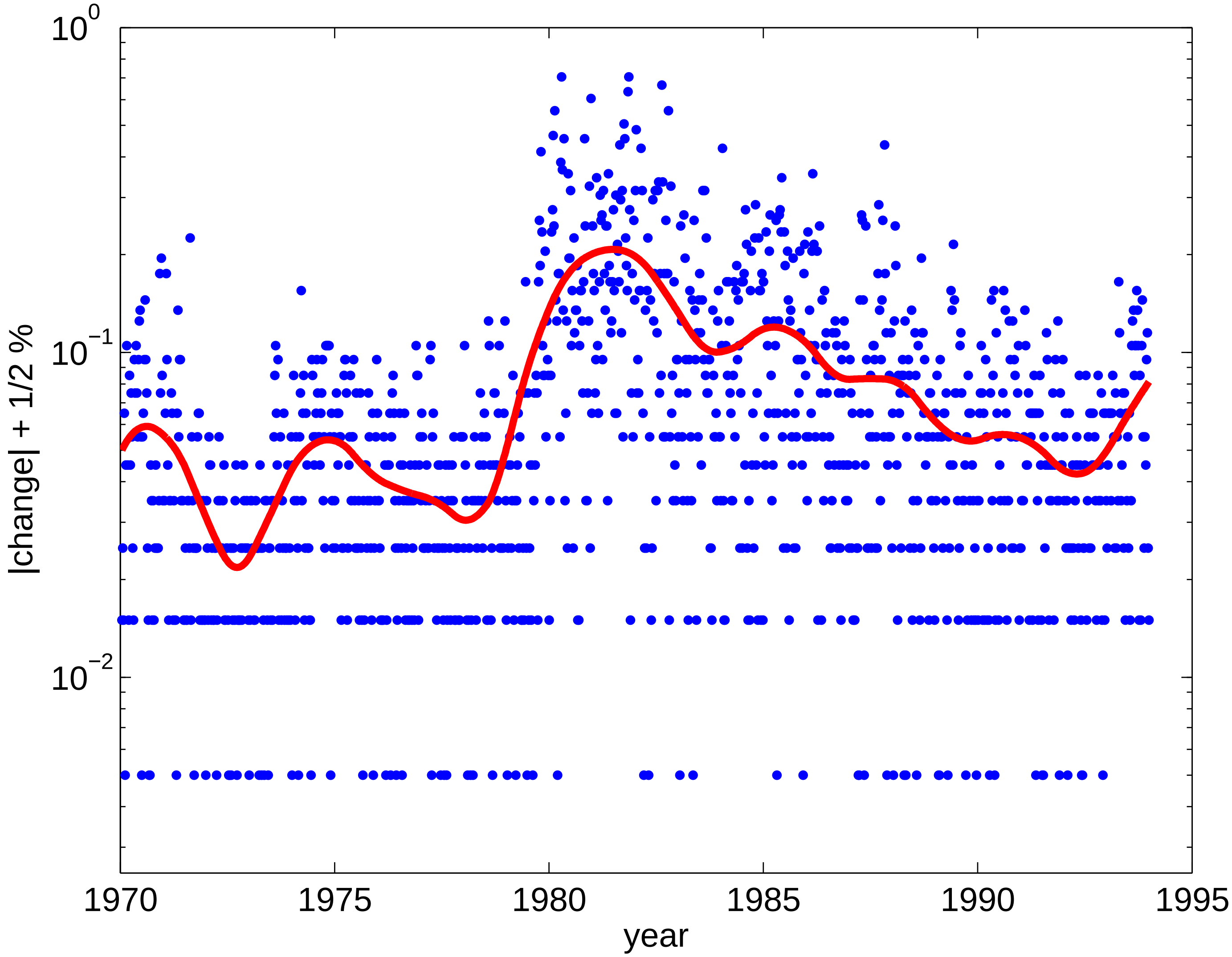
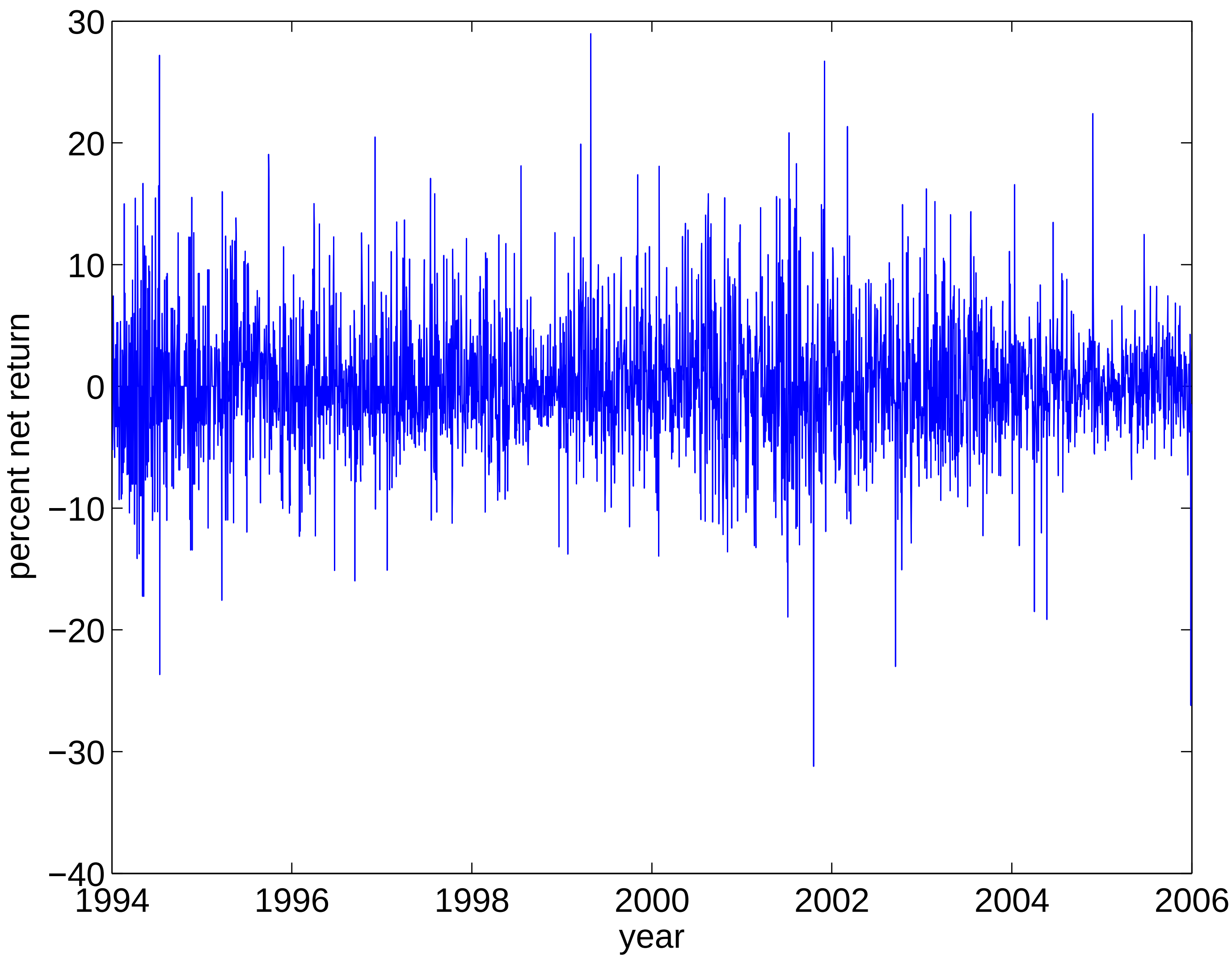
GARCH Models

