GARCH Models

Introduction

• ARMA models assume a constant volatility

• In finance, correct specification of volatility is essential

• ARMA models are used to model the conditional expectation

• They write $Y_t$ as a linear function of the past plus a white noise term
Absolute changes in weekly AAA rate
Cree Daily Returns
Cree Daily Returns
● GARCH — models of nonconstant volatility
● ARCH = AutoRegressive Conditional Heteroscedasticity
● heteroscedasticity = non-constant variance
● homoscedasticity = constant variance
• ARMA ⇒
  – unconditionally homoscedastic
  – conditionally homoscedastic

• GARCH ⇒
  – unconditionally homoscedastic, but
  – conditionally heteroscedastic

• Unconditional or marginal distribution of $R_t$ means the distribution when none of the other returns are known.
Modeling conditional means and variances

- **Idea:** If $\epsilon$ is $N(0, 1)$, and $Y = a + b\epsilon$, then $E(Y) = a$ and $\text{Var}(Y) = b^2$.

- general form for the regression of $Y_t$ on $X_{1,t}, \ldots, X_{p,t}$ is

$$Y_t = f(X_{1,t}, \ldots, X_{p,t}) + \epsilon_t \quad (1)$$

- Frequently, $f$ is linear so that

$$f(X_{1,t}, \ldots, X_{p,t}) = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_p X_{p,t}.$$

- **Principle:** To model the conditional mean of $Y_t$ given $X_{1,t}, \ldots, X_{p,t}$, write $Y_t$ as the conditional mean plus white noise.
• Let $\sigma^2(X_{1,t}, \ldots, X_{p,t})$ be the conditional variance of $Y_t$ given $X_{1,t}, \ldots, X_{p,t}$. Then the model

$$Y_t = f(X_{1,t}, \ldots, X_{p,t}) + \sigma(X_{1,t}, \ldots, X_{p,t})\epsilon_t$$  \hspace{1cm} (2)

gives the correct conditional mean and variance.

• **Principle:** To allow a nonconstant conditional variance in the model, **multiply** the white noise term by the conditional standard deviation. This product is added to the conditional mean as in the previous principle.

• $\sigma(X_{1,t}, \ldots, X_{p,t})$ must be non-negative since it is a standard deviation
ARCH(1) processes

- Let $\epsilon_1, \epsilon_2, \ldots$ be Gaussian white noise with unit variance, that is, let this process be independent $N(0,1)$.

- Then

  $$E(\epsilon_t|\epsilon_{t-1}, \ldots) = 0,$$

  and

  $$\text{Var}(\epsilon_t|\epsilon_{t-1}, \ldots) = 1. \quad (3)$$

- Property (3) is called conditional homoscedasticity.
\[ a_t = \epsilon_t \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2}. \]  

(4)

• It is required that \( \alpha_0 \geq 0 \) and \( \alpha_1 \geq 0 \)

• It is also required that \( \alpha_1 < 1 \) in order for \( a_t \) to be stationary with a finite variance.

• If \( \alpha_1 = 1 \) then \( a_t \) is stationary, but its variance is \( \infty \)

• Define

\[ \sigma_t^2 = \text{Var}(a_t|a_{t-1}, \ldots) \]
From previous slide:

\[ a_t = \epsilon_t \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2}. \]

Therefore

\[ a_t^2 = \epsilon_t^2 \{\alpha_0 + \alpha_1 a_{t-1}^2\}. \]

- Since \( \epsilon_t \) is independent of \( a_{t-1} \) and \( \text{Var}(\epsilon_t) = 1 \)

\[ E(a_t|a_{t-1},\ldots) = 0, \quad (5) \]

and

\[ \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2. \quad (6) \]
From previous slide:

\[ \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2. \]

- If \( a_{t-1} \) has an unusually large deviation
  - then the conditional variance of \( a_t \) is larger than usual
  - \( a_t \) is also expected to have an unusually large deviation
  - volatility will propagate since \( a_t \) having a large deviation makes \( \sigma_{t+1}^2 \) large so that \( a_{t+1} \) will tend to be large.
• The conditional variance tends to revert to the unconditional variance provided that $\alpha_1 < 1$ so that the process is stationary with a finite variance.

