

Amath 546/Econ 589  
Univariate GARCH Models

Eric Zivot

April 24, 2013

## Lecture Outline

- Conditional vs. Unconditional Risk Measures
- Empirical regularities of asset returns
- Engle's ARCH model
- Testing for ARCH effects
- Estimating ARCH models
- Bollerslev's GARCH model

## Conditional Expectation

Let  $\{Y_t\}$  be a sequence of random variables and let  $\{I_t\}$  be a sequence of information sets ( $\sigma$ -fields) with  $I_t \subset I$  for all  $t$  and  $I$  the universal information set. For example,

$$\begin{aligned} I_t &= \{Y_1, Y_2, \dots, Y_t\} = \text{past history of } Y_t \\ I_t &= \{(Y_s, Z_s)_{s=1}^t\}, \{Z_t\} = \text{auxiliary variables} \end{aligned}$$

Definition (Conditional Expectation). Let  $Y_t$  be a random variable with conditional pdf  $f(y_t|I_s)$ , where  $I_s$  is an information set with  $s < t$ . Then

$$\begin{aligned} \mu_{t|s} &= E[Y_t|I_s] = \int_{-\infty}^{\infty} y_t f(y_t|I_s) dy_t \\ \sigma_{t|s}^2 &= \text{var}(Y_t|I_s) = \int_{-\infty}^{\infty} (y_t - E[Y_t|I_s])^2 f(y_t|I_s) dy_t \end{aligned}$$

**Result:** Let  $X$  and  $Y$  denote random variables and let  $I$  denote an information set. If  $X \subset I$ , then

$$E[XY|I] = XE[Y|I]$$

**Proposition 1** *Law of Iterated Expectation*

Let  $I_1$  and  $I_2$  be information sets such that  $I_1 \subseteq I_2$ , and let  $Y$  be a random variable such that  $E[Y|I_1]$  and  $E[Y|I_2]$  are defined. Then

$$E[Y|I_1] = E[E[Y|I_2]|I_1] \text{ (smaller set wins)}$$

If  $I_1 = \emptyset$  (empty set) then

$$E[Y|I_1] = E[Y] \text{ (unconditional expectation)}$$

$$E[Y] = E[E[Y|I_2]|I_1] = E[E[Y|I_2]]$$

## Definition 2 *Martingale Difference Sequence*

Let  $Y_t$  and random variable and let  $I_t$  denote an information set. The pair  $(Y_t, I_t)$  is a *martingale difference sequence* (MDS) if

1.  $I_t \subset I_{t+1}$  (increasing sequence of information sets - a *filtration*)
2.  $Y_t \subset I_t$  ( $Y_t$  is *adapted* to  $I_t$ ; i.e.,  $Y_t$  is an event in  $I_t$ )
3.  $E[Y_t | I_{t-1}] = 0$

## Conditional vs. Unconditional Risk Measures

Let  $R_{t+1}$  denote an asset return between times  $t$  and  $t + 1$  (simple or continuously compounded).

**Definition 3** (*Unconditional Modeling*) *Unconditional modeling of  $R_{t+1}$  is based on the unconditional or marginal distribution of  $R_{t+1}$ . That is, risk measures are computed from the marginal distribution  $F_R$*

$$R_{t+1} \sim F_R, \quad E[R_{t+1}] = \mu, \quad \text{var}(R_{t+1}) = \sigma^2$$

Define

$$Z_{t+1} = \frac{R_{t+1} - \mu}{\sigma}, \quad Z_{t+1} \sim F_Z, \quad E[Z_{t+1}] = 0, \quad \text{var}(Z_{t+1}) = 1$$

So that

$$R_{t+1} = \mu + \sigma \times Z_{t+1}$$

Let  $I_t$  denote information known at time  $t$ . For example,  $I_t = \{R_t, R_{t-1}, \dots, R_0\}$  or  $I_t = \{(R_t, X_t), (R_{t-1}, X_{t-1}), \dots, (R_0, X_0)\}$ .

**Definition 4** (*Conditional Modeling*) Conditional modeling of  $R_{t+1}$  is based on the conditional distribution of  $R_{t+1}$  given  $I_t$ . That is, risk measures are computed from the conditional distribution  $F_{R|I_t}$

$$R_{t+1} \sim F_{R|I_t}, \quad E[R_{t+1}|I_t] = \mu_{t+1|t}, \quad \text{var}(R_{t+1}|I_t) = \sigma_{t+1|t}^2$$

Define

$$Z_{t+1} = \frac{R_{t+1} - \mu_{t+1|t}}{\sigma_{t+1|t}}, \quad Z_{t+1} \sim F_Z, \quad E[Z_{t+1}] = 0, \quad \text{var}(Z_{t+1}) = 1$$

So that

$$R_{t+1} = \mu_{t+1|t} + \sigma_{t+1|t} \times Z_{t+1}$$

## Conditional Mean, Variance and Volatility

- $E[R_{t+1}|I_t] = \mu_{t+1|t}$  = conditional mean
- $var(R_{t+1}|I_t) = \sigma_{t+1|t}^2$  = conditional variance
- $\sigma_{t+1|t}$  = conditional volatility

Intuition: As  $I_t$  changes over time so does  $\mu_{t+1|t}$  and  $\sigma_{t+1|t}$

Remark: For many daily asset returns, it is often safe to assume  $\mu_{t|t-1} = \mu \approx 0$  but it is not safe to assume  $\sigma_{t+1|t} = \sigma$



