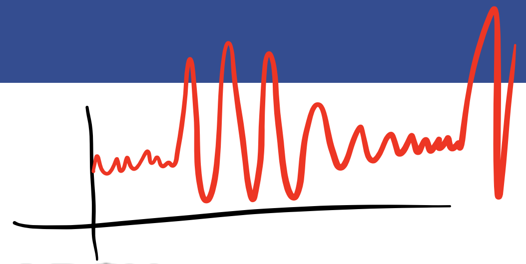
Time Series

Peter Lukianchenko

25 February 2023

Definition of GARCH



- What if p in ARCH(p) is 'too big'?
- Take more **parsimonious** version - **G**eneralized **ARCH**
- Most popular specification is GARCH(1,1)

GARCH(1,1). Bollerslev (1986)

Let the error process be such that

GARCH(1,2)

$$\varepsilon_t = \sigma_t \nu_t, \quad \nu_t \sim \text{i.i.d.} N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \cdot \sigma_{t-2}^2 \tag{5}$$

- GARCH(1,1) can be rewritten as ARCH(∞):

$$\sigma_t^2 = \alpha_0 (1 + \beta + \beta^2 + \ldots) + \alpha (1 + \beta L + \beta^2 L^2 + \ldots) \varepsilon_{t-1}^2$$

Forecasting

- Note $E_t(\sigma_{t+1}^2) = \sigma_{t+1}^2$
- Thus, 1-step-ahead forecast of variance of r_t is given directly by the model!
- Long-run variance $\sigma^2 = \frac{\alpha_0}{1-\alpha-\beta} \Rightarrow \alpha_0 = \sigma^2(1-\alpha-\beta)$
- Substitute α_0 to (5) and rewrite:

$$\sigma_{t+1}^2 = \sigma^2 + \alpha(\varepsilon_t^2 - \sigma^2) + \beta(\sigma_t^2 - \sigma^2)$$

- If $\mu_t = 0$ then $r_t = \varepsilon_t$ and tomorrow's variance is a weighted average of the long-run variance, today's squared return and today's variance
- k-step ahead forecast

$$E_t[\sigma_{t+k}^2] - \sigma^2 = (\alpha+\beta)^{k-1}(\sigma_{t+1}^2 - \sigma^2)$$

$$E_t[\sigma_{t+k}^2] \xrightarrow{k\to\infty} \sigma^2 (\text{mean reversion of volatility})$$

$\varepsilon_t \sim N(0; \sigma_t^2)$

GARCH(p,q)

Extended model:

$$\varepsilon_t = \sigma_t \nu_t, \quad \nu_t \sim \text{i.i.d.} N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

- $\sigma_t^2 = E_{t-1}(\varepsilon_t^2)$ (model for conditional variance)
- Can be expressed as ARMA$(\max(p,q), q)$ *(use $\eta_t = \varepsilon_t^2 - \sigma_t^2$)*
- Covariance stationary $\Leftrightarrow$ roots of ARMA characteristic equation lie inside unit circle (for inverse equation - outside)
- If covariance-stationary $\operatorname{Var} \varepsilon_t = \alpha_0 / (1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j)$

GARCH(1,1) vs ARMA

- Take $\eta_t = \varepsilon_t^2 - \sigma_t^2 \Rightarrow \sigma_t^2 = \varepsilon_t^2 - \eta_t$
- η_t is WN(0)
- Then (1) can be rewritten in the form:

$$\varepsilon_t^2 - \eta_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1(\varepsilon_{t-1}^2 - \eta_{t-1})$$

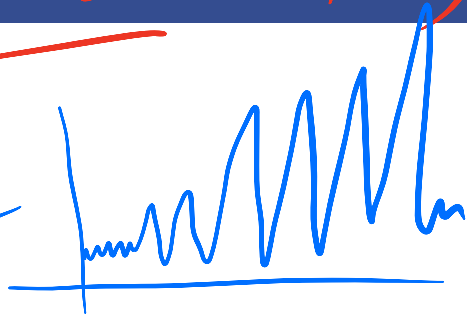
- Rearranging we get:

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + \eta_t - \beta_1 \eta_{t-1}$$

- Thus GARCH(1,1) is ARMA(1,1) for ε_t^2
- When is ARMA(1,1) covariance-stationary (= weakly stationary)?

