DOCTORAL PROGRAM
SWISS FINANCE INSTITUTE

# FINANCIAL ECONOMETRICS

**Eric Jondeau**

swiss:finance:institute

# FINANCIAL ECONOMETRICS

## *Lecture 2: Modeling Volatility - GARCH Models*

swiss:finance:institute

**Eric Jondeau**

# Objectives of the lecture

Old view of financial returns: returns are approximately **uncorrelated** (efficient market hypothesis).

$$r_t = \mu + \varepsilon_t$$

with the variance of $\varepsilon_t$ (and $r_t$) constant.

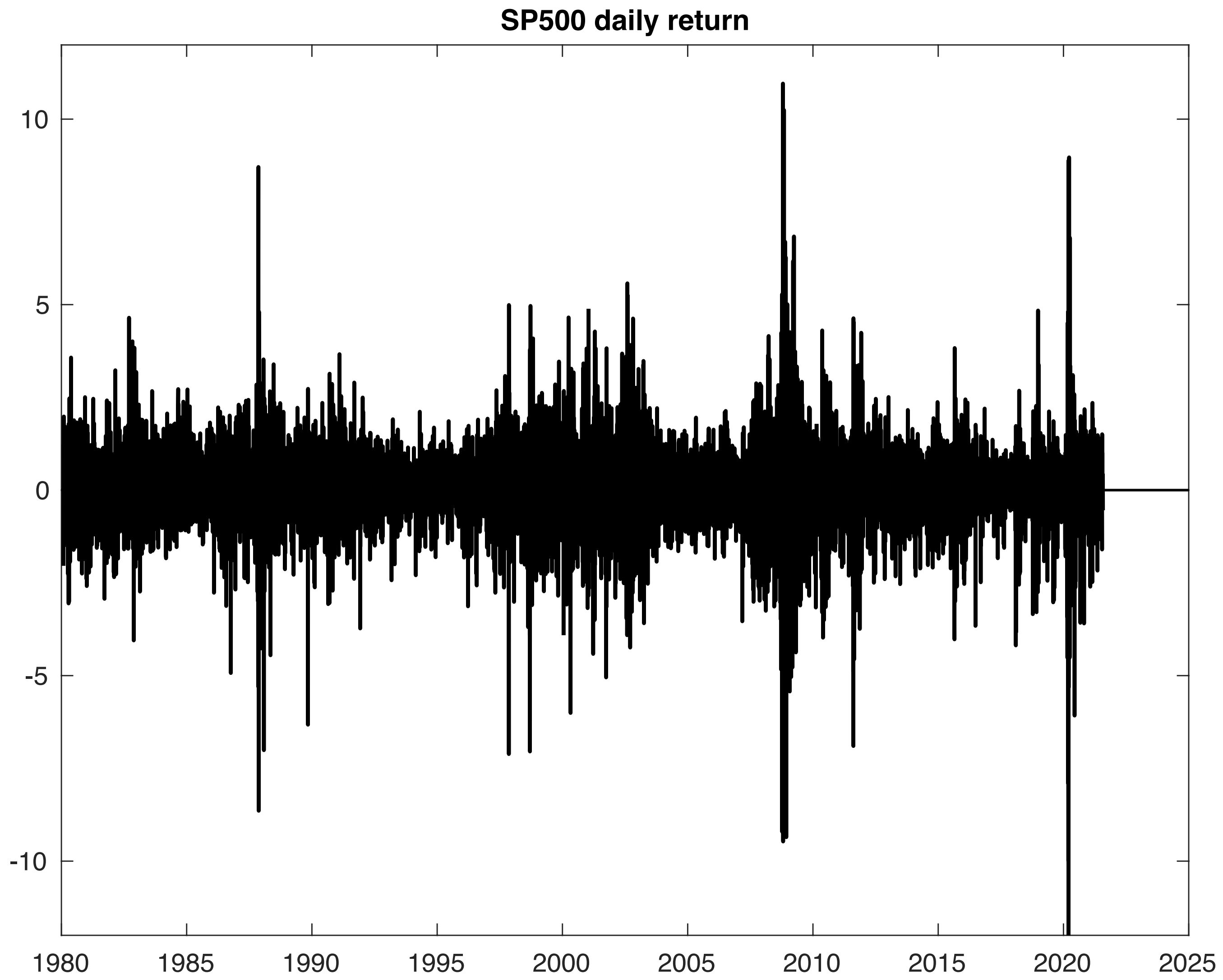
Old view mainly concerned with rationalizing the fat tails in the **unconditional** return distribution (Mandelbrot, 1963, *J. Business*, Fama, 1965, *J. Business*).

In practice, constant volatility hypothesis is clearly rejected by the data: Volatility tends to **cluster** in time

" *... large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes ...* " (Mandelbrot, 1963, *J. Business*)

Volatility clustering suggests that the **conditional** return distribution is **time-varying**.

# Objectives of the lecture − Daily returns



# Objectives of the lecture

Volatility is crucial in Finance because it is a **proxy for risk**:

- for **forecasting return**,
- for the **pricing of derivatives** such as options,
- for **asset allocation**, (trade-off between return and risk),
- for **risk management** (evaluation of the risk of a portfolio).

**Objectives of the Session:**

- Present different measures of volatility
- Identification and estimation of ARCH/GARCH models for conditional volatility
- Test of specification for volatility models.

# Objectives of the lecture

➔ **Definition of Volatility**

- **GARCH models: Presentation**
- **GARCH models: Estimation**
- **Extensions and Refinements**

# Measures of Volatility − Squared and absolute returns

Volatility is **not directly observable** from returns, and it varies through time. Therefore, we are interested in measuring the **conditional volatility** at time $t$ ($\sigma_t$).

Using **squared daily returns** to proxy daily volatility will produce a very noisy volatility estimator. Assume we have

$$r_t = \mu + \varepsilon_t \qquad \text{with} \qquad \varepsilon_t = \sigma_t z_t,\ z_t \sim N(0,1), \text{ and } \sigma_t \text{ known at date } t\text{–}1.$$

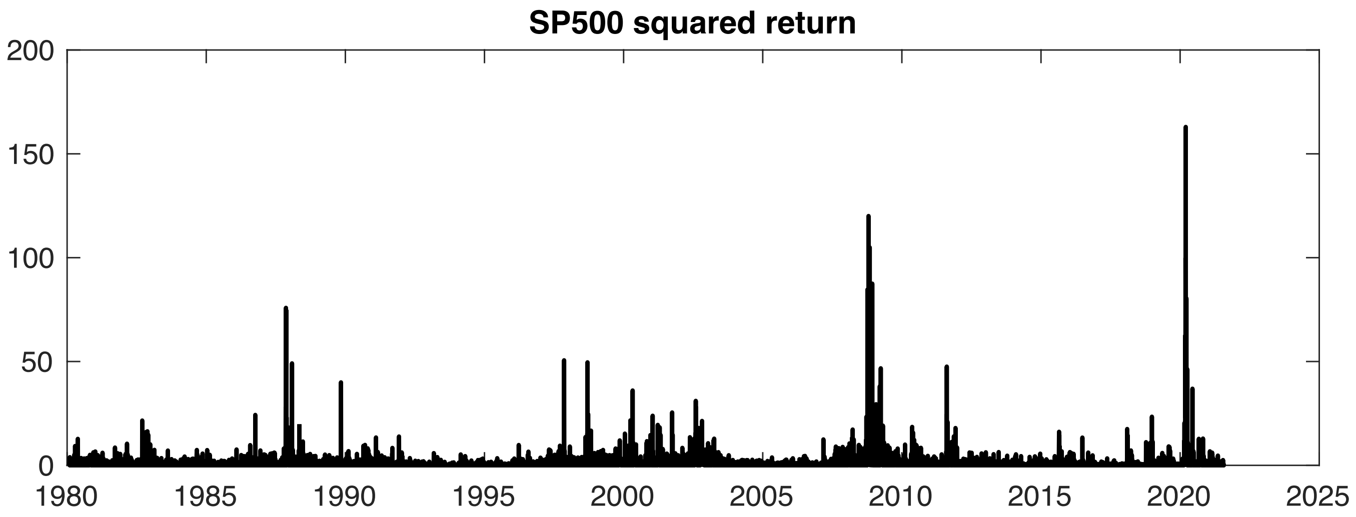
Then $E[\varepsilon_t^2 \mid I_{t-1}] = \sigma_t^2 E[z_t^2 \mid I_{t-1}] = \sigma_t^2$ because $z_t^2 \sim \chi^2(1)$, with $I_{t-1}$ the information set at time $t-1$. However, due to the form of the $\chi^2(1)$ distribution, we have

