

## Lecture 2: The ARCH Model

Zheng Tian

# Outline

- 1 The Volatility of Asset Returns
- 2 The Structure of a Volatility Model
- 3 The ARCH Model
- 4 Maximum Likelihood Estimation
- 5 Estimation of an ARCH(m) model
- 6 Model checking and forecasting
- 7 Applications of ARCH Models

# Volatility is measured with the conditional variance

- Volatility here refers to the **conditional variance** of a time series.
- For a return series  $\{r_t\}$ , we are now interested in

$$\sigma_t^2 = \text{Var}(r_t \mid F_{t-1})$$

where  $F_{t-1}$  is the information set at time  $t - 1$ .

# Characteristics of volatility (1)

- 1 There exist **volatility clusters**. That is, volatility may be high for certain time periods and low for other periods.

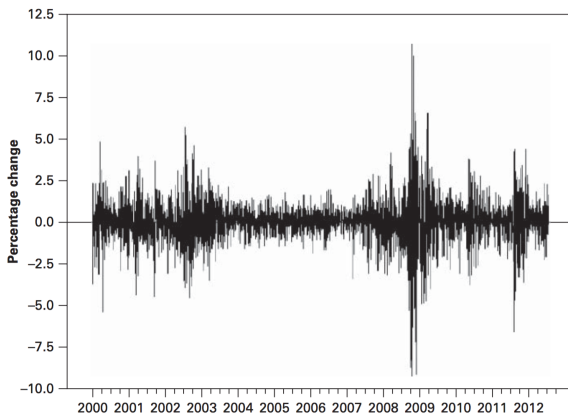


Figure: Percentage Change in the NYSE U.S.100 stock price index

## Characteristics of volatility (2)

- 1 Volatility evolves over time in a continuous manner. That is, volatility jumps are rare.

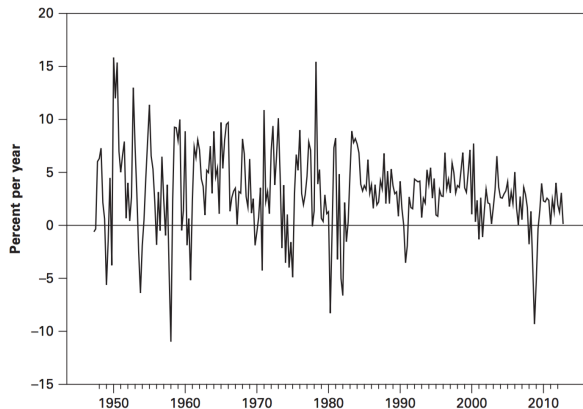


Figure: Annualized Growth Rate of Real GDP

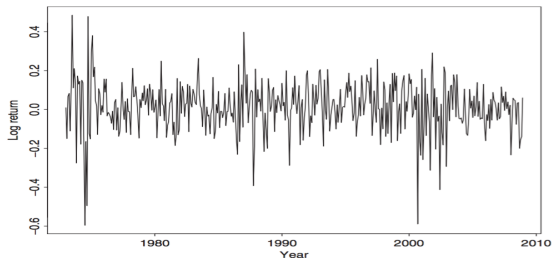
# Characteristics of volatility (3)

- 1 Volatility does not diverge to infinity. That is, volatility varies within some fixed range. Statistically speaking, this means that volatility is often stationary.
- 1 Volatility seems to react differently to a big price increase or a big price drop, referred to as the leverage effect.

# The basic idea of building a volatility model

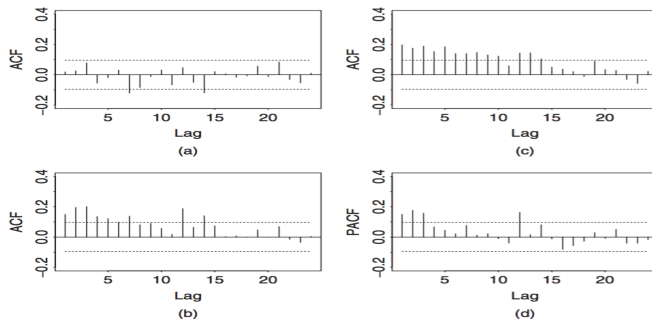
Consider the log return series  $\{r_t\}$ . The basic idea of a volatility model is

- $\{r_t\}$  may appear to be either serially uncorrelated or serially correlated with a minor order.
- However,  $\{r_t\}$  is a dependent series and the dependence arises from its conditional variance.