• The unconditional, i.e., marginal, variance of $a_t$ denoted by $\gamma_a(0)$
• The basic ARCH(1) equation is
\[ \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2. \] (7)

This gives us
\[ \gamma_a(0) = \alpha_0 + \alpha_1 \gamma_a(0). \]

• This equation has a positive solution if \( \alpha_1 < 1 \):
\[ \gamma_a(0) = \alpha_0/(1 - \alpha_1). \]
• If $\alpha_1 = 1$ then $\gamma_a(0)$ is infinite.
  – It turns out that $a_t$ is stationary nonetheless.
For an ARCH(1) process with $\alpha_1 < 1$:

$$\text{Var}(a_{t+k}|a_t, a_{t-1}, \ldots) = \gamma(0) + \alpha_1^k\{a_t^2 - \gamma(0)\} \to \gamma(0) \text{ as } k \to \infty.$$ 

In contrast, for any ARMA process:

$$\text{Var}(a_{t+k}|a_t, a_{t-1}, \ldots) = \gamma(0).$$
\[
\text{Var}(a_{t+k} \mid a_t, \ldots) \text{ for AR(1) and ARCH(1)}. \quad \gamma(0) = 1 \text{ in both cases. For ARCH(1), } \alpha_1 = .9. \quad \text{Case 1:} \\
\quad a_t^2 = 1.5. \quad \text{Case 2:} \quad a_t^2 = .5.
\]
• independence implies zero correlation but not vice versa
  – GARCH processes are good examples
  – dependence of the conditional variance on the past is the reason the process is not independent
  – independence of the conditional mean on the past is the reason that the process is uncorrelated
Example:

- $\alpha_0 = 1, \alpha_1 = .95, \mu = .1, \text{ and } \phi = .8$
Parameters: $\alpha_0 = 1$, $\alpha_1 = .95$, $\mu = .1$, and $\phi = .8$. 
Comparison of AR(1) and ARCH(1)

**AR(1)**

\[ Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t. \]

**ARCH(1)**

\[ a_t = \epsilon_t\sigma_t. \]

\[ \sigma_t = \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2}. \]
Comparison of AR(1) and ARCH(1)

**AR(1)**

\[
E(Y_t) = \mu.
\]

\[
E_t(Y_t) = \mu + \phi(Y_{t-1} - \mu).
\]

**ARCH(1)**

\[
E(a_t) = 0.
\]

\[
E_t(a_t) = 0.
\]
Comparison of AR(1) and ARCH(1)

AR(1)

\[ \sigma^2_t = \sigma^2. \]

ARCH(1)

\[ \sigma^2 = \frac{\alpha_0}{1 - \alpha_1}. \]
\[ \sigma^2_t = \alpha_0 + \alpha_1 a^2_{t-1}. \]

Recall: \( \sigma^2_t = \text{Var}(a_t|a_{t-1},\ldots). \)
The AR(1)/ARCH(1) model

• Let $a_t$ be an ARCH(1) process

• Suppose that

$$u_t - \mu = \phi(u_{t-1} - \mu) + a_t.$$ 

• $u_t$ looks like an AR(1) process, except that the noise term is not independent white noise but rather an ARCH(1) process.
• $a_t$ is not independent white noise but is uncorrelated
  – Therefore, $u_t$ has the same ACF as an AR(1) process:
    \[ \rho_u(h) = \phi^{|h|} \quad \forall \ h. \]
  – $a^2_t$ has the ARCH(1) ACF:
    \[ \rho_{a^2}(h) = \alpha_1^{|h|} \quad \forall \ h. \]
  – need to assume that both $|\phi| < 1$ and $\alpha_1 < 1$ in order for $u$ to be stationary with a finite variance
ARCH\((q)\) models

• let \(\epsilon_t\) be Gaussian white noise with unit variance
• \(a_t\) is an ARCH\((q)\) process if

\[
a_t = \sigma_t \epsilon_t
\]

and

\[
\sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2}
\]
**GARCH\((p, q)\) models**

- The GARCH\((p, q)\) model is

  \[ a_t = \epsilon_t \sigma_t \]

- Where

  \[ \sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2}. \]

- Very general time series model:
  - \(a_t\) is GARCH\((p_G, q_G)\) and
  - \(a_t\) is the noise term in an ARIMA\((p_A, d, q_A)\) model
Heavy-tailed distributions