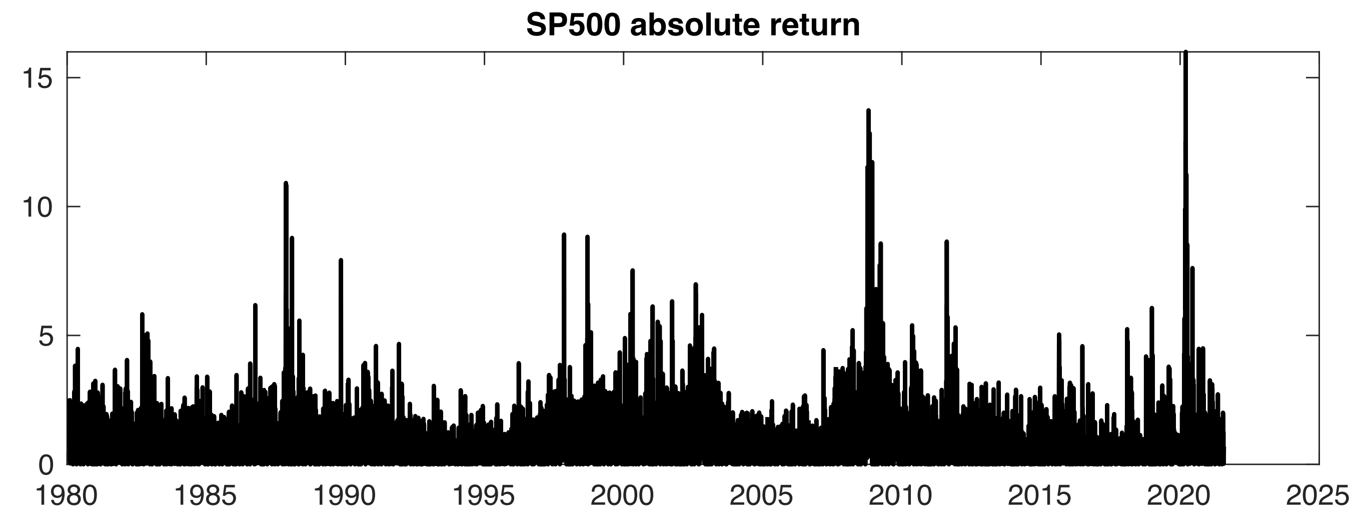
$$\Pr\left[\varepsilon_t^2 \in \left[\frac{1}{2}\sigma_t^2 ; \frac{3}{2}\sigma_t^2\right]\right] = \Pr\left[z_t^2 \in \left[\frac{1}{2};\frac{3}{2}\right]\right] = 0.2588$$

which means that $\varepsilon_t^2$ is 50% greater or smaller than $\sigma_t^2$ nearly 75% of the time.

An alternative measure of volatility is the **absolute return**. Indeed, if $\varepsilon_t \sim N(0,\sigma_t^2)$ then $E[|\varepsilon_t|] = \sigma_t \sqrt{2/\pi}$. So, $\hat{\sigma}_t = |\varepsilon_t| / \sqrt{2/\pi}$ is a proxy for $\sigma_t$.

# Measures of Volatility – Squared and absolute returns



# Measures of Volatility − Historical volatility

A very simple measure of **historical volatility** is **moving average**.

The standard moving average is computed as follows:

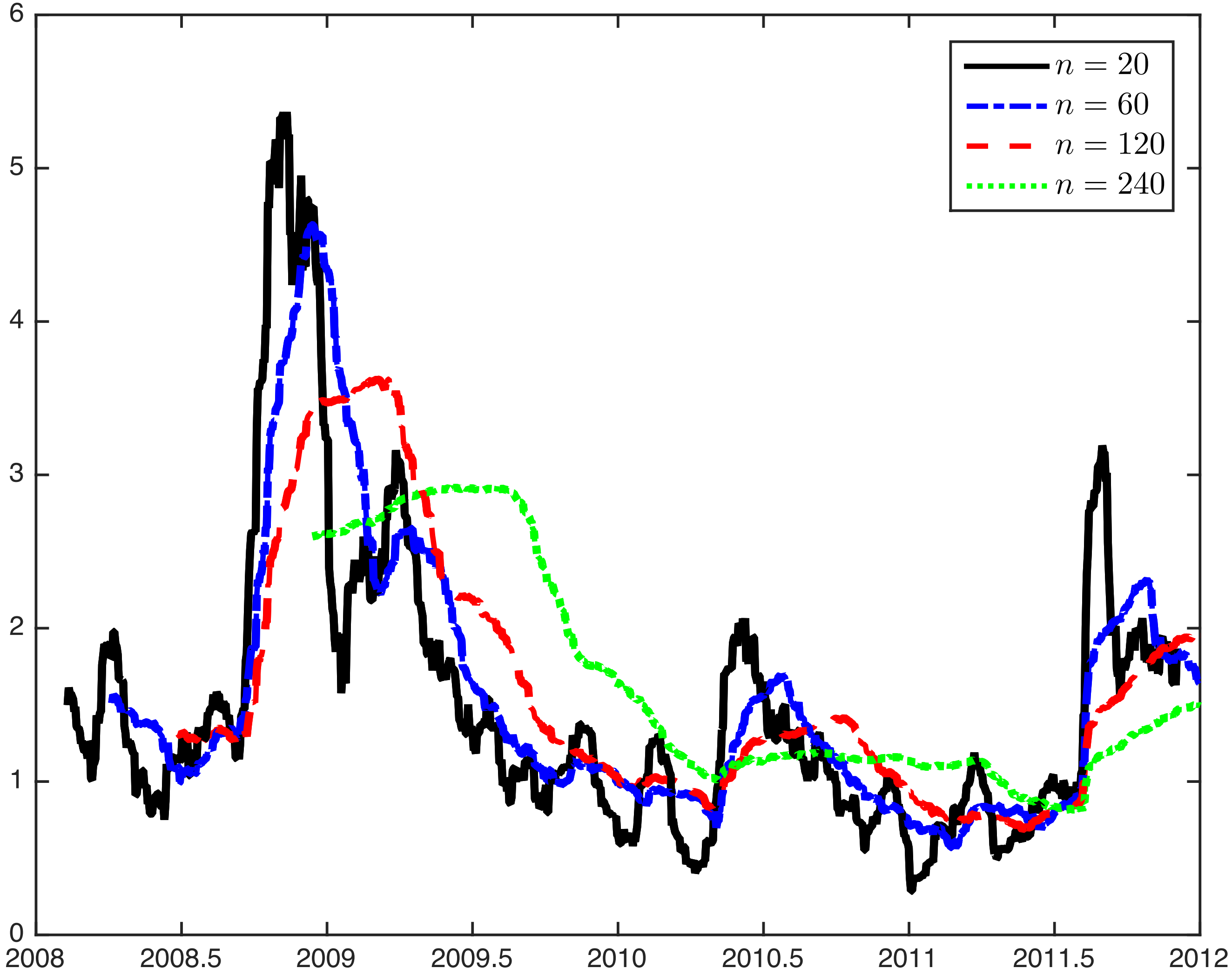
$$\sigma_t^2 = \frac{1}{n}\sum_{i=1}^{n}(r_{t-i} - \hat{\mu})^2 \qquad t = n+1,\ldots,T \qquad \text{with} \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} r_{t-i}$$

The problem with this definition is that it gives the same weight to old information ($r_{t-n}$) and to new information ($r_{t-1}$).

Drawback: Historical volatility generates **ghost features**. After an extreme event, historical volatility jumps and stays at a high level for as long as the averaging period.

Historical volatility can be useful to measure long-term volatility but is not useful for measuring short-term volatility.

# Measures of Volatility − Historical volatility



# Measures of Volatility – EWMA

An alternative measure of volatility is to overweight new information, as in the **exponentially weighted moving average (EWMA)**:

$$\sigma_t^2 = \phi\sigma_{t-1}^2 + (1-\phi)(r_{t-1} - \hat{\mu})^2 = (1-\phi)\sum_{k=1}^{t-1}\phi^{k-1}(r_{t-k} - \hat{\mu})^2 \qquad t = 2,\ldots,T$$

