VI. Time-varying variances

A. Overview

\[ y_t = \text{return on a stock in period } t \]
\[ \mu = \text{population mean return} \]
\[ y_t = \mu + u_t \]

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:

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

- $E(w_t) = 0$
- $E(w_t^2) = \lambda^2$
- $E(w_t w_{t+\tau}) = 0$ if $\tau \neq 0$

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

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

- $\varepsilon_t \sim \text{i.i.d.} \ N(0, 1)$ (e.g. $N(0, 1)$)

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

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

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

- $u_t = \sqrt{h_t} v_t$

- $v_t \sim \text{i.i.d.} \ N(0, 1)$

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

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: $\theta = (\beta', \alpha', \zeta)'$
estimate by maximum likelihood:

$$
\Omega_{t-1} = x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots
$$
$$
y_t | \Omega_{t-1} \sim N(x_t', \beta, h_t)
$$
$$
h_t = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2
$$
$$
u_t = y_t - x_t' \beta
$$
$$
f(y_t | \Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi h_t}} \exp \left[ -\frac{(y_t - x_t' \beta)^2}{2h_t} \right]
$$
$$
\mathcal{L}(\theta) = \sum_{t=1}^{T} \log f(y_t | \Omega_{t-1}; \theta)
$$
choose \( \theta \) numerically to maximize \( \mathcal{L}(\theta) \) subject to \( \zeta \geq 0, \alpha_j \geq 0 \) (e.g., set \( \alpha_j = \lambda_j^2 \)) use first \( m \) values of \( y_t \) and \( x_t \) for conditioning

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

\[ y_t | \Omega_{t-1} \sim \text{Student } t \text{ with } \nu > 2 \text{ degrees of freedom} \]

conditional mean:

\[ E(y_t | \Omega_{t-1}) = x_t \beta \]

conditional variance:

\[ E[(y_t - x_t \beta)^2 | \Omega_{t-1}] = h_t \]
\[
\log f(y_t|\Omega_{t-1};\theta) = \\
\log \left\{ \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi \Gamma(\nu/2)}} (\nu - 2)^{-1/2} \right\} - \frac{1}{2} \log(h_t) \\
-\left[ \frac{\nu + 1}{2} \right] \log \left[ 1 + \frac{(y_t - x_t \beta)^2}{h_t (\nu - 2)} \right] \\
h_t = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2
\]

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:
\[
\zeta > 0 \\
\alpha_j \geq 0 \quad j = 1, \ldots, m \\
\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1
\]
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):

\[ h_t = \alpha + \alpha(L)u_t^2 \]

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

ARCH(\infty):

\[ h_t = \alpha + \gamma(L)u_t^2 \]

\[ \gamma(L) = \gamma_1 L + \gamma_2 L^2 + \cdots + \gamma_m L^m \]

parsimony:

\[ \pi(L) = \frac{\alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_m L^m}{1 - \delta_1 L - \delta_2 L^2 - \cdots - \delta_r L^r} \]
\[(1 - \delta_1 L - \delta_2 L^2 \cdots - \delta_r L^r) h_t = (1 - \delta_1 - \delta_2 \cdots - \delta_r) \xi + (\alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_m L^m) u_t^2 \]
\[u_t \sim \text{GARCH}(r, m)\]

almost all applications use \(\text{GARCH}(1, 1)\)
\[(1 - \delta_1 L) h_t = \kappa + \alpha_1 Lu_t^2\]
\[h_t = \kappa + \delta_1 h_{t-1} + \alpha_1 u_{t-1}^2\]

