}\{n\log(2\pi)+2\log|\mathbf{D}_t|+\mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{D}_t^{-1}\mathbf{r}_t\}$

$=-(1/2)\sum_{i=1}^{n}\sum_{t=1}^{T}\{\log(2\pi)+\log(h_{it})+r_{it}^2/h_{it}\}$

$h_{it}=\omega_i+\kappa_i r_{i,t-1}^2+\lambda_i h_{i,t-1}$

can estimate $\omega_i,\kappa_i,\lambda_i$ by fitting univariate GARCH(1,1) models to series one at a time.

Second component:

$-(1/2)\sum_{t=1}^{T}\{\log|\mathbf{R}_t|+\boldsymbol{\varepsilon}_t'\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t-\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t\}$

Can maximize with respect to correlation parameters (e.g. α,β) with $\hat{\boldsymbol{\varepsilon}}_t=\hat{\mathbf{D}}_t^{-1}\mathbf{r}_t$ for $\hat{\mathbf{D}}_t$ from first step

$\hat{\mathbf{S}}=T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$

$\min_{\alpha,\beta}\sum_{t=1}^{T}\{\log|\mathbf{R}_t|+\hat{\boldsymbol{\varepsilon}}_t'\mathbf{R}_t^{-1}\hat{\boldsymbol{\varepsilon}}_t\}$

$(\hat{\alpha},\hat{\beta})$ consistent and asymptotically Normal, standard errors in Engle (2002)