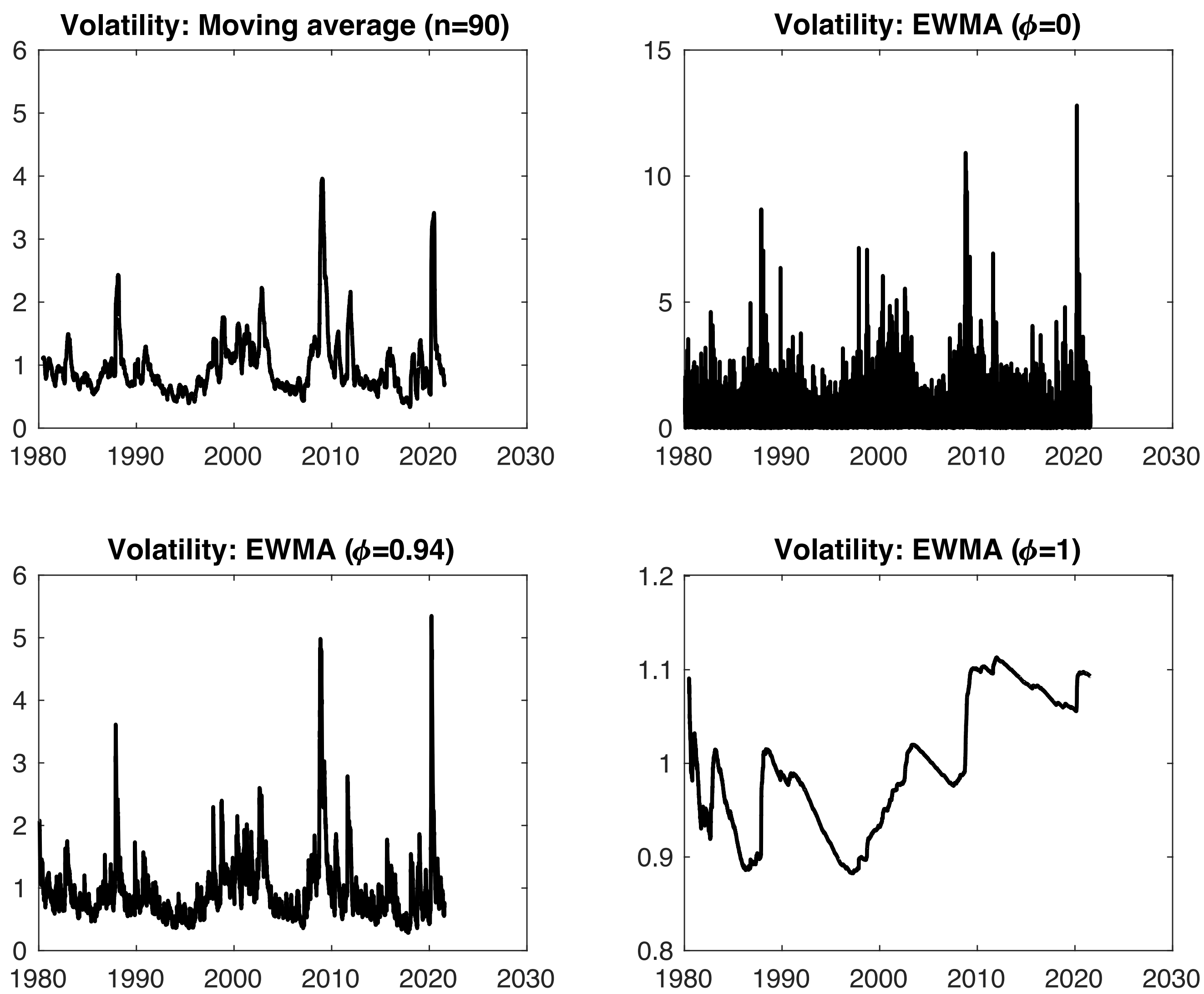
so that the weights decrease over time. $\phi \in [0;1]$.

$(1-\phi)(r_{t-1} - \hat{\mu})^2$: measures the intensity of the reaction of volatility to market events.
$\phi\sigma_{t-1}^2$: measures the persistence in volatility.

For $\phi = 0$, we have $\sigma_t^2 = (r_{t-1} - \hat{\mu})^2$ (last return)
For $\phi \to 1$, we have $\sigma_t^2 = \frac{1}{t-1}\sum_{k=1}^{t-1}(r_{t-k} - \hat{\mu})^2$ (historical volatility)

A recommended value is between $\phi = 0.75$ and 0.98.

# Measures of Volatility – EWMA



# Measures of Volatility − RiskMetrics

Special case of EWMA estimator: RiskMetrics (www.riskmetrics.com)

Basic assumption: log-returns are conditionally normal $r_{i,t} \sim N(0,\sigma_{i,t}^2)$

Daily volatility is measured with $\hat{\mu} = 0$ and $\phi = 0.94$:

$$\sigma_{i,t}^2 = \phi\sigma_{i,t-1}^2 + (1-\phi)r_{i,t-1}^2$$

The same approach is used to forecast covariances between assets:

$$\sigma_{ij,t} = \phi\sigma_{ij,t-1} + (1-\phi)r_{i,t-1}r_{j,t-1}$$

This is convenient for estimating large-dimensional covariance matrices (equity indices, interest rates, foreign exchange rate, commodities). The same $\phi$ is required to ensure a semi-definite positive covariance matrix.

The value $\phi = 0.94$ may be a good choice for some assets, but not for other assets.

# Measures of Volatility − Square Root of Time Rule

In some cases, we need to forecast the variance for different horizon.

Assume we have daily log-returns $r_t$ and we want to measure the variance of the process for a horizon $k$, i.e., for the cumulated return between $t$ and $t$+$k$:

$$r_t[k] = \log(P_{t+k}) - \log(P_t) = r_{t+1} + \cdots + r_{t+k}$$

If the return process is not serially correlated and the daily variance does not vary between $t$ and $t$+$k$ ($\sigma^2$), then

$$V[r_t[k]] = V[r_{t+1} + \cdots + r_{t+k}] = k \times V[r_{t+1}] = k \times \sigma^2$$

so that the $k$-day volatility is just $\sqrt{k}$ times the 1-day volatility.

This is the **square root of time rule**. This rule is widely used, although it may not be supported by empirical evidence.

# Objectives of the lecture

- Definition of Volatility

➔ **GARCH models: Presentation**

- GARCH models: Estimation

- Extensions and Refinements

# Structure of a volatility model

Let $r_t$ be the log-return of an asset at time $t$. Assume the following model

$$r_t = \mu_t + \varepsilon_t \qquad \text{with} \qquad z_t \sim iid(0,1)$$
$$\varepsilon_t = \sigma_t z_t$$

where

$\mu_t = E[r_t \mid I_{t-1}]$ is the conditional mean of $r_t$

$\sigma_t^2 = E[(r_t - \mu_t)^2 \mid I_{t-1}] = E[\varepsilon_t^2 \mid I_{t-1}]$ is the conditional variance of $r_t$.

A volatility model is a model that describes the evolution of $\sigma_t^2$.

There are essentially two types of models for describing the dynamics of volatility:

1. in the first category, volatility is described as an exact function of a given set of variables. This category includes **(G)ARCH models**: $\sigma_t^2 = E[\varepsilon_t^2 \mid I_{t-1}]$.

2. in the second category, volatility is described as a stochastic function. It includes **Stochastic Volatility models**: $\sigma_t^2 = E[\varepsilon_t^2 \mid I_{t-1}] + u_t$.

# ARCH Model

The ARCH (**AutoRegressive Conditional Heteroskedasticity**) model was introduced by Engle (1982, *Econometrica*).

We first describe the **initial ARCH model**.

**Basic idea**: unexpected returns $\varepsilon_t$ are serially uncorrelated but dependent. The time dependency of $\varepsilon_t^2$ is described by a quadratic function of its lagged values:

$$\varepsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \qquad \text{(known at time } t-1\text{)}$$

with $\quad z_t \sim iid\ N(0,1)$

Constraints for **strictly positive variance**: $\omega > 0$ and $\alpha_i \geq 0$ for $i = 1, \ldots, p$.

# ARCH Model

The **unconditional variance** of $r_t$ is

$$\sigma^2 = E\left[\varepsilon_t^2\right] = E\left[E_{t-1}\left[\varepsilon_t^2\right]\right] = E\left[\sigma_t^2\right] = \omega / \left(1 - \sum_{i=1}^{p} \alpha_i\right)$$

Therefore, the process $\sigma_t^2$ is **covariance stationary** iif $\sum_{i=1}^{p} \alpha_i < 1$ and $\omega < +\infty$. In this case, the process $\varepsilon_t$ has time-varying conditional variance, but its unconditional variance $\sigma^2$ is a finite constant. Therefore, $\varepsilon_t$ is also **covariance stationary**.

We also have

$$\varepsilon_t^2 = \sigma_t^2 z_t^2 = \sigma_t^2 + \sigma_t^2 (z_t^2 - 1) = \sigma_t^2 + v_t = E_t[\varepsilon_t^2] + v_t$$

where $v_t$ is the innovation of $\varepsilon_t^2$, with $E_t[v_t] = 0$.

Therefore, we have

$$\varepsilon_t^2 = \sigma_t^2 + v_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2 + v_t$$

Therefore, $\varepsilon_t^2$ is an **AR(*p*) process**.

