Nonlinear Time Series Modeling

Part II: Time Series Models in Finance

Richard A. Davis
Colorado State University

(http://www.stat.colostate.edu/~rdavis/lectures)

MaPhySto Workshop

Copenhagen

September 27 — 30, 2004

Part II: Time Series Models in Finance

1. Classification of White Noise

As we have already seen from financial data, such as log(returns), and from residuals from some ARMA model fits, one needs to consider time series models for white noise (uncorrelated) that allows for dependence.

Classification of WN (in increasing degree of "whiteness").

W1: $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, i.e.

$$EZ_t = 0,$$

$$\gamma_Z(h) = \begin{cases} \sigma^2 & \text{if } h = 0, \\ 0 & \text{otherwise.} \end{cases}$$

$(P_{\overline{sp}\{Z_s, s<t\}} Z_t = 0.)$

1. Classification of White Noise (cont)

W2: $\{Z_t\}$ is a homoscedastic martingale difference.

$$E(Z_t | Z_s, s < t) = 0$$

$$\text{Var}(Z_t) = \sigma^2, \forall t.$$

W3: $\{Z_t\}$ is conditional white noise.

$$E(Z_t | Z_s, s < t) = 0$$

$$\text{Var}(Z_t | Z_s, s < t) = \sigma^2, \forall t.$$

(An ARCH(1) process is W2 but not W3.)

W4: $\{Z_t\}$ is strict white noise.

$$\{Z_t\} \sim \text{IID}(0, \sigma^2).$$

W5: $\{Z_t\}$ is Gaussian white noise.

$$\{Z_t\} \sim \text{IID N}(0, \sigma^2).$$

2. Examples

(1) All-pass processs. Satisfies W1 and not W2.

(2) A Deterministic AR Model:

Consider the deterministic autoregression,

$$Y_t = g(Y_{t-1}),$$

where,

$$g(y) = \begin{cases} 2y, & 0 \le y \le 1/2, \\ 2(1-y), & 1/2 \le y \le 1 \end{cases}$$

and

Y_0 has density $f(x) = 1, x \in [0, 1]$.

2. Examples (cont)

Properties:

1. $\{Y_t\}$ is strictly (and weakly) stationary with marginal density $f(x)$.

2. $\{Y_t\} \sim \mathsf{WN}(0.5, \mathsf{var}(\mathsf{Y}_0))$.

3. The BLP of Y_t is 0.5 with MSE$=$ var(Y_0). The BP of Y_t is $g(Y_{t-1})$ with MSE $= 0$.

4. Plotting Y_t vs. Y_{t-1} reveals the deterministic dynamics.

2. Examples (cont)

(3) Logistic Process

The logistic equation defines a sequence $\{x_n\}$, for any given x_0, via

$$x_n = 4x_{n-1}(1 - x_{n-1}), \quad 0 < x_0 < 1.$$

The values of x_n are, for even moderately large values of n, extremely sensitive to small changes in x_0, as can be seen from the solution,

$$x_n = \sin^2(2^n \arcsin(\sqrt{x_0}), \; n = 0, 1, 2, \ldots .$$

(Clearly a very small change δ in $\arcsin(\sqrt{x_0})$ leads to a change $2^n\delta$ in the argument of the $\sin$ function defining x_n.)

2. Examples (cont)

CHAOS.TSM: The series $\{x_1, \ldots, x_{200}\}$, (correct to nine decimal places) was generated from the logistic equation with $x_0 = \pi/10$. The calculation requires specification of x_0 to at least 70 decimal places and the use of correspondingly high precision arithmetic.

The sample ACF and the AICC criterion both suggest white noise with mean .4954 as a model for the series. Under this model the best linear predictor of X_{201} would be .4954. However the best predictor of X_{201} to nine decimal places is in fact $4x_{200}(1 - x_{200}) = 0.016286669$, with zero mean-squared error.

2. Examples (cont)

(4) ARCH(1) Model:

Stationary solution $\{Z_t\}$ of

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{IID N}(0,1),$$

where

$$h_t = \alpha_0 + \alpha_1 Z_{t-i}^2,$$

with $\alpha_0 > 0$ and $\alpha_1 \geq 0$. The name ARCH signifies autoregressive conditional heteroscedasticity. h_t is the conditional variance of Z_t given $\{Z_s, s < t\}$.

2. Examples (cont)

Iterating these equations gives

$$
\begin{aligned}
Z_t^2 &= \alpha_0 e_t^2 + \alpha_1 Z_{t-1}^2 e_t^2 \\
&= \alpha_0 e_t^2 + \alpha_1 \alpha_0 e_t^2 e_{t-1}^2 + \alpha_1^2 Z_{t-2}^2 e_t^2 e_{t-1}^2 \\
&= \cdots \\
&= \alpha_0 \sum_{j=0}^{n} \alpha_1^j e_t^2 e_{t-1}^2 \cdots e_{t-j}^2 + \\
&\quad \alpha_1^{n+1} Z_{t-n-1}^2 e_t^2 e_{t-1}^2 \cdots e_{t-n}^2 \\
&= \alpha_0 \sum_{j=0}^{\infty} \alpha_1^j e_t^2 e_{t-1}^2 \cdots e_{t-j}^2,
\end{aligned}
$$

provided $0 \le \alpha_1 < 1$ and $\{Z_t\}$ is stationary and causal (i.e., Z_t is a function of $\{e_s, s \le t\}$).

2. Examples (cont)

Solution of the ARCH(1) Equations:

If $0 \le \alpha_1 < 1$, the unique causal stationary solution of the ARCH(1) equations is given by

$$Z_t = e_t \sqrt{\alpha_0 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j e_{t-1}^2 \cdots e_{t-j}^2\right)}$$

It has the properties

$$E(Z_t) = E(E(Z_t | e_s, s < t)) = 0,$$

$$\text{Var}(Z_t) = \alpha_0 / (1 - \alpha_1),$$

and

$$E(Z_{t+h} Z_t) = E(E(Z_{t+h} Z_t | e_s, s < t + h)) = 0$$

for $h > 0$.

2. Examples (cont)

Properties of ARCH(1) process:

1. Strictly stationary solution if $0 < \alpha_1 < 1$.
2. $\{Z_t\} \sim \mathrm{WN}(0, \alpha_0/(1-\alpha_1))$.
3. Not IID since

$$E(Z_t^2|Z_{t-1}) = (\alpha_0 + \alpha_1 Z_{t-1}^2)E(e_t^2|Z_{t-1}) = \alpha_0 + \alpha_1 Z_{t-1}^2.$$

4. Not Gaussian.
5. Z_t has a symmetric distribution ($Z_1 =_d - Z_1$)
6. $EZ_t^4 < \infty$ if and only if $3\alpha_1^2 < 1$. (More on moments later.)
7. If $EZ_t^4 < \infty$, then the squared process $Y_t = Z_t^2$ has the same ACF as the AR(1) process

$$W_t = \alpha_1 W_{t-1} + e_t$$

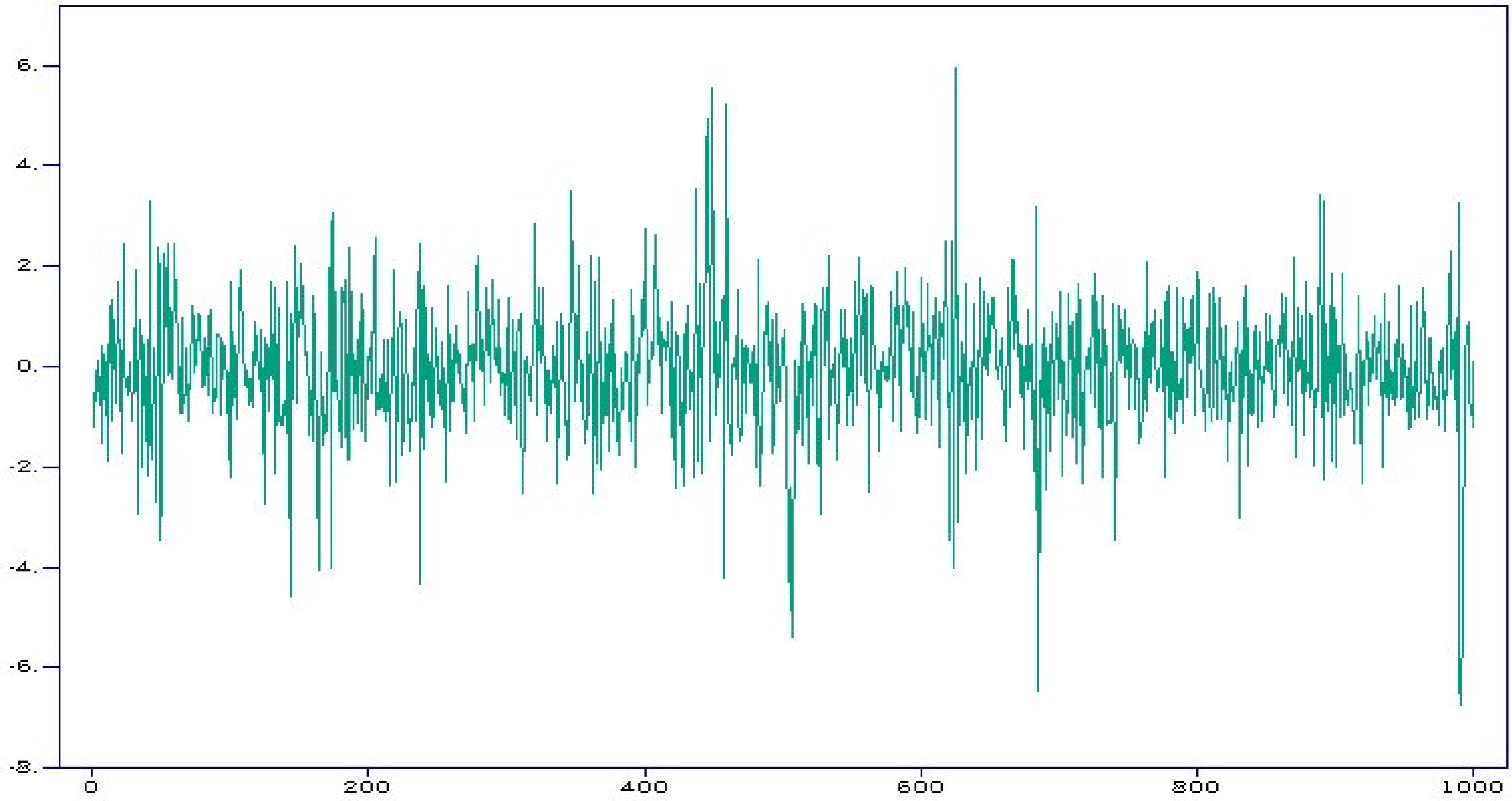
2. Examples (cont)

Likelihood function:

The ARCH(1) process is conditionally Gaussian, in the sense that for given Z_0, $\{Z_t, t \geq 1\}$ is Gaussian with known distribution. This makes it easy to write down the likelihood of $Z_1, \ldots, Z_n$ conditional on Z_0 and hence, by numerical maximization, to compute conditional maximum likelihood estimates of the parameters. For example, the conditional likelihood of the observations $\{z_1, \ldots, z_n\}$ of an ARCH(1) process given $Z_0 = z_0$ is

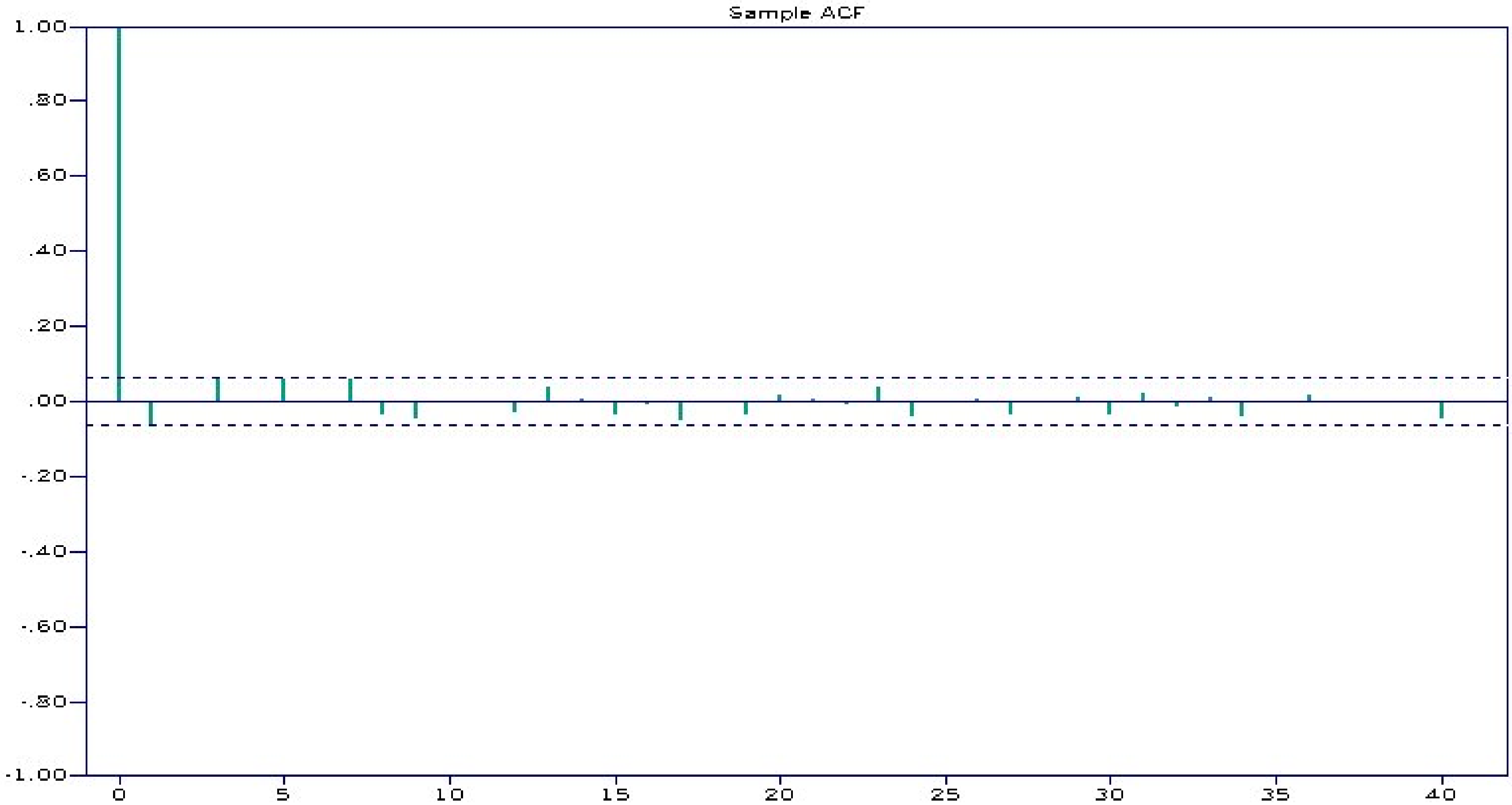
$$L = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 z_{t-1}^2)}} \exp\left\{-\frac{z_t^2}{2(\alpha_0 + \alpha_1 z_{t-1}^2)}\right\}.$$

2. Examples (cont)



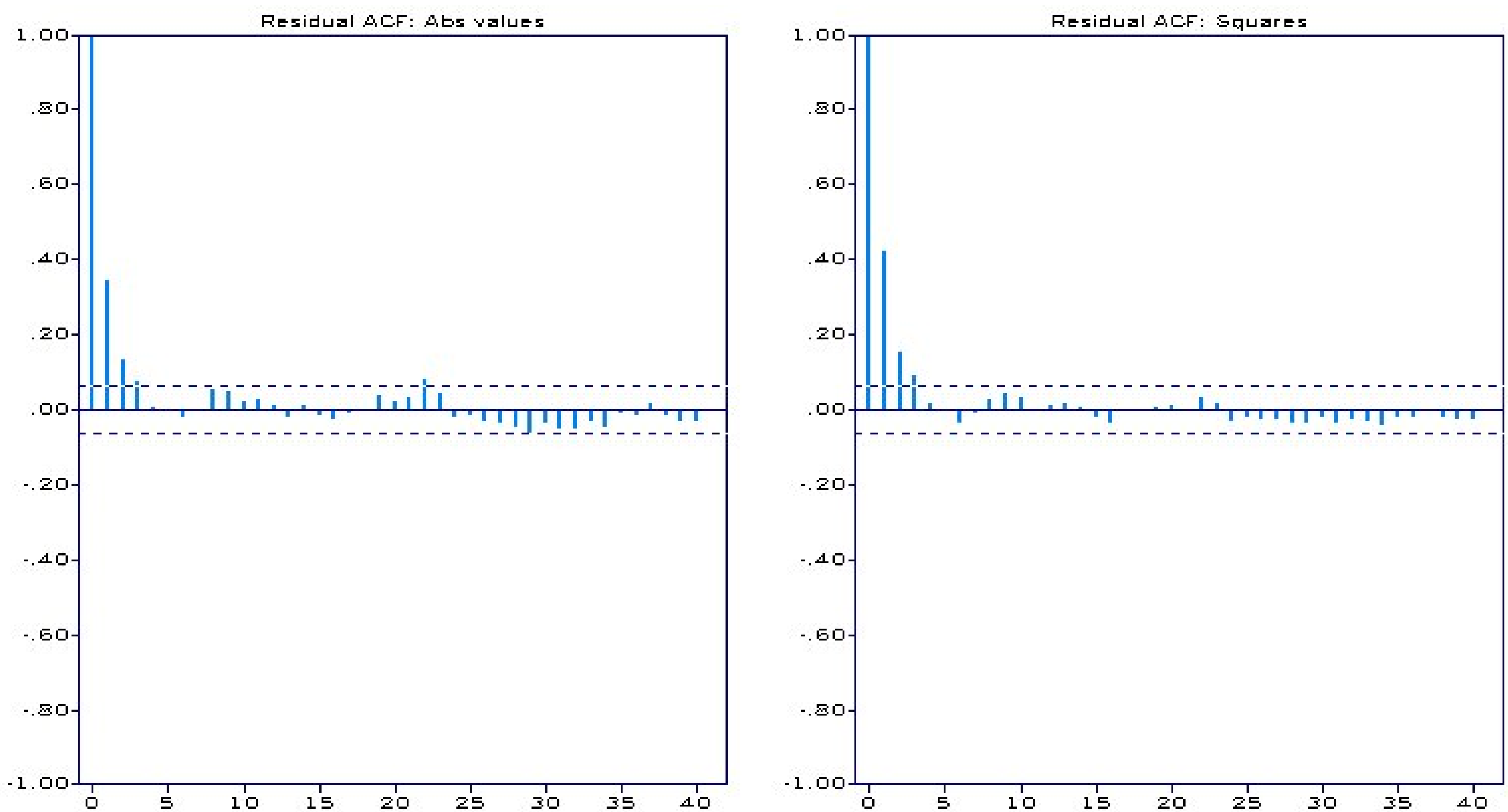
A realization of the process $Z_t = e_t\sqrt{1 + 0.5Z_{t-1}^2}$.