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

add \(u_t^2\) to both sides:
\[h_t + u_t^2 = \kappa + \delta_1 u_{t-1}^2 - \delta_1 (u_{t-1}^2 - h_{t-1}) + \alpha_1 u_{t-1}^2 + u_t^2\]
\[u_t^2 = \kappa + (\delta_1 + \alpha_1) u_{t-1}^2 + (u_t^2 - h_t) - \delta_1 (u_{t-1}^2 - h_{t-1})\]
\[E(u_t^2 | u_{t-1}, u_{t-2}, \ldots) = h_t\]
\[w_t = u_t^2 - h_t\]
\[u_t^2 = \kappa + (\delta_1 + \alpha_1) u_{t-1}^2 + w_t - \delta_1 w_{t-1}\]
\[ u_t^2 = \kappa + (\delta_1 + \alpha_1)u_{t-1}^2 + w_t - \delta_1w_{t-1} \]

**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”

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

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

However, suppose true \( \phi = 0 \)
(so \( y_t = u_t \)) and \( u_t \sim \text{GARCH}(1, 1) \)
\[
E(y_{t-1} u_t)^2 = E(u_{t-1}^2 u_t^2)
\]
\[
= \rho \left\{ E(u_t^4) - [E(u_t^2)]^2 \right\} + [E(u_t^2)]^2
\]
\[
\rho = \frac{[1-(\alpha+\delta)\delta] \alpha}{1+\delta^2-2(\alpha+\delta)\delta}
\]
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,
\[
E(u_{t-1}^2 u_t^2) > E(u_{t-1}^2) E(u_t^2)
\]
\[
t \text{ stat} \xrightarrow{L} N(0, V_{11})
\]
\[
V_{11} \geq 1
\]
\[
V_{11} \xrightarrow{p} \infty \text{ as}
\]
\[
3\alpha^2 + 2\alpha\delta + \delta^2 \xrightarrow{p} 1
\]
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 GLS $t$ test that autoregressive coefficient is zero as a function of GARCH(1,1) parameters $\alpha$ and $\delta$. Note: null hypothesis is actually true and test has nominal size of 5%.

Taylor rule:

$$\Delta r_t = \gamma_0 + \gamma_1 \pi_t + \gamma_2 y_t + \gamma_3 y_{t-1} + \gamma_4 r_{t-1} + \gamma_5 \Delta r_{t-1} + \nu_t$$

$r_t$ = fed funds rate for quarter $t$
$\pi_t$ = inflation
$y_t$ = deviation of real GDP from potential
Claim: $\gamma_1$ and $\gamma_2$ are higher now than in 1970s, which contributes to greater economic stability

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Coefficient</th>
<th>Std error (OLS)</th>
<th>Std error (White)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>0.37</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>$\pi_t$</td>
<td>0.17</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>$y_t$</td>
<td>0.18</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>$y_{t-1}$</td>
<td>0.42</td>
<td>0.31</td>
<td>0.13</td>
</tr>
<tr>
<td>$r_{t-1}$</td>
<td>-0.07</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta r_{t-1}$</td>
<td>-0.21</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>$d_t$</td>
<td>-0.50</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>$d_t \pi_t$</td>
<td>0.26</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>$d_t y_t$</td>
<td>0.05</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>$d_t y_{t-1}$</td>
<td>0.05</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>$d_t r_{t-1}$</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>$d_t \Delta r_{t-1}$</td>
<td>0.53</td>
<td>0.13</td>
<td>0.24</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Coefficient</th>
<th>Asymptotic std error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td>$\pi_t$</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>$y_t$</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>$y_{t-1}$</td>
<td>-0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>$r_{t-1}$</td>
<td>-0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>$\Delta r_{t-1}$</td>
<td>0.42</td>
<td>0.09</td>
</tr>
<tr>
<td>$d_t$</td>
<td>-0.03</td>
<td>0.12</td>
</tr>
<tr>
<td>$d_t \pi_t$</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>$d_t y_t$</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>$d_t y_{t-1}$</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>$d_t r_{t-1}$</td>
<td>-0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>$d_t \Delta r_{t-1}$</td>
<td>-0.01</td>
<td>0.11</td>
</tr>
</tbody>
</table>