# Example

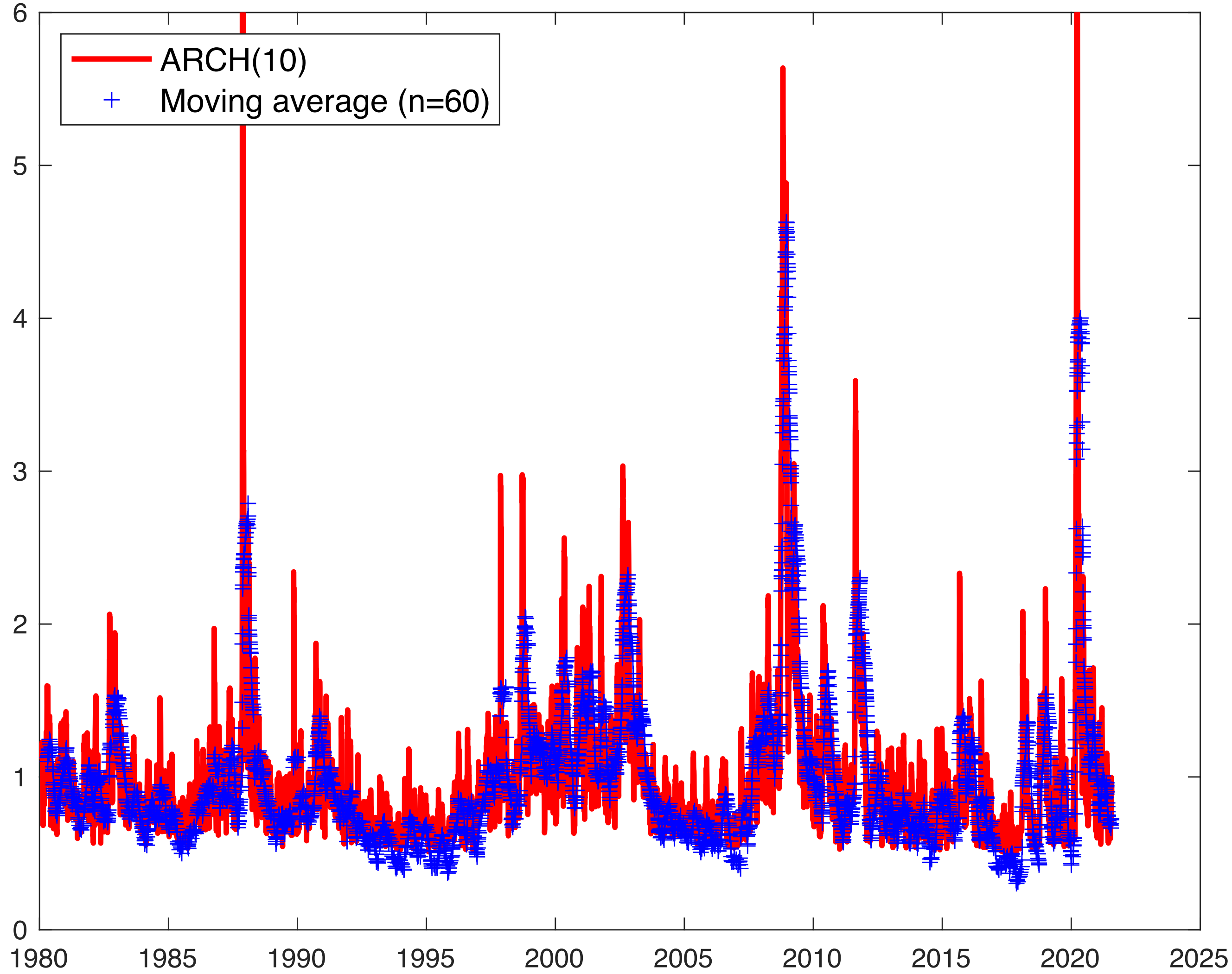
Estimation of an ARCH(10) for the S&P500 daily return between January 1980 and July 2021.

The condition mean is given by $r_t = \mu + \varepsilon_t$.

| | Estimate | Std. dev | t-stat |
|---|---|---|---|
| $a_0$ | 0.237 | 0.012 | 20.159 |
| $a_1$ | 0.081 | 0.010 | 7.961 |
| $a_2$ | 0.112 | 0.012 | 9.534 |
| $a_3$ | 0.085 | 0.011 | 7.595 |
| $a_4$ | 0.101 | 0.012 | 8.483 |
| $a_5$ | 0.087 | 0.011 | 7.637 |
| $a_6$ | 0.068 | 0.010 | 6.524 |
| $a_7$ | 0.072 | 0.011 | 6.699 |
| $a_8$ | 0.062 | 0.010 | 5.890 |
| $a_9$ | 0.080 | 0.011 | 7.079 |
| $a_{10}$ | 0.062 | 0.010 | 5.924 |

$$\log\left(L_T(\hat{\theta})\right) = -14397.164$$

# Example



ARCH(10)
Moving average (n=60)
0
1
2
3
4
5
6
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025

# GARCH Model

Large persistence in volatility: the ARCH model requires a large *p* to fit the data.

In such cases, it is more parsimonious to use the GARCH (**Generalized ARCH**) model proposed by Bollerslev (1986, *J. Econometrics*).

The GARCH($p$,$q$) model is defined as

$$\varepsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Constraints on parameters are: $\omega > 0$ and $\alpha_i \geq 0$ for $i = 1, ..., p$ and $\beta_j \geq 0$ for $j = 1, ..., q$ to ensure strictly positive variance.

The unconditional variance of $r_t$ is given by $\sigma^2 = \omega / \left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j\right)$.

Therefore, $\varepsilon_t$ is covariance stationary iif the sum of $\alpha_i$ and $\beta_j$ parameters is less than one: $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ and $\omega < +\infty$.

# Example: GARCH(1,1)

The GARCH(1,1) model is: $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

Since $\sigma_t^2 = E\left[\varepsilon_t^2 \mid \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \cdots\right]$, this gives

$$\varepsilon_t^2 = \sigma_t^2 + v_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + v_t = \omega + \gamma\varepsilon_{t-1}^2 + v_t - \beta v_{t-1}$$

with $\gamma = \alpha + \beta$ the persistence parameter and $v_t$ the innovation of $\varepsilon_t^2$ $(E[v_t] = 0)$

Therefore, $\varepsilon_t^2$ is an **ARMA(1,1) process**.

# Example

Estimation of a GARCH(1,1) for the S&P500 daily return between January 1980 and July 2021. The condition mean is given by $r_t = \mu + \varepsilon_t$.

| | Estimate | Std. dev | t-stat |
|---|---|---|---|
| $\omega$ | 0.017 | 0.002 | 9.511 |
| $\alpha_1$ | 0.090 | 0.005 | 17.899 |
| $\beta_1$ | 0.896 | 0.005 | 165.013 |

$\log\left(L_T(\hat{\theta})\right) = -14339.6831$ (−14397.164 for the ARCH(10))

$\alpha_1 + \beta_1 = 0.986$

# Excess kurtosis

GARCH models can generate excess kurtosis. Consider the **ARCH(1) model**:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \qquad \text{where} \qquad \varepsilon_t = \sigma_t z_t$$

The unconditional variance is $\sigma^2 = \omega / (1 - \alpha_1)$ so that $\alpha_1 < 1$.

Under normality, the **conditional fourth moment** is given by

$$E[\varepsilon_t^4 \mid I_{t-1}] = E[\sigma_t^4 z_t^4 \mid I_{t-1}] = 3E[\sigma_t^4 \mid I_{t-1}] = 3\left(\omega + \alpha_1 \varepsilon_{t-1}^2\right)^2$$

and the **unconditional fourth moment** is given by

$$m_4 = E[\varepsilon_t^4] = E[E[\varepsilon_t^4 \mid I_{t-1}]] = 3E[(\omega + \alpha_1 \varepsilon_{t-1}^2)^2] = 3\left(\omega^2 + 2\omega\alpha_1 E[\varepsilon_{t-1}^2] + \alpha_1^2 m_4\right)$$

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

# Excess kurtosis

Since the fourth moment of $\varepsilon_t$ is positive, we must have $(1-3\alpha_1^2)>0$ so that $\alpha_1^2<1/3$. The unconditional kurtosis is

$$\kappa = \frac{m_4}{\sigma^4} = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}\frac{(1-\alpha_1)^2}{\omega^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$$

So, **the excess kurtosis is always positive** and the tails of the distribution of $\varepsilon_t$ are fatter than those of a normal distribution, even if the conditional distribution is normal.

For more general formulation of the ARCH model, it becomes very complicated to obtain the unconditional kurtosis, but the excess kurtosis is still positive.

# Forecasting

Forecasts of the GARCH model are obtained recursively. Let $t$ be the starting date for forecasting.

The **1-step ahead forecast** for $\sigma_{t+1}^2$ is $\qquad \sigma_t^2(1) = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2$

Since $\varepsilon_t^2 = \sigma_t^2 z_t^2$, the GARCH(1,1) can be rewritten

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + (\alpha_1 + \beta_1)\sigma_{t-1}^2 + \alpha_1 \sigma_{t-1}^2 (z_{t-1}^2 - 1)$$

so that, at time $t$+2, we have

$$\sigma_{t+2}^2 = \omega + (\alpha_1 + \beta_1)\sigma_{t+1}^2 + \alpha_1 \sigma_{t+1}^2 (z_{t+1}^2 - 1) \quad \text{with } E[(z_{t+1}^2 - 1) \mid I_t] = 0$$

The **2-step ahead forecast** for $\sigma_{t+2}^2$ is $\qquad \sigma_t^2(2) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_t^2(1)$

The ***κ*-step ahead forecast** for $\sigma_{t+\kappa}^2$ is $\qquad \sigma_t^2(\kappa) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\sigma_t^2(\kappa - 1) \qquad$ for $\kappa > 1$

# Objectives of the lecture

- Definition of Volatility
- GARCH models: Presentation

➔ **GARCH models: Estimation**

- Extensions and Refinements

# ML Estimation: Principle

The MLE principle provides a means of estimating a parameter or set of parameters.