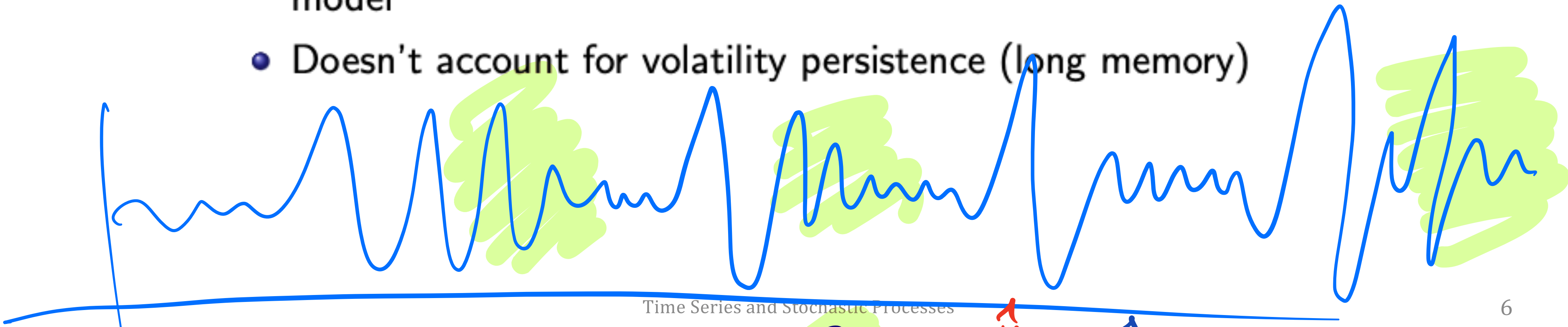
Limitations of GARCH

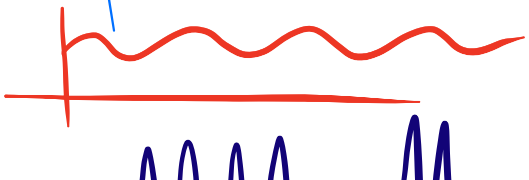
ARMA(...; ...) + Garch(..; ..)



Among others GARCH has the following limitations:

- Doesn't account for *leverage effects*
- Doesn't allow for direct feedback between conditional variance and conditional mean
- Non-negativity constraint may be violated by the estimated model
- Doesn't account for volatility persistence (long memory)





$y_t = \hat{\alpha} \cdot y_{t-1} + e_t$

$e_{t-1} = y_{t-1} - \hat{y}_{t-1}$

$\hat{y}_{t-1}$

$t-2$ $t-1$

$y_t = \alpha y_{t-1} + \varepsilon_t$

$\varepsilon_t \sim$ dist. term

$e_t = y_{t-1} - \hat{y}_{t-1}$

$\sim N(0, \sigma^2)$

e_{t-1}

IGARCH model

$\varepsilon_t \sim N(0;\sigma_t)$

$\varepsilon_t = \sigma_t \cdot v_t \sim N(0;1)$

$\boxed{\frac{\varepsilon_t}{\sigma_t} \sim N(0;1)}$

$\varepsilon_t = \sigma_t \cdot v_t$

$v_t \sim N(0;1)$

- Conditional volatility is *persistent*: if estimating GARCH(1,1) for a long time series of stock returns $\alpha_1 + \beta_1 \approx 1$

$v_t \sim N(0;1)$

$\varepsilon_t = \sigma_t \cdot v_t$

IGARCH. Nelson (1990)

Constraining $\alpha_1 + \beta_1 = 1$ in (1) yields

$$\sigma_t^2 = \alpha_0 + (1-\beta_1)\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad (2)$$

- More parsimonious representation
- Conditional variance seems to act like a process with unit root (IGARCH = **Integrated** GARCH)
- Forecast of the conditional variance:

$$\sigma_{t+1}^2 = \alpha_0 + (1-\beta_1)\varepsilon_t^2 + \beta_1\sigma_t^2$$

$$E_t[\sigma_{t+2}^2] = \hat{\alpha}_0 + \sigma_{t+1}^2$$
$$E_t[\sigma_{t+j}^2] = (j-1)\hat{\alpha}_0 + \sigma_{t+1}^2$$

$$\sigma_t^2 = \alpha_0 + (1-\beta_1)\varepsilon_{t-1}^2 + \alpha_0\beta_1 + \beta_1(1-\beta_1)\varepsilon_{t-2}^2 + \beta_1^2\sigma_{t-2}^2$$

IGARCH vs ARMA

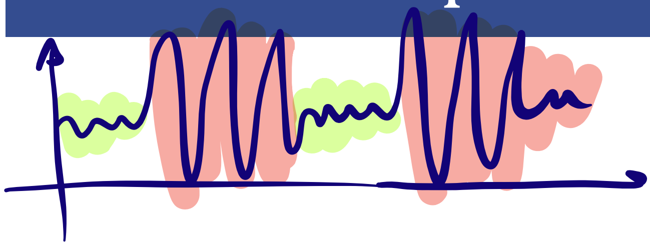
- The analogy between IGARCH and ARIMA with unit root is not perfect
- Equation (2) can be rewritten in the form