## Conditional Risk Measures based on Returns

$\sigma_{t+1|t}$  = conditional volatility

$q_{\alpha}^{R_{t+1|t}}$  = conditional quantile

$E[R_{t+1} | R_{t+1} \leq q_{\alpha}^{R_{t+1|t}}] =$  conditional shortfall

**Example:** Normal Conditional VaR and ES

$$R_{t+1} = \mu_{t+1|t} + \sigma_{t+1|t} \times Z$$
$$Z \sim N(0, 1)$$

Then

$$q_{\alpha}^{R_{t+1|t}} = \mu_{t+1|t} + \sigma_{t+1|t} \times q_{\alpha}^Z$$
$$E[R_t | R_{t+1} \leq q_{\alpha}^{-R_{t+1|t}}] = \mu_{t+1|t} + \sigma_{t+1|t} \times \frac{\phi(q_{\alpha}^Z)}{1 - \alpha}$$

**Question:** How to model  $\mu_{t+1|t}$  and  $\sigma_{t+1|t}$ ?

## Empirical Regularities of Asset Returns Related to Volatility

### 1. Thick tails

(a) Excess kurtosis decreases with aggregation

### 2. Volatility clustering.

(a) Large changes followed by large changes; small changes followed by small changes

### 3. Leverage effects

(a) changes in prices often negatively correlated with changes in volatility

#### 4. Non trading periods

- (a) Volatility is smaller over periods when markets are closed than when they are open

#### 5. Forecastable events

- (a) Forecastable releases of information are associated with high ex ante volatility

#### 6. Volatility and serial correlation

- (a) Inverse relationship between volatility and serial correlation of stock indices

## 7. Volatility co-movements

(a) Evidence of common factors to explain volatility in multiple series

## Engle's ARCH(p) Model

*Intuition:* Use an autoregressive model to capture time dependence in conditional volatility in asset returns

The ARCH(p) model for  $r_t = \ln P_t - \ln P_{t-1}$  is

$$\begin{aligned}r_t &= E[r_t|I_{t-1}] + \epsilon_t, \quad \epsilon_t|I_{t-1} \sim iid(0, \sigma_t^2) \\ \sigma_t^2 &= a_0 + a_1\epsilon_{t-1}^2 + \dots + a_p\epsilon_{t-p}^2, \quad a_0 > 0, a_i \geq 0 \\ &= a_0 + a(L)\epsilon_t^2, \quad a(L) = \sum_{j=1}^p a_j L^j\end{aligned}$$

Recall,

$$Lr_t = r_{t-1}, \quad L^2r_t = r_{t-2} \text{ etc.}$$

## Alternative error specification

$$\begin{aligned}\epsilon_t &= z_t \sigma_t \\ z_t &\sim iid(0, 1) \\ \sigma_t^2 &= a_0 + a(L)\epsilon_t^2\end{aligned}$$

### Remarks:

- The random variable  $z_t$  doesn't have to be normal. It can have a fat-tailed distribution; e.g. Student's-t
- If  $z_t$  is non-normal it must be standardized so that  $var(z_t) = E[z_t^2] = 1$ . This often involves special versions of functions in R

## Properties of ARCH Errors: ARCH(1) Process

$$r_t - E[r_t|I_{t-1}] = \epsilon_t$$

$$\epsilon_t = z_t\sigma_t, z_t \sim iid(0, 1)$$

$$\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2, a_0 > 0, a_1 \geq 0$$

- $\{\epsilon_t\}$  is covariance stationary provided  $0 \leq a_1 < 1$
- $\{\epsilon_t, I_{t-1}\}$  is a MDS with conditionally heteroskedastic errors (if  $a_1 \neq 0$ )

$$E[\epsilon_t|I_{t-1}] = E[z_t\sigma_t|I_{t-1}] = \sigma_t E[z_t|I_{t-1}] = 0$$

$$\text{var}(\epsilon_t|I_{t-1}) = E[\epsilon_t^2|I_{t-1}] = \sigma_t^2 E[z_t^2|I_{t-1}] = \sigma_t^2$$

$$E[\epsilon_t^m|I_{t-1}] = 0 \text{ for } m \text{ odd.}$$



- $\epsilon_t$  has mean zero and constant unconditional variance for all  $t$

$$\begin{aligned}
 E[\epsilon_t] &= E[E[z_t \sigma_t | I_{t-1}]] \\
 &= E[\sigma_t E[z_t | I_{t-1}]] = 0 \\
 \text{var}(\epsilon_t) &= E[\epsilon_t^2] = E[E[z_t^2 \sigma_t^2 | I_{t-1}]] \\
 &= E[\sigma_t^2 E[z_t^2 | I_{t-1}]] = E[\sigma_t^2]
 \end{aligned}$$

Under stationarity and the fact that  $E[\epsilon_t^2] = E[\sigma_t^2]$

$$\begin{aligned}
 E[\sigma_t^2] &= E[a_0 + a_1 \epsilon_{t-1}^2] = a_0 + a_1 E[\epsilon_{t-1}^2] \\
 &= a_0 + a_1 E[\epsilon_t^2] \text{ by stationarity} \\
 &= a_0 + a_1 E[\sigma_t^2] \text{ because } E[\epsilon_t^2] = E[\sigma_t^2] \\
 \Rightarrow E[\sigma_t^2] &= \bar{\sigma}^2 = \frac{a_0}{1 - a_1} > 0 \text{ provided } 0 \leq a_1 < 1
 \end{aligned}$$

- $\epsilon_t$  is an uncorrelated process:  $E[\epsilon_t \epsilon_{t-j}] = 0$  for  $j = 1, 2, \dots$

$$\begin{aligned} E[\epsilon_t \epsilon_{t-1}] &= E[E[z_t \sigma_t z_{t-1} \sigma_{t-1} | I_{t-1}]] \\ &= E[z_{t-1} \sigma_{t-1} \sigma_t E[z_t | I_{t-1}]] = 0 \end{aligned}$$