2. Examples (cont)



The sample ACF.

2. Examples (cont)



The sample ACF of the absolute values and squares.

(5) A Bilinear Model:

Consider the model,

$$Y_t = Z_t + .5Y_{t-2}Z_{t-1}, \quad \{Z_t\} \sim \mathrm{IID}(0, \sigma^2).$$

Assuming the existence of a causal stationary solution, $\{Y_t\}$, it has the following properties.

Properties:

1. $EY_t = 0$.

2. $\mathrm{Var}(Y_t) = \sigma^2(1 + .25\mathrm{Var}(Y_t))$. Hence

$$\gamma_Y(0) = \frac{\sigma^2}{1 - .25\sigma^2}, \quad \text{if } \sigma^2 < 4.$$

3. Assuming invertibility (i.e. $Z_t \in \mathcal{F}(Y_s, s \le t)$),

$$E(Y_t|Y_s, s < t) = .5Y_{t-2}Z_{t-1} \quad \text{and} \quad \mathrm{Var}(Y_t|Y_s, s < t) = \sigma^2.$$

3. "Stylized Facts" of Financial Returns

Define X_t = 100*(ln (P_t) - ln (P_{t-1})) (log returns)

- heavy tailed

$$P(|X_1| > x) \sim C\,x^{-\alpha}, \qquad 0 < \alpha < 4.$$

- uncorrelated

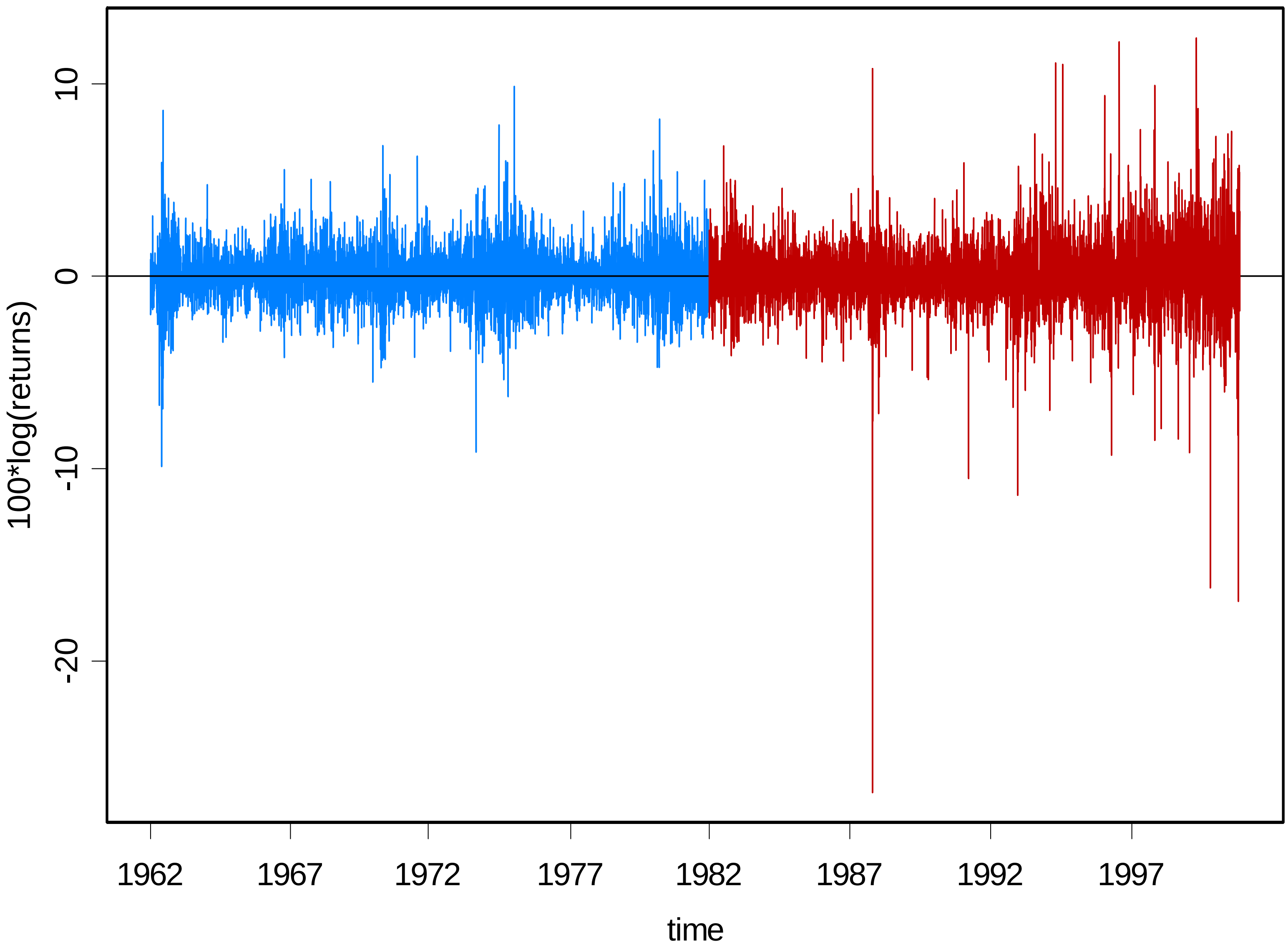
$\hat{\rho}_X(h)$ near 0 for all lags h > 0 (MGD sequence?)

- $|X_t|$ and X_t^2 have slowly decaying autocorrelations

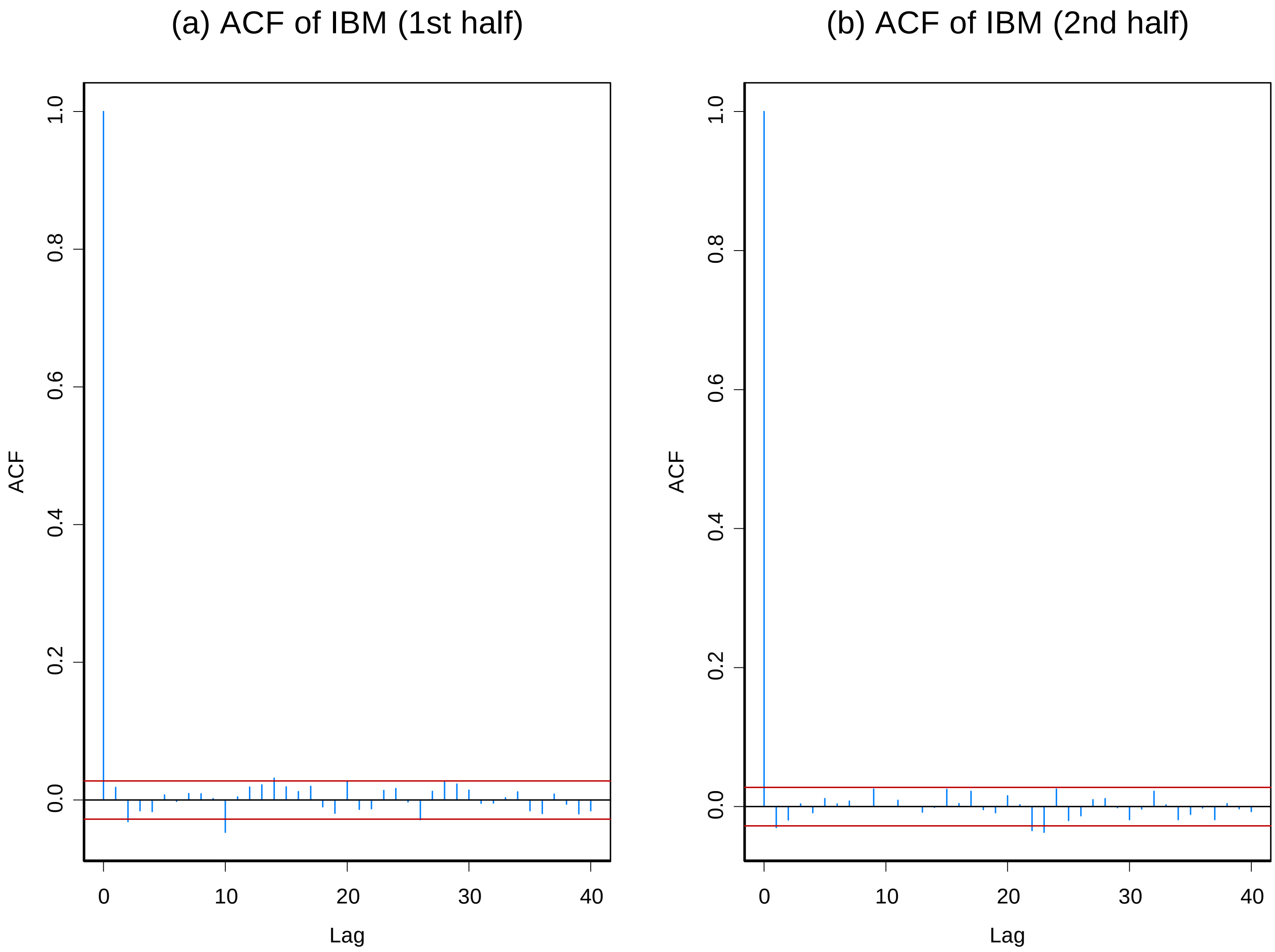
$\hat{\rho}_{|X|}(h)$ and $\hat{\rho}_{X^2}(h)$ converge to 0 *slowly* as h increases.

- process exhibits 'stochastic volatility'.

Log returns for IBM 1/3/62-11/3/00 (blue=1961-1981)

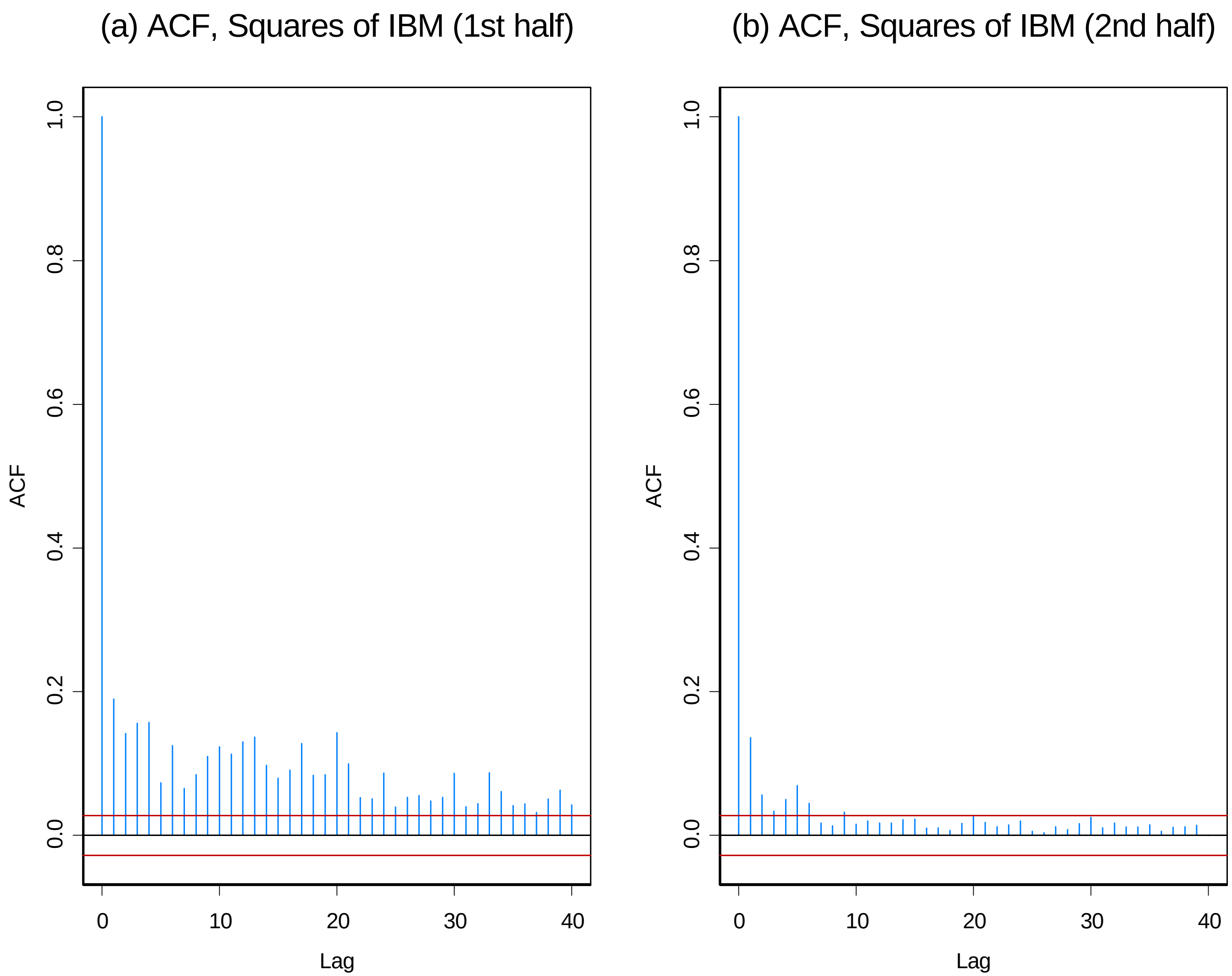


Sample ACF IBM (a) 1962-1981, (b) 1982-2000



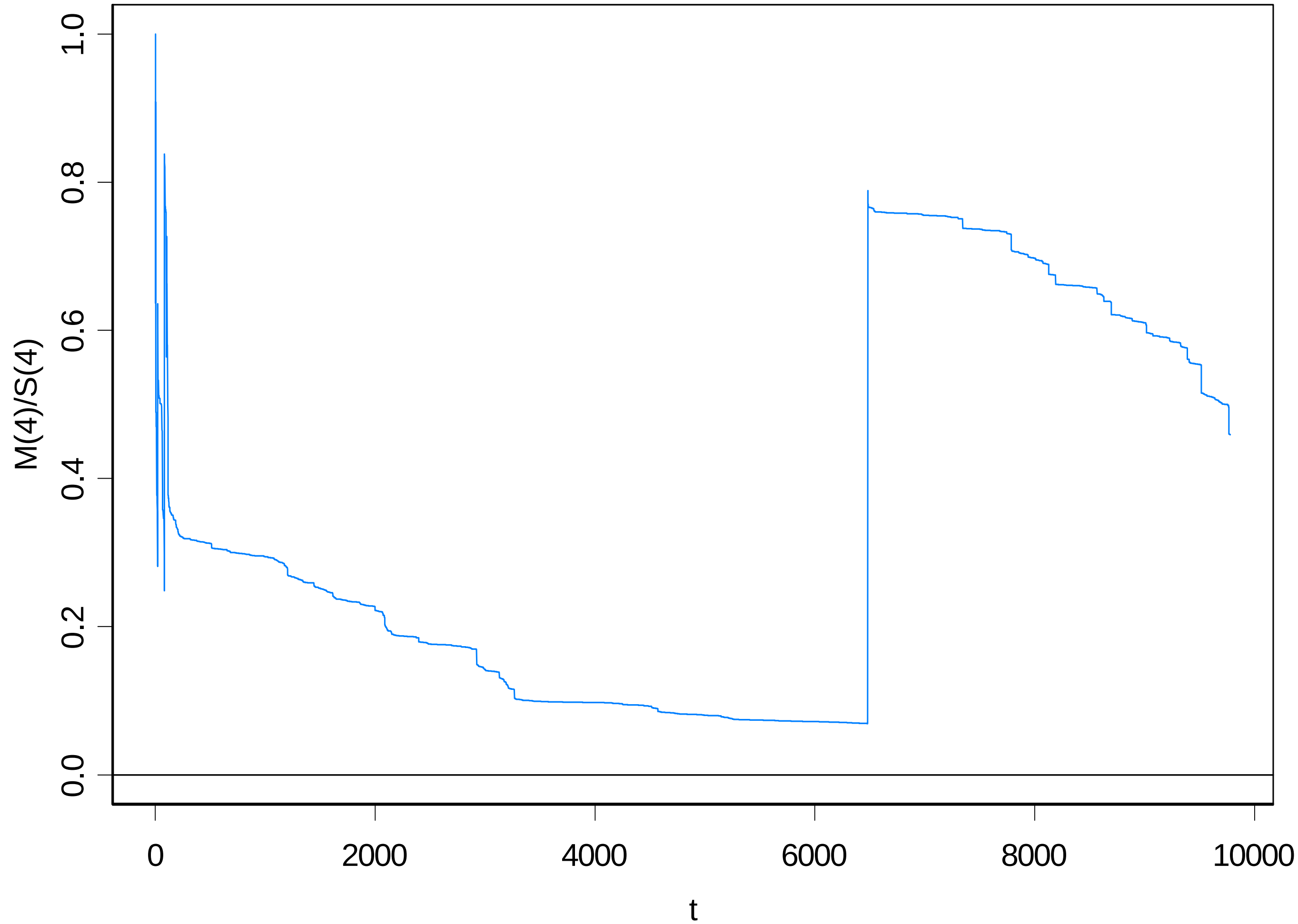
Remark: Both halves look like white noise.

ACF of squares for IBM (a) 1961-1981, (b) 1982-2000



Remark: Series are not independent white noise?