ARCH(∞)

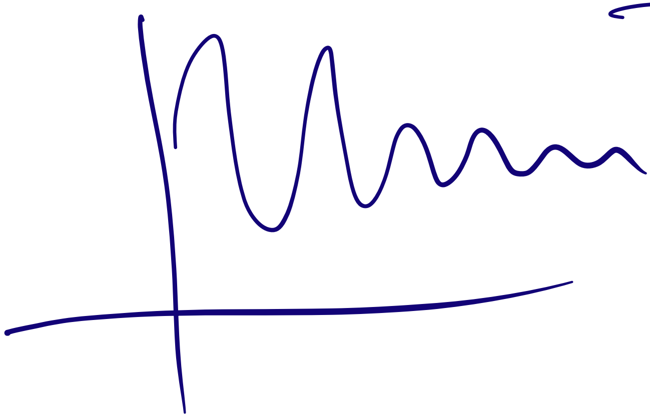
$$\Rightarrow \quad \sigma_t^2 = \frac{\alpha_0}{1-\beta_1} + (1-\beta_1)\sum_{i=0}^{\infty} \beta_1^i \varepsilon_{t-1-i}^2$$

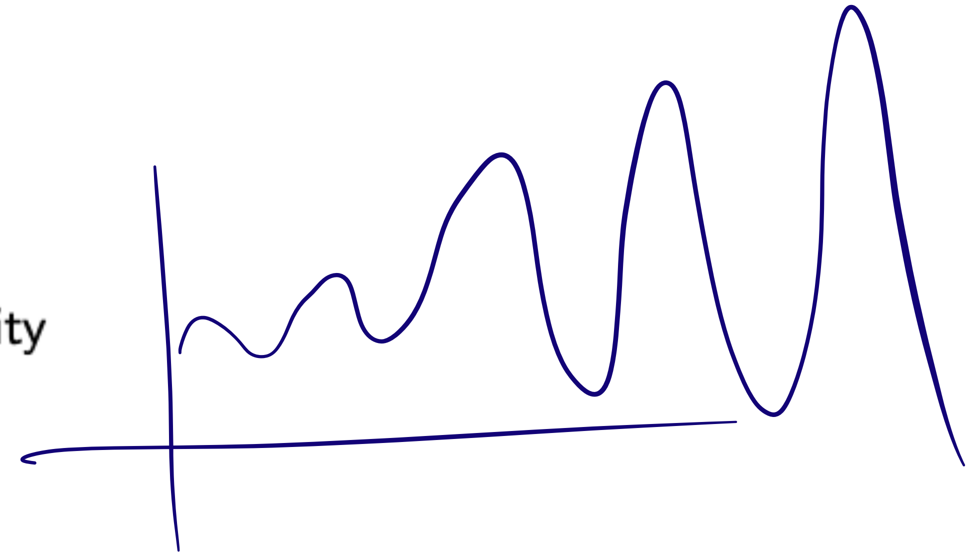
- Conditional variance is geometrically decaying function of $\{\varepsilon_t^2\}$
- Thus, IGARCH is not truly non-stationary and can be estimated like GARCH

IGARCH captures



4. Volatility clustering
9. Long memory in volatility





ARCH-M

- Risk-averse agents are compensated for holding a risky asset
- Risk premium is an increasing function of the conditional variance of returns: the greater the conditional variance, the greater the compensation for holding the asset

ARCH-M. Engle, Lilien, Robins (1987)

Excess return from holding a risky asset (wrt to a one-period T-bill)

$$r_t = c + \delta\sigma_t^2 + \varepsilon_t, \tag{3}$$

$$\varepsilon_t = \sigma_t \nu_t, \tag{4}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \tag{5}$$

- In (3) instead of σ_t^2 we can use σ_t or $\log(\sigma_t)$

Empirics

Source: AP (2015) **Model:** AR(1)-GARCH-M(1,1)

$$r_{t+1} = a_0 + a_1 r_t + \delta\sigma_{t+1} + \varepsilon_{t+1},$$
$$\varepsilon_{t+1} = \sigma_{t+1}\nu_{t+1},$$
$$\sigma^2_{t+1} = \alpha_0 + \alpha_1\varepsilon^2_t + \beta_1\sigma^2_t$$