Introduction

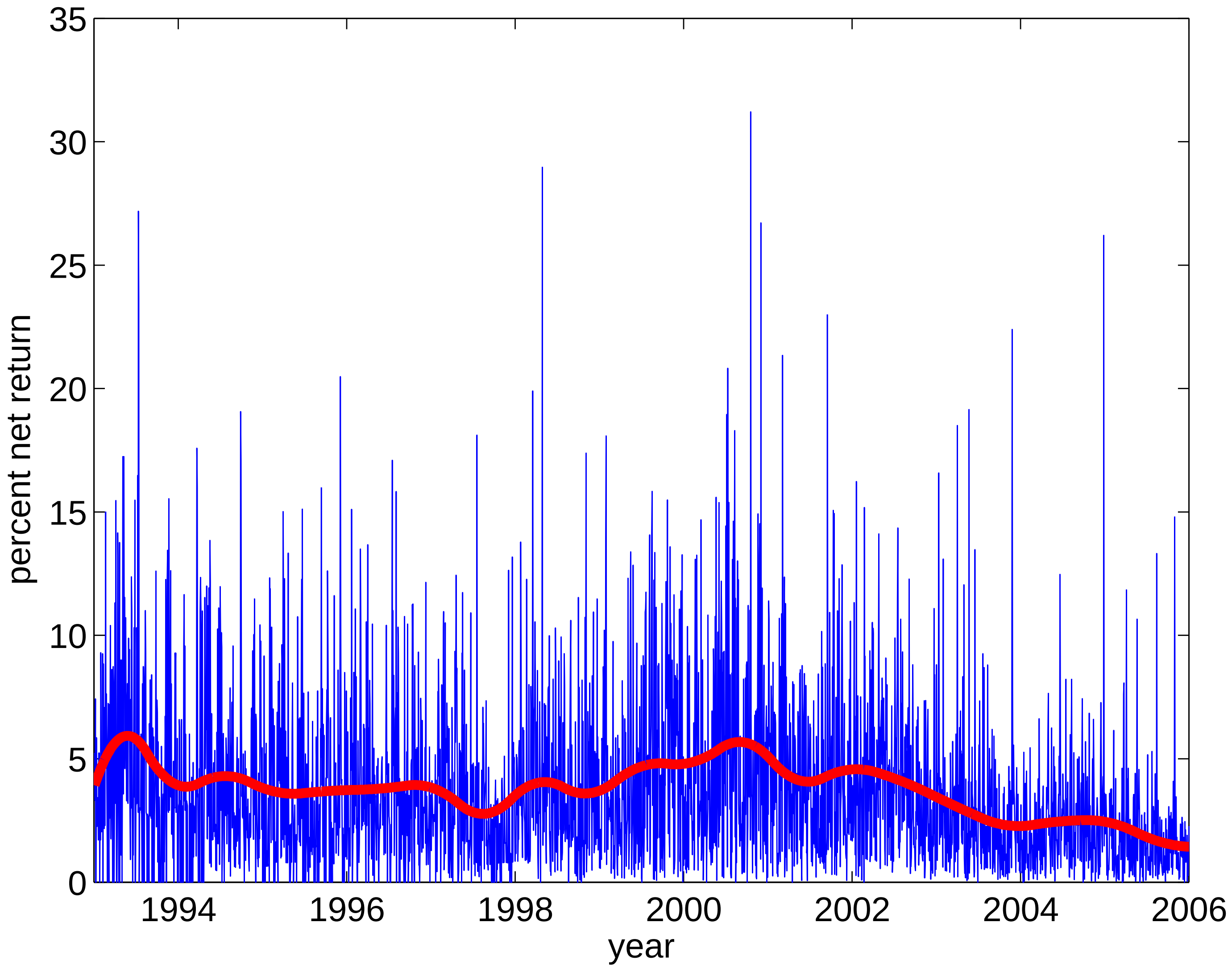
- ARMA models assume a constant volatility
- In finance, correct specification of volatility is essential
- ARMA models are used to model the conditional expectation
- They write Y_t as a linear function of the past plus a white noise term



Absolute changes in weekly AAA rate



Cree Daily Returns



Cree Daily Returns

- GARCH — models of nonconstant volatility
- ARCH = **A**uto**R**egressive **C**onditional **H**eteroscedasticity
- heteroscedasticity = non-constant variance
  phuong sai
- homoscedasticity = constant variance

- ARMA $\Rightarrow$
  - unconditionally homoscedastic
  - **conditionally** **homoscedastic**
- GARCH $\Rightarrow$
  - 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 ϵ is $N(0,1)$, and $Y = a + b\epsilon$, then $E(Y) = a$ and $\mathrm{Var}(Y) = b^2$.

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

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

- 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

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

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
$$\mathrm{Var}(\epsilon_t | \epsilon_{t-1}, \ldots) = 1. \tag{3}$$
- Property (3) is called **conditional homoscedasticity**.

$$a_t = \epsilon_t \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2}. \tag{4}$$

- 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 ∞
- Define

$$\sigma_t^2 = \mathrm{Var}(a_t | a_{t-1}, \ldots)$$

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

$$E(a_t | a_{t-1}, \ldots) = 0, \tag{5}$$

and

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2. \tag{6}$$

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 σ_{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

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2. \tag{7}$$

  This gives us

$$\gamma_a(0) = \alpha_0 + \alpha_1 \gamma_a(0).$$

- This equation has a positive solution if $\alpha_1 < 1$:

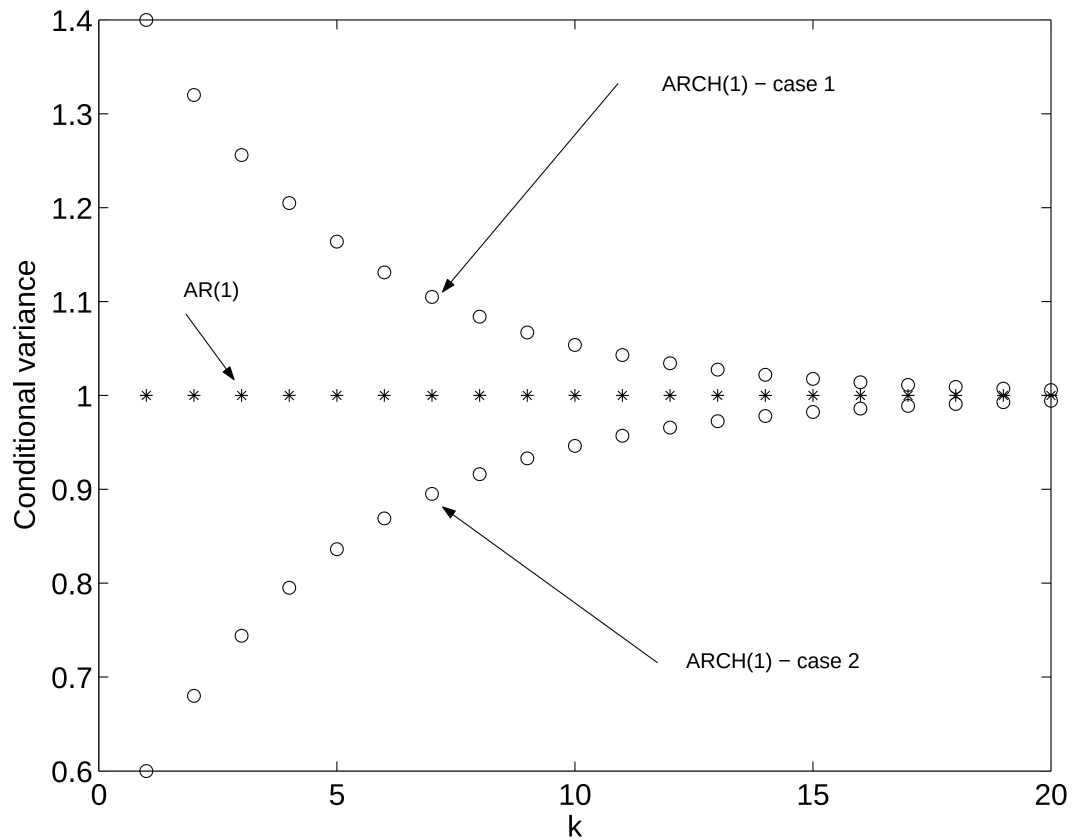
$$\gamma_a(0) = \alpha_0/(1 - \alpha_1).$$

- 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 $\alpha_1 < 1$:

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

In contrast, for any ARMA process:
$\text{Var}(a_{t+k}|a_t, a_{t-1}, \ldots) = \gamma(0)$.