Plot of $M_t(4)/S_t(4)$ for IBM



Remark: For IID data, $M_t(k)/S_t(k) \to 0$ as $t \to \infty$ iff $E|X_1|^k < \infty$, where

$$M_t = \max_{s=1,\ldots,t} |X_s|^k \quad \text{and} \quad S_t = \sum_{s=1}^{t} |X_s|^k$$

Hill's estimator of tail index

The marginal distribution F for heavy-tailed data is often modeled using *Pareto-like tails*,

$$1-F(x) = x^{-\alpha}L(x),$$

for x large, where $L(x)$ is a slowly varying function ($L(xt)/L(x) \to 1$, as $x \to \infty$). Now if

$$X \sim F, \text{ then } P(\log X > x) = P(X > \exp(x)) = \exp(-\alpha x)L(\exp(x)),$$

and hence $\log X$ has an approximate exponential distribution for large x. The spacings,

$$\log(X_{(j)}) - \log(X_{(j+1)}),\ j=1,2,\ldots,m,$$

from a sample of size n from an exponential distr are approximately independent and $Exp(\alpha j)$ distributed. This suggests estimating α^{-1} by

$$\hat{\alpha}^{-1} = \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(j+1)}\right)j$$

$$= \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(m+1)}\right)$$

Hill's estimator of tail index

Def: The *Hill estimate* of α for heavy-tailed data with distribution given by

$$1-F(x) = x^{-\alpha}L(x),$$

is

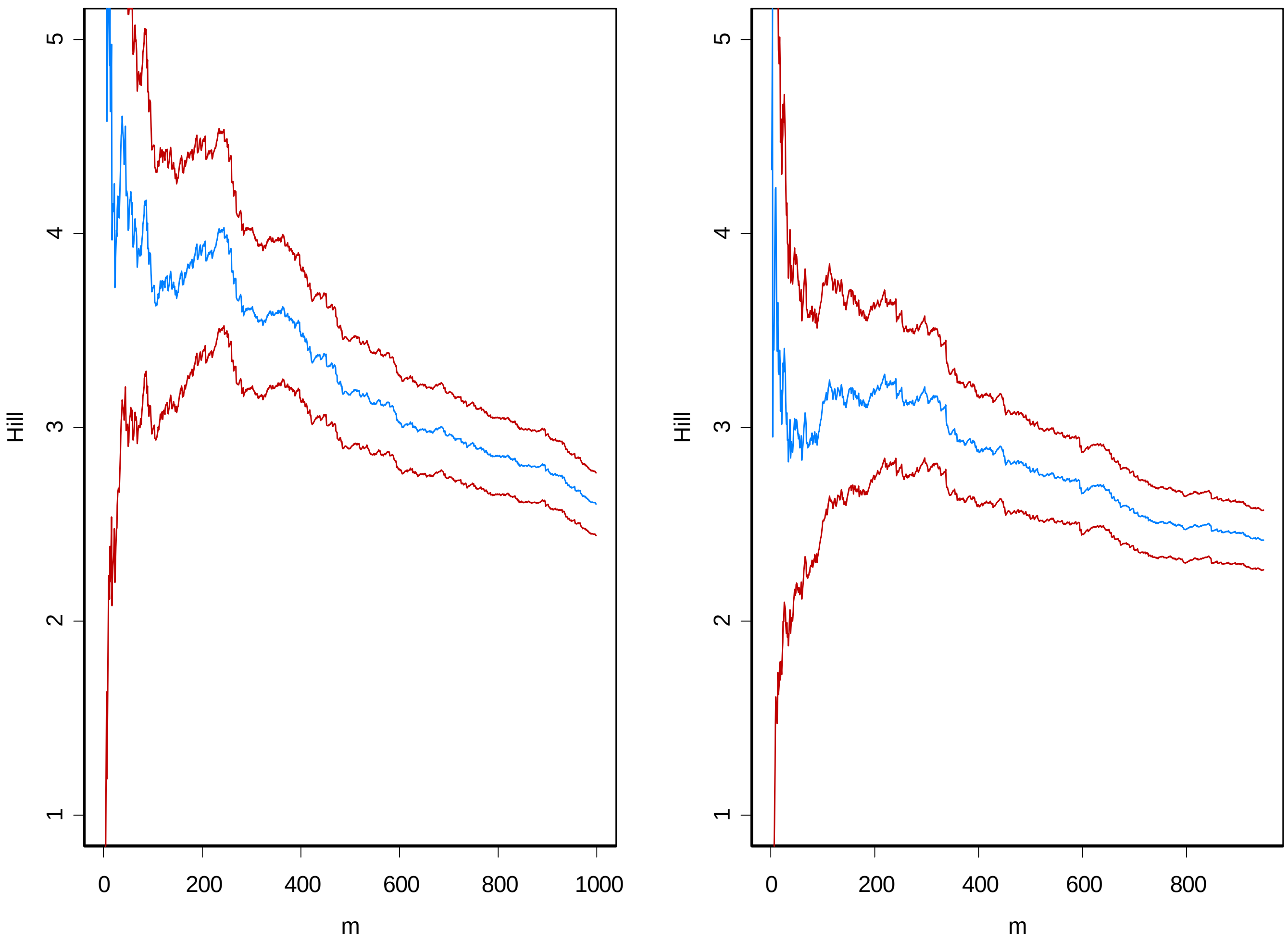
$$\hat{\alpha}^{-1} = \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(j+1)}\right)j$$
$$= \frac{1}{m}\sum_{j=1}^{m}\left(\log X_{(j)} - \log X_{(m+1)}\right)$$

The *asymptotic variance* of this estimate for α is

α^2/m and estimated by $\hat{\alpha}^2/m$.

(See also GPD=generalized Pareto distribution.)

Hill’s plot of tail index for IBM (1962-1981, 1982-2000)



4. ARCH and GARCH Models

ARCH(p) (Engle(1982))

$\{Z_t\}$ is a causal strictly and weakly stationary solution of

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \mathrm{IID}(0,1),$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2,$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0, \ i = 1, \ldots, p$.

ARCH = AutoRegressive Conditional Heteroscedasticity.

(See later for existence conditions.)

4. ARCH and GARCH Models (cont)

Properties:

If $\{Z_t\}$ is a causal stationary solution, then

1. $E(Z_t|Z_s, s < t) = E(h_t^{1/2} e_t|Z_s, s < t) = h_t^{1/2} E(e_t) = 0.$

2. $EZ_t = 0.$

3. $E(Z_s Z_t) = 0, s \neq t.$

4. $E(Z_t^2|Z_s, s < t) = h_t E(e_t^2|Z_s, s < t)$

$$= h_t E(e_t^2) = h_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2,$$

so h_t=conditional variance of Z_t given $Z_s, s < t$.

5. $E(Z_t^2) = Eh_t = \alpha_0 + \sum_{i=1}^p \alpha_i EZ_t^2, .$

so that $\quad EZ_t^2 = \alpha_0/(1 - \sum_{i=1}^p \alpha_i).$

Theorem 1 (Proof later)

(i) A NS condition for the existence of a causal SS and WS solution $\{Z_t\}$ of the ARCH(p) equations is

$$\sum_{i=1}^{p} \alpha_i < 1 \qquad (1)$$

and $\{Z_t\}$ is the unique such solution.

(ii) If (1) is satisfied and

$$(Ee_t^4)(\sum_{i=1}^{p} \alpha_i)^2 < 1, \qquad (2)$$

then $EZ_t^4 < \infty$.

Further Properties:

If conditions (1) and (2) are both satisfied then

6. $\{Z_t^2\}$ is an AR(p) process and all of its correlations are non-negative (generating persistence of volatility). To see this, note that $U_t = Z_t^2 - h_t$ is a MGD sequence and hence WN. It follows that

$$\begin{aligned} Z_t^2 &= Z_t^2 - h_t + h_t \\ &= U_t + \alpha_0 + \alpha_1 Z_{t-1}^2 + \cdots + \alpha_p Z_{t-p}^2. \end{aligned}$$

7. Z_t has heavier tails than e_t in the sense that its kurtosis $(EZ_t^4/(EZ_t^2)^2)$ is greater than or equal to that of e_t.

4. ARCH and GARCH Models (cont)

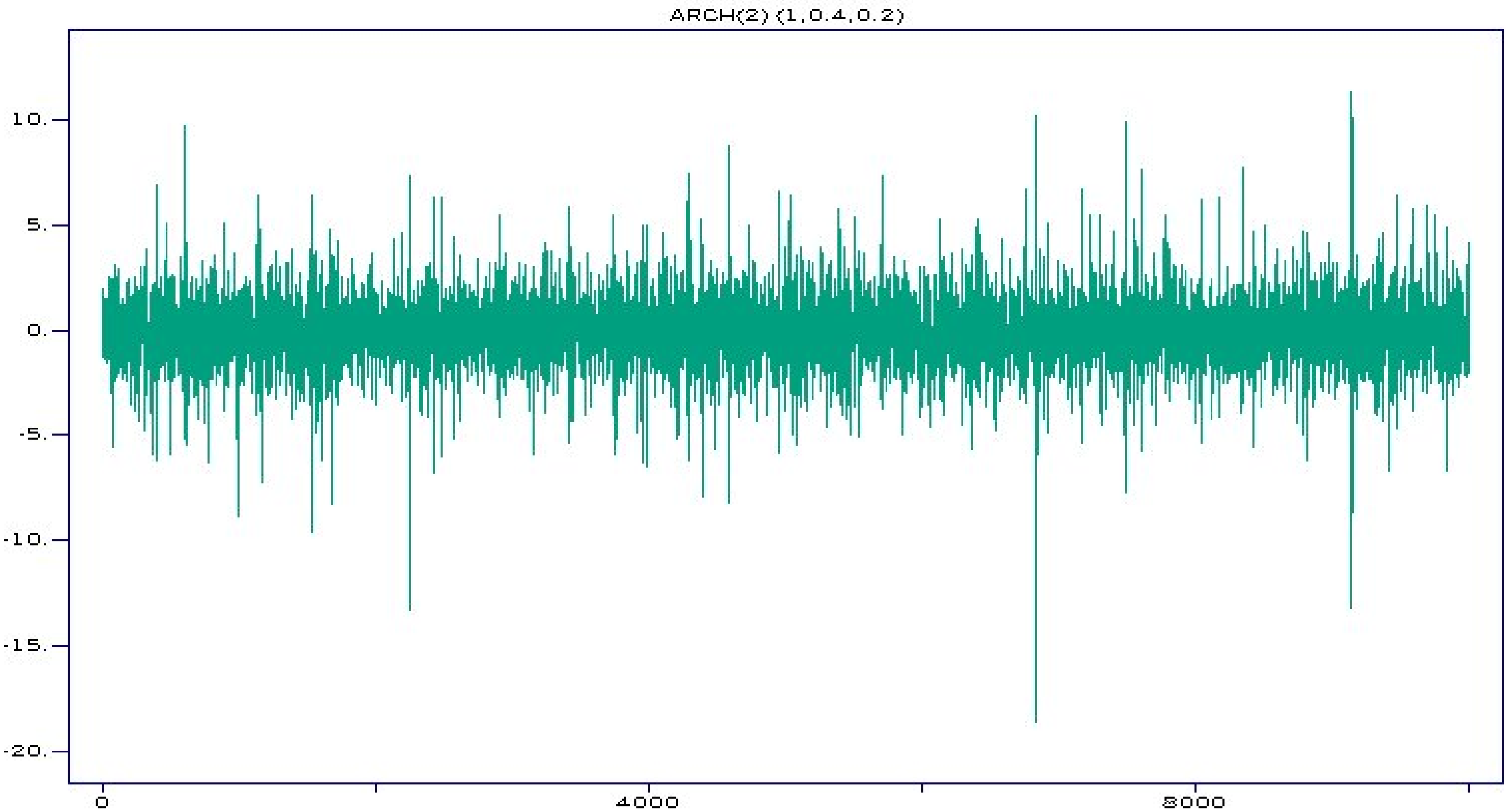
Illustration: The following graphs show the time-series plot, sample ACF and qq plot of 10000 simulated values of an ARCH(2) process $\{Z_t\}$ with

$$\alpha_0 = 1, \quad \alpha_1 = 0.4, \quad \alpha_2 = 0.2$$

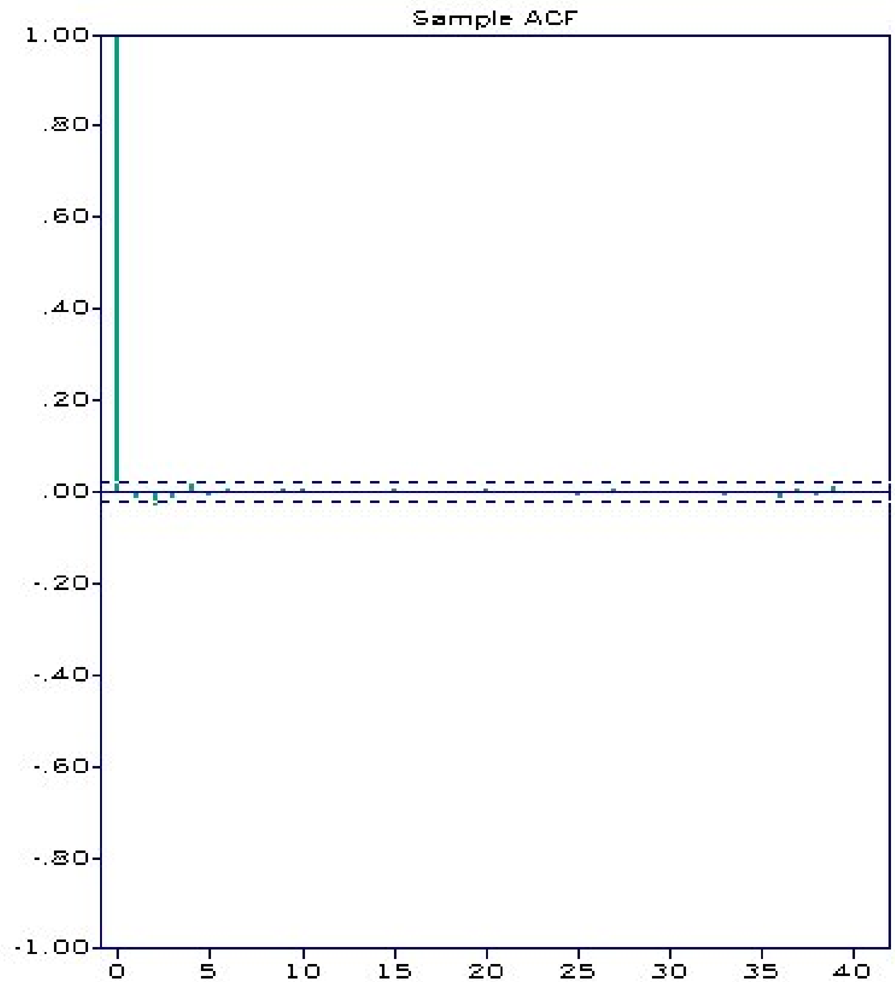
and Gaussian noise $\{e_t\}$. The sample ACF's of $\{|Z_t|\}$ and $\{Z_t^2\}$ are also shown.

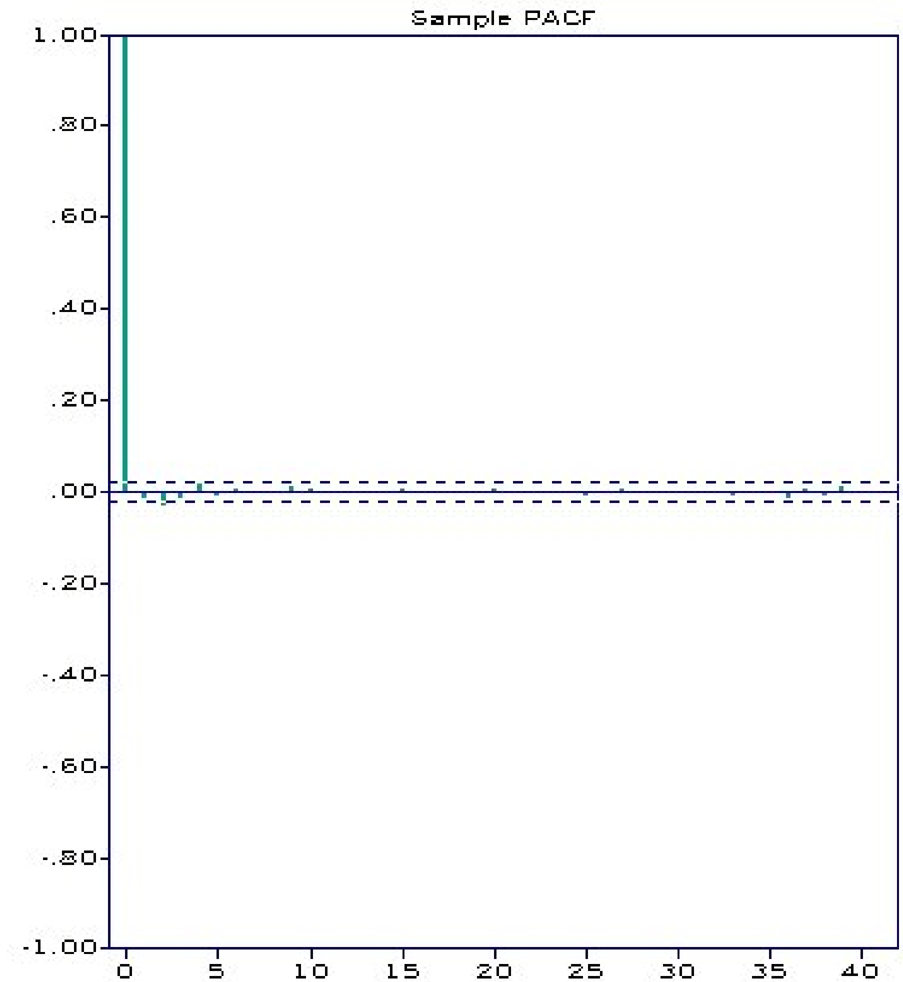
The sample ACF of $\{Z_t\}$, unlike that of $\{Z_t^2\}$, shows no evidence of the dependence in the series. For this particular ARCH process the model ACF of $\{Z_t^2\}$ has the values 0.5 at lag 1 and 0.4 at lag 2. These are in good agreement with the sample ACF shown.