Table
Parameter estimates (* – 5% significance)

	a_0	a_1	δ	$\alpha_0 \cdot 10^3$	α_1	β_1
Euro/USD ex rate	0.084	−0.003	−0.13	1.151	0.03*	0.968*
S&P 500 index	−0.043	0.016	0.08*	11.869*	0.073*	0.919*
US 3-m T-bill	−0.001	−0.063*	0.02*	0.002*	0.159*	0.841*

Leverage

- Black (1976): "bad" news have more pronounced effect on volatility of stocks than "good" news
- Negative correlation between stock returns and changes in volatility
- This is *leverage effect*
- Possible explanation:
  stock price $\downarrow \Rightarrow$ *Debt/Equity* $\uparrow \Rightarrow$ volatility $\uparrow$ (shareholders perceive their future cashflow stream as being more risky)

TGARCH

- "New information" is measured by the size of shock ε_t
- "Bad" news $\Leftrightarrow \varepsilon_t < 0$, "Good" news $\Leftrightarrow \varepsilon_t > 0$

TGARCH. Glosten, Jaganathan, Runkle (1994)

Shocks greater than a threshold have different effects than shocks below the threshold:

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \sigma_t \nu_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \lambda_1 I(\varepsilon_{t-1} < 0) \cdot \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \tag{6}$$

- In (6) $I(\varepsilon_{t-1} < 0) = \begin{cases} 1, & \text{if } \varepsilon_{t-1} < 0 \\ 0, & \text{if } \varepsilon_{t-1} \geq 0 \end{cases}$
- **TGARCH** = **Threshold** GARCH (or GJR-GARCH)

TGARCH

Source: AP (2015) **Model:** TARCH(1,1)

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \lambda_1 I(\varepsilon_{t-1} < 0) \cdot \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

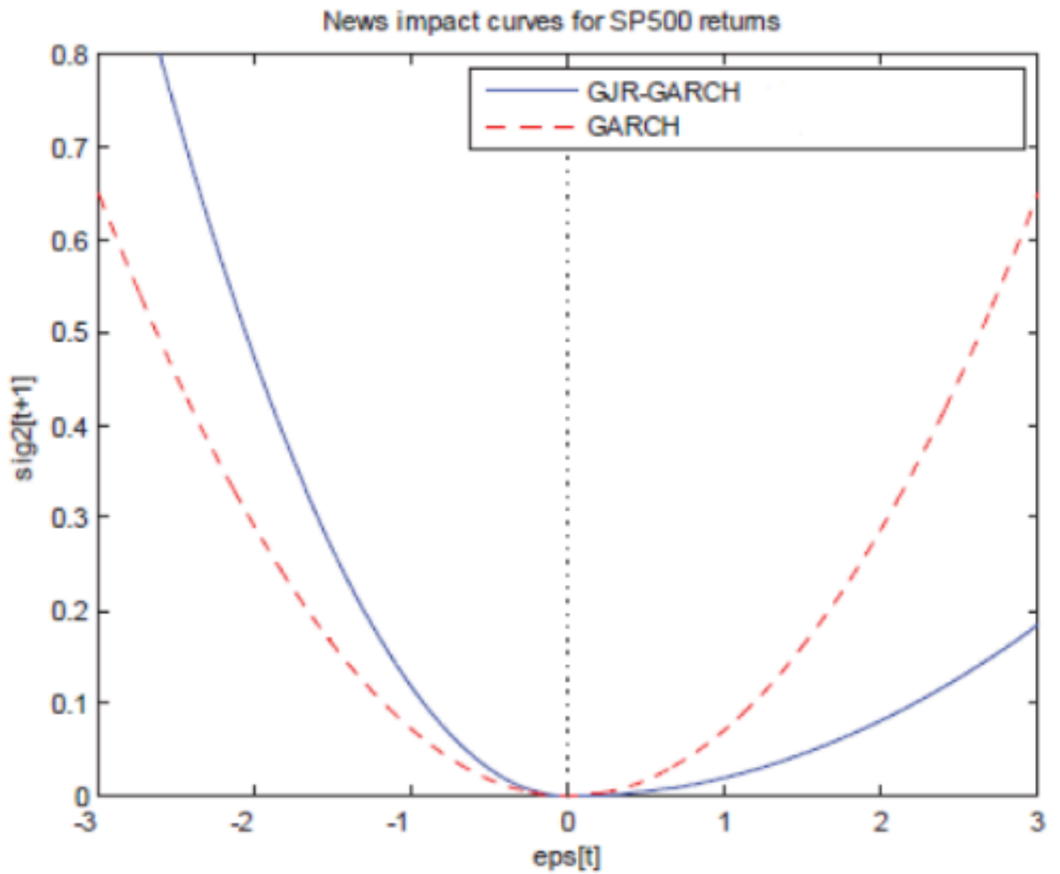
Table
Parameter estimates (* – 5% significance)