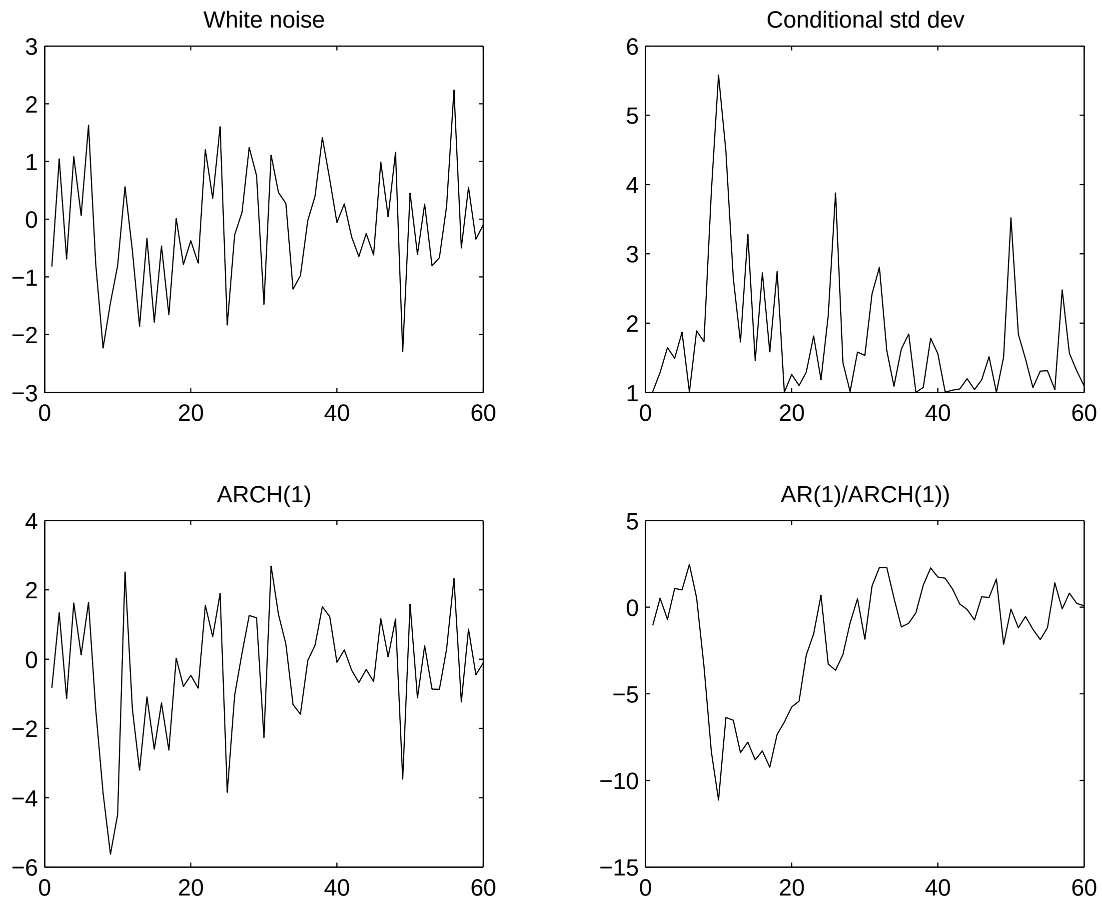


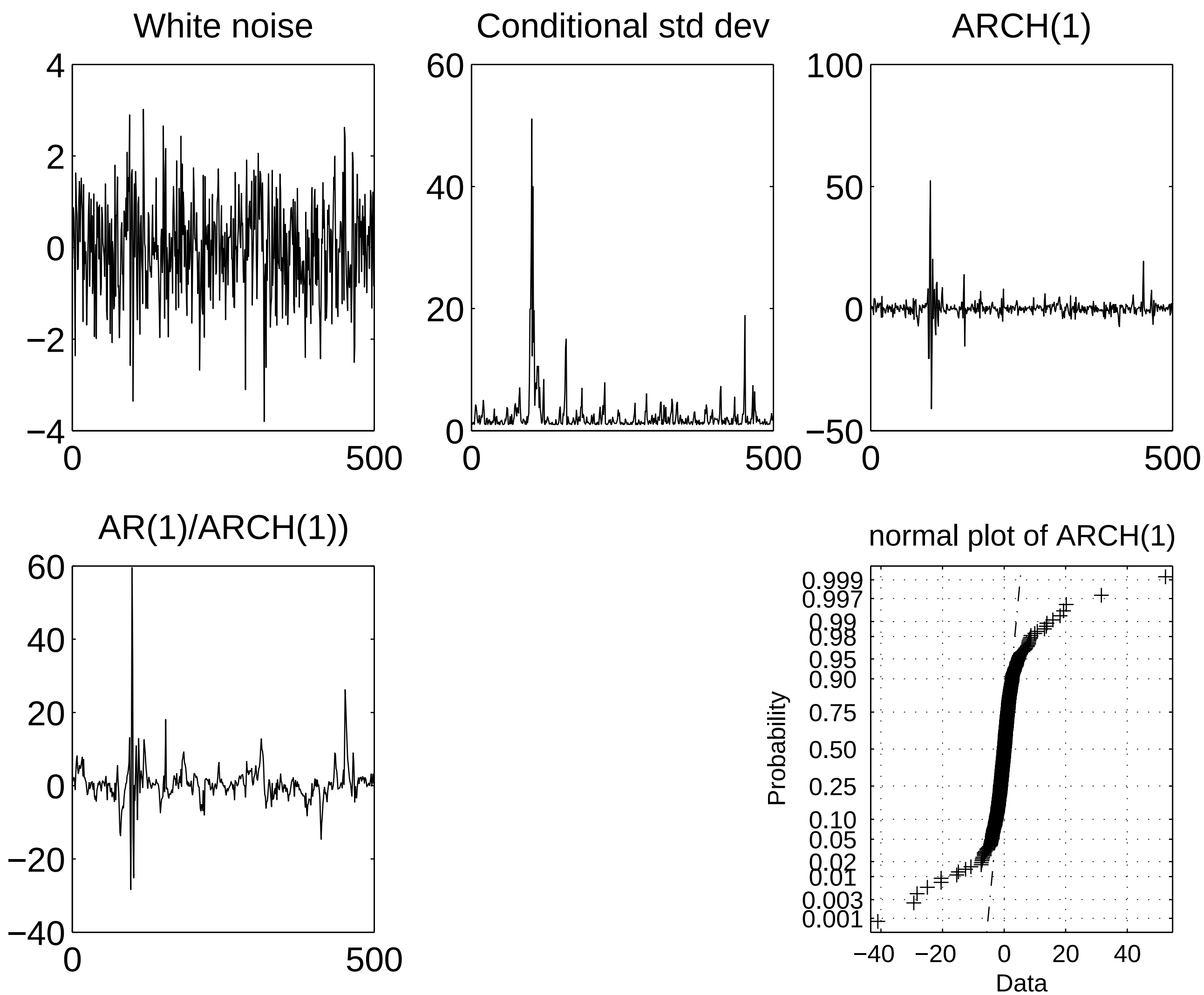
Var$(a_{t+k}|a_t,\ldots)$ **for AR(1) and ARCH(1).** $\gamma(0) = 1$ **in both cases. For ARCH(1),** $\alpha_1 = .9$**. Case 1:** $a_t^2 = 1.5$**. Case 2:** $a_t^2 = .5$**.**

- 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:

- $\alpha_0 = 1$, $\alpha_1 = .95$, $\mu = .1$, 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)

$$\sigma_t^2 = \sigma^2.$$

ARCH(1)

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1}.$$

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

Recall: $\sigma_t^2 = \mathrm{Var}(a_t | a_{t-1}, \ldots)$.

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:
$$\rho_u(h) = \phi^{|h|} \qquad \forall \quad h.$$
  - a_t^2 has the ARCH(1) ACF:
$$\rho_{a^2}(h) = \alpha_1^{|h|} \qquad \forall \quad h.$$
  - 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 ϵ_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