- $\epsilon_t$  is *leptokurtic*

$$\begin{aligned} E[\epsilon_t^4] &= E[\sigma_t^4 E[z_t^4 | I_{t-1}]] = E[\sigma_t^4] \cdot 3 \\ &\geq (E[\sigma_t^2])^2 \cdot 3 \text{ by Jensen's inequality} \\ &= (E[\epsilon_t^2])^2 \cdot 3 \\ &\Rightarrow \frac{E[\epsilon_t^4]}{(E[\epsilon_t^2])^2} > 3 \end{aligned}$$

That, is

$$\text{kurt}(\epsilon_t) > 3 = \text{kurt}(\text{normal})$$

- $\sigma_t^2$  is a serially correlated random variable

$$\begin{aligned}\sigma_t^2 &= a_0 + a_1 \epsilon_{t-1}^2, \\ E[\sigma_t^2] &= \frac{a_0}{1 - a_1} = \bar{\sigma}^2\end{aligned}$$

Using  $a_0 = (1 - a_1)\bar{\sigma}^2$ ,  $\sigma_t^2$  may be expressed as

$$\begin{aligned}\sigma_t^2 &= (1 - a_1)\bar{\sigma}^2 + a_1 \epsilon_{t-1}^2, \\ \Rightarrow \sigma_t^2 - \bar{\sigma}^2 &= a_1(\epsilon_{t-1}^2 - \bar{\sigma}^2)\end{aligned}$$

- $\{\epsilon_t^2\}$  has a stationary AR(1) representation.

$$\begin{aligned}\sigma_t^2 + \epsilon_t^2 &= a_0 + a_1 \epsilon_{t-1}^2 + \epsilon_t^2 \\ \Rightarrow \epsilon_t^2 &= a_0 + a_1 \epsilon_{t-1}^2 + (\epsilon_t^2 - \sigma_t^2)\end{aligned}$$

where  $(\epsilon_t^2 - \sigma_t^2) = v_t$  is a conditionally heteroskedastic MDS.

## Properties of ARCH Errors: ARCH(p) Process

$$\epsilon_t = z_t \sigma_t$$

$$z_t \sim iid(0, 1)$$

$$\sigma_t^2 = a_0 + a(L)\epsilon_t^2, \quad a(L) = a_0L + \dots + a_pL^p, \quad a_0 > 0, \quad a_i \geq 0$$

- $\{\epsilon_t\}$  is covariance stationary provided  $0 \leq a_1 + \dots + a_p < 1$  ( $0 \leq a(1) < 1$ )
- $E[\sigma_t^2] = \bar{\sigma}^2 = \frac{a_0}{1 - a_1 - \dots - a_p} = \frac{a_0}{a(1)}, \quad a(1) > 0$
- $\sigma_t^2$  is a serially correlated random variable  $\sigma_t^2 - \bar{\sigma}^2 = a(L)(\epsilon_t^2 - \bar{\sigma}^2)$
- $\epsilon_t^2$  has a stationary AR(p) representation  $\epsilon_t^2 = a_0 + a(L)\epsilon_t^2 + (\epsilon_t^2 - \sigma_t^2)$

## Bollerslev's GARCH Model

*Idea:* ARCH is like an AR model for volatility. GARCH is like an ARMA model for volatility.

The GARCH( $p, q$ ) model is

$$\begin{aligned}\epsilon_t &= z_t \sigma_t, \quad z_t \sim iid(0, 1) \\ \sigma_t^2 &= a_0 + a(L)\epsilon_t^2 + b(L)\sigma_t^2, \quad a_0 > 0 \\ a(L) &= a_1 L + \dots + a_p L^p, \quad a_i \geq 0 \\ b(L) &= b_1 L + \dots + b_q L^q, \quad b_j \geq 0\end{aligned}$$

Note 1: for identification of  $\beta_j$ , must have at least one ARCH coefficient  $a_i > 0$  for  $i > 0$

Note 2: GARCH(1,1) is very often the best model (See paper by Hansen and Lund, JOE)

## GARCH(1,1)

The most commonly used GARCH(p,q) model is the GARCH(1,1)

$$\begin{aligned}\sigma_t^2 &= a_0 + a_1\epsilon_{t-1}^2 + b_1\sigma_{t-1}^2 \\ a_0 &> 0, \quad a_1 \geq 0, \quad b_1 \geq 0\end{aligned}$$

Properties:

stationarity condition:  $a_1 + b_1 < 1$

$$\text{ARCH}(\infty) : a_i = a_1 b_1^{i-1}$$

$$\text{ARMA}(1,1): \epsilon_t^2 = a_0 + (a_1 + b_1)\epsilon_{t-1}^2 + u_t - b_1 u_{t-1},$$

$$u_t = \epsilon_t^2 - E_{t-1}(\epsilon_t^2)$$

$$\text{unconditional variance} : \bar{\sigma}^2 = a_0 / (1 - a_1 - b_1)$$

## Properties of GARCH model

- GARCH( $p, q$ ) is equivalent to ARCH( $\infty$ ). If  $1 - b(z) = 0$  has all roots outside unit circle then

$$\begin{aligned}\sigma_t^2 &= \frac{a_0}{1 - b(1)} + \frac{a(L)}{1 - b(L)} \epsilon_t^2 \\ &= a_0^* + \delta(L) \epsilon_t^2, \quad \delta(L) = \sum_{k=0}^{\infty} \delta_k L^k\end{aligned}$$

- $\epsilon_t$  is a stationary and ergodic MDS with finite variance provided  $a(1) + b(1) < 1$

$$\begin{aligned}E[\epsilon_t] &= 0 \\ \text{var}(\epsilon_t) &= E[\epsilon_t^2] = \frac{a_0}{1 - a(1) - b(1)} = E[\sigma_t^2] = \bar{\sigma}^2\end{aligned}$$

$$\epsilon_t^2 \sim \text{ARMA}(m, q), \quad m = \max(p, q)$$

## Conditional Mean Specification

- $E[r_t|I_{t-1}]$  is typically specified as a constant or possibly a low order ARMA process to capture autocorrelation caused by market microstructure effects (e.g., bid-ask bounce) or non-trading effects.
- If extreme or unusual market events have happened during sample period, then dummy variables associated with these events are often added to the conditional mean specification to remove these effects. The general conditional mean specification is of the form

$$E[r_t|I_{t-1}] = c + \sum_{i=1}^r \phi_i r_{t-i} + \sum_{j=1}^s \theta_j \epsilon_{t-j} + \sum_{l=0}^L \beta_l' \mathbf{x}_{t-l} + \epsilon_t,$$

where  $\mathbf{x}_t$  is a  $k \times 1$  vector of exogenous explanatory variables.