The **principle**: Given a pdf and independence of observations $(x_1,...,x_T)$

$$L_T(\theta) = \prod_{t=1}^{T} f(x_t;\theta)$$

is known as the **likelihood function**.

The likelihood equation for the maximum is:

$$\frac{\partial \log(L_T(\theta))}{\partial \theta} = 0$$

# ML Estimation: Properties

Under some regularity conditions, the ML estimator has the following properties:

- **Consistency**: $\text{plim}\ \hat{\theta}_{ML} = \theta_0$ as $T \to \infty$

- **Asymptotic normality**: $\sqrt{T}\left(\hat{\theta}_{ML} - \theta_0\right) \overset{d}{\sim} N(0, I(\theta_0)^{-1})$ as $T \to \infty$

where $I(\theta_0) = -E\left[\dfrac{\partial^2 \log(\ell_t(\theta_0))}{\partial\theta\partial\theta'}\right] = E\left[\dfrac{\partial \log(\ell_t(\theta_0))}{\partial\theta}\dfrac{\partial \log(\ell_t(\theta_0))'}{\partial\theta}\right]$

is the Fisher Information matrix of the observation at date $t$.

- **Asymptotic efficiency**: the asymptotic variance of $\hat{\theta}_{ML}$ reaches the Cramér-Rao lower bound:

$$Asy.Var[\hat{\theta}_{ML}] = (1/T)\, I(\theta_0)^{-1} \qquad \text{or} \qquad \hat{\theta}_{ML} \overset{d}{\sim} N(\theta_0, (1/T)\, I(\theta_0)^{-1})$$

- **Invariance**: if $\hat{\theta}_{ML}$ is the MLE of $\theta_0$ and if $g(\theta_0)$ is a continuous function, the MLE of $g(\theta_0)$ is $g(\hat{\theta}_{ML})$.

# ML Estimation: Case of a GARCH($p,q$) model

We consider the estimation of a GARCH($p,q$) model, with $m = \max(p,q)$.

Under normality, the **likelihood function** of the model is

$$L_T(\theta) = f(\varepsilon_{-m+1},\ldots,\varepsilon_1,\ldots,\varepsilon_T;\theta) = f(\varepsilon_T \mid I_{T-1}) \times \ldots \times f(\varepsilon_1 \mid I_0) \times f(\varepsilon_{-m+1},\ldots,\varepsilon_0;\theta)$$

$$L_T(\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \times f(\varepsilon_{-m+1},\ldots,\varepsilon_0;\theta)$$

where $\theta = (\omega,\alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q)'$ is the vector of unknown parameters. $f(\varepsilon_{-m+1},\ldots,\varepsilon_0;\theta)$ is the stationary joint pdf of $(\varepsilon_{-m+1},\ldots,\varepsilon_0)'$.

In general, we assume that this term is negligible, so that we drop it and focus on the **conditional likelihood function**:

$$L_T(\theta \mid \varepsilon_{-m+1},\ldots,\varepsilon_0) = f(\varepsilon_1,\ldots,\varepsilon_T;\theta \mid \varepsilon_{-m+1},\ldots,\varepsilon_0) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right)$$

# ML Estimation: Gradient and Hessian

The ML estimator is obtained by maximizing the likelihood, which is equivalent to maximizing the **log-likelihood function**

$$\log L_T(\theta) = \sum_{t=1}^{T} \log \ell_t(\theta) \qquad \text{where} \qquad \log(\ell_t(\theta)) = -\frac{1}{2}\left(\log(2\pi) + \log(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right)$$

is the log-likelihood of the observation $t$.

The first-order derivative vector (**Gradient**) is

$$g(\theta) = \frac{\partial \log(L_T(\theta))}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \log(\ell_t(\theta))}{\partial \theta} = \sum_{t=1}^{T} g_t(\theta) \qquad \text{with} \qquad E[g(\theta_0)] = 0$$

The second-order derivative matrix (**Hessian**) is

$$H(\theta) = \frac{\partial^2 \log(L_T(\theta))}{\partial \theta \partial \theta'} = \sum_{t=1}^{T} \frac{\partial^2 \log(\ell_t(\theta))}{\partial \theta \partial \theta'}$$

# ML Estimation: Asymptotic distribution

The Hessian is used to compute **covariance matrix** of the ML estimator, which is given by the inverse of the Fisher information matrix for an observation, $I(\theta)$.

The information matrix is

$$I(\theta) = -E\left[\frac{\partial^2 \log\big(\ell_t(\theta)\big)}{\partial\theta\partial\theta'}\right] = -\frac{1}{T}E\left[\frac{\partial^2 \log\big(L_t(\theta)\big)}{\partial\theta\partial\theta'}\right] = -\frac{1}{T}E[H(\theta)]$$

The asymptotic distribution of the ML estimator is $\sqrt{T}(\hat{\theta}-\theta_0) \overset{d}{\sim} N(0, I(\theta_0)^{-1})$

**(See the Appendix for the proof of asymptotic normality of the MLE)**

# ML Estimation: Asymptotic distribution

## Alternative measure of the information matrix

A drawback of the Hessian form of the information matrix $\hat{I}(\theta)$ is that it relies on the second-order derivatives of the log-likelihood. In practice, this measure may be very erratic. We can use the alternative estimator of $I(\theta)$ based on the first-order derivatives

$$I(\theta) = E\left[\frac{\partial \log(\ell_t(\theta))}{\partial \theta}\frac{\partial \log(\ell_t(\theta))'}{\partial \theta}\right] = \frac{1}{T}E\left[\frac{\partial \log(L_T(\theta))}{\partial \theta}\frac{\partial \log(L_T(\theta))'}{\partial \theta}\right]$$

$$= \frac{1}{T}E\left[G(\theta)G(\theta)'\right]$$

where

$$G(\theta) = \begin{bmatrix} \dfrac{\partial \log(\ell_1(\theta))}{\partial \theta} & \cdots & \dfrac{\partial \log(\ell_T(\theta))}{\partial \theta} \end{bmatrix}$$

This estimator is known as the **BHHH** estimator, or **outer product of gradients** (OPG) estimator. The two estimators are equivalent asymptotically under normality. In general, the BHHH estimator is easier to compute.

# Estimation in practice

1. Estimate the mean equation $r_t = \mu_t + \varepsilon_t$. Deduce: $\hat{\varepsilon}_t = r_t - \hat{\mu}_t$ and $\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$

2. Select initial values $\theta^{(0)} = (\omega^{(0)}, \alpha_1^{(0)}, \ldots, \alpha_p^{(0)}, \beta_1^{(0)}, \ldots, \beta_q^{(0)})'$ and $\hat{\varepsilon}_1^2 = \ldots = \hat{\varepsilon}_p^2 = \hat{\sigma}_\varepsilon^2$

3. Compute the conditional volatility $\hat{\sigma}_t^2 = \omega^{(0)} + \sum_{i=1}^{p}\alpha_i^{(0)}\hat{\varepsilon}_{t-i}^2 + \sum_{j=1}^{q}\beta_j^{(0)}\hat{\sigma}_{t-j}^2$ for each $t$

4. Compute the log-likelihood

$$\log(L_T(\theta^{(0)})) = \sum_{t=1}^{T}\log(\ell_t(\theta^{(0)})) \quad \text{where} \quad \log(\ell_t(\theta^{(0)})) = -\frac{1}{2}\left(\log(2\pi) + \log(\hat{\sigma}_t^2) + \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}_t^2}\right)$$

5. Change the value of parameters, say $\theta^{(1)}$, so that the log-likelihood increases

$$\log(L_T(\theta^{(1)})) > \log(L_T(\theta^{(0)}))$$

6. Iterate steps 3 to 5 until convergence of the log-likelihood to a fixed value.

**(See the Appendix for Algorithm for the estimation of parameters)**

# Case of a regression model

For a regression model of the type

$$r_t = X_t \delta + \varepsilon_t$$
$$\varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \qquad \text{with} \qquad z_t \sim iid\ N(0,1)$$

we should estimate parameters $\delta$ and $\theta$ simultaneously. It turns out that

$$E\left[\frac{\partial^2 \log(\ell_t(\theta,\delta))}{\partial\theta\partial\delta'}\right] = 0 \qquad \text{(See the proof in Engle, 1982)}$$

so that the information matrix is block-diagonal and the two sets of parameters can be estimated separately.

**Remark**: Two-step estimation does not work when $X_t$ depends on $\sigma_t$ (GARCH-in-Mean model)

# Testing for ARCH effects

Engle (1982) has proposed a **Lagrange-multiplier (LM) test** for ARCH effects.