$$a_t = \epsilon_t \sigma_t$$

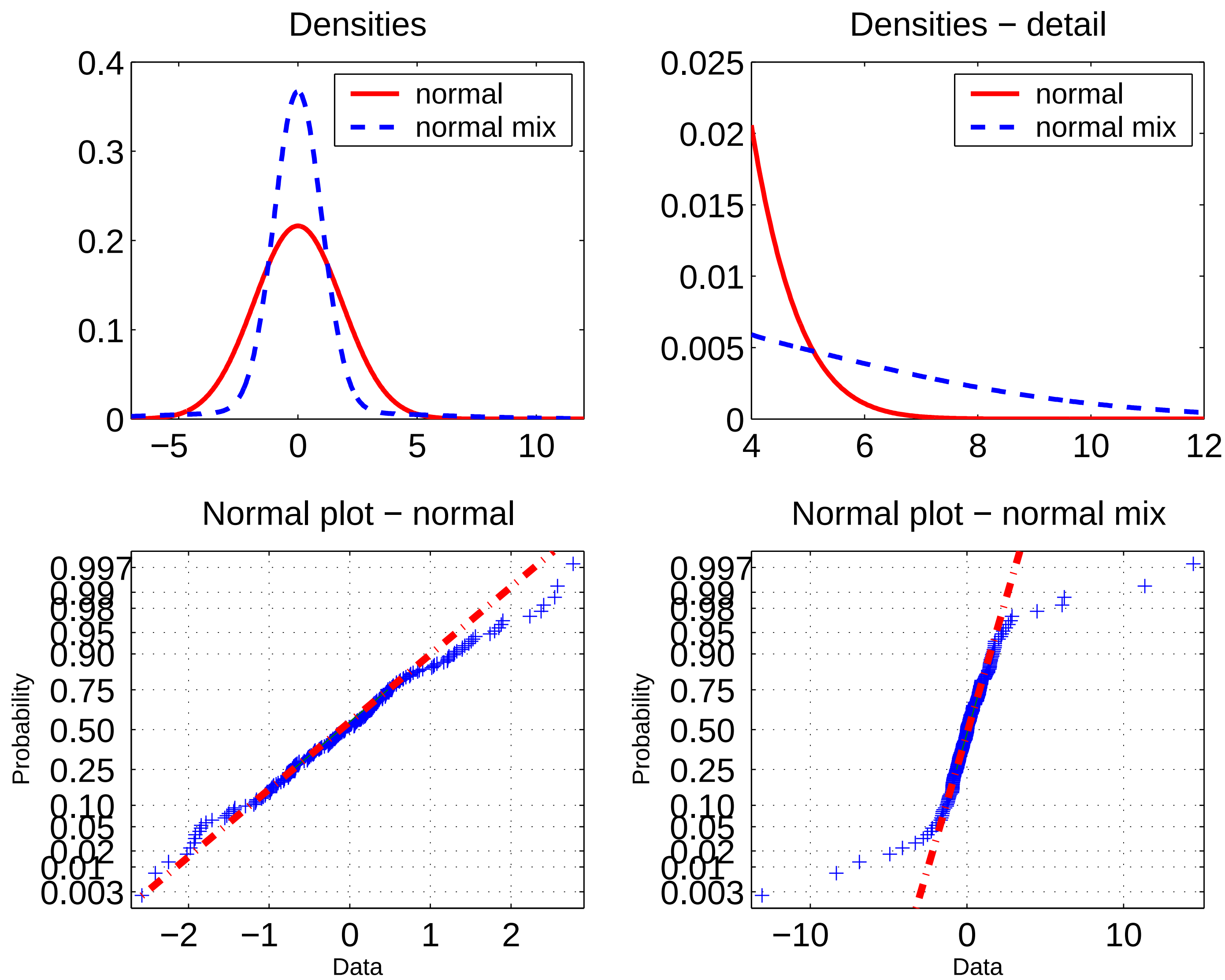
- where

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

- 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 that 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.

Property	**Gaussian WN**	**ARMA**	**GARCH**	**ARMA/ GARCH**
Cond. mean	constant	non-const	0	non-const
Cond. var	constant	constant	non-const	non-const
Cond. dist'n	normal	normal	normal	normal
Marg. mean & var.	constant	constant	constant	constant
Marg. dist'n	normal	normal	heavy-tailed	heavy-tailed

- 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 randomess" 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

Order	Q	Pr > Q	LM	Pr > LM
1	119.7578	<.0001	118.6797	<.0001
2	137.9967	<.0001	129.8491	<.0001
3	140.5454	<.0001	131.4911	<.0001
4	140.6837	<.0001	132.1098	<.0001
5	140.6925	<.0001	132.3810	<.0001
6	140.7476	<.0001	132.7534	<.0001
7	141.0173	<.0001	132.7543	<.0001
8	141.5401	<.0001	132.8874	<.0001
9	142.1243	<.0001	132.8879	<.0001
10	142.6266	<.0001	132.9226	<.0001
11	142.7506	<.0001	133.0153	<.0001
12	142.7508	<.0001	133.0155	<.0001

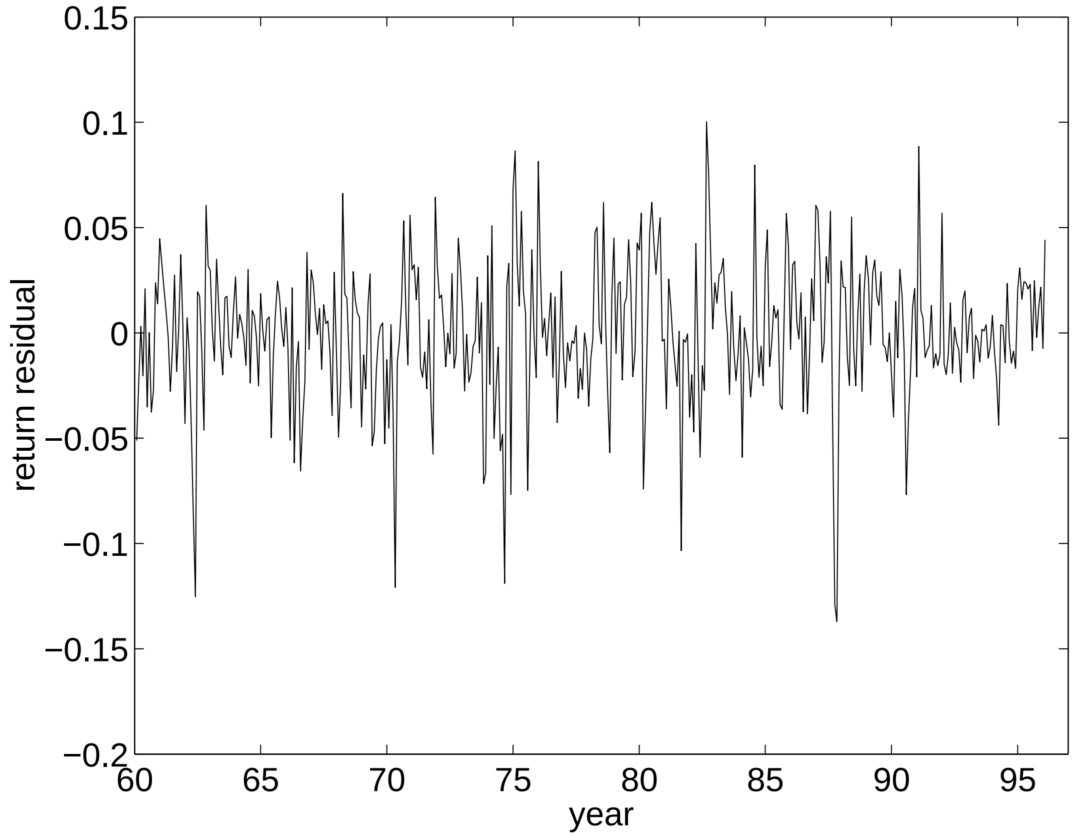
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.4810	0.3910	1.23	0.2187
AR1	1	-0.8226	0.0266	-30.92	<.0001
ARCH0	1	1.1241	0.1729	6.50	<.0001
ARCH1	1	0.6985	0.1167	5.98	<.0001

- 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 α_1 is

$$.70 \pm (2)(0.117) = (.446, .934)$$

Example: S&P 500 returns



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
  - RETURNSP = the return on the S&P 500
  - DR3 = change in the 3-month T-bill rate
  - GPW = the rate of wholesale price inflation
- RETURNSP is regressed on DR3 and GPW (factor model)

Model

$$\text{RETURNSP} = \gamma_0 + \gamma_1 \text{DR3} + \gamma_2 \text{GPW} + u_t \qquad (8)$$

- 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:

```
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	Approx 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

           Lag     Covariance     Correlation

             0        0.00108        1.000000
             1       0.000253        0.234934
```

- Also, using OLS residuals, the estimate AR parameter is:

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.234934	0.046929	-5.01

- Assuming AR(1)/GARCH(1,1) errors, the estimated parameters of the regression are:

Variable	DF	Estimate	Standard Error	t Value	Approx 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:

AR1	1	-0.2016	0.0603	-3.34	0.0008
ARCH0	1	0.000147	0.0000688	2.14	0.0320
ARCH1	1	0.1337	0.0404	3.31	0.0009
GARCH1	1	0.7254	0.0918	7.91	<.0001

- Since all p-values are small, all GARCH parameters are significant.
- GARCH1 (0.7254) >> ARCH1 (0.1337) $\Rightarrow$ 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