## Explanatory Variables in the Conditional Variance Equation

- Exogenous explanatory variables may also be added to the conditional variance formula

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i \epsilon_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2 + \sum_{k=1}^K \delta'_k \mathbf{z}_{t-k},$$

where  $\mathbf{z}_t$  is a  $m \times 1$  vector of variables, and  $\boldsymbol{\delta}$  is a  $m \times 1$  vector of positive coefficients.

- Variables that have been shown to help predict volatility are trading volume, interest rates, macroeconomic news announcements, implied volatility from option prices and realized volatility, overnight returns, and after hours realized volatility

## GARCH-in-Mean (GARCH-M)

*Idea:* Modern finance theory suggests that volatility may be related to risk premia on assets

The GARCH-M model allows time-varying volatility to be related to expected returns

$$\begin{aligned} r_t &= c + \alpha g(\sigma_t) + \epsilon_t \\ \epsilon_t &\sim \text{GARCH} \\ g(\sigma_t) &= \begin{cases} \sigma_t \\ \sigma_t^2 \\ \ln(\sigma_t^2) \end{cases} \end{aligned}$$

## Temporal Aggregation

- Volatility clustering and non-Gaussian behavior in financial returns is typically seen in weekly, daily or intraday data. The persistence of conditional volatility tends to increase with the sampling frequency.
- For GARCH models there is no simple aggregation principle that links the parameters of the model at one sampling frequency to the parameters at another frequency. This occurs because GARCH models imply that the squared residual process follows an ARMA type process with MDS innovations which is not closed under temporal aggregation.

- The practical result is that GARCH models tend to be fit to the frequency at hand. This strategy, however, may not provide the best out-of-sample volatility forecasts. For example, Martens (2002) showed that a GARCH model fit to S&P 500 daily returns produces better forecasts of weekly and monthly volatility than GARCH models fit to weekly or monthly returns, respectively.

## Testing for ARCH Effects

Consider testing the hypotheses

$$H_0 : \text{(No ARCH)} \quad a_1 = a_2 = \dots = a_p = 0$$

$$H_1 : \text{(ARCH)} \quad \text{at least one } a_i \neq 0$$

Engle derived a simple LM test

- Step 1: Compute squared residuals  $\epsilon_t$  from mean equation regression
- Step 2: Estimate auxiliary regression

$$\hat{\epsilon}_t^2 = a_0 + a_1 \hat{\epsilon}_{t-1}^2 + \dots + a_p \hat{\epsilon}_{t-p}^2 + \text{error}_t$$

- Step 3. Form the LM test statistic

$$LM_{ARCH} = T \cdot R_{AUX}^2$$

where  $T$  = sample size from auxiliary regression and  $R_{AUX}^2$  is the uncentered R-squared from the auxiliary regression. Under  $H_0$  : (No ARCH)

$$LM_{ARCH} \overset{A}{\sim} \chi^2(p)$$

Remark:

- Test has power against GARCH( $p, q$ ) alternatives

## Estimating GARCH by MLE

Consider estimating the model

$$\begin{aligned}r_t &= E[r_t|I_{t-1}] + \epsilon_t = \mathbf{x}'_t\boldsymbol{\beta} + \epsilon_t \\ \epsilon_t &= z_t\sigma_t, \quad z_t \sim iid N(0,1) \\ \sigma_t^2 &= a_0 + a(L)\epsilon_t^2 + b(L)\sigma_t^2\end{aligned}$$

*Result:* The regression parameters  $\boldsymbol{\beta}$  and GARCH parameters

$\boldsymbol{\gamma} = (a_0, a_1, \dots, a_p, b_1, \dots, b_q)'$  can be estimated separately because the information matrix for  $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$  is block diagonal.

- Step 1: Estimate  $\beta$  by OLS ignoring ARCH errors and form residuals  $\hat{\epsilon}_t = r_t - \mathbf{x}'_t \hat{\beta}$
- Step 2: Estimate ARCH process for residuals  $\hat{\epsilon}_t$  by mle.

Warning: Block diagonality of information matrix fails if

- pdf of  $z_t$  is not a symmetric density
- $\beta$  and  $\gamma$  are not variation free; e.g. GARCH-M model



## GARCH Likelihood Function Under Normality

Assume  $E[r_t|I_{t-1}] = 0$ . Let  $\boldsymbol{\theta} = (a_0, a_1, \dots, a_p, b_1, \dots, b_q)'$  denote the parameters to be estimated. Since  $\epsilon_t = z_t\sigma_t$

$$\begin{aligned} f(\epsilon_t|I_{t-1}; \boldsymbol{\theta}) &= f(z_t) \left| \frac{dz_t}{d\epsilon_t} \right| = f\left(\frac{\epsilon_t}{\sigma_t}\right) \left| \frac{1}{\sigma_t} \right| \\ &= (2\pi\sigma_t^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_t^2}\epsilon_t^2\right\} \end{aligned}$$