The null hypothesis is $H_0 : \varepsilon_t \mid I_{t-1} \sim N(0, \sigma^2)$
vs. the alternative is $H_a : \varepsilon_t \mid I_{t-1} \sim ARCH(p)$

The LM test of ARCH($p$) effects is based on the null hypothesis $H_0 : \alpha_1 = \cdots = \alpha_p = 0$ against the alternative $H_a : \alpha_1 \geq 0, \cdots, \alpha_p \geq 0$ with at least one strict inequality.

The LM test statistic is equivalent to the $T \times R^2$ test statistic, where $T$ is the sample size and $R^2$ is computed from the regression

$$\hat{\varepsilon}_t^2 = a_0 + a_1 \hat{\varepsilon}_{t-1}^2 + \cdots + a_p \hat{\varepsilon}_{t-p}^2 + v_t \qquad \text{for } t = 1, \cdots, T$$

Under the null of no ARCH effect, the $T \times R^2$ is asymptotically distributed as a $\chi^2(p)$.

**Remark:** Alternatively, we could also use the Ljung-Box statistic for $\hat{\varepsilon}_t^2$ with $p$ lags, which is asymptotically distributed as a $\chi^2(p)$.

# Testing for GARCH effects?

Assume we want to test the null hypothesis that the process is homoskedastic against the alternative that the variance is a GARCH(1,1) process.

The null hypothesis is $H_0 : \alpha_1 = \beta_1 = 0$, against $H_a : \alpha_1 \geq 0, \beta_1 \geq 0$ with at least one strict inequality.

However, assume now that $\alpha_1 = 0$, so that the GARCH(1,1) process writes

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 = \frac{\omega}{1-\beta_1} + \beta_1 \left( \sigma_{t-1}^2 - \frac{\omega}{1-\beta_1} \right) = \sigma^2 + \beta_1 (\sigma_{t-1}^2 - \sigma^2)$$

If we set, at date $t = 0$, $\sigma_0^2 = \sigma^2$ and $\sigma_t^2 = \sigma^2, \forall t$. So, $\beta_1$ is unidentified under the null.

The test of $H_0 : \alpha_1 = \beta_1 = 0$, against $H_a : \alpha_1 \geq 0$, $\beta_1 \geq 0$ with at least one strict inequality, is equivalent to the test of no ARCH(1).

The reason is that, under the null, GARCH(1,1) and ARCH(1) models are locally equivalent.

# Objectives of the lecture

- Definition of Volatility
- GARCH models: Presentation
- GARCH models: Estimation

➔ **Extensions and Refinements**

# Extensions and refinements

- Volatility asymmetries (next slides)
- Volatility persistence (see Appendix 3)
- Conditional error distributions (lecture 3)
- Multivariate models (lecture 4)
- …

# Asymmetric GARCH Model

Standard GARCH model: positive and negative news have a **symmetric effect** on the conditional volatility.

Empirically, negative returns tend to be followed by larger increases in volatility than equally large positive returns (Black, 1976, *JPE*, Christie, 1982, *IER*).

This suggests that bad news have a stronger effect on volatility than good news.

**Intuition:** Most firms have some debt. A price fall increases the probability to default, and therefore increases the risk (volatility) of the firm's equity.

Several asymmetric GARCH models have been proposed:

- GJR model
- Threshold GARCH model
- Exponential GARCH model
- …

# Asymmetric GARCH Model − GJR Model

In the **GJR** model (Glosten, Jagannathan, and Runkle, 1993, *J. Finance*), $\sigma_t^2$ depends on both the size and the sign of lagged innovations:

$$\sigma_t^2 = \omega + [\alpha_1 \underbrace{\varepsilon_{t-1}^2}_{\text{size}} + \gamma_1 \underbrace{\Pi_{t-1}^- \varepsilon_{t-1}^2}_{\text{sign}}] + \beta_1 \sigma_{t-1}^2$$

with $\Pi_t^-$ equal to 1 if $\varepsilon_t < 0$, and 0 otherwise.

The conditional volatility is positive when parameters satisfy

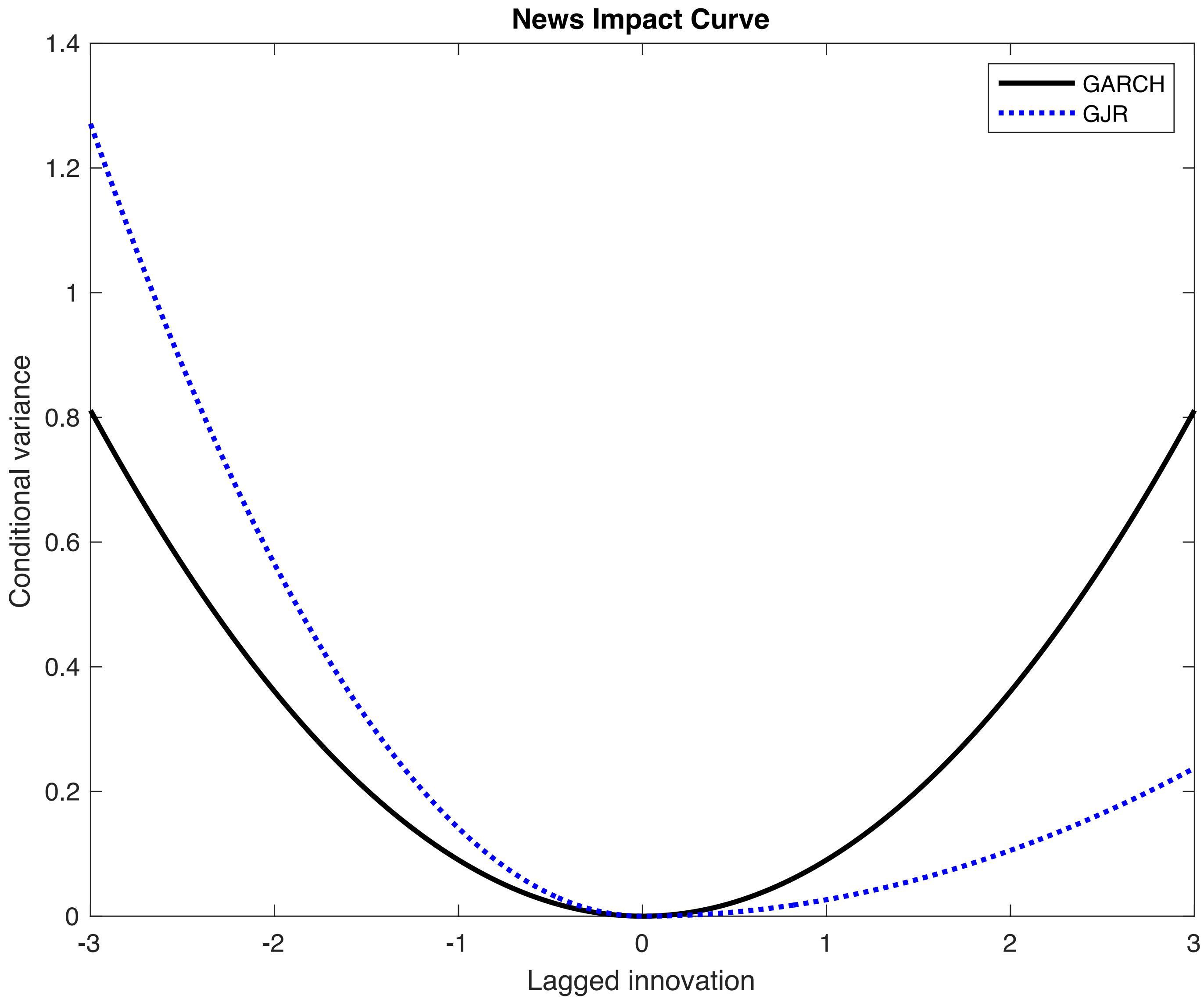
$$\omega > 0, \qquad \alpha_1 \geq 0, \qquad \alpha_1 + \gamma_1 \geq 0, \qquad \beta_1 \geq 0.$$

The process is covariance stationary iif

$$\alpha_1 + \gamma_1 / 2 + \beta_1 < 1$$

assuming that $\Pr\left[z_t < 0\right] = 0.5$ (true under symmetric distribution).