**Figure:** Time plot of monthly log returns of Intel stock from January 1973 to December 2008

# The sample ACF of $\{r_t\}$ and $\{r_t^2\}$



**Figure:** Sample ACF and PACF of various functions of monthly log stock returns of Intel Corporation from January 1973 to December 2008: (a) ACF of the log returns, (b) ACF of the squared log returns, (c) ACF of the absolute log returns, and (d) PACF of the squared log returns.



# Decompose $r_t$ into the mean and variance equations

To capture the dependence in a time series through its second moment but not the mean, we model the mean process and the variance process separately.

For a return series  $\{r_t\}$ , we can model it as

$$r_t = \mu_t + a_t \quad (1)$$

where  $\mu_t$  represents the conditional mean and  $a_t$  is modeled to capture the conditional variance.

# The mean equation

$$\mu_t = E(r_t \mid F_{t-1}) = \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i=1}^q \theta_i a_{t-i} \quad (2)$$

$$y_t = r_t - \phi_0 - \sum_{i=1}^k \beta_i x_{it}$$

$F_{t-1}$  is the information set at time  $t - 1$ .

If you combine these two equations, and let  $\mu_t = r_t - a_t$ , you will find that it is just an ARMA( $p, q$ ) model with additional regressors  $x_{it}$ .

# The variance equation

Denote the conditional variance of  $r_t$  with  $\sigma_t^2$ .

$$\begin{aligned}\sigma_t^2 &= \text{Var}(r_t \mid F_{t-1}) = E((r_t - E(r_t \mid F_{t-1})))^2 \mid F_{t-1}) \\ &= E((r_t - \mu_t)^2 \mid F_{t-1}) \\ &= \text{Var}(a_t \mid F_{t-1})\end{aligned}$$

# The variance equation (cont'd)

- If we assume that  $E(a_t | F_{t-1}) = 0$ , we can see that  $\sigma_t^2 = E(a_t^2 | F_{t-1})$ .
- We can use the lagged value of  $a_t^2$  to represent the information set  $F_{t-1}$
- The simplest model is a linear model, like the following

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2$$

# The procedure of building a volatility model

Building a volatility model for an asset return series consists of four steps:

- 1 Specify a mean equation by testing for serial dependence in the data and, if necessary, building an econometric model (e.g., an ARMA model) for the return series to remove any linear dependence.
- 2 Use the squared residuals of the mean equation to test for ARCH effects.
- 3 Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations.
- 4 Check the fitted model carefully and refine it if necessary.

# Testing for the presence of ARCH effect

## The Ljung-Box test for the series of $a_t^2$

Upon obtaining the residuals from the estimation of an adequate mean equation, we can use the squared residuals  $\{\hat{a}_t^2\}$  to test the existence of autocorrelation.

- The Ljung-Box test is used to test the null hypothesis  
 $H_0 : \rho_1 = \dots = \rho_m = 0$ .
- The  $Q(m)$  statistic is calculated and compared with the critical value from  $\chi^2(m)$  distribution at the desired significance level.
- The rejection of the null hypothesis implies that there is autoregressive conditional heteroskedastic (ARCH) effect.

# The LM test

## An auxiliary regression

We estimate a  $AR(m)$  model regarding  $\{\hat{a}_t^2\}$ , that is,

$$\hat{a}_t^2 = \alpha_0 + \alpha_1 \hat{a}_{t-1}^2 + \cdots + \alpha_m \hat{a}_{t-m}^2 + e_t$$

## The LM test

With this model, we test the joint hypothesis

$$H_0 : \alpha_1 = \cdots = \alpha_m = 0$$

- The LM statistic is  $NR^2$  where  $N$  is the sample size of this regression and  $R^2$  is the coefficient of the determination of this regression.
- Given the null hypothesis is true, this statistic follows a  $\chi^2(m)$  distribution.

# The LM test (cont'd)

Alternatively, we can use F statistic to test the joint hypothesis.