For a sample of size  $T$ , the prediction error decomposition gives

$$\begin{aligned} &f(\epsilon_T, \epsilon_{T-1}, \dots, \epsilon_1; \boldsymbol{\theta}) \\ &= \left( \prod_{t=m+1}^T f(\epsilon_t|I_{t-1}; \boldsymbol{\theta}) \right) \cdot f(\epsilon_1, \dots, \epsilon_m; \boldsymbol{\theta}) \\ &= \left( \prod_{t=m+1}^T (2\pi\sigma_t^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_t^2}\epsilon_t^2\right\} \right) \cdot f(\epsilon_1, \dots, \epsilon_m; \boldsymbol{\theta}) \end{aligned}$$

## Remarks

- $\sigma_t^2 = a_0 + a(L)\epsilon_t^2 + b(L)\sigma_t^2$  is evaluated recursively given  $\theta$  and starting values for  $\sigma_t^2$  and  $\epsilon_t^2$ . For example, consider GARCH(1,1)

$$\sigma_1^2 = a_0 + a_1\epsilon_0^2 + b_1\sigma_0^2$$

Need to specify starting values values  $\epsilon_0^2$  and  $\sigma_0^2$ . Then all other  $\sigma_t^2$  values can be calculated recursively.

- The log-likelihood function is

$$\begin{aligned} & \frac{-(T-m)}{2} \ln(2\pi) - \sum_{t=m+1}^T \left[ \frac{1}{2} \ln(\sigma_t^2) + \frac{1}{2} \frac{\epsilon_t^2}{\sigma_t^2} \right] \\ & + \ln(f(\epsilon_1, \dots, \epsilon_m; \gamma)) \end{aligned}$$

*Problem:* the marginal density for the initial values  $f(\varepsilon_1, \dots, \varepsilon_m; \theta)$  does not have a closed form expression so exact mle is not possible. In practice, the initial values  $\varepsilon_1, \dots, \varepsilon_m$  are set equal to zero and the marginal density  $f(\varepsilon_1, \dots, \varepsilon_m; \theta)$  is ignored. This is conditional mle.

## Practical issues

- To initialize the log-likelihood starting values for the model parameters  $a_i$  ( $i = 0, \dots, p$ ) and  $b_j$  ( $j = 1, \dots, q$ ) need to be chosen and an initialization of  $\epsilon_t^2$  and  $\sigma_t^2$  must be supplied.
- Zero values are often given for the conditional variance parameters other than  $a_0$  and  $a_1$ , and  $a_0$  is set equal to the unconditional variance of  $r_t$ . For the initial values of  $\sigma_t^2$ , a popular choice is

$$\sigma_t^2 = \epsilon_t^2 = \frac{1}{T} \sum_{s=m+1}^T r_s^2, \quad t \leq m,$$

- Variance targeting can be used to eliminate estimation of  $a_0$  and improve numerical stability. For example, variance targeting in the GARCH(1,1) model is

$$\begin{aligned}\sigma_t^2 &= a_0 + a_1\epsilon_{t-1}^2 + b_1\sigma_{t-1}^2 \\ \bar{\sigma}^2 &= \frac{a_0}{1 - a_1 - b_1} \Rightarrow a_0 = \bar{\sigma}^2(1 - a_1 - b_1) \\ &\Rightarrow \sigma_t^2 = \bar{\sigma}^2(1 - a_1 - b_1) + a_1\epsilon_{t-1}^2 + b_1\sigma_{t-1}^2 \\ &= \bar{\sigma}^2 + a_1(\epsilon_{t-1}^2 - \bar{\sigma}^2) + b_1(\sigma_{t-1}^2 - \bar{\sigma}^2)\end{aligned}$$

In estimation,  $\bar{\sigma}^2$  is fixed at the sample variance of returns and is not a parameter to be estimated.

- Once the log-likelihood is initialized, it can be maximized using numerical optimization techniques. The most common method is based on a Newton-Raphson iteration of the form

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \lambda_n \mathbf{H}(\hat{\theta}_n)^{-1} \mathbf{s}(\hat{\theta}_n),$$

- For GARCH models, the BHHH algorithm is often used. This algorithm approximates the Hessian matrix using only first derivative information

$$-\mathbf{H}(\boldsymbol{\theta}) \approx \mathbf{B}(\boldsymbol{\theta}) = \sum_{t=1}^T \frac{\partial l_t}{\partial \boldsymbol{\theta}} \frac{\partial l_t}{\partial \boldsymbol{\theta}'}$$

- Under suitable regularity conditions, the ML estimates are consistent and asymptotically normally distributed and an estimate of the asymptotic covariance matrix of the ML estimates is constructed from an estimate of the final Hessian matrix from the optimization algorithm used.

## Numerical Accuracy of GARCH Estimates

- GARCH estimation is widely available in a number of commercial software packages (e.g. EVIEWS, GAUSS, MATLAB, Ox, RATS, S-PLUS, TSP) and there are a few free open source implementations (fGarch and rugarch in R). Can even use Excel! (See S. Taylor's book *Asset Price Dynamics, Volatility and Prediction*)
- Starting values, optimization algorithm choice, and use of analytic or numerical derivatives, and convergence criteria all influence the resulting numerical estimates of the GARCH parameters.

- The GARCH log-likelihood function is not always well behaved, especially in complicated models with many parameters, and reaching a global maximum of the log-likelihood function is not guaranteed using standard optimization techniques. Poor choice of starting values can lead to an ill-behaved log-likelihood and cause convergence problems.



## Quasi-Maximum Likelihood Estimation

- The assumption of conditional normality is not always appropriate.
- However, even when normality is inappropriately assumed, maximizing the Gaussian log-likelihood results in *quasi-maximum likelihood estimates* (QMLEs) that are consistent and asymptotically normally distributed provided the conditional mean and variance functions of the GARCH model are correctly specified.