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

A. Overview

B. Extensions

1. Exponential GARCH (EGARCH)

\[ u_t = \sqrt{h_t} \ v_t \]

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

\[ v_t \sim \text{i.i.d.} \ (0, 1) \]

\[ \pi_j > 0 \Rightarrow |v_{t-j}| \uparrow, \text{ then } h_t \uparrow \]

\[ x = 0 \Rightarrow \text{positive } v_{t-j} \text{ and negative } v_{t-j} \text{ has identical effects on variance} \]
\[ \log h_t = \zeta + \sum_{j=1}^{\infty} \pi_j [v_{t-j}] - E[v_{t-j}] + \chi v_{t-j} ] \]

\( \chi < 0 \Rightarrow \) 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 $c_\eta$ and $\lambda_\eta$ 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

2. 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 \(i = n\): \([t-\Delta, t]\)

\(r_i = \text{return over segment } i\)
\[= p(t-h+i\Delta) - p(t-h+(i-1)\Delta)\]
\[\sim N(\mu \Delta, \sigma^2 \Delta)\]

Question 1: Can we get better inference about \(\mu\) 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 \( \Delta = h/n \)

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

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

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 \( \sigma^2 \) 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 $\Delta = h/n$
\[
\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n r_i^2 \Delta^{-1} = h^{-1} \sum_{i=1}^n r_i^2
\]
\[
\hat{\sigma}_n^2 = h^{-1} \sum_{i=1}^n [p(t - h + i\Delta) - p(t - h + (i - 1)\Delta)]^2
\]
\[
= h^{-1} \sum_{i=1}^n (\mu \Delta + \sigma \sqrt{\Delta} x_i)^2
\]
$x_i \sim \text{i.i.d. } N(0, 1)$

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:
\[
\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2 \text{ as } n \to \infty \text{ for any } h
\]
More generally, if 
\[ dp(t) = \mu(t)dt + \sigma(t)dW(t) \]
\[ \forall \xi > 0 \exists h > 0 : \]
\[ \sup_{t-h \leq \tau \leq t} |\sigma^2(\tau) - \sigma^2(t)| < \xi \quad (a.s.) \]
then \[ \lim_{n \to \infty, h \to 0} \hat{\sigma}_{n,h,t}^2 = \sigma^2(t) \]

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

![Graph 1](image)

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

![Graph 2](image)
3. Dynamic conditional correlation
Consider a collection of zero-mean GARCH(1,1) processes:
\[ r_{it} = \sqrt{h_{it}} \varepsilon_{it} \]
\[ \varepsilon_{it} \sim \text{i.i.d. } (0, 1) \]
\[ h_{it} = \omega_i + \kappa_i r_{i,t-1}^2 + \lambda_i h_{i,t-1} \]
\[ i = 1, \ldots, n \]

\[ q_{ijt} = s_{ij} + \alpha (r_{i,t-1} r_{j,t-1} - s_{ij}) + \beta (q_{ij,t-1} - s_{ij}) \]
\[ \alpha + \beta \leq 1 \]
If \( \alpha + \beta = 1 \), amounts to forecast \( \varepsilon_{it} \varepsilon_{jt} \)
by exponential smoothing.
\[ s_{ij} = E(\varepsilon_{it} \varepsilon_{jt}) \quad \text{(unconditional correlation)} \]
\[ Q_t = (1 - \alpha - \beta)S + \alpha s_{ij} \varepsilon_{i,t-1} + \beta Q_{t-1} \]
If \( Q_0 \) is positive definite then so is \( \{Q_t\}_{t=1}^T \)

Define \( \rho_{ijt} = \frac{q_{ijt}}{\sqrt{q_{ii}q_{jj}}} \)
\[ R_t = \begin{bmatrix} \rho_{11t} & \cdots & \rho_{1nt} \\ \vdots & \cdots & \vdots \\ \rho_{nt} & \cdots & \rho_{nnt} \end{bmatrix} \]
positive definite with ones along diagonal (a correlation matrix)
More generally, could consider
\[
Q_t = S \circ (11' - A - B) \\
+ A \circ \varepsilon_{t-1} \varepsilon_{t-1}' + B \circ Q_{t-1}
\]
so each correlation gets its own 
\(\alpha_{ij}, \beta_{ij}\) instead of \(\alpha_{ij} = \alpha, \beta_{ij} = \beta\).