4. ARCH and GARCH Models (cont)

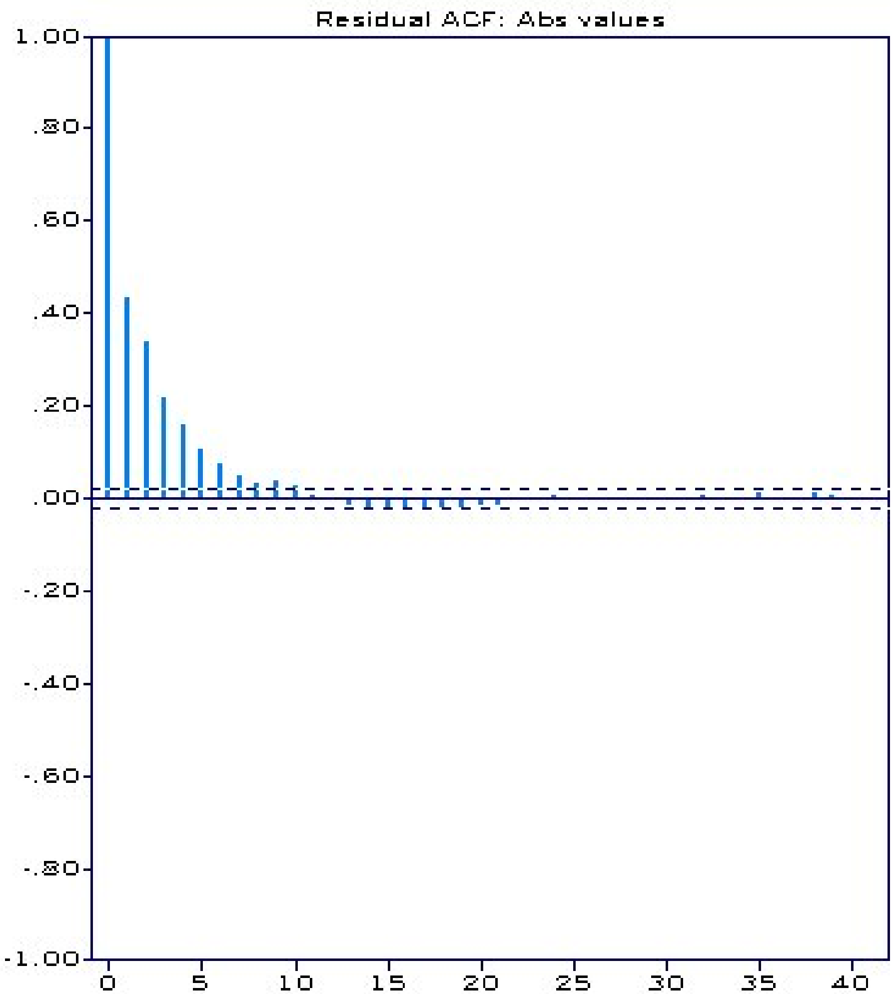


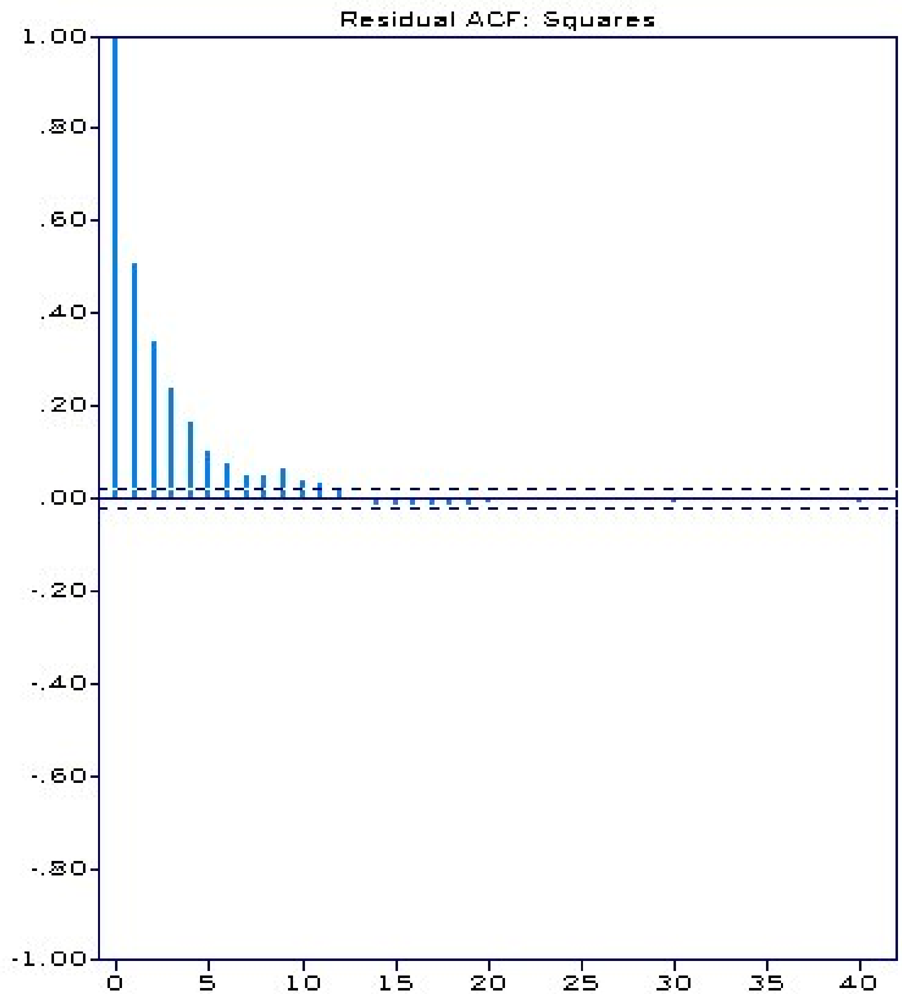
4. ARCH and GARCH Models (cont)



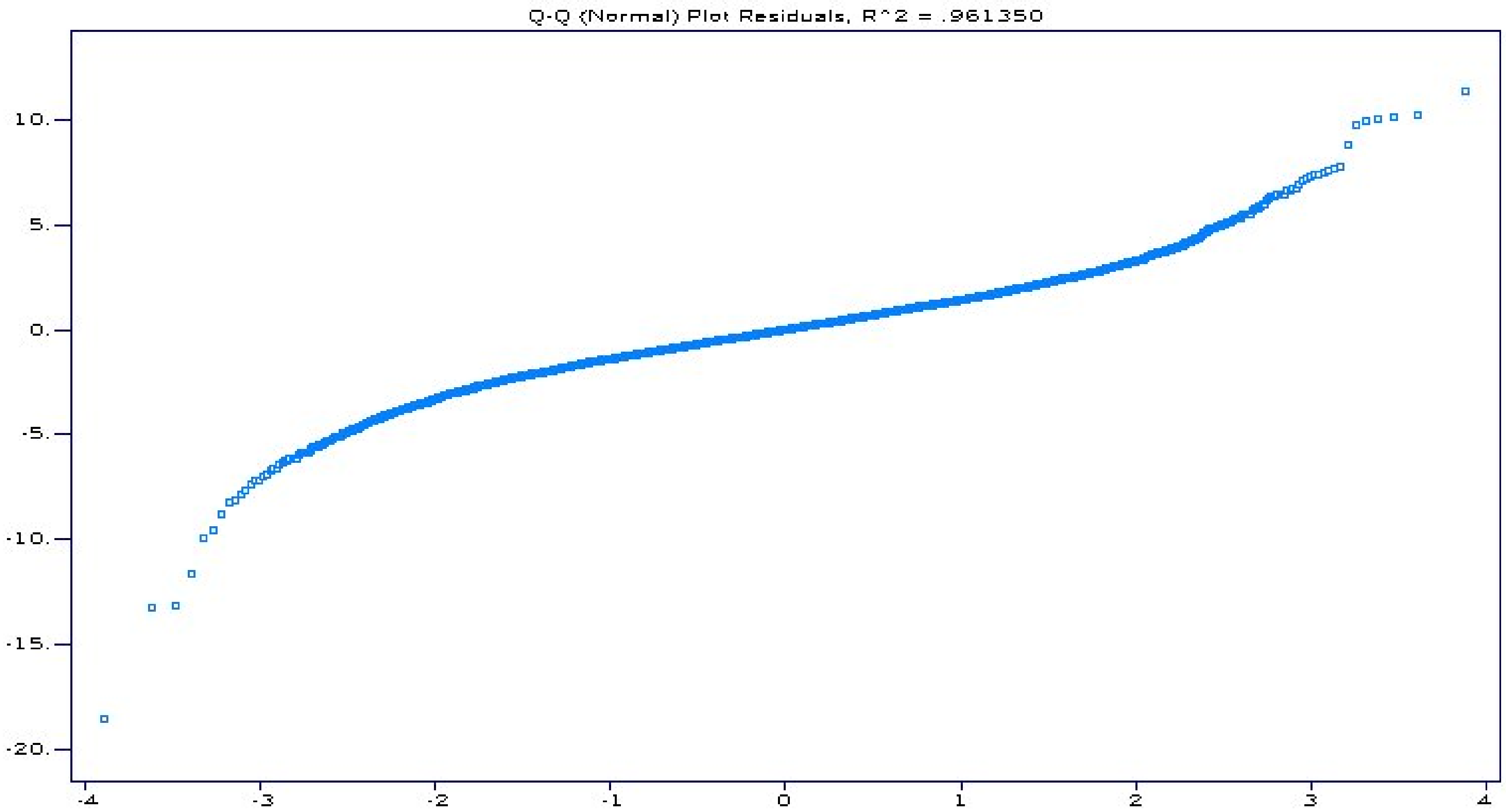


4. ARCH and GARCH Models (cont)





4. ARCH and GARCH Models (cont)



4. ARCH and GARCH Models (cont)

GARCH(p,q) (Bollerslev(1986))

$\{Z_t\}$ is a causal strictly and weakly stationary solution of

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \mathrm{IID}(0,1),$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i},$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for each i.

GARCH = Generalized ARCH.

(See later for existence conditions.)

4. ARCH and GARCH Models (cont)

If $\Sigma\beta_i < 1$ and $EZ_t^2 < \infty$:

then $1 - \sum_{i=1}^{q} \beta_i z^i \neq 0 \quad \forall |z| \leq 1$ and so

$$\frac{1}{1 - \sum \beta_i z^i} = \sum_{i=0}^{\infty} \xi_i z^i, \quad |z| \leq 1,$$

where $\quad \xi_0 = 1, \quad \xi_j \geq 0 \quad \forall j, \quad \sum |\xi_j| < \infty.$

Hence

$$\begin{aligned} h_t &= \xi(B)(\alpha_0 + \sum^{p} \alpha_i B^i) Z_{t-i}^2 \\ &= \psi_0 + \sum_{j=1}^{\infty} \psi_j Z_{t-j}^2, \end{aligned}$$

where $\psi_0 = \xi(B)\alpha_0 = \alpha_0/(1 - \sum \beta_i)$. In particular we see that $h_t \in \sigma(e_s, s < t)$ and $\{Z_t\}$ is an ARCH(∞) process.

4. ARCH and GARCH Models (cont)

Properties:

If $\{Z_t\}$ is causal and stationary and $\sum_{i=1}^{q} \beta_i < 1$, then

1. $E(Z_t | Z_s, s < t) = E[h_t^{1/2} e_t | Z_s, s < t] = h_t^{1/2} E(e_t) = 0.$

2. $EZ_t = 0.$

3. $E(Z_s Z_t) = 0, s \neq t.$

4. $E(Z_t^2 | Z_s, s < t) = E(h_t e_t^2 | Z_s, s < t)$

$$= h_t E e_t^2 = h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i},$$

so h_t=conditional variance of Z_t given $Z_s, s < t$.

5. $E(Z_t^2) = Eh_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i EZ_t^2 + \sum_{i=1}^{q} \beta_i Eh_t,$

so that $\quad EZ_t^2 = Eh_t = \alpha_0 / (1 - \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{q} \beta_i).$

4. ARCH and GARCH Models (cont)

6. If $EZ_t^4 < \infty$, then $\{Z_t^2\}$ follows an ARMA(m, q) process with $m = \max(p, q)$.

$$
\begin{aligned}
Z_t^2 &= Z_t^2 - h_t + h_t \\
&= U_t + \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \\
&= U_t + \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) Z_{t-i}^2 - \sum_{i=1}^{q} \beta_i (Z_{t-i}^2 - h_{t-i}) \\
&= U_t + \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) Z_{t-i}^2 - \sum_{i=1}^{q} \beta_i U_{t-i}
\end{aligned}
$$

Theorem 2 (ARCH(∞)) (see Giraitis, Kokoszka and Leipus (2000) *Economic Theory*, 16, 3-22).

For the equations,

$$Z_t^2 = h_t e_t^2, \quad \{e_t\} \sim \mathrm{IID}(0,1),$$

$$h_t = \psi_0 + \sum_{i=1}^{\infty} \psi_i Z_{t-i}^2, \quad \psi_j \geq 0 \ \forall j,$$

(i) there is a unique causal finite-mean SS solution $\{Z_t^2\}$

(with $EZ_t^2 = \psi_0/(1 - \sum_{i=1}^{\infty} \psi_i)$) if

$$\sum_{i=1}^{\infty} \psi_i < 1. \qquad (1)$$

(ii) If (1) is satisfied and

$$(Ee_t^4)(\sum_{i=1}^{\infty} \psi_i)^2 < 1, \qquad (2)$$

 then $EZ_t^4 < \infty$.

Theorem 3 (GARCH(p,q)) (see Bollerslev (1986)). The equations,

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \mathrm{IID}(0,1),$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i},$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for each i,

have a causal weakly stationary solution if and only if

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

There is exactly one such solution.

SS but not WS GARCH Processes

Stochastic recurrence equations

To examine conditions for existence of SS but not necessarily WS causal solutions of the GARCH equations it is convenient to consider a general class of equations defined by

$$\mathbf{Y}_t = A_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \qquad (1)$$

where $\{(A_t, \mathbf{B}_t)\}$ is an iid sequence, $\mathbf{Y}_t$, $\mathbf{B}_t$ are $d \times 1$ random vectors, and A_t is a $d \times d$ random matrix.

Embrechts, Klüppelberg and Mikosch, *Modelling Extremal Events*, Springer (1997)

Davis and Mikosch, in *Nonlinear and Nonstationary Signal Processing*, eds Fitzgerald et al., Cambridge University Press, 2000.

4. ARCH and GARCH Models (cont)

Examples

(i) ARCH(1)

$$Z_t^2 = (\alpha_0 + \alpha_1 Z_{t-1}^2)e_t^2 = (\alpha_1 e_t^2) Z_{t-1}^2 + \alpha_0 e_t^2.$$

These equations is equivalent to the stochastic recurrence equation (1)

$$Y_t = A_t Y_{t-1} + B_t,$$

with

$$d = 1,$$

$$Y_t = Z_t^2,$$

$$A_t = \alpha_1 e_t^2,$$

$$B_t = \alpha_0 e_t.$$

(ii) GARCH(1,1)

$$
\begin{aligned}
Z_t^2 &= h_t e_t^2 = (\alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 h_{t-1}) e_t^2 \\
h_t &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 h_{t-1}.
\end{aligned}
$$

The required SRE is two-dimensional which can be expressed as

$$
\begin{bmatrix} Z_t^2 \\ h_t \end{bmatrix} = \begin{bmatrix} \alpha_1 e_t^2 & \beta_1 e_t^2 \\ \alpha_1 & \beta_1 \end{bmatrix} \begin{bmatrix} Z_{t-1}^2 \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_0 e_t^2 \\ \alpha_0 \end{bmatrix}
$$

Alternatively, a one-dimensional SRE can be developed for the volatility process h_t. Note that

$$
h_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 h_{t-1} = \alpha_0 + \alpha_1 h_{t-1} e_{t-1}^2 + \beta_1 h_{t-1}
$$

so that the SRE has the form

$$
h_t = (\alpha_1 e_{t-1}^2 + \beta_1) h_{t-1} + \alpha_0.
$$

In this case,

$$
\mathbf{Y}_t = h_t, \quad A_t = \alpha_1 e_{t-1}^2 + \beta_1, \quad \mathbf{B}_t = \alpha_0.
$$

4. ARCH and GARCH Models (cont)

(iii) GARCH(p, q)

$$Z_t = \sqrt{h_t} e_t, \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}.$$

These equations are equivalent to $Z_t = \sqrt{h_t} e_t$ and the stochastic recurrence equation (1) with

$$d = p + q,$$

$$\mathbf{Y}_t = (Z_t^2, \ldots, Z_{t-p+1}^2, h_t, \ldots, h_{t-q+1})'$$

$$\mathbf{B}_t = (\alpha_0 e_t^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)'.$$

$$A_t = \begin{bmatrix} \alpha_1 e_t^2 & \cdots & \alpha_p e_t^2 & \beta_1 e_t^2 & \cdots & \beta_q e_t^2 \\ 1 & & & & & \\ & \ddots & & & & \\ & & 1 \quad 0 & & & \\ \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \quad 0 \end{bmatrix}$$

(iv) A Bilinear Model

$$X_t = aX_{t-1} + bX_{t-1}Z_{t-1} + Z_t, \quad \{Z_t\} \sim \mathrm{IID}(0, \sigma^2).$$

Define $Y_t = aX_t + bX_tZ_t$. Then $\{Y_t\}$ satisfies the SRE (1) with

$$d = 1,$$

$$A_t = a + bZ_t,$$

$$\mathbf{B}_t = (a + bZ_t)Z_t.$$

Theorem 4 (Causal SS solution of (1))

If the following three conditions are satisfied:

(i) $E \ln^+ ||A_1|| < \infty$,

(ii) $E \ln^+ |\mathbf{B}_1| < \infty$,

(iii) $\gamma := \inf\{n^{-1} E \ln ||A_1 \cdots A_n||, n \in \mathbb{N}\} < 0$,

then

$$\mathbf{Y}_t = \mathbf{B}_t + \sum_{k=1}^{\infty} A_t \cdots A_{t-k+1} \mathbf{B}_{t-k} \qquad (2)$$

is the unique (in distribution) causal strictly stationary solution of (1) (and the series converges with probability one).

Note: $||A|| := \sup_{|\mathbf{x}|=1} |A\mathbf{x}|$, where $|\mathbf{x}|$ is the Euclidean norm of the vector $\mathbf{x}$.

Under condition (i), γ for the sequence $\{A_n\}$ can also be expressed as $\gamma = \lim_{n\to\infty} \frac{1}{n} \ln ||A_1 \cdots A_n||$.

Corollary (Case $d = 1$)

If $d = 1$ and $E \ln^+ |A_1| < \infty$, then $E \ln |A_1| < \infty$ and so, by the strong law of large numbers,

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln |A_1 \cdots A_n| = E \ln |A_1|.$$

(Unfortunately there is no simple explicit expression for γ when $d > 1$.)

Hence if $d = 1$, $-\infty \leq E \ln |A_1| < 0$, and $E \ln^+ |B_1| < \infty$, then the unique SS causal solution of (1) is given by (2).

Theorem 5 (GARCH(1,1) with $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$, and $P(e_t = 0) = 0$)

A necessary and sufficient condition for $\gamma < 0$, and hence the existence of a unique causal SS solution of the GARCH equations, is

$$E \ln(\alpha_1 e_t^2 + \beta_1) < 0. \tag{3}$$

Remarks:

(i) Jensenizing the preceding inequality, we see that

$$E \ln(\alpha_1 e_t^2 + \beta_1) \leq \ln E(\alpha_1 e_t^2 + \beta_1) = \ln(\alpha_1 + \beta_1) < 0.$$

This latter condition is equivalent to the NS condition for weak stationarity, $\alpha_1 + \beta_1 < 1$.

(ii) For an ARCH(1) model, the NS condition for $\gamma < 0$ (and hence the existence of a strictly stationary condition) is

$$E \ln(\alpha_1 e_t^2) < 0\,,$$

which is equivalent to

$$\alpha_1 < \exp\{-E(\ln e_t^2)\} = 2e^E \approx 3.568\ldots\,,$$

where E is Euler's constant. Jensenizing, we recover the NS condition for WS, i.e., $\alpha_1 < 1$.

Parameter Estimation for Finite-Variance GARCH Models

Our model is

$$Z_t = \sqrt{h_t}\ e_t, \quad \{e_t\} \sim \mathrm{IID}(0,1),$$

with

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i},$$

$\alpha_0 > 0$, $\alpha_j, \beta_j \geq 0$ for $j \geq 1$, and

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

For modeling purposes it is usually assumed in addition that either

$$e_t \ \sim \ N(0,1),$$

or that

$$\sqrt{\frac{\nu}{\nu-2}} e_t \ \sim \ t_\nu, \quad \nu > 2,$$

where t_ν denotes Student's t-distribution with ν degrees of freedom. Other distributions for e_t can however be used.