# Asymmetric GARCH Model – News Impact Curve (S&P)



# GARCH Model − Illustration

## *Daily index returns between January 1980 and July 2021*

### 1) Test for ARCH effects

The test for ARCH effects of order $p$ is based on the regression

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \ldots + \alpha_p \hat{\varepsilon}_{t-p}^2 + u_t \qquad (\hat{\varepsilon}_t \text{ pre-filtered for 10 lags (daily) or 1 lag)}$$

The $T \times R^2$ test statistic is distributed as a $\chi^2(p)$ under the null.

| | S&P500 | | DAX | | FTSE | |
|---|---|---|---|---|---|---|
| | **statistic** | **p-value** | **statistic** | **p-value** | **statistic** | **p-value** |
| **Daily returns** | | | | | | |
| ARCH(1) | 254.717 | 0 | 446.973 | 0 | 319.447 | 0 |
| ARCH(10) | 1167.623 | 0 | 1549.211 | 0 | 671.859 | 0 |
| **Weekly returns** | | | | | | |
| ARCH(1) | 170.266 | 0 | 91.075 | 0 | 246.099 | 0 |
| ARCH(10) | 219.404 | 0 | 190.877 | 0 | 287.374 | 0 |
| **Monthly returns** | | | | | | |
| ARCH(1) | 13.947 | 0 | 30.359 | 0 | 5.944 | 0.015 |
| ARCH(10) | 20.517 | 0.025 | 37.997 | 0 | 8.740 | 0.557 |

# GARCH Model − Illustration

## 2) Estimation of a GARCH(1,1) model

We now estimate the following GARCH(1,1) model

$$\hat{\varepsilon}_t = \sigma_t z_t \qquad z_t \sim N(0,1)$$
$$\sigma_t^2 = \omega + \alpha \hat{\varepsilon}_{t-1}^2 + \beta \sigma_{t-1}^2$$

| | S&P500 | | DAX | | FT All Shares | |
|---|---|---|---|---|---|---|
| | statistic | std error | statistic | std error | statistic | std error |
| **Daily returns** | | | | | | |
| $\omega$ | 0.0173 | 0.0018 | 0.0166 | 0.0020 | 0.0099 | 0.0010 |
| $\alpha$ | 0.0901 | 0.0050 | 0.1119 | 0.0056 | 0.0260 | 0.0015 |
| $\beta$ | 0.8955 | 0.0054 | 0.8874 | 0.0050 | 0.9650 | 0.0019 |
| **weekly returns** | | | | | | |
| $\omega$ | 0.3055 | 0.0593 | 0.2431 | 0.0565 | 0.2956 | 0.0614 |
| $\alpha$ | 0.1646 | 0.0205 | 0.1314 | 0.0152 | 0.0987 | 0.0136 |
| $\beta$ | 0.7819 | 0.0250 | 0.8475 | 0.0164 | 0.8503 | 0.0192 |
| **Monthly returns** | | | | | | |
| $\omega$ | 0.6955 | 0.3745 | 3.1900 | 1.4747 | 0.9597 | 0.4848 |
| $\alpha$ | 0.1195 | 0.0330 | 0.1338 | 0.0408 | 0.1269 | 0.0367 |
| $\beta$ | 0.8531 | 0.0378 | 0.7712 | 0.0690 | 0.8361 | 0.0429 |

# GARCH Model − Illustration

## 3) Estimation of a GJR model

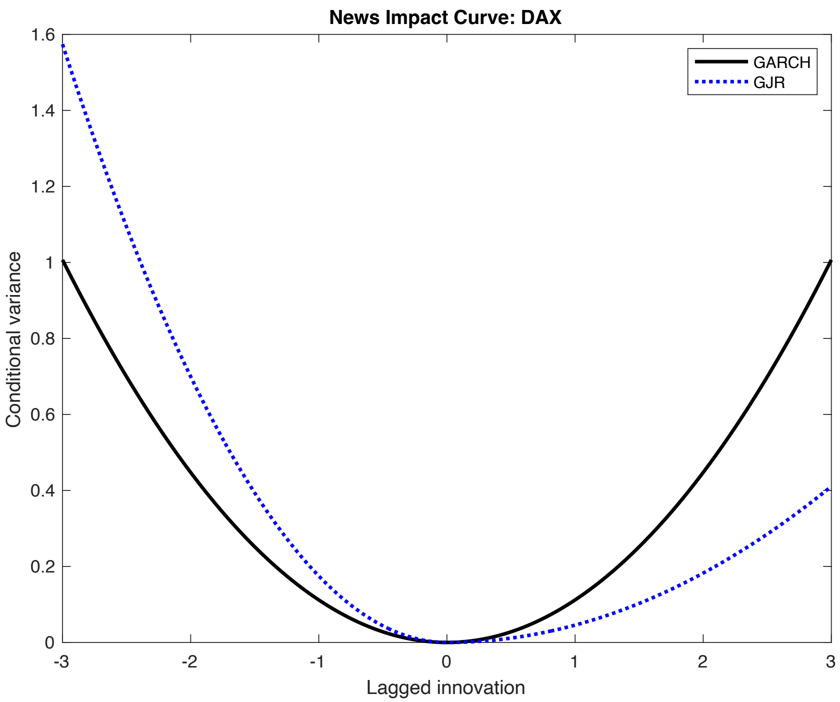
We now estimate the following GJR model

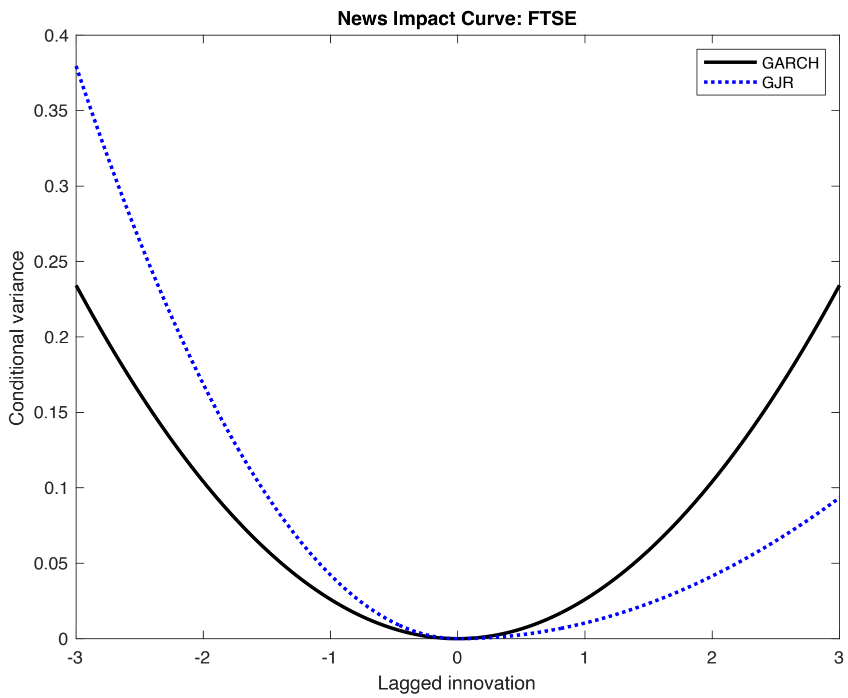
$$\sigma_t^2 = \omega + \alpha \hat{\varepsilon}_{t-1}^2 + \gamma\, \Pi_{t-1}^{-} \hat{\varepsilon}_{t-1}^2 + \beta \sigma_{t-1}^2$$

| | S&P500 | | DAX | | FT All Shares | |
|---|---|---|---|---|---|---|
| | **statistic** | **std error** | **statistic** | **std error** | **statistic** | **std error** |
| **Daily returns** | | | | | | |
| $\omega$ | 0.0205 | 0.0048 | 0.0213 | 0.0054 | 0.0197 | 0.0332 |
| $\alpha$ | 0.0264 | 0.0065 | 0.0456 | 0.0093 | 0.0104 | 0.0215 |
| $\gamma$ | 0.1148 | 0.0213 | 0.1293 | 0.0287 | 0.0318 | 0.0642 |
| $\beta$ | 0.8960 | 0.0159 | 0.8849 | 0.0170 | 0.9537 | 0.0479 |
| **Weekly returns** | | | | | | |
| $\omega$ | 0.5058 | 0.2198 | 0.4995 | 0.2206 | 0.7149 | 0.5937 |
| $\alpha$ | 0.0228 | 0.0210 | 0.0575 | 0.0209 | 0.0000 | 0.0590 |
| $\gamma$ | 0.2902 | 0.1099 | 0.1995 | 0.0680 | 0.1993 | 0.1407 |
| $\beta$ | 0.7227 | 0.0861 | 0.7866 | 0.0480 | 0.7558 | 0.1411 |
| **Monthly returns** | | | | | | |
| $\omega$ | 6.2115 | 48.6694 | 3.8768 | 15.8958 | 0.8738 | 0.7722 |
| $\alpha$ | 0.0000 | 3.4965 | 0.0000 | 0.2950 | 0.1655 | 0.0884 |
| $\gamma$ | 0.4329 | 3.2239 | 0.1660 | 0.2507 | -0.0570 | 0.0913 |
| $\beta$ | 0.4551 | 1.4664 | 0.7862 | 0.8794 | 0.8350 | 0.0617 |

# GARCH Model − Illustration

## 4) The news impact curves for daily returns



# GARCH Model − Illustration

## 5) Normality tests for residuals

The normality test is based on the Jarque-Bera statistic:

| | S&P500 | | DAX | | FT All Shares | |
|---|---|---|---|---|---|---|
| | statistic | p-value | statistic | p-value | statistic | p-value |
| **Daily returns** | | | | | | |
| Returns | 327566 | 0 | 26865 | 0 | 171354 | 0 |
| GARCH residuals | 7777 | 0 | 16312 | 0 | 532547 | 0 |
| GJR residuals | 5554 | 0 | 10131 | 0 | 873382 | 0 |
| **Weekly returns** | | | | | | |
| Returns | 3251 | 0 | 3293 | 0 | 5890 | 0 |
| GARCH residuals | 1099 | 0 | 1152 | 0 | 5622 | 0 |
| GJR residuals | 543 | 0 | 779 | 0 | 7982 | 0 |
| **Monthly returns** | | | | | | |
| Returns | 265.8 | 0 | 225.4 | 0 | 578.6 | 0 |
| GARCH residuals | 207.3 | 0 | 38.6 | 0 | 324.6 | 0 |
| GJR residuals | 205.4 | 0 | 28.7 | 0 | 324.8 | 0 |

# Appendix 1: Proof of asymptotic normality of the MLE

We know that $g(\hat{\theta}) = 0$.