- stock returns have "heavy-tailed" or "outlier-prone" distributions
- reason for the outliers may be that the conditional variance is not constant
- GARCH processes exhibit heavy-tails
- Example — 90% $N(0, 1)$ and 10% $N(0, 25)$
- variance of this distribution is $(.9)(1) + (.1)(25) = 3.4$
  - standard deviation is 1.844
- distribution is **MUCH** different than a $N(0, 3.4)$ distribution
Comparison on normal and heavy-tailed distributions.
• For a $N(0, \sigma^2)$ random variable $X$,

$$P\{|X| > x\} = 2(1 - \Phi(x/\sigma)).$$

• Therefore, for the normal distribution with variance 3.4,

$$P\{|X| > 6\} = 2(1 - \Phi(6/\sqrt{3.4})) = .0011.$$
• For the normal mixture population which has variance 1 with probability .9 and variance 25 with probability .1 we have that

\[ P\{|X| > 6\} = 2\{.9(1 - \Phi(6)) + .1(1 - \Phi(6/5))\} \]

\[ = (.9)(0) + (.1)(.23) = .023. \]

• Since .023/.001 \approx 21, the normal mixture distribution is 21 times more likely to be in this outlier range than the normal distribution.
<table>
<thead>
<tr>
<th>Property</th>
<th>Gaussian WN</th>
<th>ARMA</th>
<th>GARCH</th>
<th>ARMA/GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. mean</td>
<td>constant</td>
<td>non-const</td>
<td>0</td>
<td>non-const</td>
</tr>
<tr>
<td>Cond. var</td>
<td>constant</td>
<td>constant</td>
<td>non-const</td>
<td>non-const</td>
</tr>
<tr>
<td>Cond. dist’n</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
</tr>
<tr>
<td>Marg. mean &amp; var.</td>
<td>constant</td>
<td>constant</td>
<td>constant</td>
<td>constant</td>
</tr>
<tr>
<td>Marg. dist’n</td>
<td>normal</td>
<td>normal</td>
<td>heavy-tailed</td>
<td>heavy-tailed</td>
</tr>
</tbody>
</table>
• All of the processes are stationary $\Rightarrow$ marginal means and variances are constant

• Gaussian white noise is the “baseline” process.
  – conditional distribution $=$ marginal distribution
  – conditional means and variances are constant
  – conditional and marginal distributions are normal

• Gaussian white noise is the “source of randomness” for the other processes
  – therefore, they all have normal conditional distributions
Fitting GARCH models

Fit to 300 observation from a simulated AR(1)/ARCH(1)

Listing of the SAS program for the simulated data

options linesize = 65 ;
data arch ;
infile 'c:\courses\or473\sas\garch02.dat' ;
input y ;
run ;
title 'Simulated ARCH(1)/AR(1) data' ;
proc autoreg ;
model y =/nlag = 1 archtest garch=(q=1); run ;
### SAS output

Q and LM Tests for ARCH Disturbances

<table>
<thead>
<tr>
<th>Order</th>
<th>Q</th>
<th>Pr &gt; Q</th>
<th>LM</th>
<th>Pr &gt; LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>119.7578</td>
<td>&lt; .0001</td>
<td>118.6797</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>2</td>
<td>137.9967</td>
<td>&lt; .0001</td>
<td>129.8491</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>3</td>
<td>140.5454</td>
<td>&lt; .0001</td>
<td>131.4911</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>4</td>
<td>140.6837</td>
<td>&lt; .0001</td>
<td>132.1098</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>5</td>
<td>140.6925</td>
<td>&lt; .0001</td>
<td>132.3810</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>6</td>
<td>140.7476</td>
<td>&lt; .0001</td>
<td>132.7534</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>7</td>
<td>141.0173</td>
<td>&lt; .0001</td>
<td>132.7543</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>8</td>
<td>141.5401</td>
<td>&lt; .0001</td>
<td>132.8874</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>9</td>
<td>142.1243</td>
<td>&lt; .0001</td>
<td>132.8879</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>10</td>
<td>142.6266</td>
<td>&lt; .0001</td>
<td>132.9226</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>11</td>
<td>142.7506</td>
<td>&lt; .0001</td>
<td>133.0153</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>12</td>
<td>142.7508</td>
<td>&lt; .0001</td>
<td>133.0155</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Variable</td>
<td>DF</td>
<td>Estimate</td>
<td>Error</td>
<td>t Value</td>
</tr>
<tr>
<td>-----------</td>
<td>----</td>
<td>----------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.4810</td>
<td>0.3910</td>
<td>1.23</td>
</tr>
<tr>
<td>AR1</td>
<td>1</td>
<td>-0.8226</td>
<td>0.0266</td>
<td>-30.92</td>
</tr>
<tr>
<td>ARCH0</td>
<td>1</td>
<td>1.1241</td>
<td>0.1729</td>
<td>6.50</td>
</tr>
<tr>
<td>ARCH1</td>
<td>1</td>
<td>0.6985</td>
<td>0.1167</td>
<td>5.98</td>
</tr>
</tbody>
</table>