4. ARCH and GARCH Models (cont)

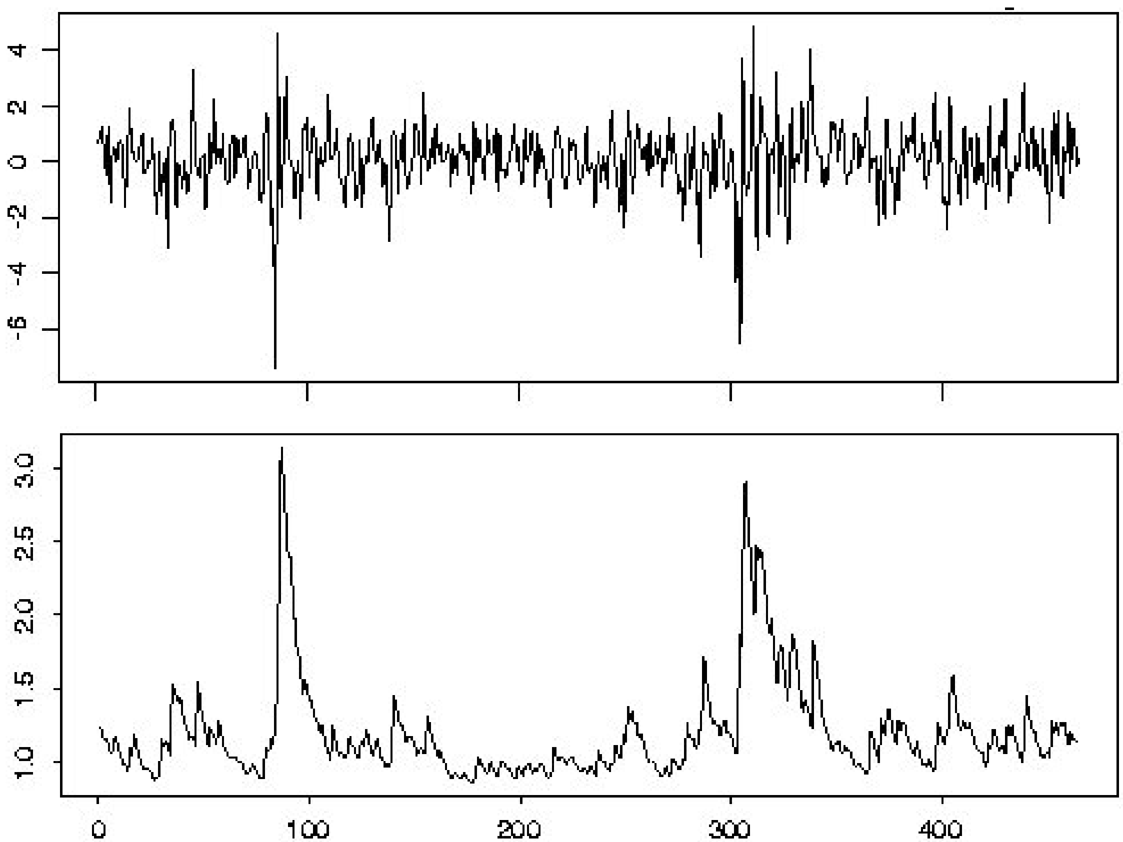
Note:

In the analysis of empirical financial data such as percentage daily stock returns (defined as $100 \ln(P_t/P_{t-1})$, where P_t is the closing price on trading day t), it is often found that better fits to the data are obtained by using the heavier-tailed Student's t-distribution for the distribution of Z_t given $\{Z_s, s < t\}$.

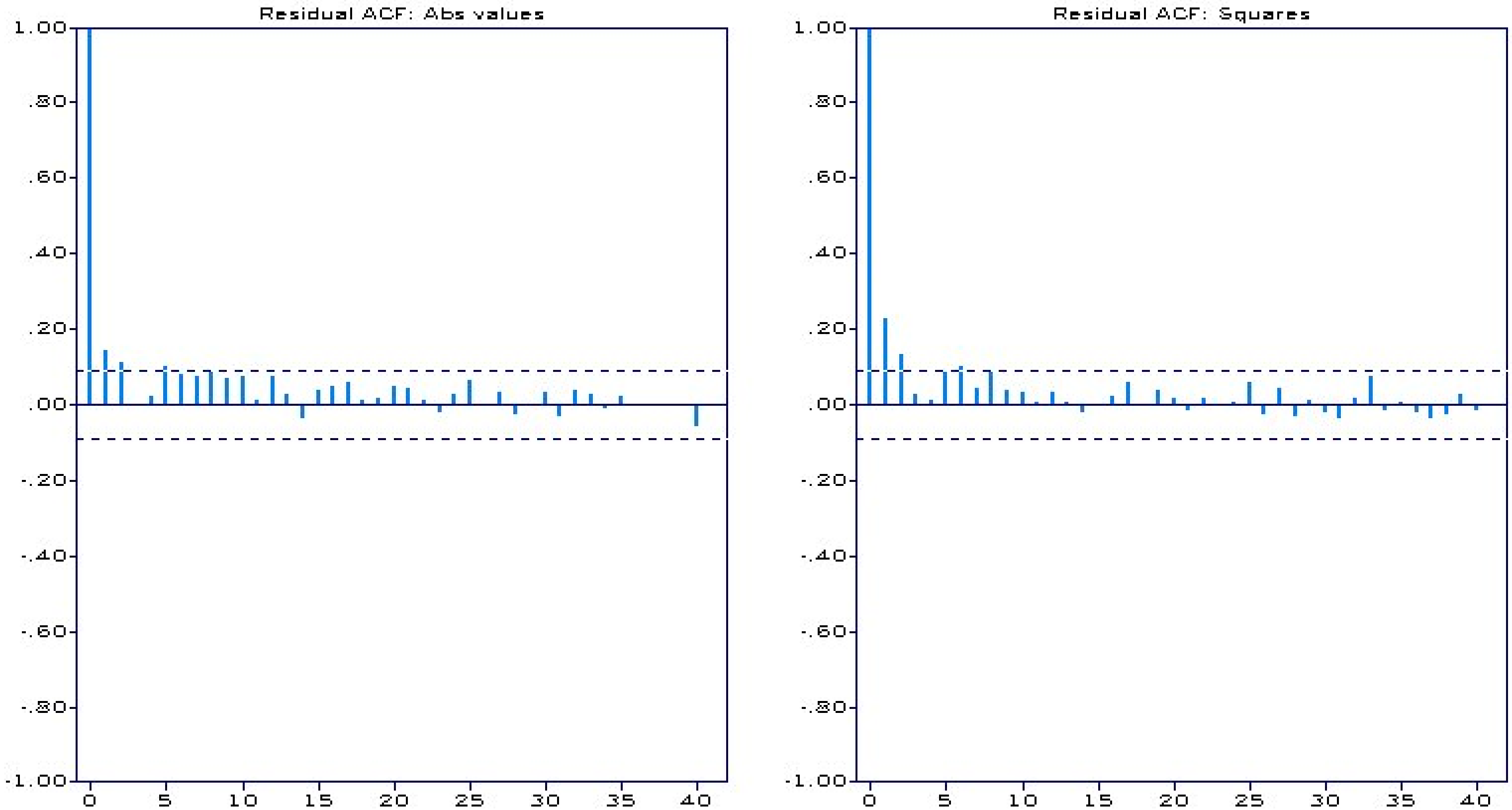
The "persistence of volatility" (large (small) fluctuations in the data tend to be followed by fluctuations of comparable magnitude) is reflected by GARCH models through the correlation in the sequence $\{h_t\}$ of conditional variances.

Example (Fitting a GARCH model to stock returns.)

The top graph shows the percentage daily returns of the Dow-Jones Industrial Index for the period July 1st, 1997, through April 9th, 1999, contained in the file E1032.TSM. The graph suggests that there are sustained periods of both high volatility (in October, 1997, and August, 1998) and of low volatility in between.



The sample autocorrelation function of this series, has very small values, however the sample autocorrelations of the absolute values and squares of the data are significantly different from zero, indicating dependence in spite of the lack of autocorrelation. These properties suggest that an ARCH or GARCH model might be appropriate for this series.



4. ARCH and GARCH Models (cont)

The model,

$$Y_t = a + Z_t,$$

where $\{Z_t\}$ is Gaussian-driven GARCH(p, q) can be fitted using ITSM as follows.

Open the project E1032.TSM and click on the red button labeled GAR at the top of the ITSM screen. In the resulting dialog box enter the desired values of p and q, e.g. 1 and 1 for GARCH(1,1).

With `Use normal noise` selected, click on `OK` and then click on the red MLE button. Subract the sample mean, which will be used as the estimate of a (unless you wish to assume that the parameter a is zero).

The `GARCH Maximum Likelihood Estimation` box will then open. When you click on `OK` the conditional likelihood maximization will proceed.

4. ARCH and GARCH Models (cont)

Denoting by $\{\hat{Z}_t\}$ the (possibly) mean-corrected observations, the program ITSM maximizes the likelihood of $\hat{Z}_{p+1}, \ldots, \hat{Z}_n$ conditional on the known values $\hat{Z}_1, \ldots, \hat{Z}_p$, and with assumed values 0 for each $\hat{Z}_t$, $t \leq 0$, and $\hat{\sigma}^2$ for each h_t, $t \leq 0$, where $\hat{\sigma}^2$ is the sample variance of $\{\hat{Z}_1, \ldots, \hat{Z}_n\}$. In other words the program maximizes

$$L(\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) = \prod_{t=p+1}^{n} \frac{1}{\sigma_t} \phi\left(\frac{\hat{Z}_t}{\sigma_t}\right),$$

with respect to the coefficients $\alpha_0, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$, where ϕ denotes the standard normal density, and the standard deviations $\sigma_t = \sqrt{h_t}$, $t \geq 1$, are computed from the GARCH recursions with Z_t replaced by $\hat{Z}_t$, and with $\hat{Z}_t = 0$ and $h_t = \hat{\sigma}_t^2$ for $t \leq 0$.

4. ARCH and GARCH Models (cont)

Comparison of models with different orders p and q can be made with the aid of the AICC, which is defined in terms of the conditional likelihood L as

$$\text{AICC} := -2\frac{n}{n-p}\ln L + 2(p+q+2)n/(n-p-q-3).$$

The factor $\frac{n}{n-p}$ multiplying the first term on the right has been introduced to correct for the fact that the number of factors in he conditional likelihood is only $n-p$. Notice also that the GARCH(p,q) model has $p+q+1$ coefficients.

Estimated mean:

$$\hat{a} = 0.0608$$

Minimum-AICC Gaussian GARCH model for $\hat{Z}_t = Y_t - \hat{a}$:
GARCH(1,1) with

$$\hat{\alpha}_0 = 0.1300, \hat{\alpha}_1 = 0.1266, \hat{\beta}_1 = 0.7922,$$

AICC value = 1469.02.

The bottom graph shown earlier shows the corresponding estimated conditional standard deviations, $\hat{\sigma}_t = \sqrt{\hat{h}_t}$, which clearly reflect the changing volatility of the series $\{Y_t\}$. This graph is obtained from ITSM by clicking on the red SV (stochastic volatility) button.

Model-checking:

Under the fitted model, the GARCH residuals, $\{\hat{Z}_t/\hat{\sigma}_t\}$, should be approximately IID N(0,1).

Check independence: Sample ACF's of the absolute values and squares of the residuals (fifth red button at the top of the ITSM window) look OK.

Check normality: `Garch>Garch residuals> QQ-Plot(normal)` should give approximately a straight line through the origin with slope 1. But deviations are large for large values of $|\hat{Z}_t|$, suggestng a heavier-tailed model, e.g. one with conditional t-distribution. Jarque-Bera test for normality has p-value=.00000 to 5 decimal places - reject normality!

Fitting a t-GARCH Model:

To fit a t-GARCH model the conditional likelihood is replaced by

$$L(\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \nu) = \prod_{t=p+1}^{n} \frac{\sigma_t^{-1}\sqrt{\nu}}{\sqrt{\nu - 2}} \; t_\nu \left(\widehat{Z}_t \frac{\sigma_t^{-1}\sqrt{\nu}}{\sqrt{\nu - 2}} \right).$$

Maximization is now carried out with respect to the coefficients $\alpha_0, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ and the degrees of freedom ν of the t-density, t_ν.

Proceed as before but select `t-distribution for noise` in each of the dialog boxes where it appears.

Good idea to initialize the coefficients by first fitting a Gaussian GARCH model and then optimizing with t-distributed noise.

Estimated mean:

$$\widehat{a} = 0.0608$$

Minimum-AICC t-GARCH model for $\widehat{Z}_t = Y_t - \widehat{a}$:

t-GARCH(1,1) with

$$\widehat{\alpha}_0 = 0.1324, \widehat{\alpha}_1 = 0.0672, \widehat{\beta}_1 = 0.8400, \widehat{\nu} = 5.714$$

AICC value = 1437.89.

Model-checking:

Under the fitted model, the GARCH residuals, $\{\hat{Z}_t/\hat{\sigma}_t\}$, should be approximately IID and t-distributed with 5.714 degrees of freedom.

Check independence: Sample ACF's of the absolute values and squares of the residuals (fifth red button at the top of the ITSM window) look OK.

Check t-distribution: Selecting the 6th red button at the top of the ITSM window will give a qq plot using quantiles of the t-distribution with the fitted degrees of freedom (5.714 in this case). The graph is closer to linear than for the Gaussian model.

The improvement in AICC strongly suggests the superiority of the t-driven model.

The estimated mean is $\hat{a} = 0.0608$ as before, and the minimum-AICC GARCH model for the residuals, $\hat{Z}_t = Y_t - \hat{a}$, is the GARCH(1,1) with estimated parameter values

$$\hat{\alpha}_0 = 0.1324, \hat{\alpha}_1 = 0.0672, \hat{\beta}_1 = 0.8400, \hat{\nu} = 5.714,$$

and an AICC value (as in (10.3.17) with q replaced by $q+1$) of 1437.89. Thus from the point of view of AICC, the model with conditional t-distribution is substantially better than the conditional Gaussian model. The sample ACF of the absolute values and squares of the GARCH residuals are much the same as those found using Gaussian noise, but the qq plot (obtained by clicking on the red QQ button) is closer to the expected line than was the case for the model with Gaussian noise. $\square$

4. ARCH and GARCH Models (cont)

ARMA and regression models with GARCH errors

ITSM can be used to fit an ARMA or regression model with GARCH errors by using the procedure described in the last lecture to fit a GARCH model to the residuals $\{\hat{Z}_t\}$ from the ARMA (or regression) fit.

Example Open the file SUNSPOTS.TSM, subtract the mean and use the option `Model>Estimation>Autofit` with the default ranges for p and q.

This gives an ARMA(3,4) model for the mean-corrected data.

Clicking on the second green button at the top of the ITSM window, we see that the sample ACF of the ARMA residuals is compatible with iid noise.

However the sample ACF's of the absolute and squared residuals suggest dependence.

To fit a GARCH(1,1) model to the ARMA residuals:

(i) Click on the red GAR button, enter the value 1 for both p and q and click `OK`.

(ii) Click on the red MLE button, click `OK` in the dialog box, and the `GARCH ML Estimates` window will open, showing the estimated parameter values.

(iii) Repeat the steps in the previous sentence two more times and the window will display the following ARMA(3,4) model for the mean-corrected sunspot data and the fitted GARCH model for the ARMA noise process $\{Z_t\}$.

$$X_t = 2.463 Z_{t-1} - 2.248 Z_{t-2} + .757 Z_{t-3} + Z_t - .948 Z_{t-1}$$

$$-.296 Z_{t-2} + .313 Z_{t-3} + .136 Z_{t-4},$$

where

$$Z_t = \sqrt{h_t} e_t$$

and

$$h_t = 31.152 + .223 Z_{t-1}^2 + .596 h_{t-1}.$$

4. ARCH and GARCH Models (cont)

The AICC value for the GARCH fit (805.12) should be used for comparing alternative GARCH models for the ARMA residuals. The AICC value adjusted for the ARMA fit (821.70) should be used for comparison with alternative ARMA models (with or without GARCH noise). Standard errors of the estimated coefficients are also displayed.

Simulation using the fitted ARMA(3,4) model driven by GARCH(1,1) noise can be carried out by selecting the option n `Model>Simulate`. If you retain the default settings in the `ARMA Simulation` dialog box and click `OK` you will see a simulated realization of the model for the *original* data in SUNSPOTS.TSM.

4. ARCH and GARCH Models (cont)

The AICC value for the GARCH fit (805.12) should be used for comparing alternative GARCH models for the ARMA residuals. The AICC value adjusted for the ARMA fit (821.70) should be used for comparison with alternative ARMA models (with or without GARCH noise). Standard errors of the estimated coefficients are also displayed.

Simulation using the fitted ARMA(3,4) model driven by GARCH(1,1) noise can be carried out by selecting the option n `Model>Simulate`. If you retain the default settings in the `ARMA Simulation` dialog box and click `OK` you will see a simulated realization of the model for the *original* data in SUNSPOTS.TSM.

5. Forecasting with GARCH

Since the GARCH process $\{Z_t\}$ is a martingale difference sequence,

$$E(Z_{t+m}|Z_s, s \le t) = 0 \quad \forall m. \tag{1}$$

The past is therefore of no help in predicting Z_{t+h} and the best (in terms of MSE) predictor is the same as the best linear predictor.

However

$$E(Z_{t+1}^2|Z_s, s \le t) = h_{t+1} = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t+1-i}^2 + \sum_{i=1}^{q} \beta_i h_{t+1-i},$$

so that the past is valuable for forecasting the future variance of $\{Z_t\}$. (This is in contrast with the case $\{Z_t\} \sim \mathrm{IID}(0, \sigma^2)$, when $E(Z_{t+1}^2|Z_s, s \le t) = E(Z_{t+1}^2) = \sigma^2$.)

5. Forecasting with GARCH (cont)

Calculation of $E(Z_{t+j}^2|\mathcal{F}_t)$ for an ARCH(1) process:

$$E(Z_{t+1}^2|\mathcal{F}_t) = h_{t+1} = \alpha_0 + \alpha_1 Z_t^2$$

$$\begin{aligned}
E(Z_{t+2}^2|\mathcal{F}_t) &= E[h_{t+2}e_{t+2}^2|\mathcal{F}_t] \\
&= E[E(h_{t+2}e_{t+2}^2|e_s, s \le t+1)|\mathcal{F}_t] \\
&= E[h_{t+2}|\mathcal{F}_t] \\
&= E[\alpha_0 + \alpha_1 Z_{t+1}^2|\mathcal{F}_t] \\
&= \alpha_0(1+\alpha_1) + \alpha_1^2 Z_t^2.
\end{aligned}$$

Repeating this argument gives

$$E(Z_{t+k}^2|\mathcal{F}_t) = \alpha_0(1+\alpha_1+\cdots+\alpha_1^{k-1}) + \alpha_1^k Z_t^2.$$

6. IGARCH

IGARCH

If

$$(1-\beta(z)-\alpha(z)) = (1-B)\phi(z) \text{ and } \phi(1) \neq 1,$$

i.e.

$$\phi(B)\nabla Z_t^2 = \alpha_0 + (1-\beta(B))W_t. \qquad (1)$$

then $\{Z_t\}$ is said to be IGARCH(p,q) (Engle and Bollerslev, Econometric Reviews 5, 1-50, 81-87, 1986).

Bougerol and Picard (J.Econometrics 52, 115-128, 1992) showed that if the distribution of e_t has unbounded support and no atom at zero then there is a unique stationary causal solution for $\{Z_t\}$ in this case, but $EZ_t^2 = \infty$.