- Let  $SSR_0 = \sum_{t=m+1}^T (\hat{a}_t^2 - \bar{\omega})^2$ , where  $\bar{\omega} = (1/T) \sum_{t=1}^T \hat{a}_t^2$ .
- Let  $SSR_1 = \sum_{t=m+1}^T \hat{e}_t^2$  where  $\hat{e}_t$  is the residuals from the regression.
- The F statistic is

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)} \sim F(m, T - 2m - 1)$$

- Rejecting the null hypothesis motivates us to model the possible ARCH effect.



# An example

Go back to Figure 4. Since the return series is already stationary, we directly test the squared return series to check the ARCH effect.

- In the LM test of the ARCH effect,  $F = 53.62$  and the p value is close to zero.
- The Ljung–Box statistics of the  $a_t^2$  series also shows strong ARCH effects with  $Q(12) = 89.85$ , the p value of which is close to zero.
- Therefore, we can confirm that the return series of Intel stock has an ARCH effect, and next we need to model such an effect.

# The basic idea of an ARCH model

Consider a series of shocks  $\{a_t\}$  in a return series  $\{r_t\}$ . The basic idea of an Autoregressive Conditional Heteroskedasticity (ARCH) model is

- 1 the shock  $a_t$  of the return series is serially uncorrelated but dependent;
- 2 the dependence of  $a_t$  can be modeled through an autoregressive process of  $a_t^2$ .

# The ARCH(m) model

An ARCH(m) model takes the following form

$$a_t = \sigma_t \epsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 \quad (3)$$

where  $\epsilon_t \sim i.i.d.(0, 1)$ ,  $\alpha_0 > 0$  and  $\alpha_i \geq 0$  for  $i = 1, \dots, m$ .

- The assumption of  $\text{Var}(\epsilon_t) = 1$  is to make the analysis regarding the properties of the ARCH(m) model easy;
- The assumption of  $\alpha_0 > 0$  and  $\alpha_i \geq 0$  is to ensure the conditional variance of  $a_t$  is positive.
- $\alpha_1, \dots, \alpha_m$  should also satisfy some regularity conditions to ensure the unconditional variance of  $a_t$  is finite.

# The Properties of an ARCH Model

- Let's take an ARCH(1) model as an example to discuss the properties of ARCH model.
- The goal is to see how such a model can capture the basic idea mentioned above and the stylized fact that highly volatile periods tend to be followed by high volatility periods.

Assume an ARCH(1) model as follows

$$a_t = \sigma_t \epsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2, \epsilon_t \sim i.i.d.(0, 1) \quad (4)$$

where  $\alpha_0 > 0$  and  $\alpha_1 \geq 0$ .

# The unconditional mean and variance of $a_t$

## The unconditional mean

$$E(a_t) = E(\sigma_t \epsilon_t) = E(\sigma_t)E(\epsilon_t) = 0$$

The second equality is ensured because  $\sigma_t$  and  $\epsilon_t$  are independent, and the third equality comes from the assumption of  $E(\epsilon_t) = 0$ .

## The unconditional variance

$$\begin{aligned}\text{Var}(a_t) &= E(a_t^2) = E(\sigma_t^2 \epsilon_t^2) \\ &= E(\alpha_0 + \alpha_1 a_{t-1}^2) \cdot 1 = \alpha_0 + \alpha_1 \text{Var}(a_{t-1})\end{aligned}$$

Assuming the unconditional mean of  $a_t$  is a constant(why?), we can have

$$\text{Var}(a_t) = \frac{\alpha_0}{1 - \alpha_1}$$

Since the variance should be positive and finite, we must have  $0 \leq \alpha_1 < 1$ .

# The unconditional covariance of $a_t$

Since  $\epsilon_t$  and  $\epsilon_{t-i}$  for  $i \neq 0$  are independent,

$$\begin{aligned}\text{Cov}(a_t, a_{t-i}) &= E(a_t a_{t-i}) = E(\sigma_t \epsilon_t \sigma_{t-i} \epsilon_{t-i}) \\ &= E(\sigma_t \sigma_{t-i}) E(\epsilon_t \epsilon_{t-i}) = 0\end{aligned}$$

- What we get now?
  - $a_t$  has constant unconditional mean and variance,
  - $a_t$  is serially uncorrelated.