- AR parameter: $\hat{\phi} = -.8226$
  - this is $+.8226$ in our notation
  - close to the true value of 0.8

- estimates of the ARCH parameters:
  - $\hat{\alpha}_0 = 1.12$ (true value = 1)
  - $\hat{\alpha}_1 = .70$ (true value = .95)
• standard errors of the ARCH parameters are rather large

• approximate 95% confidence interval for $\alpha_1$ is

\[ .70 \pm (2)(0.117) = (0.466, 0.934) \]
Residuals when the S&P 500 returns are regressed against the change in the 3-month T-bill rates and the rate of inflation.
• This analysis uses
  – $\text{RETURNSP} = \text{the return on the S&P 500}$
  – $\text{DR3} = \text{change in the 3-month T-bill rate}$
  – $\text{GPW} = \text{the rate of wholesale price inflation}$

• $\text{RETURNSP}$ is regressed on $\text{DR3}$ and $\text{GPW}$ (factor model)
Model

\[ \text{RETURNSP} = \gamma_0 + \gamma_1 \text{DR3} + \gamma_2 \text{GPW} + u_t \quad (8) \]

- \( u_t \) is an AR(1)/GARCH(1,1) process
- Therefore,
  \[ u_t = \phi_1 u_{t-1} + a_t, \]
- \( a_t \) is a GARCH(1,1) process:
  \[ a_t = \epsilon_t \sigma_t \]
- where
  \[ \sigma_t = \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2}. \]
SAS Program

Key command:

```sas
proc autoreg;
model returnsp = DR3 gpw/nlag = 1 archtest garch=(p=1,q=1);
```

- “returnsp = DR3 gpw” specifies the regression model
- “nlag = 1” specifies the AR(1) structure.
- “garch=(p=1,q=1)” specifies the GARCH(1,1) structure.
- “archtest” specifies that tests of conditional heteroscedasticity be performed
SAS output

• The p-values of the Q and LM tests are all very small, less than .0001. Therefore, the errors in the regression model exhibit conditional heteroscedasticity.

• Ordinary least squares estimates of the regression parameters are:

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|-------|-----|
| Intercept| 1  | 0.0120   | 0.001755       | 6.86    | <.0001|
| DR3      | 1  | -0.8293  | 0.3061         | -2.71   | 0.0070|
| GPW      | 1  | -0.8550  | 0.2349         | -3.64   | 0.0003|
• Using residuals from the OLS estimates, the estimated residual autocorrelations are:

Estimates of Autocorrelations

<table>
<thead>
<tr>
<th>Lag</th>
<th>Covariance</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00108</td>
<td>1.000000</td>
</tr>
<tr>
<td>1</td>
<td>0.000253</td>
<td>0.234934</td>
</tr>
</tbody>
</table>
• Also, using OLS residuals, the estimate AR parameter is:

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.234934</td>
<td>0.046929</td>
<td>-5.01</td>
</tr>
</tbody>
</table>
- Assuming AR(1)/GARCH(1,1) errors, the estimated parameters of the regression are:

| Variable | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|----------|----------------|---------|------|-----|
| Intercept| 1  | 0.0125   | 0.001875       | 6.66    | <.0001|
| DR3      | 1  | -1.0665  | 0.3282         | -3.25   | 0.0012|
| GPW      | 1  | -0.7239  | 0.1992         | -3.63   | 0.0003|