- An asymptotic covariance matrix for the QMLEs that is robust to conditional non-normality is estimated using

$$\mathbf{H}(\hat{\boldsymbol{\theta}}_{QML})^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}}_{QML})\mathbf{H}(\hat{\boldsymbol{\theta}}_{QML})^{-1},$$

where  $\hat{\boldsymbol{\theta}}_{QML}$  denotes the QMLE of  $\boldsymbol{\theta}$ , and is often called the “sandwich” estimator.

## Determining lag length

- Use model selection criteria (AIC or BIC)
- For GARCH( $p, q$ ) models, those with  $p, q \leq 2$  are typically selected by AIC and BIC.
- Low order GARCH( $p, q$ ) models are generally preferred to a high order ARCH( $p$ ) for reasons of parsimony and better numerical stability of estimation (high order GARCH( $p, q$ ) processes often have many local maxima and minima).
- For many applications, it is hard to beat the simple GARCH(1,1) model.

## Model Diagnostics

Correct model specification implies

$$\hat{z}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t} \sim iid N(0, 1)$$

- Test for normality - Jarque-Bera, QQ-plot
- Test for serial correlation - Ljung-box, SACF, SPACF
- Test for ARCH effects - serial correlation in squared standardized residuals, LM test for ARCH

## GARCH and Forecasts for the Conditional Mean

Assume

$$\begin{aligned}r_t &= E[r_t|I_{t-1}] + \varepsilon_t = \mu + \varepsilon_t, \quad t = 1, \dots, T \\ \varepsilon_t &= z_t \sigma_t, \quad z_t \sim N(0, 1) \\ \sigma_t^2 &= a_0 + a_1 \varepsilon_{t-1}^2 + b_1 \sigma_{t-1}^2\end{aligned}$$

- Suppose one is interested in forecasting future values  $r_{T+h}$  based on  $I_T$ . Here, the minimum mean squared error  $h$ -step ahead forecast of  $r_{T+h}$  is

$$E[r_{T+h}|I_T] = \mu \text{ and } \varepsilon_{T+h} = r_{T+h} - \mu$$

which does not depend on the GARCH parameters

- The conditional variance of this forecast error is

$$\text{var}(\epsilon_{T+h}|I_T) = E[\sigma_{T+h}^2|I_T],$$

which does depend on the GARCH parameters. Therefore, in order to produce confidence bands for the  $h$ –step ahead forecast the  $h$ –step ahead volatility forecast  $E[\sigma_{T+h}^2|I_T]$  is needed.

## Forecasting From GARCH Models

Consider the basic GARCH(1,1) model

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2$$

from  $t = 1, \dots, T$ . The best linear predictor of  $\sigma_{T+1}^2$  using information at time  $T$  is

$$\begin{aligned} E[\sigma_{T+1}^2 | I_T] &= a_0 + a_1 E[\epsilon_T^2 | I_T] + b_1 E[\sigma_T^2 | I_T] \\ &= a_0 + a_1 \epsilon_T^2 + b_1 \sigma_T^2 \end{aligned}$$

Using the chain-rule of forecasting and  $E[\epsilon_{T+1}^2 | I_T] = E[\sigma_{T+1}^2 | I_T]$

$$\begin{aligned} E[\sigma_{T+2}^2 | I_T] &= a_0 + a_1 E[\epsilon_{T+1}^2 | I_T] + b_1 E[\sigma_{T+1}^2 | I_T] \\ &= a_0 + (a_1 + b_1) E[\sigma_{T+1}^2 | I_T] \end{aligned}$$

In general, for  $k \geq 2$

$$\begin{aligned} E[\sigma_{T+k}^2 | I_T] &= a_0 + (a_1 + b_1)E[\sigma_{T+k-1}^2 | I_T] \\ &= a_0 \sum_{i=0}^{k-1} (a_1 + b_1)^i + (a_1 + b_1)^{k-1} (a_1 \epsilon_T^2 + b_1 \sigma_T^2). \end{aligned}$$

Note: If  $|a_1 + b_1| < 1$ , then as  $k \rightarrow \infty$

$$E[\sigma_{T+k}^2 | I_T] \rightarrow E[\sigma_T^2] = \frac{a_0}{1 - a_1 - b_1}$$

An alternative representation of the forecasting equation starts with the mean-adjusted form

$$\sigma_{T+1}^2 - \bar{\sigma}^2 = a_1(\epsilon_T^2 - \bar{\sigma}^2) + b_1(\sigma_T^2 - \bar{\sigma}^2),$$

where  $\bar{\sigma}^2 = a_0 / (1 - a_1 - b_1)$  is the unconditional variance. Then by recursive substitution

$$E_T[\sigma_{T+k}^2] - \bar{\sigma}^2 = (a_1 + b_1)^{k-1} (E[\sigma_{T+1}^2] - \bar{\sigma}^2).$$



## Remarks

- The forecast of volatility is defined as

$$\begin{aligned} E[\sigma_{T+k}|I_T] &= \left( E[\sigma_{T+k}^2|I_T] \right)^{1/2} \\ &\neq E[\sigma_{T+k}|I_T] \text{ (by Jensen's inequality)} \end{aligned}$$

- Standard errors for  $E[\sigma_{T+k}|I_T]$  are not available in closed form but may be computed using simulation methods. See ruGARCH documentation.