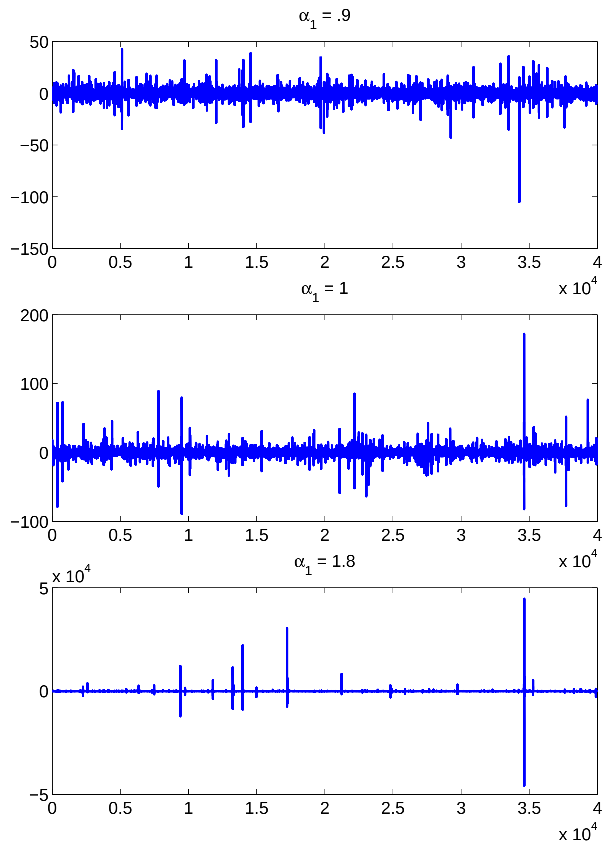
$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1.$$

A GARCH(p, q) process is called an I-GARCH process if

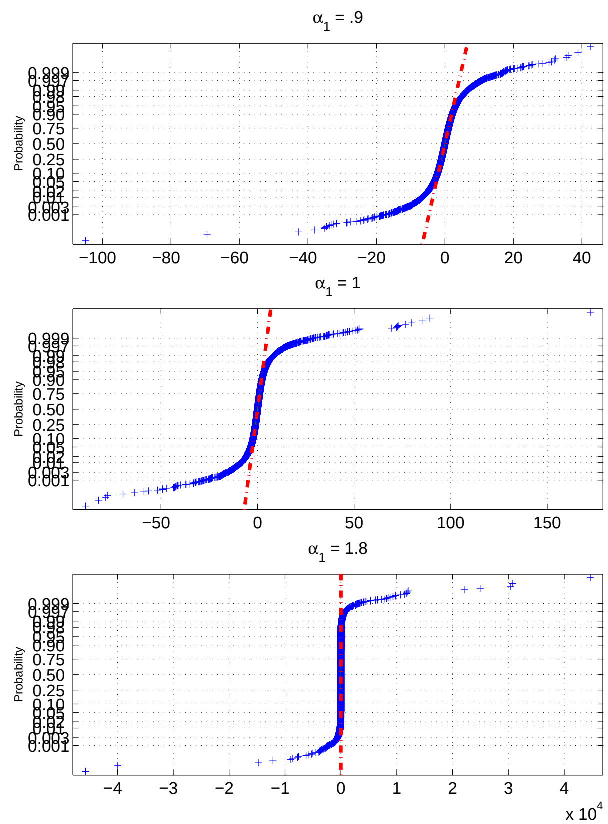
$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1.$$

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

Here are some simulations of ARCH(1) processes:



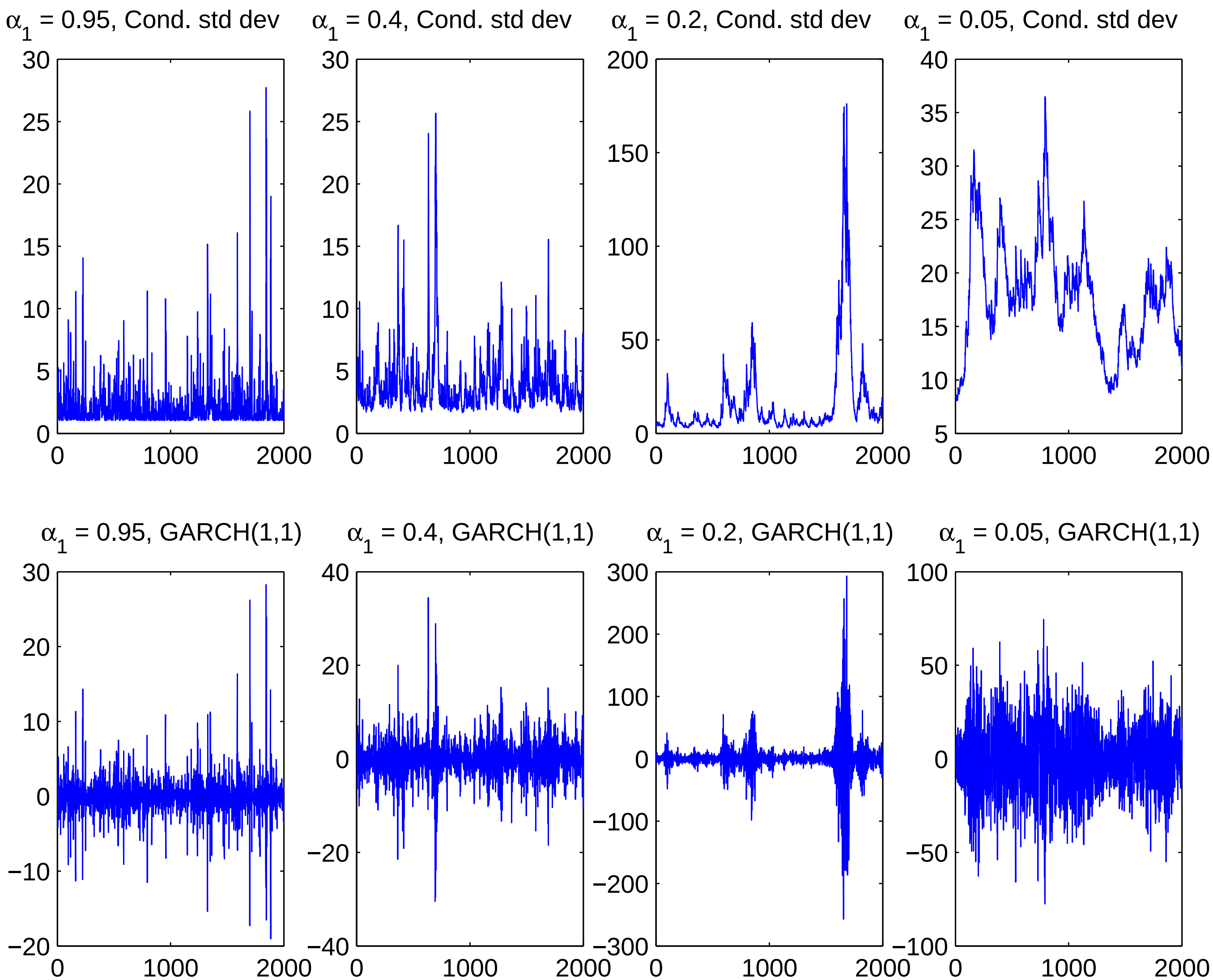
Simulated ARCH(1) processes with $\alpha_1 = .9$, 1, and 1.8.



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 α_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

$$\int_{-\infty}^{0} x f_X(x) dx = -\infty \tag{9}$$

and

$$\int_{0}^{\infty} x f_X(x) dx = \infty \tag{10}$$

then the expectation is, formally, $-\infty + \infty \Rightarrow$ not defined

- if both integrals are finite, then the expectation is the sum of these two integrals

- **Exercise:** $f_X(x) = 1/4$ if $|x| < 1$ and $f_X(x) = 1/(4x^2)$ if $|x| \geq 1$
  - f_X is a density since