- Notice that these differ slightly from OLS estimates.
- Since all p-values are small, both independent variables are significant.
- However, the Total R-square value is only 0.0551, so the regression has little predictive value.
• The estimated GARCH parameters are:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>-0.2016</td>
<td>0.0603</td>
<td>-3.34</td>
<td>0.0008</td>
</tr>
<tr>
<td>ARCH0</td>
<td>0.000147</td>
<td>0.0000688</td>
<td>2.14</td>
<td>0.0320</td>
</tr>
<tr>
<td>ARCH1</td>
<td>0.1337</td>
<td>0.0404</td>
<td>3.31</td>
<td>0.0009</td>
</tr>
<tr>
<td>GARCH1</td>
<td>0.7254</td>
<td>0.0918</td>
<td>7.91</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

• Since all p-values are small, all GARCH parameters are significant.

• GARCH1 (0.7254) >> ARCH1 (0.1337) ⇒ reasonably long persistence of volatility.
I-GARCH models

- I-GARCH or integrated GARCH processes designed to model persistent changes in volatility

- A GARCH\((p, q)\) process is stationary with a finite variance if

\[
\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1.
\]
A GARCH\((p, q)\) process is called an I-GARCH process if
\[
\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1.
\]

- I-GARCH processes are either non-stationary or have an infinite variance.

Here are some simulations of ARCH(1) processes:
Simulated ARCH(1) processes with $\alpha_1 = .9$, 1, and 1.8.
Normal plots of ARCH(1) processes with $\alpha_1 = .9$, 1, and 1.8.
Comments on the figures

- all three processes do revert to their mean, 0
- larger the value of $\alpha_1$ the more the volatility comes in sharp bursts
- processes with $\alpha_1 = .9$ and $\alpha_1 = 1$ looks similar
- none of the processes in the figure show much persistence of higher volatility
- to model persistence of higher volatility, one needs an I-GARCH($p, q$) process with $q \geq 1$
- Next figure shows simulations from I-GARCH(1,1) processes
Simulations of I-GARCH(1,1) processes. $\alpha_1 + \beta_1 = 1$
To fit I-GARCH in SAS:

```
proc autoreg;
model returnsp =/nlag = 1 garch=(p=1,q=1,type=integrated);
run;
```

- The default value of “type” is “nonneg” which only constrains the GARCH coefficients to be non-negative.

- “type=integrated” in addition imposes the sum-to-one constraint of the I-GARCH model
What does infinite variance mean?

- let $X$ be a random variable with density $f_X$
- the expectation of $X$ is
  \[
  \int_{-\infty}^{\infty} x f_X(x) \, dx
  \]
  provided that this integral is defined.
If
\[ \int_{-\infty}^{0} x f_X(x) dx = -\infty \]  \hspace{1cm} (9)
and
\[ \int_{0}^{\infty} x f_X(x) dx = \infty \] \hspace{1cm} (10)
then the expectation is, formally, \(-\infty + \infty \Rightarrow \text{not defined}\)

- if both integrals are finite, then the expectation is the sum of these two integrals
• **Exercise:** \( f_X(x) = \frac{1}{4} \) if \(|x| < 1\) and \( f_X(x) = \frac{1}{4x^2} \) if \(|x| \geq 1\)

- \( f_X \) is a density since 
\[
\int_{-\infty}^{\infty} f_X(x) \, dx = 1
\]

- Then expectation does not exist since
\[
\int_{-\infty}^{0} x f_X(x) \, dx = -\infty
\]

- and
\[
\int_{0}^{\infty} x f_X(x) \, dx = \infty
\]
What are the implications of having no expectation?

- assume sample of iid sample from $f_X$
- law of large numbers $\Rightarrow$ sample mean will converge to the expectation
- law of large numbers doesn’t apply if expectation is not defined
- there is no point to which the sample mean can converge
  - it will just wander without converging
Sample means of ARCH(1) processes with $\alpha_1 = .9$, 1, and 1.8.
What are the implications of having infinite variance

- now suppose that the expectation of $X$ exists and equals $\mu_X$
- the variance
  \[
  \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx
  \]
- if this integral is $+\infty$ then the variance is infinite
- law of large numbers $\Rightarrow$ sample variance will converge to the variance
- variance of $X$ is infinity $\Rightarrow$ the sample variance will converge to infinity
Sample variances: ARCH(1) with $\alpha_1 = .9$, 1, and 1.8.
GARCH-M processes