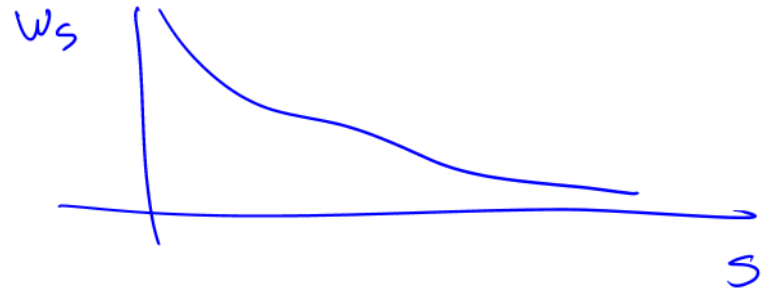
## EWMA Forecasts

- The GARCH(1,1) forecasting algorithm is closely related to an exponentially weighted moving average (EWMA) of past values of  $\epsilon_t^2$ . This type of forecast was proposed by the RiskMetrics group at J.P. Morgan.
- The EWMA forecast of  $\sigma_{T+1}^2$  has the form

$$\sigma_{T+1,EWMA}^2 = (1 - \lambda) \sum_{s=0}^{\infty} \lambda^s \epsilon_{T-s}^2 = \sum_{s=0}^{\infty} w_s \epsilon_{T-s}^2, \quad \lambda \in (0, 1)$$

$$w_s = (1 - \lambda) \lambda^s, \quad \sum_{s=0}^{\infty} w_s = 1$$

$$\sum_{k=0}^{\infty} \lambda^k = \frac{1}{1-\lambda}$$



- The EWMA forecast of  $\sigma_{T+1}^2$  is based on geometrically declining weights determined by  $\lambda$

$\lambda \approx 1 \Rightarrow$  very slowly declining weights

$\lambda \approx 0 \Rightarrow$  very quickly declining weights

- The half-life of the EWMA forecast is the number of periods in which the weights decline by half

$$\text{half-life} = \frac{\ln(0.5)}{\ln(\lambda)}$$

- For daily data, J.P. Morgan found that  $\lambda = 0.94$  gave sensible short-term forecasts

$$\lambda = 0.94 \Rightarrow \text{half-life} = \frac{\ln(0.5)}{\ln(0.94)} = 11.2 \text{ days}$$

$$\sigma_{T+1}^2 = a_0 + a_1 \epsilon_T^2 + b_1 \sigma_T^2, \quad a_0 = 0, \quad a_1 = 1 - \lambda, \quad b_1 = \lambda$$

- By recursive substitution, the EWMA forecast may be re-expressed as

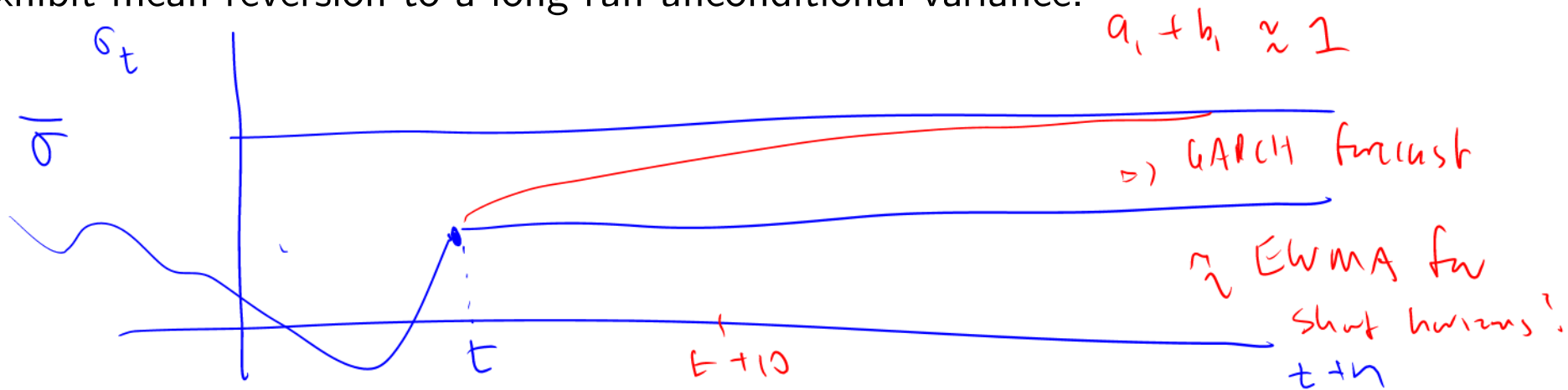
$$\sigma_{T+1,EWMA}^2 = (1 - \lambda)\epsilon_T^2 + \lambda\sigma_{T,EWMA}^2$$

which is of the form of a GARCH(1,1) model with  $a_0 = 0$ ,  $a_1 = 1 - \lambda$  and  $b_1 = \lambda$ . This is an integrated GARCH(1,1) or IGARCH(1,1) model.

- In the IGARCH(1,1),  $a_1 + b_1 = 1 - \lambda + \lambda = 1$  so  $\epsilon_t$  is not covariance stationary and so  $E[\epsilon_t^2] = E[\sigma_t^2]$  is not defined!
- EWMA/IGARCH(1,1) forecasts behave like random walk forecasts - best forecast is current value

$$\begin{aligned} \sigma_{T+1,EWMA}^2 &= (1 - \lambda)\epsilon_T^2 + \lambda\sigma_{T,EWMA}^2 \\ \sigma_{T+2,EWMA}^2 &= (1 - \lambda)\epsilon_{T+1}^2 + \lambda\sigma_{T+1,EWMA}^2 \\ &= (1 - \lambda)\sigma_{T+1,EWMA}^2 + \lambda\sigma_{T+1,EWMA}^2 = \sigma_{T+1,EWMA}^2 \\ &\vdots \\ \sigma_{T+h,EWMA}^2 &= \sigma_{T+1,EWMA}^2 \end{aligned}$$

As a result, unlike the GARCH(1,1) forecast, the EWMA forecast does not exhibit mean reversion to a long-run unconditional variance.