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

  - Then expectation does not exist since

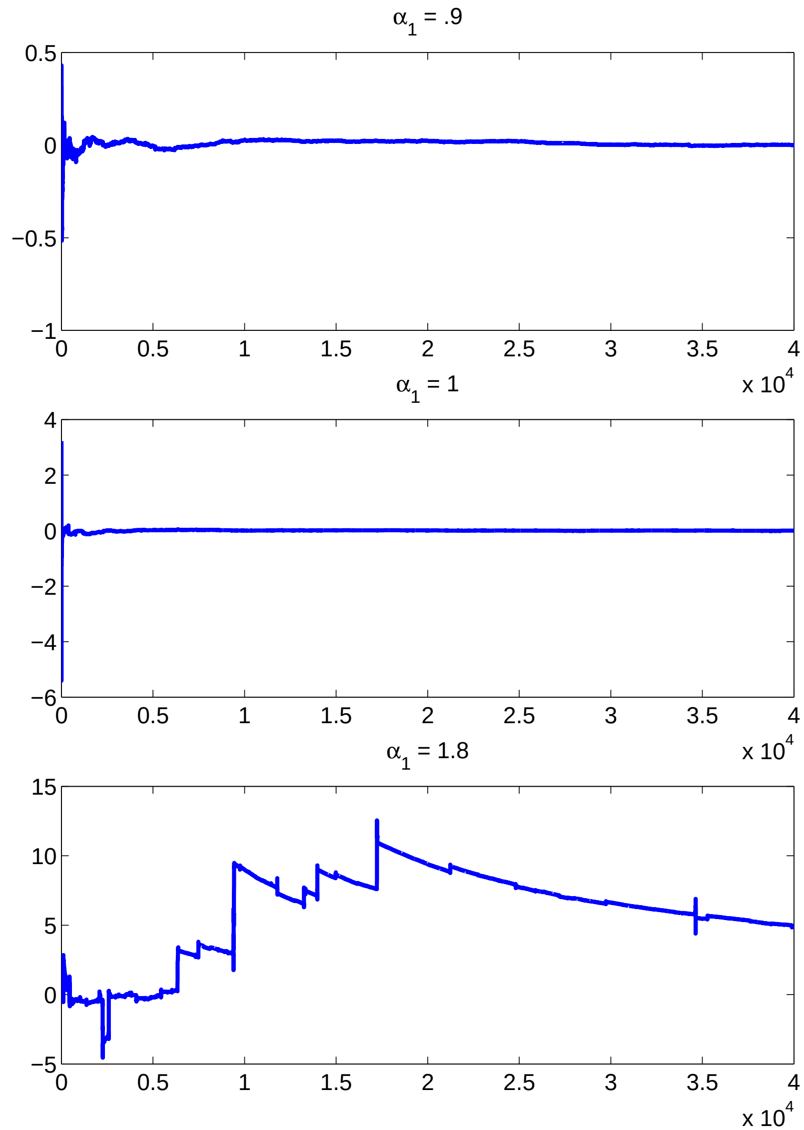
$$\int_{-\infty}^{0} x f_X(x)dx = -\infty$$

  - 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



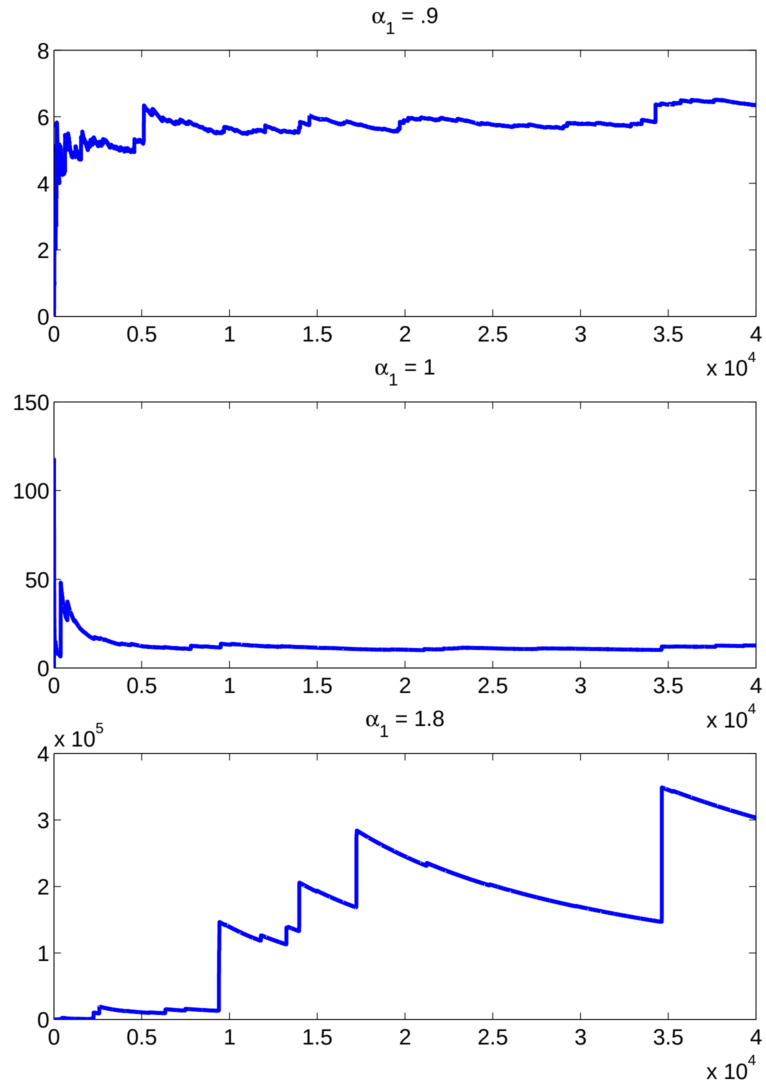
Sample means of ARCH(1) processes with $\alpha_1 = .9$, 1, and 1.8.

What are the implications of having infinite variance

- now suppose that the expectation of X exists and equals μ_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 (σ_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
- δ is the regression coefficient for σ_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

$$\log(\sigma_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_1 g(\epsilon_{t-i}) + \sum_{i=1}^{p} \beta_i \log(\sigma_{t-i}),$$

- 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:

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

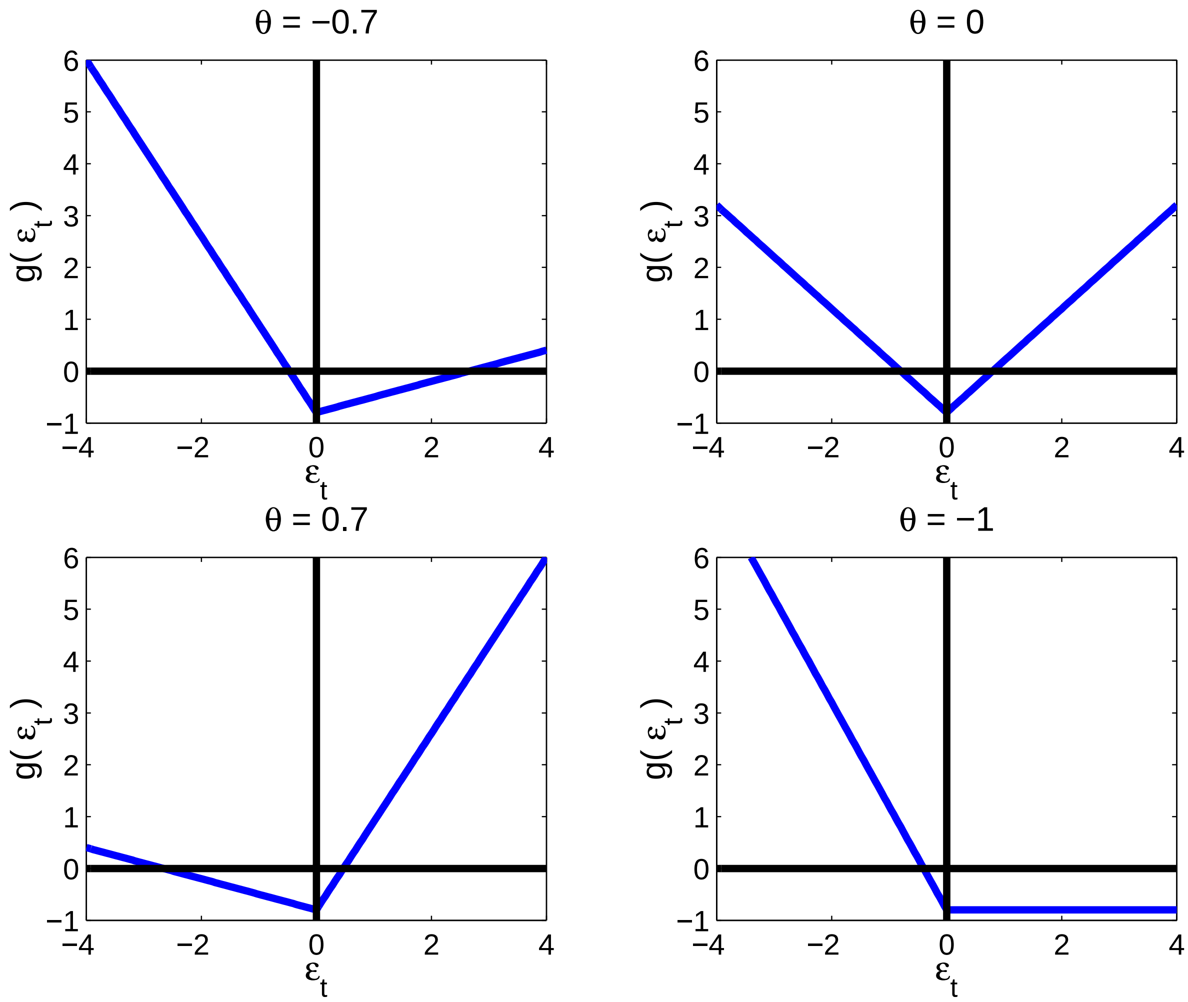
- To understand g note that

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

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 $0 < \gamma + \theta < \gamma - \theta$
- $\hat{\theta} = -.7$ 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 θ.