• if we fit a regression model with GARCH errors
  – could use the conditional standard deviation \( (\sigma_t) \)
    as one of the regression variables

• when the dependent variable is a return
  – the market demands a higher risk premium for higher risk
  – so higher conditional variability could cause higher returns
• GARCH-M models in SAS — add keyword "mean,"
  e.g.,
  
  proc autoreg ;
  model returnsp =/nlag = 1 garch=(p=1,q=1,mean);
  run ;

• or for I-GARCH-M
  
  proc autoreg ;
  model returnsp =/nlag = 1 garch=(p=1,q=1,mean,type=integrated);
  run ;
GARCH-M example: S&P 500

- GARCH(1,1)-M was fit in SAS
- $\delta$ is the regression coefficient for $\sigma_t$
- $\hat{\delta} = .5150$
  - standard error = .3695
- t-value = 1.39
- p-value = .1633
• since p-value = .1633
  – could accept the null hypothesis that \( \delta = 0 \)
  – no strong evidence that there are higher returns during times of higher volatility.

• volatility of S&P 500 is market risk so this is somewhat surprising (think of CAPM)

• may be that the effect is small but not 0 (\( \hat{\delta} \) is positive, after all)

• AIC criterion does select the GARCH-M model
E-GARCH

• E-GARCH models are used to model the “leverage effect”
  - prices become more volatile as prices decrease

• E-GARCH, model is

\[
\log(\sigma_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i g(\epsilon_{t-i}) + \sum_{i=1}^{p} \beta_i \log(\sigma_{t-i}),
\]

• where

\[
g(\epsilon_t) = \theta \epsilon_t + \gamma \{|\epsilon_t| - E(|\epsilon_t|)\}
\]

• \(\log(\sigma_t)\) can be negative \(\Rightarrow\) no constraints on parameters
From the previous page:

\[ g(\epsilon_t) = \theta \epsilon_t + \gamma \{|\epsilon_t| - E(|\epsilon_t|)\} \]

To understand \( g \) note that

\[ g(\epsilon_t) = -\gamma E(|\epsilon_t|) + (\gamma + \theta)|\epsilon_t| \quad \text{if} \quad \epsilon_t > 0, \]

and

\[ g(\epsilon_t) = -\gamma E(|\epsilon_t|) + (\gamma - \theta)|\epsilon_t| \quad \text{if} \quad \epsilon_t < 0, \]

typically, \(-1 < \hat{\theta} < 0\) so that \( 0 < \gamma + \theta < \gamma - \theta \)

\( \hat{\theta} = -0.7 \) in the S&P 500 example

\( E(|\epsilon_t|) = \sqrt{2/\pi} = .7979 \) (good calculus exercise)
The $g$ function for the S&P 500 data (top left panel) and several other values of $\theta$. 
• SAS fits the E-GARCH model
  – $\gamma$ fixed as 1
  – $\theta$ estimated

• E-GARCH model is specified by using “type=exp” as in:

```sas
proc autoreg ;
model returnsp =/nlag = 1 garch=(p=1,q=1,mean,type=exp);
run ;
```
Back to the S&P 500 example

- SAS can fit six different AR(1)/GARCH(1,1) models
  - “type” = “integrated,” “exp,” or “nonneg”
  - GARCH-in-mean effect can be included or not

- following table contains the AIC statistics
  - models ordered by AIC (best fitting to worse)
<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Δ AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-GARCH-M</td>
<td>−1783.9</td>
<td>0</td>
</tr>
<tr>
<td>E-GARCH</td>
<td>−1783.1</td>
<td>0.8</td>
</tr>
<tr>
<td>GARCH-M</td>
<td>−1764.6</td>
<td>19.3</td>
</tr>
<tr>
<td>GARCH</td>
<td>−1764.1</td>
<td>19.8</td>
</tr>
<tr>
<td>I-GARCH-M</td>
<td>−1758.0</td>
<td>25.9</td>
</tr>
<tr>
<td>I-GARCH</td>
<td>−1756.4</td>
<td>27.5</td>
</tr>
</tbody>
</table>

AIC statistics for six AR(1)/GARCH(1,1) models fit to the S&P 500 returns data. Δ AIC is change in AIC between a given model and E-GARCH-M.
• AR(2) and E-GARCH(1,2)-M, E-GARCH(2,1)-M, and E-GARCH(2,2)-M models were tried
  – none of these lowered AIC
  – none had all parameters significant at $p = .1$
Listing of SAS output for the E-GARCH-M model:

The AUTOREG Procedure
Estimates of Autoregressive Parameters

<table>
<thead>
<tr>
<th>Lag</th>
<th>Coefficient</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.234934</td>
<td>0.046929</td>
<td>-5.01</td>
</tr>
</tbody>
</table>

Algorithm converged.