Likelihood function for \(\varepsilon_i \sim N(0, I_n)\)
\[
\Omega = \{r, \varepsilon_{t-1}, \ldots, r_t\} \\
r \|= \Omega_{t-1} \sim N(0, D_tR_tD_t) \\
D_t = \text{diag} \{\sqrt{h_{ii}}\} \\
h_{ii} = \omega_i + \kappa_i r_{t-1}^r + \lambda_i h_{i, t-1} \\
\varepsilon_i = D_t^{-1} r_t \\
Q_t = S \circ (11' - A - B) + A \circ \varepsilon_{t-1} \varepsilon_{t-1}' + B \circ Q_{t-1} \\
Q_t^* = \text{diag} \{\sqrt{q_{ii}}\} \\
R_t = Q_t^{-1} Q_t^{-1}
\]

\[
\mathcal{L} = -(1/2) \sum_{t=1}^T \{n \log(2\pi) \\
+ \log|D_tR_tD_t' + r_t' D_t'^{-1} R_t'^{-1} D_t^{-1} r_t\} \\
= -(1/2) \sum_{t=1}^T \{n \log(2\pi) + 2 \log|D_t| \\
+ \log|R_t| + \varepsilon_i' R_t^{-1} \varepsilon_i\} \\
= -(1/2) \sum_{t=1}^T \{n \log(2\pi) + 2 \log|D_t| + r_t' D_t'^{-1} D_t^{-1} r_t \\
- \varepsilon_i' \varepsilon_i + \log|R_t| + \varepsilon_i' R_t^{-1} \varepsilon_i\}
\]
First component:
\[-(1/2) \sum_{t=1}^T \left\{ n \log(2\pi) + 2 \log|\mathbf{D}_t| + \mathbf{r}_t' \mathbf{D}_t^{-1} \mathbf{r}_t \right\} \]
\[= -(1/2) \sum_{i=1}^n \sum_{t=1}^T \{ \log(2\pi) + \log(h_{ii}) + r_{ti}^2 / h_{ii} \} \]
\[h_{ii} = \omega_i + \kappa_i r_{ti-1}^2 + \lambda_i h_{i,i-1} \]
can estimate \( \omega_i, \kappa_i, \lambda_i \) by fitting univariate GARCH(1,1) models to series one at a time.

Second component:
\[-(1/2) \sum_{t=1}^T \{ \log|\mathbf{R}_t| + \mathbf{e}_t' \mathbf{R}_t^{-1} \mathbf{e}_t - \mathbf{e}_t' \mathbf{e}_t \} \]
Can maximize with respect to correlation parameters (e.g. \( \alpha, \beta \))
with \( \hat{\mathbf{e}}_i = \hat{\mathbf{D}}_t^{-1} \mathbf{r}_t \) for \( \hat{\mathbf{D}}_t \) from first step
\[\hat{\mathbf{S}} = T^{-1} \sum_{t=1}^T \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \]

\[\min_{\alpha, \beta} \sum_{t=1}^T \{ \log|\mathbf{R}_t| + \hat{\mathbf{e}}_t' \mathbf{R}_t^{-1} \hat{\mathbf{e}}_t \} \]
(\( \hat{\alpha}, \hat{\beta} \)) consistent and asymptotically Normal, standard errors in Engle (2002)
VII. Time-varying variances