- SAS fits the E-GARCH model
  - γ fixed as 1
  - θ estimated
- E-GARCH model is specified by using "type=exp" as in:

```
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)

Model	**AIC**	**Δ AIC**
E-GARCH-M	−1783.9	0
E-GARCH	−1783.1	0.8
GARCH-M	−1764.6	19.3
GARCH	−1764.1	19.8
I-GARCH-M	−1758.0	25.9
I-GARCH	−1756.4	27.5

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
                                          Standard
          Lag      Coefficient             Error    t Value
            1        -0.234934          0.046929      -5.01
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
```

Variable	DF	Estimate	Standard Error	t Value	Approx 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

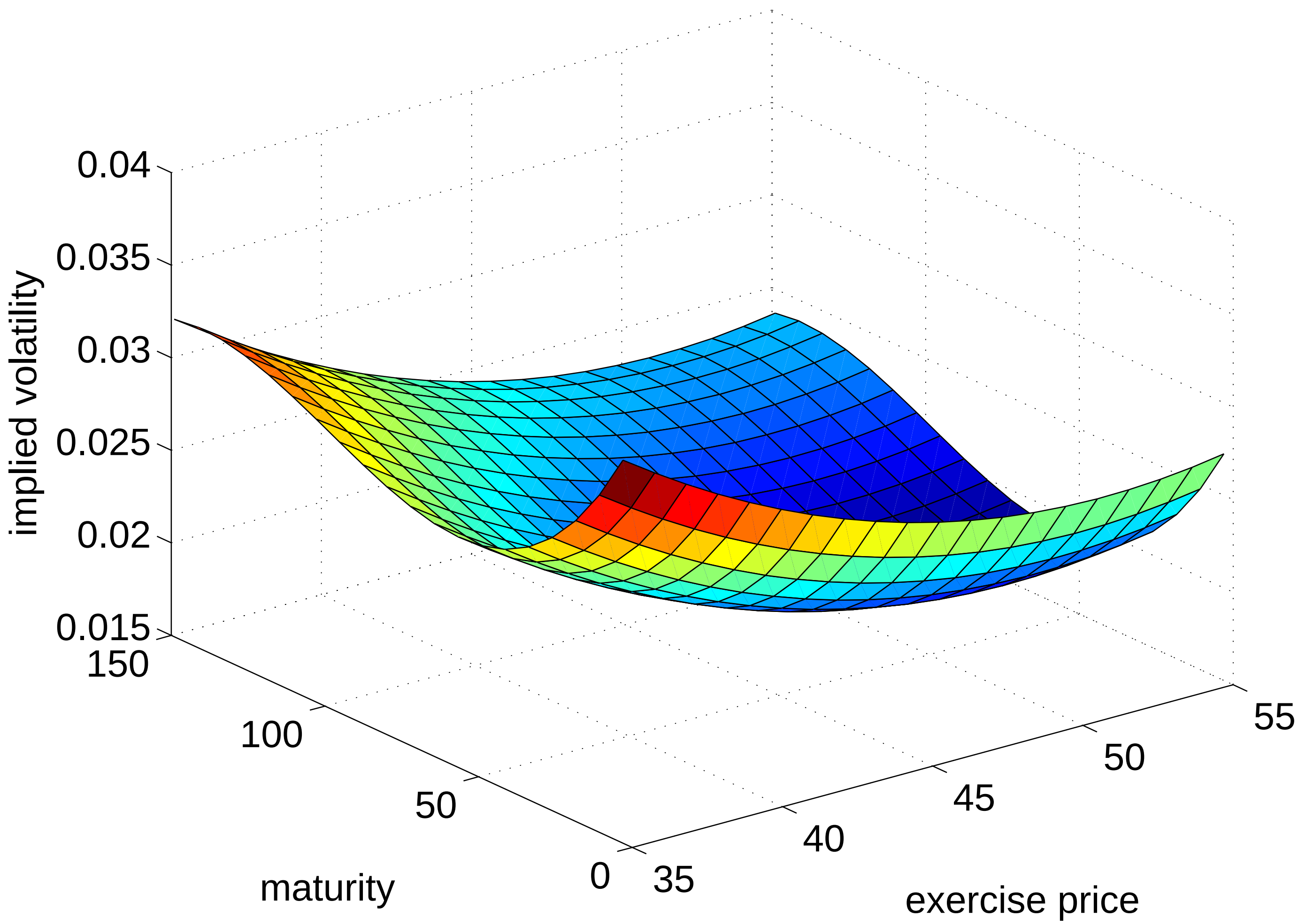
Here's a sample of other GARCH models mentioned in Bollerslev, Engle, and Nelson (1994):

- 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**



0.04
0.035
0.03
0.025
0.02
0.015
implied volatility
150
100
50
0
maturity
35
40
45
50
55
exercise price

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 ν_t is WhiteNoise$(0, 1)$.

  - $c = 0$ **is an ordinary GARCH model**
  - λ **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 ϵ_t is WhiteNoise$(0, 1)$.

Now there are **five unknown parameters**:

- h_0
- β_1, β_2, and β_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.”

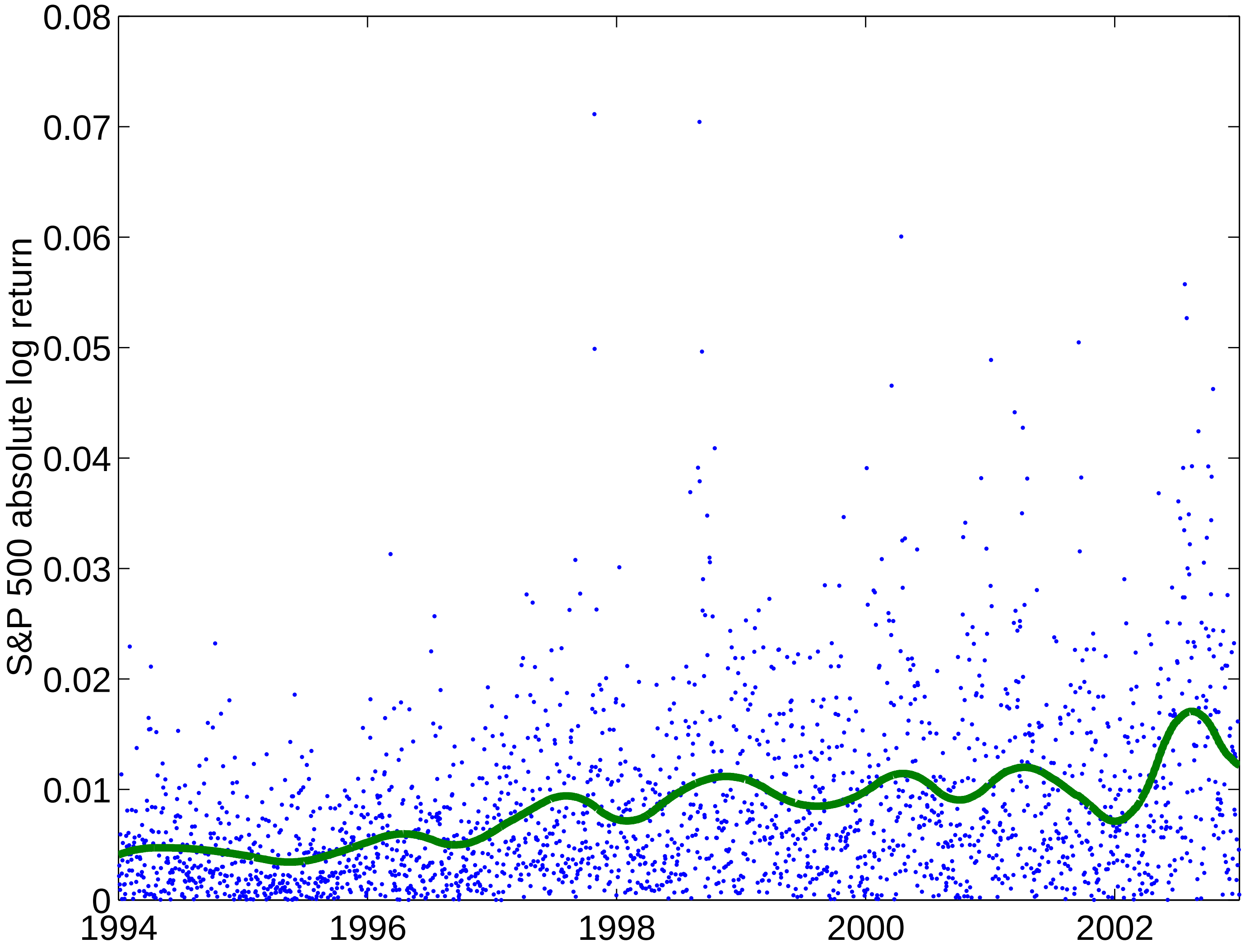
“The stock market, for instance, typically varies by about 15 percent to 20 percent a year”.