A Taylor approximation around the true parameter $\theta_0$ gives

$$g(\hat{\theta}) = g(\theta_0) + H(\bar{\theta})(\hat{\theta} - \theta_0) = 0 \qquad \text{with } \bar{\theta} \text{ between } \theta_0 \text{ and } \hat{\theta}.$$

Rearranging terms gives asymptotically (so that $\text{plim}(\bar{\theta} - \theta_0) = 0$)

$$\sqrt{T}(\hat{\theta} - \theta_0) = \left[-H(\theta_0)\right]^{-1}\left[\sqrt{T} g(\theta_0)\right] = \left[-\frac{1}{T} H(\theta_0)\right]^{-1}\left[\frac{1}{\sqrt{T}} g(\theta_0)\right]$$

with $V\left[\frac{1}{\sqrt{T}} g(\theta_0)\right] = \frac{1}{T} E\left[g(\theta_0) g(\theta_0)'\right]$

# Appendix 1: Proof of asymptotic normality of the MLE

Therefore, we have the following asymptotic distribution:

$$\sqrt{T}(\hat{\theta}-\theta_0) \overset{d}{\sim} N\left(0, \left(-\frac{1}{T}E\left[H(\theta_0)\right]\right)^{-1}\left(\frac{1}{T}E\left[g(\theta_0)g(\theta_0)'\right]\right)\left(-\frac{1}{T}E\left[H(\theta_0)\right]\right)^{-1}\right)$$

Under normality of the innovations, we have

$$-\frac{1}{T}E\left[H(\theta_0)\right] = \frac{1}{T}E\left[g(\theta_0)g(\theta_0)'\right] = I(\theta_0)$$

so that

$$\sqrt{T}(\hat{\theta}-\theta_0) \overset{d}{\sim} N\left(0, I(\theta_0)^{-1}\right)$$

# Appendix 2: Algorithm for the estimation of parameters

As there is no analytical solution to the ML estimation, the optimal parameter estimates are found using iterative algorithms. The general idea is to update the parameter estimates using the scheme:

$$\theta^{(i+1)} = \theta^{(i)} + \lambda^{(i)} \delta^{(i)}$$

where $\lambda^{(i)}$ is a step length and $\delta^{(i)}$ is a direction vector.

The direction $\delta^{(i)}$ is chosen so that the likelihood under $\theta^{(i+1)}$ is greater than the likelihood under $\theta^{(i)}$. In the **Newton-Raphson algorithm**, it is based on the idea to maximize the second-order Taylor approximation of the likelihood around $\theta^{(0)}$, i.e.

$$\log\left(L_T(\theta)\right) = \log\left(L_T\left(\theta^{(0)}\right)\right) + g\left(\theta^{(0)}\right)'\left(\theta - \theta^{(0)}\right) + \frac{1}{2}\left(\theta - \theta^{(0)}\right)' H\left(\theta^{(0)}\right)\left(\theta - \theta^{(0)}\right)$$

with $g(\theta) = \frac{\partial \log\left(L_T(\theta)\right)}{\partial \theta}$ the gradient vector and $H(\theta) = \frac{\partial^2 \log\left(L_T(\theta)\right)}{\partial \theta \partial \theta'}$ the Hessian matrix

# Appendix 2: Algorithm for the estimation of parameters

Setting the derivative of this equation with respect to $\theta$ equal to 0 gives

$$g\left(\theta^{(0)}\right) + H\left(\theta^{(0)}\right)\left(\theta - \theta^{(0)}\right) = 0 \qquad \text{or} \quad \theta^{(1)} = \theta^{(0)} - H\left(\theta^{(0)}\right)^{-1} g\left(\theta^{(0)}\right)$$

If the log-likelihood is quadratic, the Taylor approximation is perfect and $\theta^{(1)}$ is the MLE. If the quadratic approximation is reasonably good, the Newton-Raphson algorithm should converge to the maximum, using the iteration

$$\theta^{(i+1)} = \theta^{(i)} - H\left(\theta^{(i)}\right)^{-1} g\left(\theta^{(i)}\right)$$

In practice, this optimization is often replaced by

$$\theta^{(i+1)} = \theta^{(i)} - \lambda\, H\left(\theta^{(i)}\right)^{-1} g\left(\theta^{(i)}\right)$$

where we compute $\theta^{(i+1)}$ and the associated log likelihood $\log\left(L_T\left(\theta^{(i+1)}\right)\right)$ for various step lengths $\lambda$. We choose as the estimate $\theta^{(i+1)}$ the value that produces the largest value for the log likelihood.

# Appendix 3: Volatility Persistence

What about **strict stationarity**? A process $X_t$ is strictly stationary if the distribution of $(X_{t_1},\cdots,X_{t_n})$ is the same as the distribution of $(X_{t_1+k},\cdots,X_{t_n+k})$ $\forall n, k$.

**Theorem** (Nelson, *Econometric Theory*, 1990):

If $\omega = 0$, then $\sigma_t^2 \overset{as}{\rightarrow} \infty$ iff $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] > 0$

$\sigma_t^2 \overset{as}{\rightarrow} 0$ iff $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] < 0$

If $\omega > 0$, then $\sigma_t^2 \overset{as}{\rightarrow} \infty$ iff $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] \geq 0$

$\sigma_t^2$ strictly stationary iff $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] < 0$

Due to the Jensen's inequality, we have

$$E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] < \log\left(E\left[\alpha_1 z_t^2 + \beta_1\right]\right) = \log\left(\alpha_1 + \beta_1\right)$$

Therefore, even when $\alpha_1 + \beta_1 = 1$, the process $\sigma_t^2$ is still strictly stationary. $\varepsilon_t$ is also strictly stationary.

# Appendix 3: Volatility Persistence

## Sketch of the proof

### 1. Case $\omega = 0$:

$$\sigma_t^2 = \sigma_0^2 \prod_{k=1}^{t} \left(\alpha_1 z_{t-k}^2 + \beta_1\right) \qquad \text{or} \qquad \log(\sigma_t^2) = \log(\sigma_0^2) + \sum_{k=1}^{t} \log(\alpha_1 z_{t-k}^2 + \beta_1)$$

$\log(\sigma_t^2)$ is a random walk with a drift given by $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right]$

since $\frac{1}{t}\sum_{k=1}^{t} \log(\alpha_1 z_{t-k}^2 + \beta_1) \xrightarrow{as} E\left[\log\left(\alpha_1 z_{t-k}^2 + \beta_1\right)\right]$ (strong law of large numbers)

- If $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] > 0$ then $\sigma_t^2 \xrightarrow{as} \infty$
- If $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] < 0$ then $\sigma_t^2 \xrightarrow{as} 0$
- If $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] = 0$ then $\log(\sigma_t^2)$ random walk with no drift

# Appendix 3: Volatility Persistence

**2. Case $\omega > 0$:**

$$\sigma_t^2 = \sigma_0^2 \prod_{k=1}^{t} \left(\alpha_1 z_{t-k}^2 + \beta_1\right) + \omega \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} \left(\alpha_1 z_{t-i}^2 + \beta_1\right)\right]$$

- If $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] \geq 0$, then the results of the previous case apply and $\sigma_t^2 \overset{as}{\to} \infty$.
- If $E\left[\log\left(\alpha_1 z_t^2 + \beta_1\right)\right] < 0$, then $\prod_{k=1}^{t} \left(\alpha_1 z_{t-k}^2 + \beta_1\right) = \exp\left(\sum_{k=1}^{t} \log(\alpha_1 z_{t-k}^2 + \beta_1)\right) \overset{as}{\to} 0$

so that the second term in $\sigma_t^2$ is positive but not infinite.

In this case, the process $\sigma_t^2$ is said to be **strictly stationary**, as the effect of past variances vanishes in the long term. $\varepsilon_t$ is also strictly stationary.

**Remark:** IGARCH process may reflect other dynamics for volatility. For instance, if the true model is a regime-switching model for volatility, estimating a GARCH model will generally result in a nearly integrated volatility process.