A. Overview
B. Extensions
C. Markov-switching GARCH

Options for Markov-switching GARCH:
Replace
\[ h_t = \gamma_s + \alpha_s u_{t-1}^2 + \beta_s h_{t-1} \]
with
\[ h_t = \gamma_s + \alpha_s u_{t-1}^2 + \beta_s \tilde{h}_{t-1} \]
\[ \tilde{h}_{t-1} = \sum_{i=1}^{N} \hat{\epsilon}_{t-i-1} \gamma_i + \alpha_i \tilde{y}_{t-2} + \beta_i \tilde{h}_{t-2} \]

Options for Markov-switching GARCH:
(2) Haas, Mittnik, and Paolella
\[ h_{jt} = \gamma_j + \alpha_j u_{t-1}^2 + \beta_j h_{jt-1} \]
\[ y_t = \sqrt{h_{s_i,t}} u_t \]
Options for Markov-switching GARCH:
(3) Bauwens, Preminger, and Rombouts, 
(\textit{Econometrics Journal}, 2010)-- numerical Bayesian methods

VI. Time-varying variances
A. Overview
B. Extensions
C. Markov-switching GARCH
D. Stochastic volatility

GARCH family:
\[ y_t = x_t^\prime \beta + u_t \]
\[ u_t = \sqrt{h_t} \ v_t \]
\[ v_t \sim \text{i.i.d.} \ (0, 1) \ (\text{e.g. } N(0, 1)) \]
\[ h_t = h(u_{t-1}, u_{t-2}, \ldots) \]
Implication:
the difference between
the realized value $y_t$ and its
conditional expectation $x_t$ 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
\[ E(u_t^2|x_t, y_{t-1}, x_{t-1}, y_{t-2}, \ldots) \]
will be some function of
\( \langle x_t, y_{t-1}, x_{t-1}, y_{t-2}, \ldots \rangle \)
so why not take this function
as a primitive of the model?

Note sv model above implies
\[ y_t^2 = \exp(v_t^2) \]
\[ \log y_t^2 = h_t + \log v_t^2 \]
\[ \log y_t^2 = \mu + (h_t - \mu) + \log v_t^2 \]

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
solution: auxiliary particle filter
\[ \psi = (\mu, \phi, \sigma)' \]
\[ \Omega_t = \{y_t, y_{t-1}, \ldots, y_1\} \]
goal: approximate
\[ p(\xi_t|\Omega_t, \psi) \]
\[ p(y_t|\Omega_{t-1}, \psi) \]

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 \( y_{t+1} \)
\[ \hat{h}_{t+1}^{(i)} = \mu + \phi(h_t^{(i)} - \mu) \]
\[ \hat{\xi}_t^{(i)} = \frac{1}{\sqrt{2\pi \exp[h_{t+1}^{(i)}/2]}} \exp\left(\frac{-y_{t+1}^2}{2\exp[h_{t+1}^{(i)}/2]}\right) \]
(2) Set $\tilde{\omega}_t^{(i)} = \frac{\xi_t^{(i)}}{\sum_{i=1}^{D} \xi_t^{(i)}}$
and resample $\Lambda_t^{(i)}$ with prob $\tilde{\omega}_t^{(i)}$:
$\Lambda_t^{(i)} = \begin{cases} 
\Lambda_t^{(i)} & \text{with probability } \tilde{\omega}_t^{(i)} \\
\vdots \\
\Lambda_t^{(D)} & \text{with probability } \tilde{\omega}_t^{(D)} 
\end{cases}$