In practice when fitting GARCH models it is often found that $\alpha(1)+\beta(1) \approx 1$, supporting the practical relevance of the IGARCH model even though $EZ_t^2 = \infty$.

7. Stochastic Volatility Models

These differ from GARCH models by taking h_t to depend on unobserved or "latent" variables. The most popular is the log-normal SV model (Taylor, 1986, Modelling financial time series, Wiley).

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{IID N}(0,1),$$

$$\ln h_t = \gamma_0 + \gamma_1 \ln h_{t-1} + \eta_{t-1}, \quad \{\eta_t\} \sim \text{IID N}(0, \sigma^2),$$

where $\{e_t\}$ and $\{\eta_t\}$ are independent.

Assume $|\gamma_1| < 1$. Then $\{\ell_t := \ln h_t\}$ is a Gaussian AR(1) process with

$$\mu_\ell := E\ell_t = \frac{\gamma_0}{1-\gamma_1}$$

and

$$\sigma_\ell^2 := \text{Var}(\ell_t) = \frac{\sigma^2}{1-\gamma_1^2}.$$

Properties of Z_t

(i) $\{Z_t\}$ is strictly stationary.

(ii) $E(Z_t^r) = E(e_t^r) E \exp(r\ell_t/2)$

$$= \begin{cases} 0 & \text{if } r \text{ is odd,} \\ \left[\prod_{i=1}^m (2i-1)\right] \exp\left(\frac{m\gamma_0}{1-\gamma_1} + \frac{m^2\sigma^2}{2(1-\gamma_1^2)}\right) & \text{if } r = 2m. \end{cases}$$

(iii) Kurtosis:

$$\frac{EZ_t^4}{(EZ_t^2)^2} = 3\exp\left(\frac{\sigma^2}{1-\gamma_1^2}\right) \geq 3.$$

As in Gaussian GARCH models, the tails are heavier than normal.

(iv) ACVF of $\{Z_t^2\}$:

Define $\mathcal{F}_t = \sigma(e_s, \eta_s,\ s \le t)$. Then

$\{Z_t\}$ is an $\mathcal{F}_t$-martingale-difference sequence since
$$E(Z_t|\mathcal{F}_{t-1}) = \sqrt{h_t}E(e_t|\mathcal{F}_{t-1}) = 0.$$

$\{Z_t\}$ has finite fourth moments by (ii).

If $t > s$,

$$E(Z_t^2 Z_s^2|\mathcal{F}_{t-1}) = h_s h_t e_s^2 E(e_t^2|\mathcal{F}_{t-1}) = h_s h_t e_s^2.$$

and so

$$E(Z_t^2 Z_s^2) = \exp(\ell_s + \ell_t),$$

where $\{\ell_t\}$ is the causal stationary AR(1) process defined by

$$\ell_t = \gamma_0 + \gamma_1 \ell_{t-1} + \eta_t.$$

So for $h > 0$,

$$\text{Cov}(Z_{t+h}^2, Z_t^2) = E\exp(\ell_{t+h}\ell_t) - E\exp(\ell_{t+h})E\exp(\ell_t).$$

$$= \exp[2\mu_\ell + \sigma_\ell^2(1+\gamma_1^h)] - \exp[2\mu_\ell + \sigma_\ell^2].$$

From (ii),

$$\text{Var}(Z_t^2) = 3\exp(2\mu_\ell + 2\sigma_\ell^2) - \exp[2\mu_\ell + \sigma_\ell^2].$$

Hence, for $h \neq 0$,

$$\rho_{Z^2}(h) = \frac{\exp(\sigma_\ell^2\gamma_1^{|h|}) - 1}{3\exp(\sigma_\ell^2) - 1}$$

$$\approx \frac{\sigma_\ell^2\exp(\sigma_\ell^2)}{3\exp(\sigma_\ell^2) - 1}\gamma_1^{|h|}.$$

This approximation is the ACF of an ARMA(1,1) so the SV model has some resemblance to a GARCH(1,1) process. Notice that if $\gamma_1 < 0$ the ACF of a SV model can have negative values.

7. Stochastic Volatility Models (cont)

[The ACVF of an ARMA(1,1) process:

$$Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}(0.\sigma^2.$$

Multiplying each side by Y_{t-j}, $j = 0, 1, 2, \ldots$, taking expectations and using the expansion,

$$Y_t = Z_t + (\theta + \phi) Z_{t-1} + \cdots,$$

gives the equations,

$$\gamma(0) - \phi\gamma(1) = \sigma^2(1 + \theta(\theta + \phi)),$$

$$\gamma(1) - \phi\gamma(0) = \sigma^2\theta,$$

$$\gamma(j) - \phi\gamma(j-1) = 0, \quad j = 2, 3, \ldots,$$

with solution

$$\gamma(0) = \sigma^2 \left[1 + \frac{(\theta+\phi)^2}{1-\phi^2}\right], \quad \gamma(1) = \sigma^2 \left[\theta + \phi + \frac{(\theta+\phi)^2\phi}{1-\phi^2}\right],$$

and

$$\gamma(j) = \gamma(1)\phi^{j-1}, \quad j = 2, 3, \ldots.]$$

7. Stochastic Volatility Models (cont)

(v) The process $\{\ln Z_t^2\}$:

$$\ln Z_t^2 = \ell_t + \ln e_t^2$$

Since $\{\ln Z_t^2\}$ is the sum of the iid sequence $\{\ln e_t^2\}$ and the independent AR(1) process $\{\ell_t\}$, it is an ARMA(1,1) process. If $e_t \sim \mathrm{N}(0,1)$ then $E \ln e_t^2 = -1.27$ and $\mathrm{Var}(\ln e_t^2)$ $=4.93$. (The distribution of $\ln e_t^2$ has a very long left tail.) Hence,

$$\rho_{\ln Z^2}(h) = \frac{\sigma_\ell^2 \gamma_1^{|h|}}{\sigma_\ell^2 + 4.93}, \quad h \neq 0.$$

7. Stochastic Volatility Models (cont)

[AR(1) with observation noise

If $X_t = Y_t + Z_t$, where $\{Y_t\}$ is an AR(1) process, $\{Z_t\}$ is white noise and $Y_s \perp Z_t\ \forall s,t$,

then the spectral density of $\{X_t\}$ is

$$f_X(\omega) = \frac{\sigma_1^2}{2\pi(1+\phi^2-2\phi\cos\omega)} + \frac{\sigma_2^2}{2\pi}$$

$$= \frac{a+b\cos\omega}{1+\phi^2-2\phi\cos\omega},$$

showing that $\{X_t\}$ is an ARMA(1,1) process.]

7. Stochastic Volatility Models (cont)

Estimation for SV Models

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{IID N}(0,1),$$

$$\ln h_t = \gamma_0 + \gamma_1 \ln h_{t-1} + \eta_{t-1}, \quad \{\eta_t\} \sim \text{IID N}(0,\sigma^2),$$

The parameters to be estimated are σ^2, γ_0 and γ_1. We know that $\{Z_t^2\}$ resembles an ARMA(1,1) process and we have explicit formulae for the moments, $\mu = EZ_t^2$ and $\gamma(h) =$Cov(Z_{t+h}^2, Z_t^2) in terms of σ^2, γ_0 and γ_1.

A method of moments procedure for estimating the parameters would be to equate μ, $\gamma(0)$ and $\gamma(1)$ to the corresponding sample estimates $\hat{\mu}$, $\hat{\gamma}(0)$ and $\hat{\gamma}(1)$ respectively and to solve the resulting three nonlinear equations for σ^2, γ_0 and γ_1. However this method has low asymptotic efficiency (see TSTM p.253).

(i) Generalized Method of Moments (GMM)

This method was proposed by Hansen (1982), Econometrica 50, 1029-1054, in an attempt to improve the efficiency of the moment method.

Instead of using exactly p moment equations, where p is the number of parameters to be estimated, he suggested specifying a larger number of equations.

Since these cannot all be simultaneously satisfied, the parameters are estimated by minimizing a specified norm of the vector of errors.

In our particular case this method can be implemented by considering the s-dimensional vector,

$$\mathbf{g}(\boldsymbol{\theta}, \mathbf{X}_n) = \begin{bmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma}(0) - \gamma(0) \\ \vdots \\ \widehat{\gamma}(s-2) - \gamma(s-2) \end{bmatrix}.$$

7. Stochastic Volatility Models (cont)

The GMM estimate of $\boldsymbol{\theta} = (\sigma^2, \gamma_0, \gamma_1)'$ is the value which minimizes

$$\mathbf{g}' W_n \mathbf{g},$$

where W_n is an $s \times s$ matrix of weights. Hansen computes the asymptotic distribution of the estimators and it turns out that the optimal weighting matrix is the inverse of the covariance matrix $S(\boldsymbol{\theta})$ of $\mathbf{g}$.

He proposes the iterative scheme:

1. Minimize $\mathbf{g}' W_n \mathbf{g}$ with $W_n = I_s$ (the $s \times s$ identity matrix).

2. Use the estimated $\boldsymbol{\theta}$ to compute $S(\boldsymbol{\theta})$.

3. Obtain a new estimate of $\boldsymbol{\theta}$ by minimizing $\mathbf{g}' W_n \mathbf{g}$ $W_n = S(\boldsymbol{\theta})^{-1}$ and return to Step 2.

7. Stochastic Volatility Models (cont)

Theorem (Asymptotic behavior of GMM)
(see Hamilton, Time Series Analysis, chap.14). If

$$\mathbf{g} \text{ is differentiable in } \boldsymbol{\theta},$$

$$\sqrt{n}\mathbf{g} \xrightarrow{d} \mathsf{N}(\mathbf{0}, S),$$

and if the GMM estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is computed using a positive definite weighting matrix S_n^{-1} such that $S_n \xrightarrow{p} S$ and if for any $\boldsymbol{\theta}_n$ converging in probability to $\boldsymbol{\theta}_0$,

$$\left[\frac{\partial g_i}{\partial \theta_j}\right]_{\boldsymbol{\theta}_n} \xrightarrow{p} \left[\frac{\partial g_i}{\partial \theta_j}\right]_{\boldsymbol{\theta}_0} = D',$$

and the $s \times p$ matrix D' has linearly independent columns, then the GMM estimators satisfy

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathsf{N}(\mathbf{0}, V),$$

where

$$V = [DS^{-1}D']^{-1}.$$

7. Stochastic Volatility Models (cont)

(ii) Maximization of the Gaussian (quasi-)Likelihood (MGL)

The equations defining the SV model,

$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{IID N}(0,1),$$

$$\ln h_t = \gamma_0 + \gamma_1 \ln h_{t-1} + \eta_{t-1}, \quad \{\eta_t\} \sim \text{IID N}(0,\sigma^2),$$

can be rewritten in the state-space form,

$$Y_t^* := \ln(Z_t^2) = \ell_t + \ln(e_t^2), \quad \{e_t\} \sim \text{IID N}(0,1),$$

$$\ell_t = \gamma_0 + \gamma_1 \ell_{t-1} + \eta_{t-1}, \quad \{\eta_t\} \sim \text{IID N}(0,\sigma^2).$$

If $\{Y_t^*\}$ were Gaussian, we could write the likelihood of $Y_1^*, \ldots, Y_n^*$ in terms of the best linear one-step predictors $\hat{Y}_t^*$ and their mean square errors $\{v_{t-1}\}$. These can be computed (for any given parameter values γ_0, γ_1 and σ^2) using the Kalman recursions, and maximized with respect to the parameters to get the MGL estimators of γ_0, γ_1 and σ^2.

(iii) Estimation using the Whittle likelihood

Another estimation procedure is based on the Whittle approximation to the likelihood for $Y_t = \ln Z_t^2$ was considered by Breidt, Crato, Lima (1998). Instead of maximizing the MGL, they suggest minimizing

$$\sum_{j=1}^{n-1} I_n(\omega_j)/g(\omega_j),$$

where $I_n(\omega_j)$ and $g(\omega_j)$ are the periodogram and model spectral density of Y_t at the Fourier frequency $\omega_j = 2j\pi/n$ respectively. In this case the spectral density of Y_t has a rather simple form,

$$g(\omega) = f_{\ln h_t}(\omega) + 4.93/(2\pi)$$

where $f_{\ln h_t}(\omega)$ is the spectral density of the $\ln h_t$ process. Notice that in the case $\ln h_t$ is a long memory process, such as a fractionally integrated process, then this spectral density has a very simple form.

7. Stochastic Volatility Models (cont)

(iv) Simulation Based Estimation

Since computation of the likelihood function function requires an n-fold integration over the latent process, an explicit formula for the likelihood does not exist. However, one can compute this integral using simulation based methods such as MCMC and importance sampling. The latter will be discussed in more detail in the section on parameter-driven state-space models.

8. Regular variation and application to financial TS models

8.1 Regular variation — univariate case

Def: The random variable X is *regularly varying with index* α if

$$P(|X|> t\,x)/P(|X|>t) \to x^{-\alpha} \quad \text{and} \quad P(X> t)/P(|X|>t) \to p,$$

or, equivalently, if

$$P(X> t\,x)/P(|X|>t) \to p x^{-\alpha} \quad \text{and} \quad P(X< -t\,x)/P(|X|>t) \to q x^{-\alpha},$$

where $0 \le p \le 1$ and $p+q=1$.

Equivalence:

X is RV(α) *if and only if* $P(X \in t \bullet) / P(|X|>t) \to_v \mu(\bullet)$

($\to_v$ vague convergence of measures on $R\backslash\{0\}$). In this case,

$$\mu(dx) = \big(p\alpha\, x^{-\alpha-1}\, I(x>0) + q\alpha\, (-x)^{-\alpha-1}\, I(x<0)\big)\, dx$$

Note: $\mu(tA) = t^{-\alpha}\, \mu(A)$ for every t and A bounded away from 0.

8.1 Regular variation — univariate case (cont)

Another formulation (polar coordinates):

Define the ± 1 valued rv θ, $P(\theta = 1) = p$, $P(\theta = -1) = 1 - p = q$.

Then

X is RV(α) *if and only if*

$$\frac{P(|X| > t\,x,\, X/|X| \in S)}{P(|X| > t)} \to x^{-\alpha} P(\boldsymbol{\theta} \in S)$$

or

$$\frac{P(|X| > t\,x,\, X/|X| \in \bullet)}{P(|X| > t)} \to_v x^{-\alpha} P(\boldsymbol{\theta} \in \bullet)$$

($\to_v$ vague convergence of measures on $S^0 = \{-1,1\}$).

8.2 Regular variation — multivariate case

Multivariate regular variation of $\mathbf{X}=(X_1, \ldots, X_m)$: There exists a random vector $\theta \in S^{m-1}$ such that

$$P(|\mathbf{X}|> t\,x,\ \mathbf{X}/|\mathbf{X}| \in \bullet\,)/P(|\mathbf{X}|>t) \to_v x^{-\alpha} P(\,\theta \in \bullet\,)$$

($\to_v$ vague convergence on S^{m-1}, unit sphere in R^m) .

- $P(\,\theta \in \bullet)$ is called the spectral measure
- α is the index of $\mathbf{X}$.

Equivalence:

$$\frac{P(\mathbf{X} \in t\bullet)}{P(|\mathbf{X}|>t)} \to_v \mu(\bullet)$$

μ is a measure on R^m which satisfies for $x > 0$ and A bounded away from 0,

$$\mu(xB) = x^{-\alpha}\mu(xA).$$

8.2 Regular variation — multivariate case (cont)

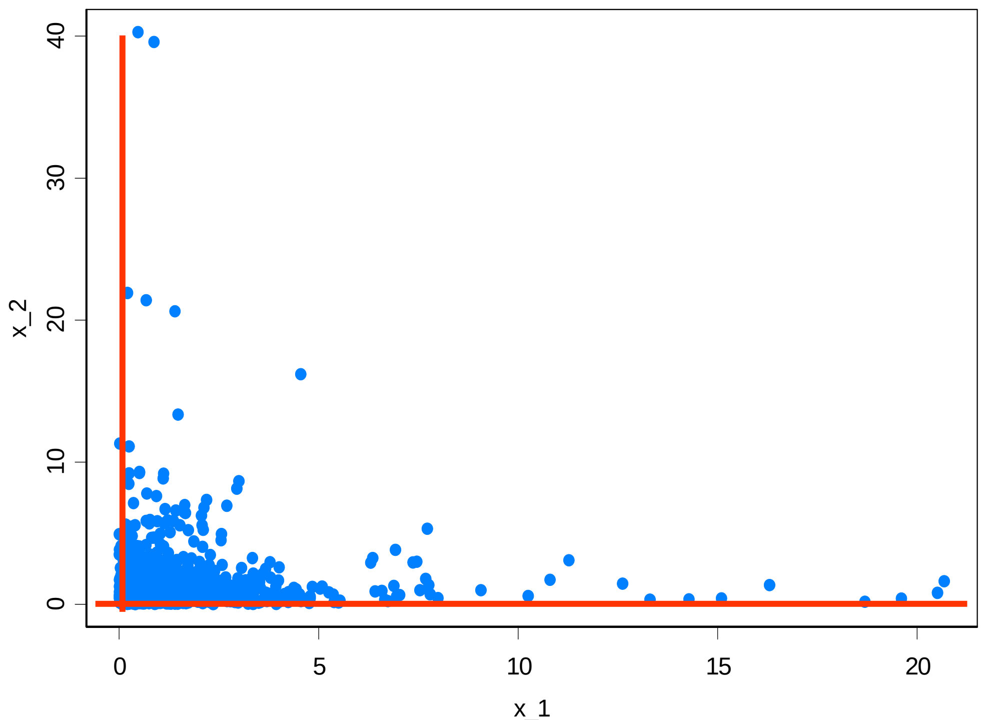
Examples:

1. If $X_1 > 0$ and $X_2 > 0$ are iid RV(α), then $\mathbf{X} = (X_1, X_2)$ is multivariate regularly varying with index α and *spectral distribution*

$$P(\theta = (0,1)) = P(\theta = (1,0)) = .5 \quad \text{(mass on axes).}$$

Interpretation: Unlikely that X_1 and X_2 are very large at the same time.

Figure: plot of (X_{t1}, X_{t2}) for realization of 10,000.