	$\alpha_0 \cdot 10^3$	α_1	β_1	λ_1	λ_1 t-stat
Euro/USD ex rate	1.266*	0.023*	0.969*	0.011	1.375
S&P 500 index	15.963*	0.020*	0.917*	0.097*	3.593
US 3-m T-bill	0.002*	0.120*	0.839*	−0.083*	−2.306

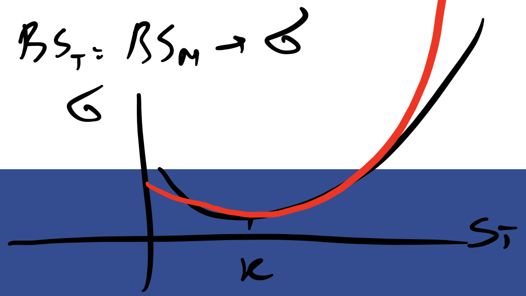
- $\lambda_1 > 0$ for stock returns, $\lambda_1 < 0$ for interest rates, $\lambda_1 \approx 0$ for exchange rates

Pagan and Schwert

- Graph of σ_t^2 as a function of ε_{t-1}, everything else fixed



$BS \to BS_T = BS_M \to \sigma$



EGARCH

- Another model accounting for asymmetric effect is **E**GARCH =**Exponential** GARCH

EGARCH. Nelson (1991)

The equation for the conditional variance is in log-linear form:

$$\ln(\sigma_t^2) = \alpha_0 + \beta \ln(\sigma_{t-1}^2) + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \tag{7}$$

- Doesn't require non-negativity constraints
- In (7) parameter γ is “in charge” of asymmetry
- Standardized shocks $\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right)$ are used – unit-free measure
- BUT: difficult to forecast conditional variance

$Var_t = Loc_t \cdot Stoch_t$

Garch

Empirics

Source: AP (2015) **Model:** EGARCH(1,1)

$$\ln(\sigma_t^2) = \alpha_0 + \beta \ln(\sigma_{t-1}^2) + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

Table
Parameter estimates (* – 5% significance)

	$\alpha_0 \cdot 10^3$	α	β	γ	γ t-stat
Euro/USD ex rate	−0.056*	0.068*	0.996*	−0.010	1.429
S&P 500 index	−0.099*	0.128*	0.984*	−0.080*	−4.706
US 3-m T-bill	−0.481*	0.295*	0.970*	0.061*	3.813

- $\gamma < 0$ for stock returns, $\gamma > 0$ for interest rates, $\gamma \approx 0$ for exchange rates
- $\gamma < 0$ implies higher future volatility following negative shock

Testing

- Estimate TARCH (or EGARCH) model and perform t-test for $\lambda_1 = 0 \quad (\gamma = 0)$
- Or perform diagnostic test for leverage effects in residuals:
  1. Estimate (G)ARCH model
  2. Form standardized residuals $s_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$, form $d_t = I(\hat{\varepsilon}_t < 0)$
  3. Estimate a regression:
     - $s_t^2 = a_0 + a_1 s_{t-1} + a_2 s_{t-2} + \dots$
       no leverage effects $\Leftrightarrow$ not reject $H_0: a_1 = a_2 = \dots = 0$
     - or Sign Bias regression:
       $s_t^2 = a_0 + a_1 d_{t-1} + u_t$
       No sign helpfulness in predicting $\sigma_t \Leftrightarrow$ not reject $H_0: a_1 = 0$
     - or Generalized regression:
       $s_t^2 = a_0 + a_1 d_{t-1} + a_2 d_{t-1} \cdot s_{t-1} + a_3 (1 - d_{t-1}) \cdot s_{t-1} + u_t$
       no leverage effects $\Leftrightarrow$ not reject $H_0: a_1 = a_2 = a_3 = 0$

More

- PARCH (Ding, Granger and Engle, 1993):

$$\sigma_t^\gamma = \alpha_0 + \alpha\varepsilon_{t-1}^\gamma + \beta\sigma_{t-1}^\gamma$$

- APARCH (Ding, Granger and Engle, 1993):

$$\sigma_t^\gamma = \alpha_0 + \alpha(\varepsilon_{t-1} - \delta|\varepsilon_{t-1}|)^\gamma + \beta\sigma_{t-1}^\gamma$$