Exponential GARCH Estimates

| SSE       | 0.44211939 | Observations | 433   |
| MSE       | 0.00102    | Uncond Var   | .     |
| Log Likelihood | 900.962569 | Total R-Square | 0.1050 |
| SBC       | -1747.2885 | AIC          | -1783.9251 |
| Normality Test | 24.9607 | Pr > ChiSq   | <.0001 |

SSE: Sum of Squared Errors
MSE: Mean Squared Error
SBC: Schwarz Bayesian Criterion
AIC: Akaike Information Criterion
Pr > ChiSq: Probability of Chi-Square
| Variable   | DF | Estimate | Standard Error | t Value | Pr > |t| |
|------------|----|----------|----------------|---------|------|---|
| Intercept  | 1  | -0.003791 | 0.0102         | -0.37   | 0.7095 |
| DR3        | 1  | -1.2062   | 0.3044         | -3.96   | <.0001 |
| GPW        | 1  | -0.6456   | 0.2153         | -3.00   | 0.0027 |
| AR1        | 1  | -0.2376   | 0.0592         | -4.01   | <.0001 |
| EARCH0     | 1  | -1.2400   | 0.4251         | -2.92   | 0.0035 |
| EARCH1     | 1  | 0.2520    | 0.0691         | 3.65    | 0.0003 |
| EGARCH1    | 1  | 0.8220    | 0.0606         | 13.55   | <.0001 |
| THETA      | 1  | -0.6940   | 0.2646         | -2.62   | 0.0087 |
| DELTA      | 1  | 0.5067    | 0.3511         | 1.44    | 0.1490 |
The GARCH zoo

Here’s a sample of other GARCH models mentioned in Bollerslev, Engle, and Nelson (1994):

- QARCH = quadratic ARCH
- TARCH = threshold ARCH
- STARCH = structural ARCH
- SWARCH = switching ARCH
- QTARCH = quantitative threshold ARCH
- vector ARCH
- diagonal ARCH
- factor ARCH
GARCH Models in Finance

Remember the problem of implied volatility — it depended on $K$ and $T$!

• Black-Scholes assumes a constant variance

• But GARCH effects are common so the Black-Scholes model is not adequate

• Having a volatility function (smile) is a quick fix
  – but not logical
Options can be priced assuming the log-returns are a GARCH process (rather than a random walk)

- Multinomial (not binomial) tree — to have different levels of volatility
- Need to keep track of price and conditional variance
Ritchken and Trevor use an **NGARCH (nonlinear asymmetric GARCH)** model:

$$\log(S_{t+1}/S_t) = r + \lambda \sqrt{h_t} - h_t/2 + \sqrt{h_t} \nu_{t+1}$$

$$h_{t+1} = \beta_0 + \beta_1 h_t + \beta_2 h_t (\nu_{t+1} - c)^2$$

where $\nu_t$ is WhiteNoise$(0, 1)$.

- $c = 0$ is an ordinary GARCH model
- $\lambda$ is a “risk premium”

Under the risk-neutral (martingale) measure:

$$\log(S_{t+1}/S_t) = (r - h_t/2) + \sqrt{h_t} \epsilon_{t+1},$$

$$h_{t+1} = \beta_0 + \beta_1 h_t + \beta_2 h_t (\epsilon_{t+1} - c^*)^2,$$

where $c^* = c + \lambda$ and $\epsilon_t$ is WhiteNoise$(0, 1)$. 
Now there are five unknown parameters:

• $h_0$

• $\beta_1$, $\beta_2$, and $\beta_3$

• $c^*$

These parameters are estimated by nonlinear least-squares –

“Implied GARCH parameters”
Volatility smile is “explained” as due to GARCH effects:

- nonconstant variance
- nonnormal marginal distribution
From Chapter 7, “Bank of Volatility,” of “When Genius Failed” by Roger Lowenstein:

“Early in 1998, Long-Term began to short large amounts of equity volatility.