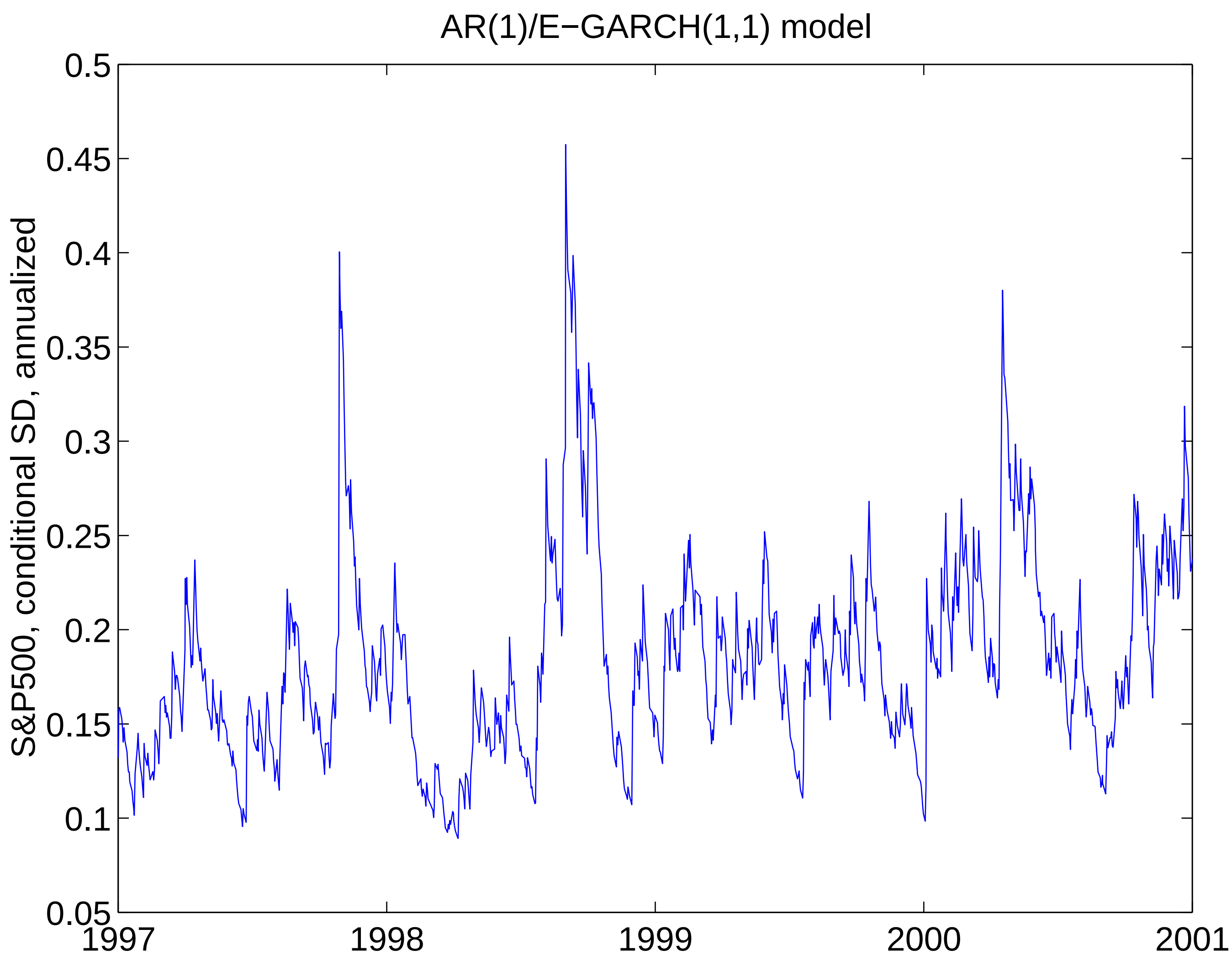
“And when the model told them that the markets were mispricing equity vol, they were willing to bet the firm on it.”

“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."


S&P 500 absolute log return
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
1994
1996
1998
2000
2002

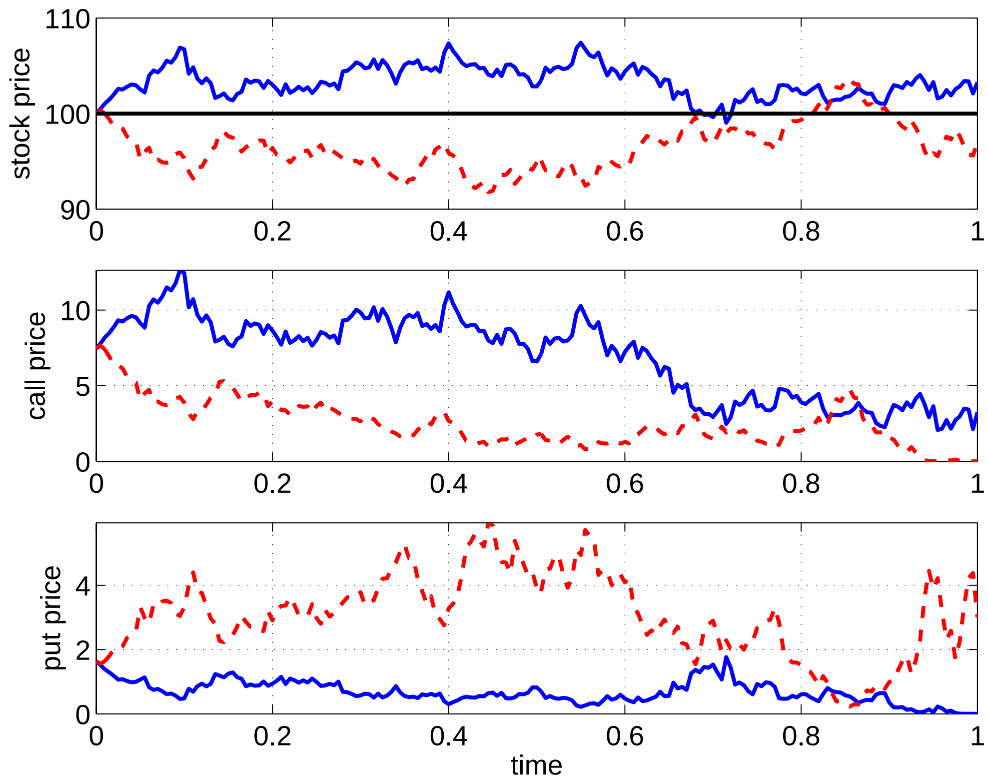


$$\textbf{Annualized SD} = \sqrt{253 * \textbf{daily variance}}$$

```
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 ;
```

nlag	p	q	E-GARCH	M-GARCH	AIC
1	2	2	no	no	−14,366
1	1	2	no	no	−15,002
1	1	2	yes	no	−15,105
1	1	2	yes	yes	−15,103
1	2	1	yes	no	−15,105
1	1	1	yes	no	−15,107
0	1	1	yes	no	−14,366
2	2	2	yes	no	−15,112
1	2	2	yes	no	−15,104
2	2	3	yes	no	−15,100
2	3	2	yes	no	−15,100
1	3	2	yes	no	−15,101

In fact, it sold insurance (options) both ways—against a sharp downturn and against a sharp rise.



The "Greeks"

$C(S, T, t, K, \sigma, r)$ = price of an option

$$\Delta = \frac{\partial}{\partial S} C(S, T, t, K, \sigma, r) \qquad \text{"Delta"}$$

$$\Theta = \frac{\partial}{\partial t} C(S, T, t, K, \sigma, r) \qquad \text{"Theta"}$$

$$\mathcal{R} = \frac{\partial}{\partial r} C(S, T, t, K, \sigma, r) \qquad \text{"Rho"}$$

$$\mathcal{V} = \frac{\partial}{\partial \sigma} C(S, T, t, K, \sigma, r) \qquad \text{"Vega"}$$

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."