- SQR-GARCH (Taylor, 1986 and Schwert, 1989):

$$\sigma_t = \alpha_0 + \alpha|\varepsilon_{t-1}| + \beta\sigma_{t-1}$$

- QARCH (Sentana, 1991):

$$\sigma_t^2 = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \delta\varepsilon_{t-1} + \beta\sigma_{t-1}^2$$

- NARCH (Higgins and Berra, 1992)

$$\sigma_t^2 = (\phi_0\alpha_0^\delta + \phi_1\varepsilon_{t-1}^{2\delta} + \phi_2\varepsilon_{t-2}^{2\delta} + ... + \phi_p\varepsilon_{t-p}^{2\delta})^{1/\delta}$$

- All-in-the-family GARCH (Hentschel, 1995)

$$\frac{\sigma_t^\gamma - 1}{\gamma} = \alpha_0 + \alpha\sigma_{t-1}^\gamma \left[|\nu_{t-1} - \delta| - \lambda(\nu_{t-1} - \delta)\right] + \beta\frac{\sigma_{t-1} - 1}{\gamma}$$

- To be continued...

Anything better than GARCH(1,1)

Paper by Hansen and Lunde (2005):
330 different ARCH-type models considered for Deutsche mark - US dollar exchange rate and for IBM equity returns.

Conclusions:

- For exchange rates: no evidence against GARCH(1,1)
- For equity returns: APARCH(2,2)

$$\sigma_t^\gamma = \alpha_0 + \sum_{i=1}^{2} \alpha_i (|\varepsilon_{t-1}| - \delta_i \varepsilon_{t-1})^\gamma + \sum_{i=1}^{2} \beta_i \sigma_{t-1}^\gamma$$

- Allows for leverage effect (when $\delta_i \neq 0$)
- Allows for stronger serial correlation in absolute values of returns in power $\gamma < 2$ than in squared returns

Likelyhood function

- Suppose $\{\varepsilon_t\} \sim$ i.i.d. $N(0, \sigma^2)$,
- Likelihood of observation ε_t is

$$l_t(\sigma^2) := f(\varepsilon_t|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\varepsilon_t^2}{2\sigma^2})$$

- $\{\varepsilon_t\}$ are independent $\Rightarrow$ likelihood of joint realizations is:

$$\mathcal{L}(\sigma^2|\varepsilon_1, ..., \varepsilon_T) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\varepsilon_t^2}{2\sigma^2})$$

- We want to find the *most likely* parameters given actual data
- So $\mathcal{L}(\sigma^2|\varepsilon_1, ..., \varepsilon_T) \to \max$

Likelyhood function

- Better take natural log:

$$\ln \mathcal{L} = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\varepsilon_t)^2 \rightarrow \max$$

Example

$$\text{if } y_t = bx_t + \varepsilon_t \ (\Leftrightarrow \varepsilon_t = y_t - bx_t)$$

$$\ln \mathcal{L} = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - bx_t)^2 \rightarrow \max_{b,\sigma^2}$$

- FOC: $\frac{\partial \ln \mathcal{L}}{\partial b} = 0$, $\frac{\partial \ln \mathcal{L}}{\partial \sigma^2} = 0$

MLE

- Solutions for FOC:

$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$$

$$\hat{\sigma^2} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2$$

- MLEs are
  - Consistent
  - Asymptotically normal
  - Asymptotically efficient
- Food for thought: what's the connection to OLS?

Example (AR(1) - GARCH(1,1))

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t,$$
$$\varepsilon_t = \sigma_t \nu_t, \quad \nu_t \sim \text{i.i.d. } N(0,1)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- Denote $\boldsymbol{\theta} = (a_0, a_1, \alpha_0, \alpha_1, \beta_1)$
- Then joined pdf

$$\mathcal{L}(\boldsymbol{\theta}|y_1,\ldots,y_T) = f(y_1, \ldots, y_T|\boldsymbol{\theta}) =$$
$$f_{y_1}(y_1) \cdot f_{y_2|y_1}(y_2|y_1) \cdot \ldots \cdot f_{y_T|y_{T-1},\ldots,y_1}(y_T|y_{T-1},\ldots,y_1) \qquad (3)$$

GARCH case

- In (3) we don't know the unconditional distribution of y_1
- So take conditional likelihood

$$f(y_2,\dots,y_T|y_1,\boldsymbol{\theta}) = f_{y_2|y_1}(y_2|y_1)\cdot\dots\cdot f_{y_T|y_{T-1},\dots,y_1}(y_T|y_{T-1},\dots,y_1) = \prod_{t=2}^{T}\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp(-\frac{\varepsilon_t^2}{2\sigma_t^2})$$