## Forecasting the Volatility of Multiperiod Returns

- Let  $r_t = \ln(P_t) - \ln(P_{t-1})$ . The GARCH forecasts are for daily volatility at different horizons  $h$ .
- For risk management and option pricing with stochastic volatility, volatility forecasts are needed for multiperiod returns. With continuously compounded returns, the  $h$ -day return between days  $T$  and  $T + h$  is simply the sum of  $h$  single day returns

$$r_{T+h}(h) = \sum_{j=1}^h r_{T+j}.$$

- Assuming returns are uncorrelated, the conditional variance of the  $h$ –period return is then

$$\begin{aligned}\text{var}(r_{T+h}(h)|I_T) &= \sigma_T^2(h) = \sum_{j=1}^h \text{var}(r_{T+j}|I_T) \\ &= E[\sigma_{T+1}^2|I_T] + \cdots + E[\sigma_{T+h}^2|I_T].\end{aligned}$$

- If returns have constant variance  $\bar{\sigma}^2$ , then

$$\sigma_T^2(h) = h\bar{\sigma}^2 \text{ and } \sigma_T(h) = \sqrt{h}\bar{\sigma}$$

This is known as the “square root of time” rule as the  $h$ –day volatility scales with  $\sqrt{h}$ . In this case, the  $h$ –day variance per day,  $\sigma_T^2(h)/h$ , is constant.

- If returns are described by a GARCH model then the square root of time rule does not necessarily apply. Plugging the GARCH(1,1) model forecasts for  $E[\sigma_{T+1}^2|I_T], \dots, E[\sigma_{T+h}^2|I_T]$  into  $\text{var}(r_{T+h}(h)|I_T)$  gives

$$\sigma_T^2(h) = h\bar{\sigma}^2 + (E[\sigma_{T+1}^2] - \bar{\sigma}^2) \left[ \frac{1 - (a_1 + b_1)^h}{1 - (a_1 + b_1)} \right]$$

- For the GARCH(1,1) process the square root of time rule only holds if  $E[\sigma_{T+1}^2] = \bar{\sigma}^2$ . Whether  $\sigma_T^2(h)$  is larger or smaller than  $h\bar{\sigma}^2$  depends on whether  $E[\sigma_{T+1}^2]$  is larger or smaller than  $\bar{\sigma}^2$ .
- The *term structure of volatility* is a plot of  $\sigma_T^2(h)/h$  versus  $h$ .
  - If the square root of time rule holds then the term structure of volatility is flat



## VaR Forecasts

### *Unconditional VaR Forecasts*

Let  $r_t$  denote the continuously compounded daily return on an asset/portfolio and let  $\alpha$  denote confidence level (e.g.  $\alpha = 0.95$ ). Then the 1-day unconditional value-at-risk,  $VaR_\alpha$ , is usually defined as the negative of the  $(1 - \alpha)$ -quantile of the unconditional daily return distribution:

$$\begin{aligned} VaR_\alpha &= -q_{1-\alpha}^R = -F_r^{-1}(1 - \alpha) \\ F_r &= CDF \text{ of } r_t \end{aligned}$$

Example: Let  $r_t \sim iid N(\mu, \sigma^2)$ . Then

$$\begin{aligned} q^\alpha &= \mu + \sigma \times q_{1-\alpha}^z, \quad q_{1-\alpha}^z = \Phi^{-1}(1 - \alpha) \\ \widehat{VaR}_\alpha &= -(\hat{\mu} + \hat{\sigma} \times q_{1-\alpha}^z) \end{aligned}$$

Example cont'd: Consider the  $h$ -day return,

$$r_{t+h}(h) = r_t + r_{t+1} + \cdots + r_{t+h}.$$

If  $r_t \sim iid N(\mu, \sigma^2)$  then  $r_{t+h}(h) \sim N(h\mu, h\sigma^2)$ . Then the  $h$ -day unconditional value-at-risk,  $VaR_\alpha^h$ , is

$$VaR_\alpha^h = -q_\alpha^{r(h)} = -(h\mu + \sqrt{h}\sigma q_{1-\alpha}^z)$$

## Conditional VaR Forecasts

Now assume  $r_t$  follows a GARCH process:

$$\begin{aligned}r_t &= \mu + \sigma_t z_t \\ \sigma_t^2 &\sim GARCH(p, q) \\ z_t &\sim N(0, 1)\end{aligned}$$

Then the 1-day conditional VaR,  $VaR_{\alpha,t}$ , is

$$VaR_{\alpha,t} = -q_{1-\alpha}^r = -(\mu + \sigma_t q_{1-\alpha}^z)$$

Note that  $VaR_{\alpha,t}$  is time varying because  $\sigma_t$  is time varying. The unconditional VaR,  $VaR_{\alpha}$ , is constant over time.

The estimated/forecasted VaR is

$$\widehat{VaR}_{\alpha,t} = -(\hat{\mu} + \hat{\sigma}_t q_{1-\alpha}^z), \quad \hat{\sigma}_t = \text{GARCH forecast volatility}$$

For a GARCH process, the  $h$ -day conditional value-at-risk,  $VaR_{\alpha,t}^h$ , is

$$VaR_{\alpha,T}^h = -(h\mu + \sigma_T(h)q_{1-\alpha}^z)$$

where

$$\sigma_T^2(h) = E[\sigma_{T+1}^2|I_T] + \dots + E[\sigma_{T+h}^2|I_T]$$

and  $E[\sigma_{T+1}^2|I_T], \dots, E[\sigma_{T+h}^2|I_T]$  are the GARCH  $h$ -step ahead forecasts of conditional variance.

The estimated/forecasted VaR is

$$\widehat{VaR}_{\alpha,T}^h = -(h\hat{\mu} + \hat{\sigma}_T(h)q_{1-\alpha}^z)$$