# The kurtosis of $a_t$

- Assume that  $\epsilon \sim N(0, 1)$ , implying that  $E(\epsilon_t^4) = 3$ . Thus, we have

$$\begin{aligned} E(a_t^4) &= E(\sigma_t^4 \epsilon_t^4) = E(\sigma_t^4) E(\epsilon_t^4) = 3E(\sigma_t^4) \\ &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 E(a_{t-1}^2) + \alpha_1^2 E(a_{t-1}^4)) \end{aligned}$$

- Assume that  $a_t$  is fourth-order stationary so that we can define  $m_4 = E(a_t^4) = E(a_{t-1}^4)$ . Then, using the fact that  $E(a_t^2) = \alpha_0/(1 - \alpha_1)$ , we can solve  $m_4$  from the above equation.

$$m_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}$$

# The kurtosis of $a_t$ (cont'd)

This result regarding  $m_4$  has two important implications:

- 1 Since the fourth moment of  $a_t$  is positive, we see that  $\alpha_1$  must also satisfy the condition  $1 - 3\alpha_1^2 > 0$ , that is,  $0 \leq \alpha_1^2 < \frac{1}{3}$ .
- 2 The kurtosis of  $a_t$  is

$$\text{kurtosis} = \frac{E(a_t^4)}{E(a_t^2)^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2} > 3$$

Thus, the the excess kurtosis of  $a_t$  is positive and the tail distribution of  $a_t$  is heavier than that of a normal distribution.



# The conditional mean

- Let's write  $E_{t-1}(a_t)$  to represent the conditional mean given the information set  $F_{t-1}$ , i.e.,  $E(E(a_t | F_{t-1}))$ .
- Since  $\epsilon_t$  is i.i.d, we have  $E_{t-1}(\epsilon_t) = E(\epsilon_t) = 0$ . Thus,

$$\begin{aligned} E_{t-1}(a_t) &= E_{t-1}(\sigma_t \epsilon_t) = E_{t-1} \left( (\alpha_0 + \alpha_1 a_{t-1}^2)^{1/2} \epsilon_t \right) \\ &= (\alpha_0 + \alpha_1 a_{t-1}^2)^{1/2} E_{t-1}(\epsilon_t) = 0 \end{aligned}$$

# The conditional variance

- The conditional variance of  $a_t$  is

$$\begin{aligned}\text{Var}_{t-1}(a_t) &= E_{t-1}(a_t^2) = E_{t-1}(\sigma_t^2 \epsilon_t^2) \\ &= E_{t-1}((\alpha_0 + \alpha_1 a_{t-1}^2) \epsilon_t^2) \\ &= E_{t-1}(\alpha_0 + \alpha_1 a_{t-1}^2) E_{t-1}(\epsilon_t^2) \\ &= (\alpha_0 + \alpha_1 a_{t-1}^2) E(\epsilon_t^2) \\ &= \alpha_0 + \alpha_1 a_{t-1}^2 = \sigma_t^2\end{aligned}$$

- How does the conditional variance capture the stylized fact?

# The definition of the likelihood function

The likelihood function is the joint density function  $f(\mathbf{y}|\boldsymbol{\theta})$  when we consider it as a function of the parameters  $\boldsymbol{\theta}$  given a set of data  $\mathbf{y}$ .

- $\mathbf{y} = (y_1, \dots, y_T)$  represents the observations, which are assumed to be identically independently distributed.
- $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  represents all parameters in the model (i.e., data generating process) that generates  $\mathbf{y}$ .
- For each observation  $y_t$ , its (marginal) PDF is  $f(y_t|\boldsymbol{\theta})$ .

# The likelihood function when $y_t$ is independent

The likelihood function

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{t=1}^T f(y_t|\boldsymbol{\theta}) \quad (5)$$

The log-likelihood function

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{t=1}^T \ell_t(\boldsymbol{\theta}|y_t) \quad (6)$$

where  $\ell_t(\boldsymbol{\theta}|y_t) = \ln(f(y_t|\boldsymbol{\theta}))$  is the **contribution** to the loglikelihood function.