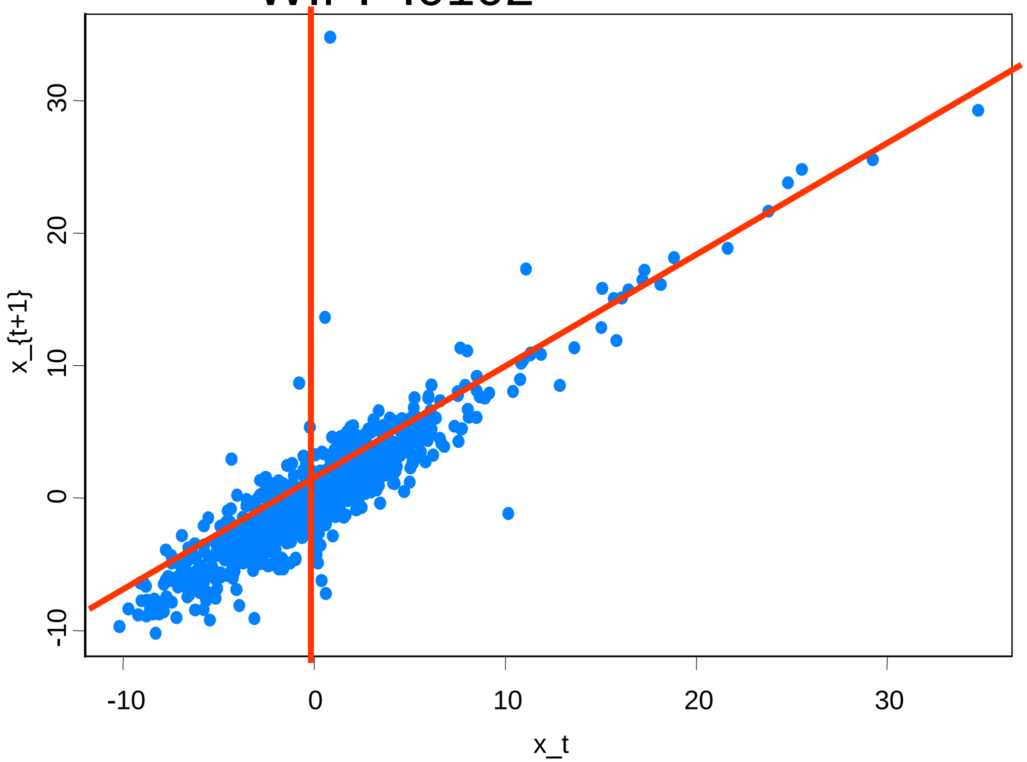
2. If $X_1 = X_2 > 0$, then $\mathbf{X} = (X_1, X_2)$ is multivariate regularly varying with index α and *spectral distribution*

$$P(\,\theta = (1/\sqrt{2},\, 1/\sqrt{2})\,) = 1.$$

3. AR(1): $X_t = .9\, X_{t-1} + Z_t$, $\{Z_t\}$~IID symmetric stable (1.8)

$$\text{Distr of } \theta: \begin{cases} \pm(1,.9)/\text{sqrt}(1.81), & \text{W.P. } .9898 \\ \pm(0,1), & \text{W.P. } .0102 \end{cases}$$

Figure: plot of (X_t, X_{t+1}) for realization of 10,000.



8.3 Applications of multivariate regular variation

- Domain of attraction for *sums of iid random vectors* (Rvaceva, 1962). That is, when does the partial sum

$$a_n^{-1}\sum_{t=1}^{n}\mathbf{X}_t$$

converge for some constants a_n?

- *Spectral measure* of multivariate stable vectors.
- *Domain of attraction* for componentwise maxima of iid random vectors (Resnick, 1987). Limit behavior of

$$a_n^{-1}\bigvee_{t=1}^{n}\mathbf{X}_t$$

- Weak convergence of *point processes* with iid points.
- Solution to *stochastic recurrence equations*, $\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t$
- Weak convergence of *sample autocovariances*.

8.3 Applications of multivariate regular variation (cont)

Linear combinations:

X ~RV(α) $\Rightarrow$ all linear combinations of **X** are regularly varying

i.e., there exist α and slowly varying fcn L(.), s.t.

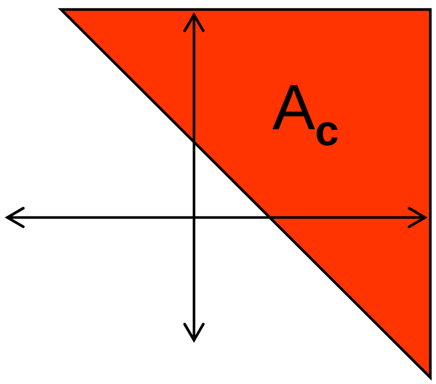
$$P(\mathbf{c}^T\mathbf{X} > t)/(t^{-\alpha}L(t)) \to w(\mathbf{c}),$$ exists for all real-valued **c**,

where

$$w(t\mathbf{c}) = t^{-\alpha}w(\mathbf{c}).$$

Use vague convergence with $A_\mathbf{c}=\{\mathbf{y}: \mathbf{c}^T\mathbf{y} > 1\}$, i.e.,

$$\frac{P(\mathbf{X} \in tA_\mathbf{c})}{t^{-\alpha}L(t)} = \frac{P(\mathbf{c}^T\mathbf{X} > t)}{P(|\mathbf{X}| > t)} \to \mu(A_\mathbf{c}) =: w(\mathbf{c}),$$



where $t^{-\alpha}L(t) = P(|\mathbf{X}| > t)$.

8.3 Applications of multivariate regular variation (cont)

Converse?

X ~RV(α) $\Leftarrow$ all linear combinations of **X** are regularly varying?

There exist α and slowly varying fcn $L(.)$, s.t.

(LC) $P(\mathbf{c}^T\mathbf{X} > t)/(t^{-\alpha}L(t)) \to w(\mathbf{c})$, exists for all real-valued **c.**

Theorem (Basrak, Davis, Mikosch, `02). Let **X** be a random vector.

1. If **X** satisfies (LC) with α non-integer, then **X** is RV(α).
2. If **X** > 0 satisfies (LC) for non-negative **c** and α is non-integer, then **X** is RV(α).
3. If **X** > 0 satisfies (LC) with α an odd integer, then **X** is RV(α).

8.4 Applications of theorem

1. Kesten (1973). Under general conditions, (LC) holds with L(t)=1 for stochastic recurrence equations of the form

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \qquad (\mathbf{A}_t, \mathbf{B}_t) \sim \text{IID},$$

$\mathbf{A}_t$ $d \times d$ random matrices, $\mathbf{B}_t$ random d-vectors.

It follows that the distributions of $\mathbf{Y}_t$, and in fact all of the finite dim'l distrs of $\mathbf{Y}_t$ are regularly varying (if α is non-even).

2. GARCH processes. Since squares of a GARCH process can be embedded in a SRE, the *finite dimensional distributions* of a *GARCH* are regularly varying.

8.5 Examples

Example of ARCH(1): $X_t=(\alpha_0+\alpha_1 X^2_{t-1})^{1/2}Z_t$, $\{Z_t\}$~IID.

α found by solving $E|\alpha_1 Z^2|^{\alpha/2} = 1$.

α_1	.312	.577	1.00	1.57
α	8.00	4.00	2.00	1.00

Distr of θ:

$$P(\theta \in \bullet) = E\left\{||(B,Z)||^\alpha I(\arg((B,Z)) \in \bullet)\right\}/ E||(B,Z)||^\alpha$$

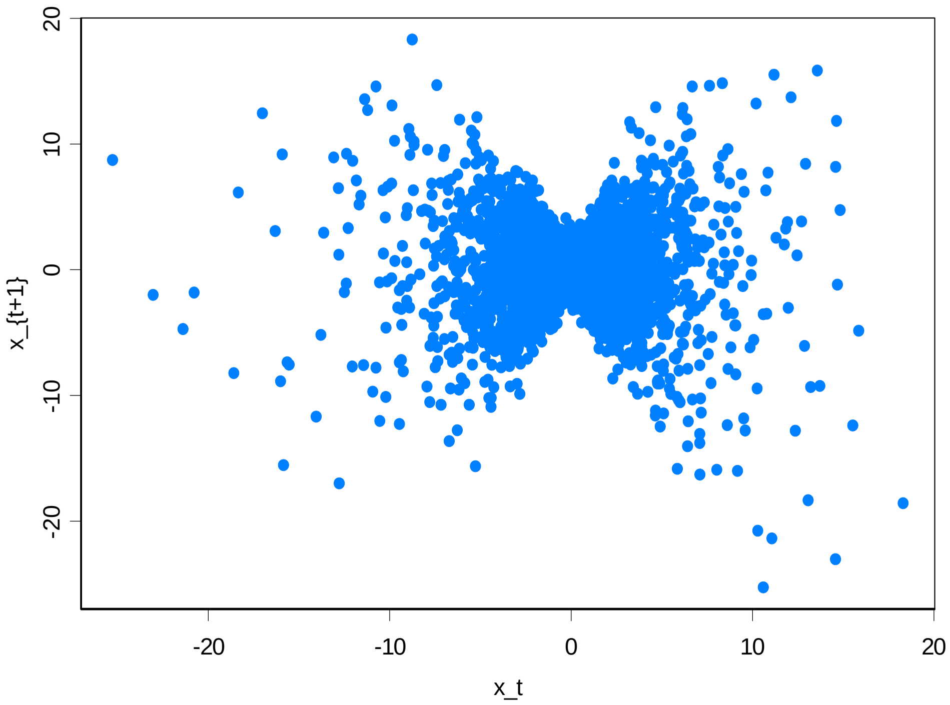
where

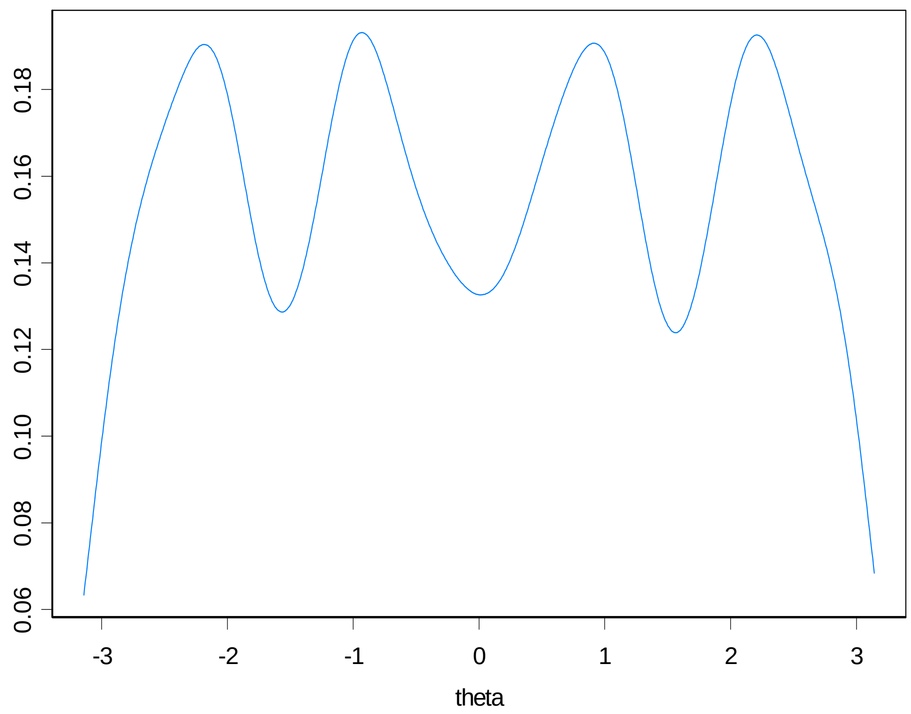
$$P(B = 1) = P(B = -1) =.5$$

8.4 Examples (cont)

Example of ARCH(1): $\alpha_0=1$, $\alpha_1=1$, $\alpha=2$, $X_t=(\alpha_0+\alpha_1 X^2_{t-1})^{1/2}Z_t$, $\{Z_t\}$~IID

Figures: plots of (X_t, X_{t+1}) and estimated distribution of θ for realization of 10,000.





8.4 Applications of theorem (cont)

Example: SV model $X_t = \sigma_t Z_t$

Suppose $Z_t \sim RV(\alpha)$ and

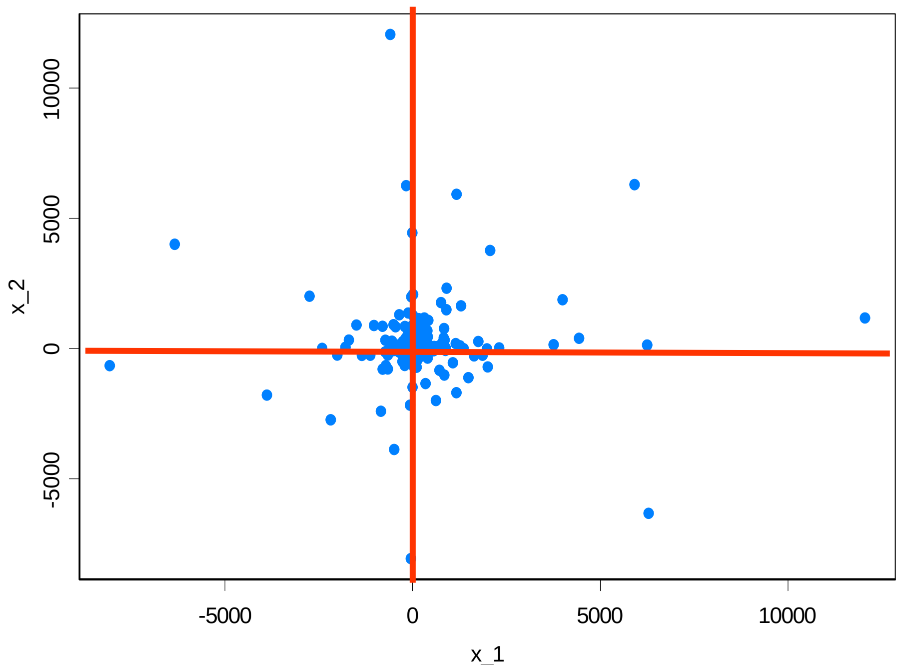
$$\log \sigma_t^2 = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty, \{\varepsilon_t\} \sim \mathrm{IID\,N}(0,\sigma^2).$$

Then $\mathbf{Z}_n=(Z_1,\ldots,Z_n)'$ is regulary varying with index α and so is

$$\mathbf{X}_n = (X_1,\ldots,X_n)' = \mathrm{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{Z}_n$$

with spectral distribution concentrated on $(\pm 1,0)$, $(0,\pm 1)$.

Figure: plot of (X_t,X_{t+1}) for realization of 10,000.



8.6 Extremes for GARCH and SV processes

Setup

- $X_t = \sigma_t Z_t$, $\{Z_t\} \sim$ IID (0,1)
- X_t is RV (α)
- Choose $\{b_n\}$ s.t. $nP(X_t > b_n) \to 1$

Then

$$P^n(b_n^{-1} X_1 \le x) \to \exp\{-x^{-\alpha}\}.$$

Then, with $M_n = \max\{X_1, . . . , X_n\}$,

(i) GARCH:

$$P(b_n^{-1} M_n \le x) \to \exp\{-\gamma x^{-\alpha}\},$$

γ is extremal index ($0 < \gamma < 1$).

(ii) SV model:

$$P(b_n^{-1} M_n \le x) \to \exp\{-x^{-\alpha}\},$$

extremal index $\gamma = 1$ no clustering.

8.6 Extremes for GARCH and SV processes (cont)

(i) GARCH: $P(b_n^{-1}M_n \le x) \to \exp\{-\gamma x^{-\alpha}\}$

(ii) SV model: $P(b_n^{-1}M_n \le x) \to \exp\{-x^{-\alpha}\}$

Remarks about extremal index.

(i) $\gamma < 1$ implies clustering of exceedances

(ii) Numerical example. Suppose c is a threshold such that

$$P^n(b_n^{-1}X_1 \le c) \sim .95$$

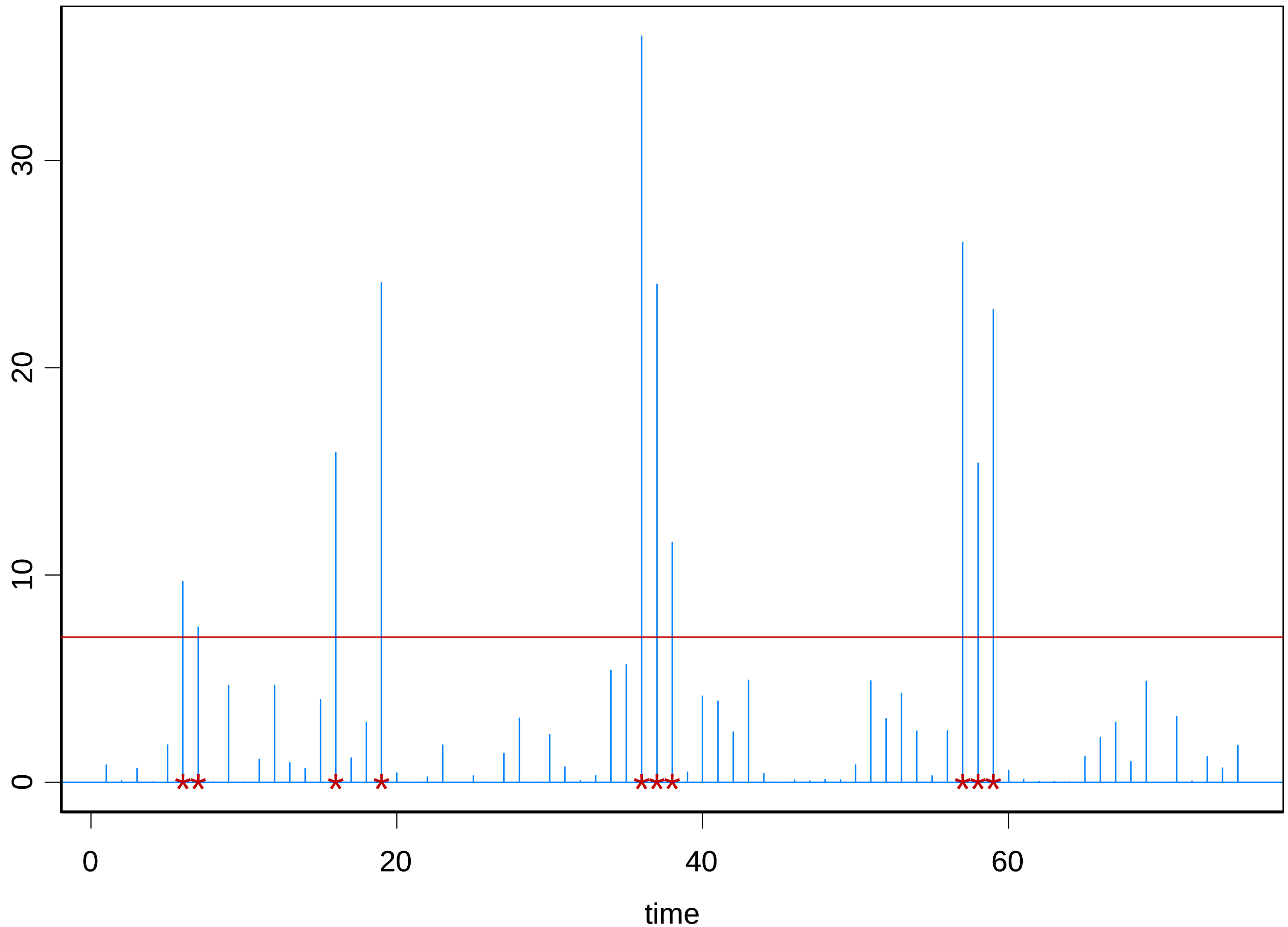
Then, if $\gamma = .5$, $P(b_n^{-1}M_n \le c) \sim (.95)^{.5} = .975$

(iii) $1/\gamma$ is the *mean cluster size* of exceedances.