- Take log:

$$\ln\mathcal{L}(\boldsymbol{\theta}) = -\frac{T-1}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\ln(\sigma_t^2) - \frac{1}{2}\sum_{t=2}^{T}(\frac{\varepsilon_t^2}{\sigma_t^2}) \quad (4)$$

- Note that now $\sigma_t^2 \neq const$

- In (3) we don't know the unconditional distribution of y_1
- So take conditional likelihood

$$f(y_2,\dots,y_T|y_1,\boldsymbol{\theta}) = f_{y_2|y_1}(y_2|y_1)\cdot\ldots\cdot f_{y_T|y_{T-1},\dots,y_1}(y_T|y_{T-1},\dots,y_1) = \prod_{t=2}^{T}\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp(-\frac{\varepsilon_t^2}{2\sigma_t^2})$$

- Take log:

$$\ln\mathcal{L}(\boldsymbol{\theta}) = -\frac{T-1}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\ln(\sigma_t^2) - \frac{1}{2}\sum_{t=2}^{T}(\frac{\varepsilon_t^2}{\sigma_t^2}) \quad (4)$$

- Note that now $\sigma_t^2 \neq const$

- in (4) $\varepsilon_t = y_t - a_0 - a_1 y_{t-1}$ and $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$
- So

$$\ln \mathcal{L}(\boldsymbol{\theta}) \to \max_{\boldsymbol{\theta}=(a_0,a_1,\alpha_0,\alpha_1,\beta_1)}$$

Issues:

1. What are *initial values* ε_1 and σ_1^2?
   (often $\varepsilon_1 := 0, \sigma_1^2 := \frac{\alpha_0}{(1-\alpha_1-\beta_1)}$)
2. No simple analytical solutions for $\hat{\boldsymbol{\theta}} \Rightarrow$ *use numerical methods*
   (in R, Matlab, EVIews,...)
3. Different algorithms can find *different local maxima*
   (BHHH/Marquardt algorithms - see Press et.al, 1992)

Food for thought: Why OLS can't be used?

If two models are nested then $\Rightarrow$ statistical tests for spare parameters can be used.

Example (Nested models)

TGARCH 'nests' GARCH

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \lambda_1 I(\varepsilon_{t-1} < 0) \cdot \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Thus,

$$H_0 \colon \lambda_1 = 0,$$
$$H_1 \colon \lambda_1 \neq 0$$

- t-tests, F-tests for simple restrictions

Likelihood ratio (LR) test involves estimation of two models

1. Restricted model (when H_0 is true) $\Rightarrow$ likelihood $\mathcal{L}_r$
2. Unrestricted model (when H_1 is true) $\Rightarrow$ likelihood $\mathcal{L}_u$

If H_0 is true

$$LR := -2\ln\left(\frac{\mathcal{L}_r}{\mathcal{L}_u}\right) \longrightarrow \chi^2(m),$$

where m – number of restrictions

LR-test

- Is more flexible for non-linear models
- Can be used for non-linear restrictions

If number of parameters in the model $\uparrow \Rightarrow$ estimation error also $\uparrow$

Information criteria (T is # of observations, k is # of parameters):

$$AIC := -\frac{2}{T}\ln\mathcal{L} + \frac{2k}{T}$$
$$BIC := -\frac{2}{T}\ln\mathcal{L} + \frac{2k}{T}\ln(\sqrt{T})$$
$$HQIC := -\frac{2}{T}\ln\mathcal{L} + \frac{2k}{T}\ln(\ln T)$$

- First term represents the goodness-of-fit
- Second terms represent penalty for extra parameters
- We want to choose model with the smallest info criterion

Let f_t be the forecast of the realized value y_t

Example

Let $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$, then $f_t := \hat{y}_t = \hat{a}_0 + \hat{a}_1 y_{t-1}$

$$MSE := \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

- We want model with the lowest MSE ($\Leftrightarrow$ highest R^2)
- Here loss function $L(y_t, \hat{y}_t) = (y_t - \hat{y}_t)^2$
- OLS estimation chooses parameter values to *minimize MSE*

- Denote variance forecast by h_t

Example

For GARCH(1,1) model $h_t := \sigma_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \varepsilon_{t-1}^2 + \widehat{\beta}_1 \sigma_{t-1}^2$