# The joint density when $y_t$ is dependent

If  $y_1, y_2, \dots, y_T$  are dependent, their joint density can be written as

$$f(y_1, y_2, \dots, y_T) = f(y_1)f(y_2|y_1) \cdots f(y_T|y_1, \dots, y_{T-1})$$

or conveniently denoted as

$$f(\mathbf{y}^T) = \prod_{t=1}^T f(y_t|\mathbf{y}^{t-1})$$

# The likelihood and log-likelihood functions when $y_t$ is dependent

When  $\mathbf{y}$  is also dependent on  $\boldsymbol{\theta}$ , the likelihood function is then

$$L(\boldsymbol{\theta}|\mathbf{y}^T) = \prod_{t=1}^T f(y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}) \quad (7)$$

And the log-likelihood function is

$$\ell(\boldsymbol{\theta}|\mathbf{y}^T) = \sum_{t=1}^T \ell_t(\boldsymbol{\theta}|\mathbf{y}^t) \quad (8)$$

# The maximum likelihood (ML) estimator

- The ML estimator maximizes the log-likelihood function over the parameter space in which  $\theta$  lies in.

$$\max_{\{\theta \in \Theta\}} \ell(\theta | \mathbf{y}^T) = \sum_{t=1}^T \ell_t(\theta | \mathbf{y}^t) \quad (9)$$

- The ML estimator is usually obtained by computational methods, like the Newton or quasi-Newton method.

# The assumption of the distribution of $\epsilon_t$

Consider an ARCH(m) model

$$a_t = \sigma_t \epsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2$$

- Assume that  $\epsilon_t \sim N(0, 1)$  and  $\epsilon_t$  is i.i.d.
- How can we get the likelihood function of an ARCH(m) model based on this assumption?



# The distribution of $a_t$

The conditional distribution of  $a_t$

$$\epsilon_t \sim N(0, 1) \Rightarrow a_t | F_{t-1} \sim N(0, \sigma_t^2)$$

where  $\sigma_t^2$  is given by the ARCH(m) model. The conditional PDF of each  $a_t$  for  $t = 1, \dots, T$  is

The conditional PDF of  $a_t$

$$f(a_t | F_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right)$$

where  $F_{t-1}$  is the information set represented by  $(a_1, \dots, a_{t-1})$ .

The joint density for  $a_1, \dots, a_T$  given the parameter vector  $\alpha$

- For convenience, we suppress  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m)$  in the conditional joint density  $f(a_1, \dots, a_T \mid \alpha)$ .
- Since  $a_1, \dots, a_T$  are not independent, we have the joint density as

$$\begin{aligned} f(a_1, a_2, \dots, a_T) &= f(a_T \mid F_{T-1}) f(a_{T-1} \mid F_{T-2}) \cdots f(a_{m+1} \mid F_m) f(a_1, \dots, a_m) \\ &= \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right) \times f(a_1, \dots, a_m) \end{aligned} \quad (10)$$

# The conditional likelihood function

Dropping  $f(a_1, \dots, a_m)$  because the exact form of it is often complicated, we get the conditional likelihood function

$$L(\alpha|a_1, \dots, a_T) = \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right) \quad (11)$$

- $\alpha$  enters the likelihood function through

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2$$

- Maximizing the likelihood function results in the maximum likelihood estimator under the normality assumption.

# The conditional log-likelihood function

Taking logarithm of the likelihood function yields the log-likelihood function

$$\ell(\alpha|a_1, \dots, a_T) = \sum_{t=m+1}^T \left[ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_t^2) - \frac{1}{2} \frac{a_t^2}{\sigma_t^2} \right] \quad (12)$$

The log-likelihood function can be simplified as

$$\ell(\alpha|a_1, \dots, a_T) = -\frac{1}{2} \sum_{t=m+1}^T \left[ \ln(\sigma_t^2) + \frac{a_t^2}{\sigma_t^2} \right]$$

# Order determination

- Before estimating an ARCH(m) model, we need to determine the order  $m$ .
- The basic idea is that we treat an ARCH(m) model as an AR process of  $\{a_t^2\}$ , and apply the partial autocorrelation function (PACF) to determine  $m$ .

# Why using the PACF?

We justify the use of the PACF of  $\{a_t^2\}$  to determine  $m$  through two perspectives.