(iv) Use γ to *discriminate* between GARCH and SV models.

(v) Even for the light-tailed SV model (i.e., $\{Z_t\}$ ~IID N(0,1), the *extremal index* is 1 (see Breidt and Davis `98)

8.6 Extremes for GARCH and SV processes (cont)



8.7 Summary of results for ACF of GARCH(p,q) and SV models

GARCH(p,q)

$\alpha \in (0,2)$:

$$(\hat{\rho}_X(h))_{h=1,\ldots,m} \xrightarrow{d} (V_h / V_0)_{h=1,\ldots,m},$$

$\alpha \in (2,4)$:

$$\left(n^{1-2/\alpha}\hat{\rho}_X(h)\right)_{h=1,\ldots,m} \xrightarrow{d} \gamma_X^{-1}(0)\left(V_h\right)_{h=1,\ldots,m}.$$

$\alpha \in (4,\infty)$:

$$\left(n^{1/2}\hat{\rho}_X(h)\right)_{h=1,\ldots,m} \xrightarrow{d} \gamma_X^{-1}(0)\left(G_h\right)_{h=1,\ldots,m}.$$

Remark: Similar results hold for the sample ACF based on $|X_t|$ and X_t^2.

8.7 Summary of results for ACF of GARCH(p,q) and SV models (cont)

SV Model

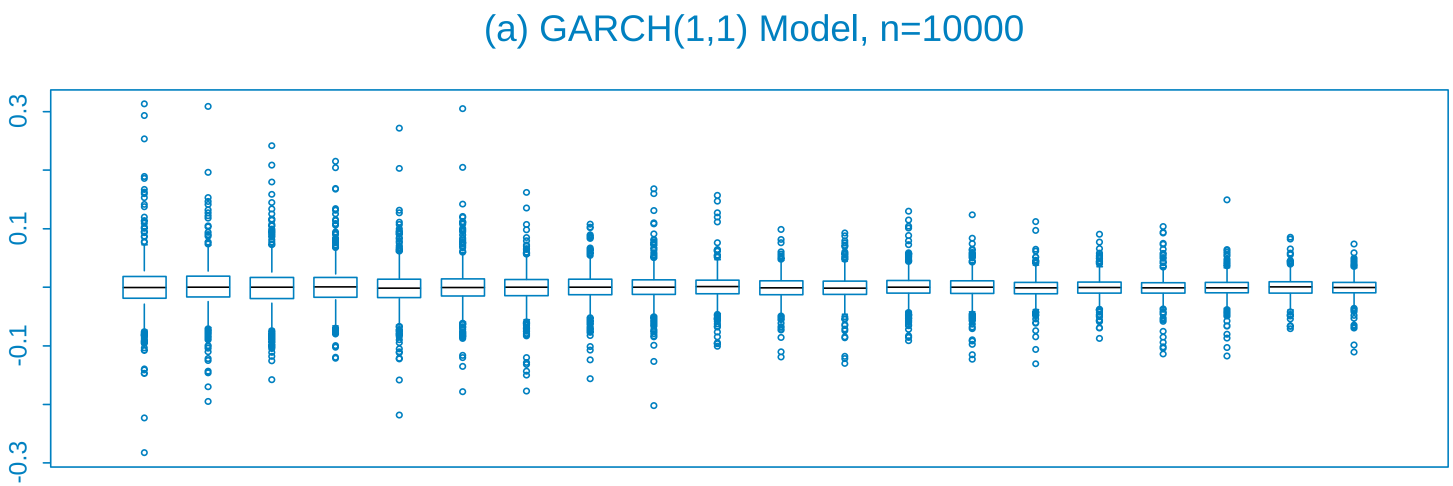
$\alpha \in (0,2)$:

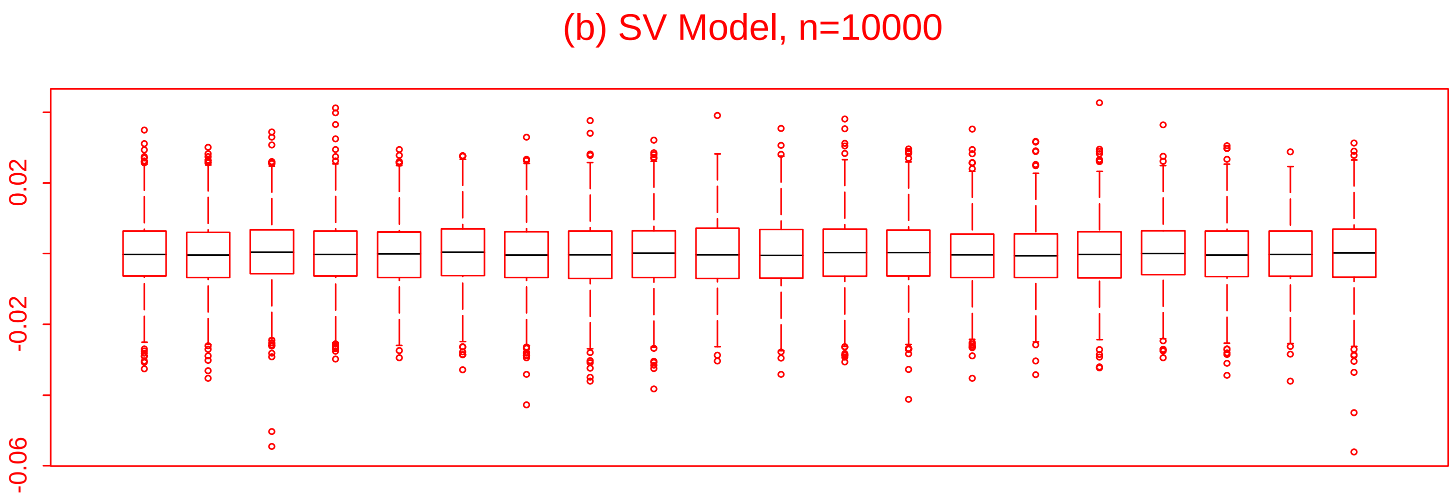
$$(n/\ln n)^{1/\alpha}\hat{\rho}_X(h) \xrightarrow{d} \frac{\|\sigma_1\sigma_{h+1}\|_\alpha}{\|\sigma_1\|_\alpha^2}\frac{S_h}{S_0}.$$

$\alpha \in (2, \infty)$:

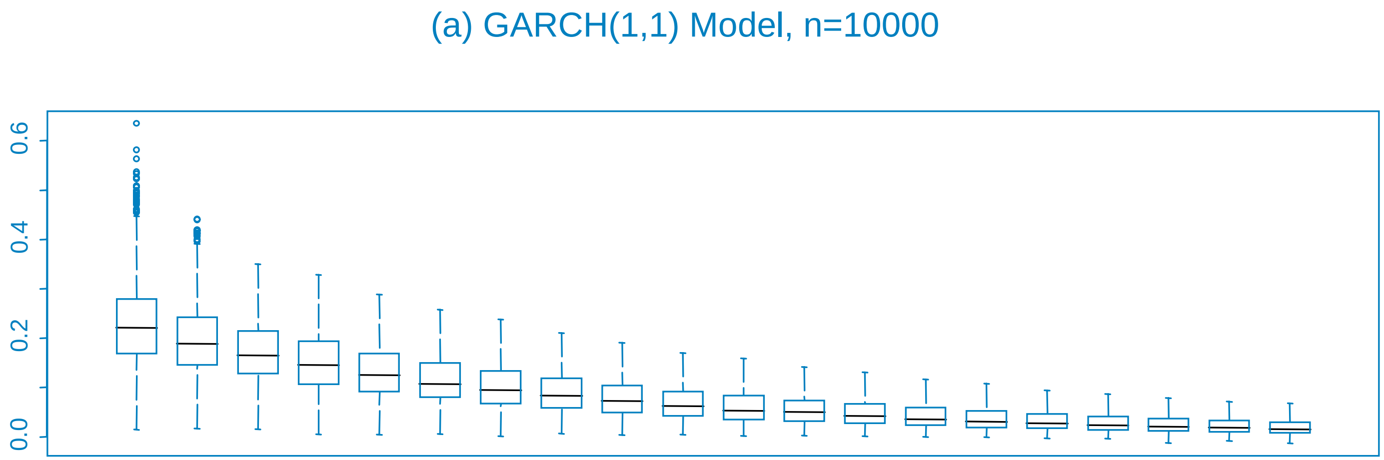
$$\left(n^{1/2}\hat{\rho}_X(h)\right)_{h=1,\ldots,m} \xrightarrow{d} \gamma_X^{-1}(0)\left(G_h\right)_{h=1,\ldots,m}.$$

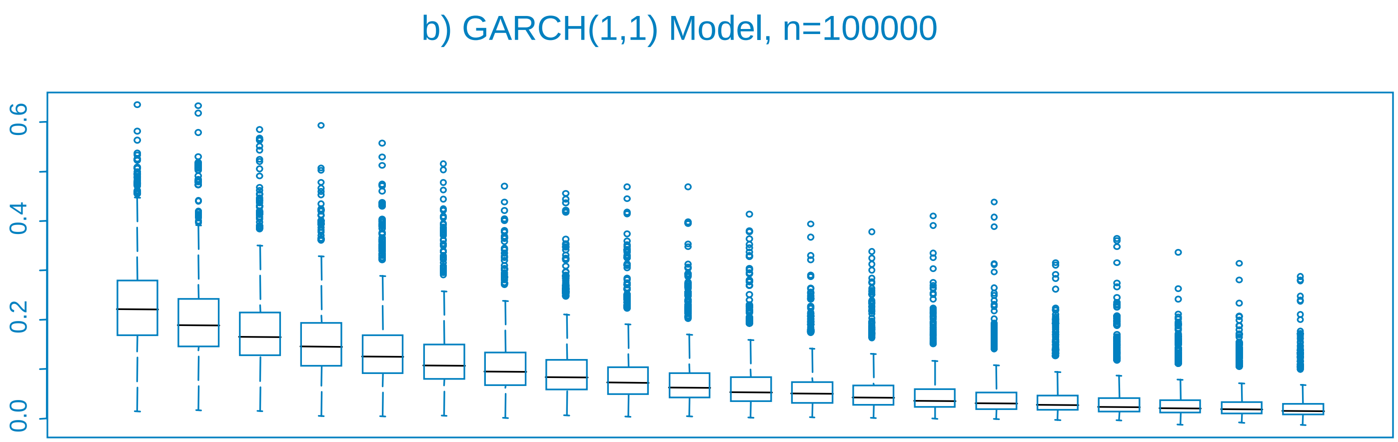
Sample ACF for GARCH and SV Models (1000 reps)



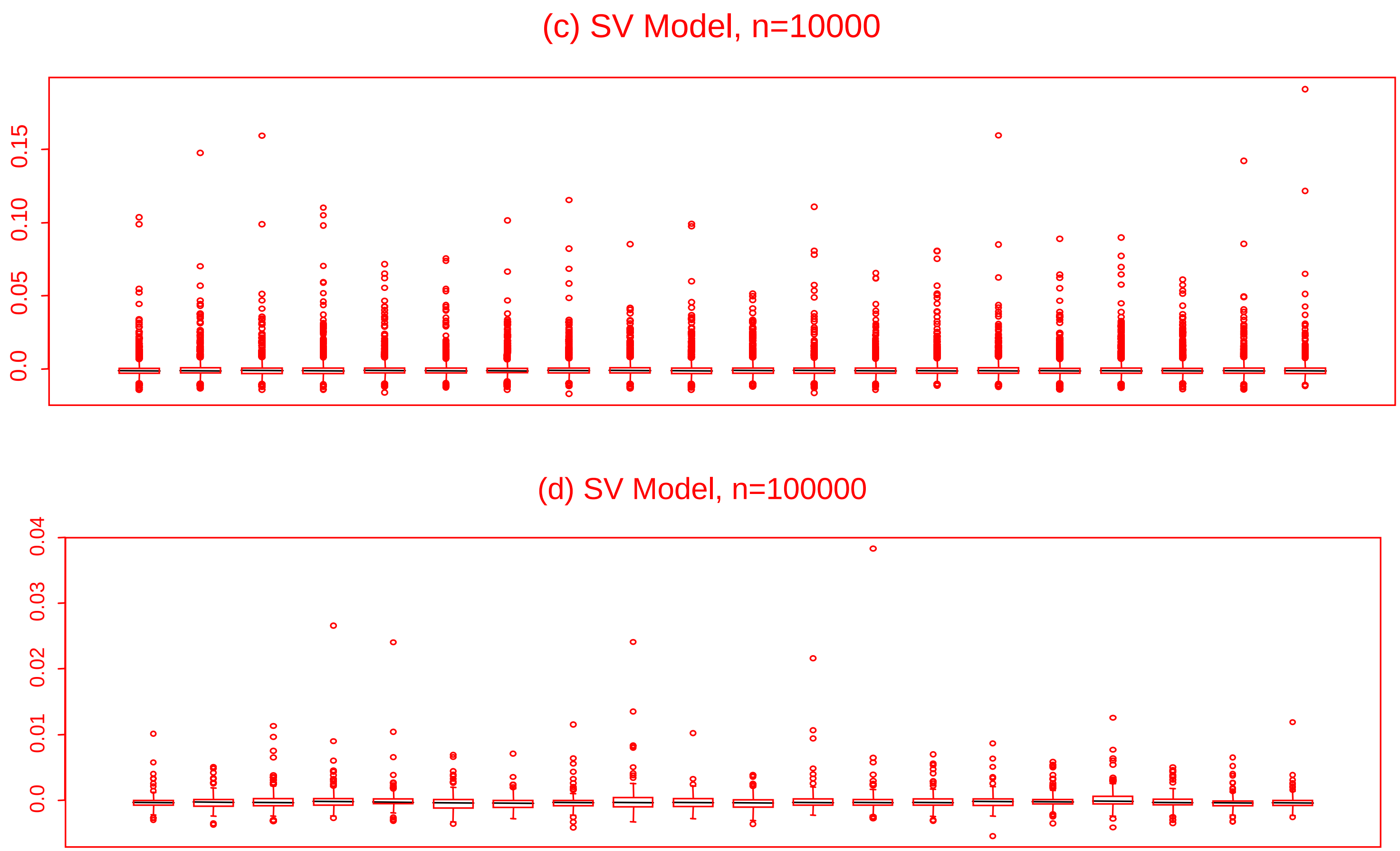


Sample ACF for Squares of GARCH (1000 reps)

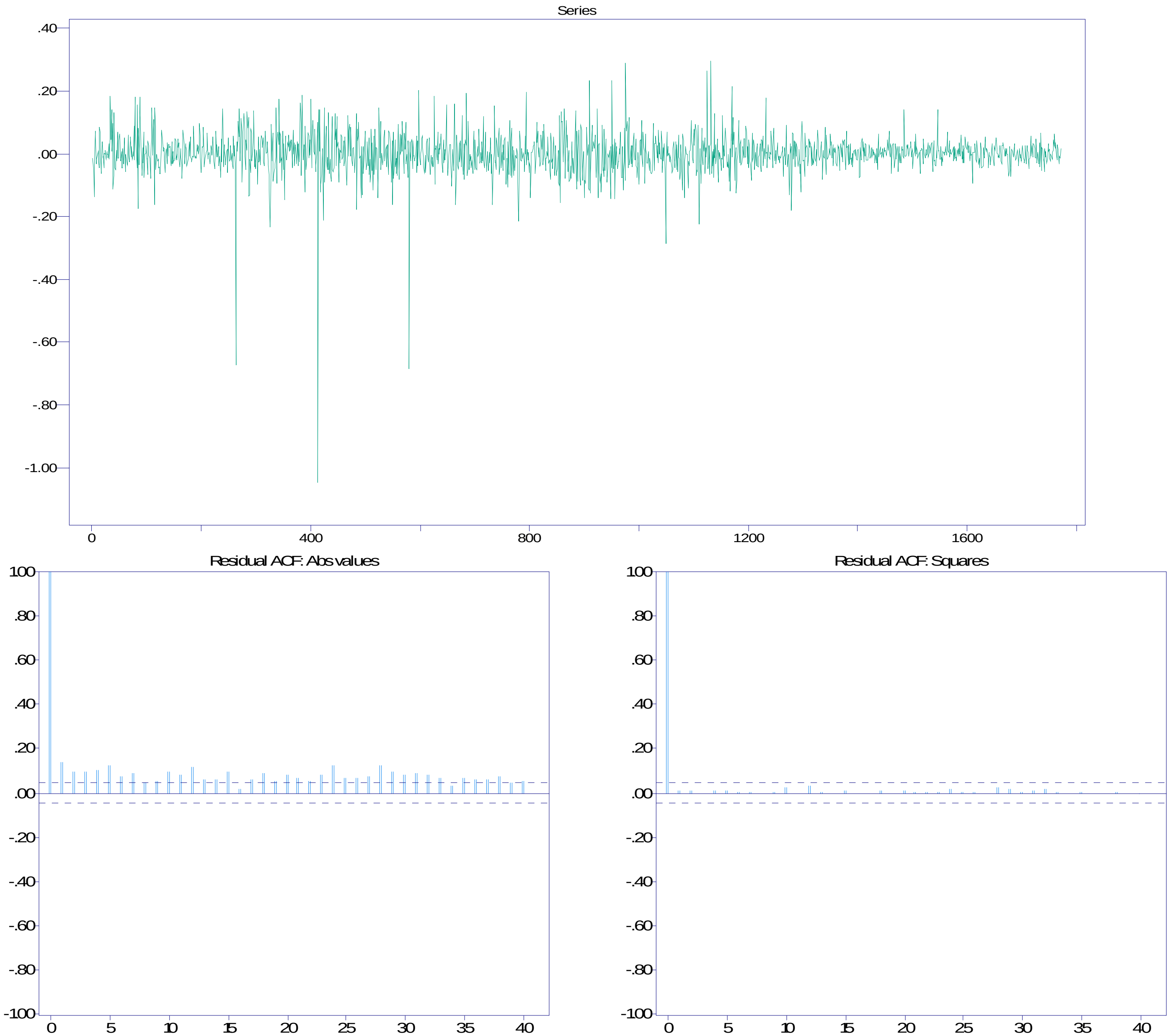




Sample ACF for Squares of SV (1000 reps)



Example: Amazon returns May 16, 1997 to June 16, 2004.



Wrap-up

- *Regular variation* is a flexible tool for modeling both *dependence* and *tail heaviness.*

- Useful for establishing *point process convergence* of heavy-tailed time series.

- *Extremal index* $\gamma < 1$ for GARCH and $\gamma = 1$ for SV.

Unresolved issues related to RV$\Leftrightarrow$ (LC)

- α = 2n?

- there is an example for which $\mathbf{X}_1, \mathbf{X}_2 > 0$, and $(\mathbf{c}, \mathbf{X}_1)$ and $(\mathbf{c}, \mathbf{X}_2)$ have the same limits for all $\mathbf{c} > \mathbf{0}$.

- $\alpha = 2n-1$ and $\mathbf{X} \not> 0$ (not true in general).