This simple trade, second nature to Rosenfeld and David Modest, would be indecipherable to 999 out of 1,000 Americans.

Equity vol comes straight from the Black-Scholes model.”
“The stock market, for instance, typically varies by about 15 percent to 20 percent a year”.

“And when the model told them that the markets were mispricing equity vol, they were willing to bet the firm on it.”

“The options market was anticipating volatility in the stock market of roughly 20 percent. Long-term viewed this as incorrect ... Thus, it figured that options prices would sooner or later fall.”
Long-Term began to **short options** on the S&P 500 are similar European options. “They were ‘**selling volatility.’”

“In fact, it sold insurance (options) both ways—against a sharp downturn and against a sharp rise.”
Annualized SD = $\sqrt{253} \times \text{daily variance}$
data capm;
set Sasuser.capm;
logR_sp500 = dif(log(close_sp500)) ;
logR_msft = dif(log(close_msft)) ;
run;
proc gplot;
plot logR_sp500*date ;
run ;
proc autoreg ;
model logR_sp500 = /nlag=1 garch=(p=1,q=1,type=exp) ;
output out=outdata cev=cev ;
run ;
proc gplot ;
plot cev*date ;
run ;
<table>
<thead>
<tr>
<th>nlag</th>
<th>p</th>
<th>q</th>
<th>E-GARCH</th>
<th>M-GARCH</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>no</td>
<td>no</td>
<td>$-14,366$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>no</td>
<td>no</td>
<td>$-15,002$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>$-15,105$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>yes</td>
<td>yes</td>
<td>$-15,103$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>$-15,105$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>$-15,107$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>$-14,366$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>$-15,112$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>$-15,104$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>yes</td>
<td>no</td>
<td>$-15,100$</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>$-15,100$</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>$-15,101$</td>
</tr>
</tbody>
</table>
In fact, it sold insurance (options) both ways—against a sharp downturn and against a sharp rise.
The “Greeks”

\[ C(S, T, t, K, \sigma, r) = \text{price of an option} \]

\[ \Delta = \frac{\partial}{\partial S} C(S, T, t, K, \sigma, r) \quad \text{“Delta”} \]

\[ \Theta = \frac{\partial}{\partial t} C(S, T, t, K, \sigma, r) \quad \text{“Theta”} \]

\[ \mathcal{R} = \frac{\partial}{\partial r} C(S, T, t, K, \sigma, r) \quad \text{“Rho”} \]

\[ \mathcal{V} = \frac{\partial}{\partial \sigma} C(S, T, t, K, \sigma, r) \quad \text{“Vega”} \]
Put-call parity

Put and call prices with same $K$ and $T$ are related:

$$P(S, T, t, K, \sigma, r) = C(S, T, t, K, \sigma, r) + e^{-r(T-t)}K - S.$$ 

Therefore, the call and put have the same vegas

$$\frac{\partial}{\partial \sigma} P(S, T, t, K, \sigma, r) = \frac{\partial}{\partial \sigma} C(S, T, t, K, \sigma, r)$$

but difference deltas

$$\frac{\partial}{\partial S} P(S, T, t, K, \sigma, r) = \frac{\partial}{\partial S} C(S, T, t, K, \sigma, r) - 1$$

$$0 < \Delta(\text{call}) = \Phi(d_1) < 1$$

$$-1 < \Delta(\text{put}) = \Phi(d_1) - 1 < 0$$
From previous slide:

\[-1 < \Delta(\text{put}) = \Phi(d_1) - 1 < 0\]

Suppose we buy \(N_1\) call options and \(N_2\) put options.

Delta of the portfolio is

\[N_1 \Phi(d_1) + N_2(\Phi(d_1) - 1)\]

which is zero if

\[\frac{N_1}{N_2} = \frac{1 - \Phi(d_1)}{\Phi(d_1)}\]

Hedging is possible with a put and call with different \(K\) and \(T\) – each then has its own value of \(d_1\)
Selling equity vol was very clever, but as Lowenstein remarks:

“This was—so unlike the partners’ credo—rank speculation.”