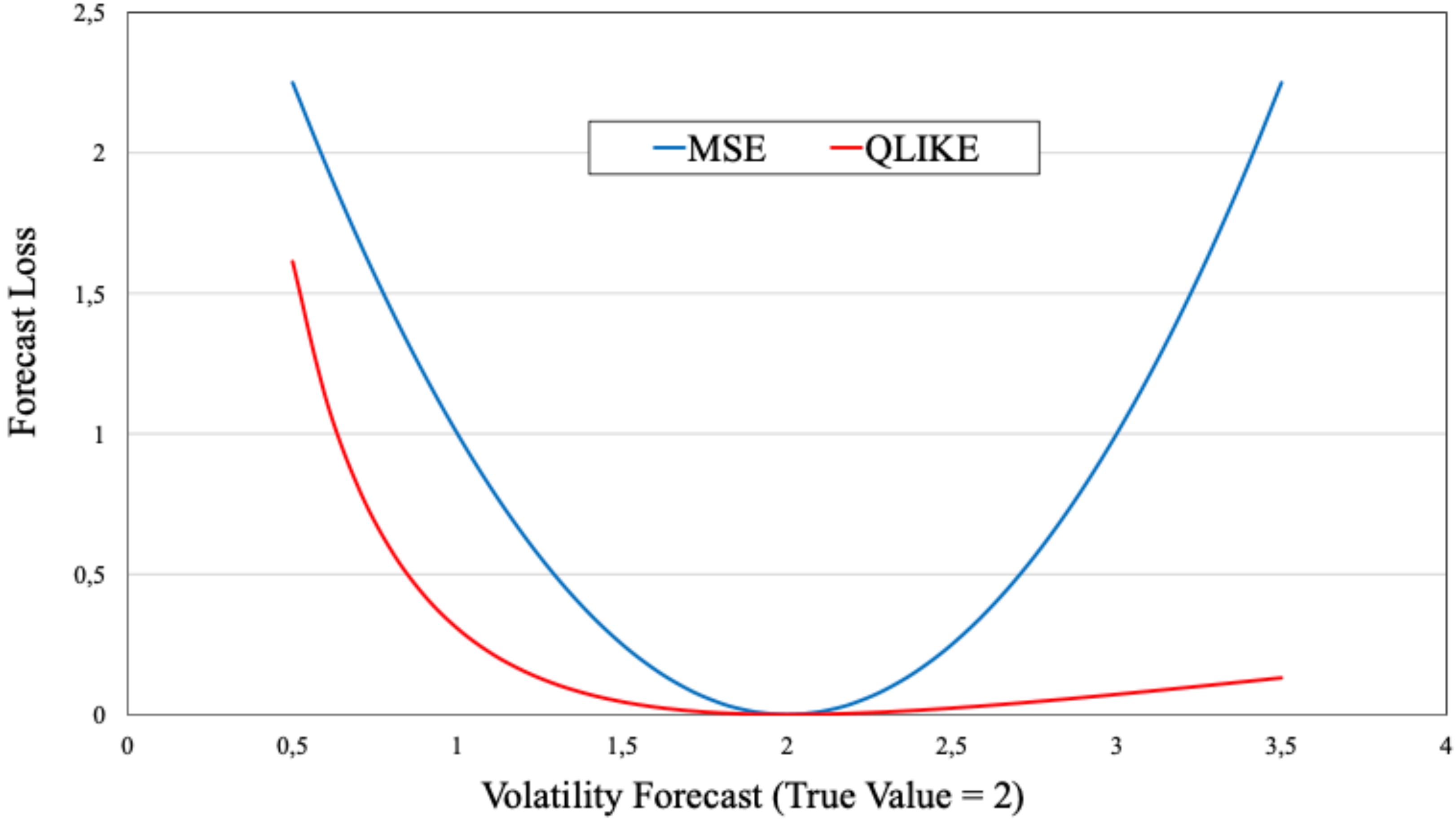
- Conditional variance is not observable ex post $\Rightarrow$ instead of unknown realized value we use proxy $\tilde{\sigma}_t$

Loss functions:

1. Squared error: $L(\tilde{\sigma}_t^2, h_t) = (\tilde{\sigma}_t^2 - h_t)$
2. QLIKE: $L(\tilde{\sigma}_t^2, h_t) = \frac{\tilde{\sigma}_t^2}{h_t} - \ln \frac{\tilde{\sigma}_t^2}{h_t} - 1$

To choose volatility model rank them according to the average loss

$$\bar{L}_i := \frac{1}{T} \sum_{t=1}^{T} L(\tilde{\sigma}_t^2, h_t^i) \qquad \text{for model } i$$


MSE
QLIKE
Forecast Loss
2,5
2
1,5
1
0,5
0
0
0,5
1
1,5
2
2,5
3
3,5
4
Volatility Forecast (True Value = 2)