- ① We can consider  $a_t^2$  as an unbiased estimator of  $\sigma_t^2$  given the sample data because  $E_{t-1}(a_t^2) = \sigma_t^2$ . Therefore, we use  $a_t^2$  as an approximate to  $\sigma_t^2$ .
- ② We can define  $\eta_t = a_t^2 - \sigma_t^2$ . It can be shown that
  - $E(\eta_t) = 0$  and  $E(\eta_t \eta_{t-s}) = 0$  for  $s > 0$ .
  - But  $\eta_t$  is not i.i.d. because  $a_t^2$  is dependent.

So an ARCH(m) model is essentially an AR(m) model, except that  $\eta_t$  is not i.i.d. That is,

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 + \eta_t$$

# Alternative distribution assumptions of $\epsilon_t$

- $\epsilon_t$  has a Student-t distribution: capture the heavy-tail of volatility.
- $\epsilon_t$  has a skew-Student-t distribution: capture the skewness of asset return.
- $\epsilon_t$  has a general error distribution: enclose the normal distribution and heavy-tail distributions.

# Model checking

## The standardized residuals

Compute the standardized residuals

$$\tilde{a}_t = \frac{\hat{a}_t}{\hat{\sigma}_t}$$

which should mimic the behavior of  $\epsilon_t$ .

## Check the mean equation

- Use the Ljung-Box statistic for  $\{\tilde{a}_t\}$ .

## Check the volatility equation

- Use the Ljung-Box statistic for  $\{\tilde{a}_t\}$ .
- Use the QQ plot or the Shapiro-Wilk test for normality assumption.



# Forecasting

## 1-step-ahead forecast

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \cdots + \alpha_m a_{h+1-m}^2$$

## 2-step-ahead forecast

$$\sigma_h^2(2) = \alpha_0 + \alpha_1 \sigma_h^2(1) + \alpha_2 a_h^2 + \cdots + \alpha_m a_{h+2-m}^2$$

## $\ell$ -step-ahead forecast

$$\sigma_h^2(\ell) = \alpha_0 + \sum_{i=1}^m \alpha_i \sigma_h^2(\ell - i)$$

# An ARCH model for the monthly log returns of Intel stock

## The proposed model

$$r_t = \mu + \alpha_t, a_t = \sigma_t \epsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \alpha_3 a_{t-3}^2$$

## The estimated model

$$r_t = \underset{(0.0057)}{0.0122} + a_t, \sigma_t^2 = \underset{(0.0010)}{0.0106} + \underset{(0.0757)}{0.2131} a_{t-1}^2 + \underset{(0.0480)}{0.0770} a_{t-2}^2 + \underset{(0.0688)}{0.0599} a_{t-3}^2$$

Since  $\alpha_2$  and  $\alpha_3$  are statistically insignificant, we drop the last two terms and re-estimate the model.

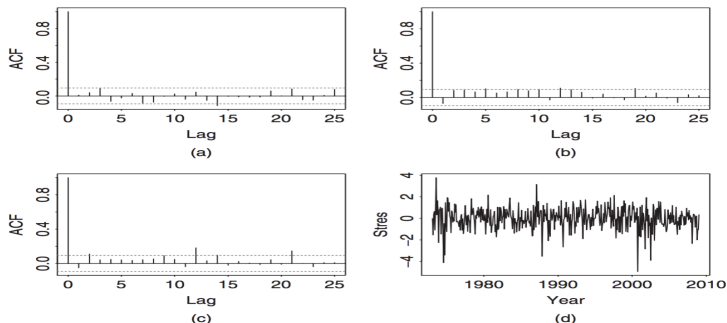
$$r_t = \underset{(0.0053)}{0.0126} + a_t, \sigma_t^2 = \underset{(0.0010)}{0.0111} + \underset{(0.0761)}{0.3560} a_{t-1}^2$$

# Model checking

## The Ljung-Box test

- $Q(10)$  for  $\{\tilde{a}_t\}$  is 12.64,  $p\text{-value} = 0.24 \Rightarrow$  no autocorrelation.
- $Q(10)$  for  $\{\tilde{a}_t^2\}$  is 14.75,  $p\text{-value} = 0.14 \Rightarrow$  no autocorrelation.

## The sample ACF



**Figure 3.5** Model checking statistics of Gaussian ARCH(1) model in Eq. (3.12) for monthly log returns of Intel stock from January 1973 to December 2008: Parts (a), (b), and (c) show sample ACF of standardized residuals, their squared series, and absolute series, respectively; part (d) is time plot of standardized residuals.