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

(4) Calculate weights
\[
\omega_{t+1}^{(j)} = \frac{1}{\tilde{\omega}_t^{(j)}} \frac{1}{\sqrt{2\pi \exp[\lambda_t^{(j)/2}]}} \exp\left(\frac{-\gamma_t^{(j)}}{2\exp[\lambda_t^{(j)/2}]}\right)
\]
\[
\hat{p}(y_{t+1}|\Omega_t; \psi) = D^{-1} \sum_{j=1}^{D} \omega_{t+1}^{(j)}
\]
\[
\tilde{\omega}_{t+1}^{(j)} = \frac{\omega_{t+1}^{(j)}}{D^{-1} \sum_{j=1}^{D} \omega_{t+1}^{(j)}}
\]
\[
\tilde{E}(h_{t+1}|\Omega_{t+1}; \psi) = \sum_{j=1}^{D} \tilde{\omega}_{t+1}^{(j)} h_{t+1}^{(j)}
\]
(5) Resample
\[ \Lambda_{t+1}^{(i)} = \begin{cases} 
\Lambda_{t+1}^{(1)} \text{ with probability } \hat{\phi}_{t+1}^{(1)} \\
\vdots \\
\Lambda_{t+1}^{(D)} \text{ with probability } \hat{\phi}_{t+1}^{(D)} 
\end{cases} \]

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

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

\[ y_t = x_t' \beta + \exp(h_t/2)v_t \]

\[ v_t \sim \text{Student } t \left( 0, 1, \eta \right) \]

Just replace \( N(0, \exp(h_t/2)) \)
densities above with
Student \( t \left( x_t' \beta, \exp(h_t/2), \eta \right) \)
Alternatively, other tricks could allow us to use linear state-space methods as building blocks.

Return to original problem:
\[ \xi_t = \phi \xi_{t-1} + \sigma \eta_t \]
\[ \log y_t^2 = \mu + \xi_t + \log v_t^2 \]
\[ v_t \sim N(0, 1) \]
\[ \log v_t^2 = \log \chi^2(1) \]

\[ z_t = \log v_t^2 = \log \chi^2(1) \]
can approximate this density arbitrarily well with a mixture of Normals.
\[ p(z_t) = \sum_{i=1}^{K} \frac{\pi_i}{\sqrt{2\pi \tau_i}} \exp\left[-\frac{(z_t - \delta_i)^2}{2\tau_i^2}\right] \]

\( K = 7 \) gives excellent approximation values of \( \pi_i, \tau_i, \delta_i \) are numerically known (not a function of data, instead function of \( \log[\chi^2(1)] \) distribution)

We could generate a value for \( z_t \) from this distribution in two steps:

- **step 1:** generate \( s_t \in \{1, 2, \ldots, K\} \)
  
  \[ \text{Prob}(s_t = i) = \pi_i \]

- **step 2:** generate \( z_t|s_t \sim N(\delta_{s_t}, \tau_{s_t}^2) \)

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

log \( y_t^2 = \mu + \xi_t + z_t \)

\( z_t|s_t \sim N(\delta_{s_t}, \tau_{s_t}^2) \)

Conditional on \( s = (s_1, \ldots, s_T)' \) this is Gaussian linear state-space model
\[ \psi = (\mu, \phi, \sigma)' \]
Sampling from \( p(y|\xi, s, \psi) \)
or \( p(\xi|y, s, \psi) \) are standard

\[
p(s|y, \xi, \psi) = \prod_{t=1}^{T} p(s_t|z_t) \\
p(s_t = j|z_t) = \frac{(\pi_t/j_t) \exp \left[ -\frac{(y_t-j)^2}{2\sigma^2} \right]}{\sum_{j=1}^{K} (\pi_t/j_t) \exp \left[ -\frac{(y_t-j)^2}{2\sigma^2} \right]} \\
(\text{recall } \pi_j, \delta_j, \tau_j \text{ are all known properties of } \log \chi^2(1) \text{ distribution})
\]

Summary of linear Gaussian representation:
\[ \xi_t = \phi \xi_{t-1} + \sigma \eta_t \]
\[ y^*_t = \mu + \xi_t + z_t \]
\[ z_t \sim N(0, \tau^2_{x_t}) \]
\[ y^*_t = \log y^2_t - \delta_{s_t} \]