# Financial Econometrics

## Slides-09: Volatility Modelling



Dr. Rachida Ouysse
School of Economics[1]

## Lecture Plan



- Motivation for modeling return volatility
- Measures of return volatility
- Conditional volatility via smoothing
- ARCH
  - Conditional variance is a function of info set;
  - It captures "clustering" in return series;
  - It explains non-normality of return, to some extent;
  - It can be used to improve interval forecasts and VaR (Value at Risk);
  - Estimation and testing.

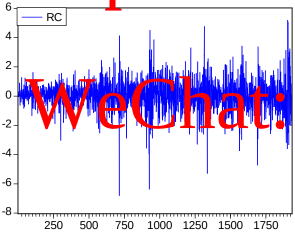
## Introduction and Motivation



eg. Volatility in NYSE Composite index return

- Clustering.
- Squared returns are strongly autocorrelated.








Correlogram of RC2

Sample: 1 1931
Included observations: 1930

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.179 | 0.179 | 62.044 | 0.000 |
| | | 2 | 0.168 | 0.140 | 116.47 | 0.000 |
| | | 3 | 0.175 | 0.131 | 176.00 | 0.000 |
| | | 4 | 0.109 | 0.044 | 198.89 | 0.000 |
| | | 5 | 0.184 | 0.130 | 264.80 | 0.000 |
| | | 6 | [illegible] | 0.035 | [illegible] | 0.000 |
| | | 7 | [illegible] | 0.067 | 325.24 | 0.000 |
| | | 8 | [illegible] | [illegible] | 349.60 | 0.000 |
| | | 9 | 0.108 | 0.039 | 372.11 | 0.000 |
| | | 10 | 0.095 | 0.015 | 389.80 | 0.000 |
| | | 11 | 0.107 | 0.042 | 411.90 | 0.000 |
| | | 12 | 0.102 | 0.028 | 432.21 | 0.000 |
| | | 13 | 0.066 | -0.006 | 440.72 | 0.000 |
| | | 14 | 0.053 | -0.016 | 446.26 | 0.000 |
| | | 15 | 0.041 | -0.016 | 449.58 | 0.000 |
| | | 16 | 0.097 | 0.052 | 467.86 | 0.000 |
| | | 17 | 0.040 | -0.018 | 470.96 | 0.000 |
| | | 18 | 0.068 | 0.024 | 479.94 | 0.000 |
| | | 19 | 0.063 | 0.014 | 487.77 | 0.000 |
| | | 20 | 0.045 | 0.005 | 491.70 | 0.000 |

## Motivation



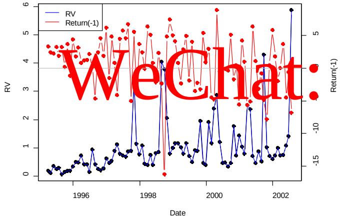
e.g. Volatility in NYSE Composite Index Return

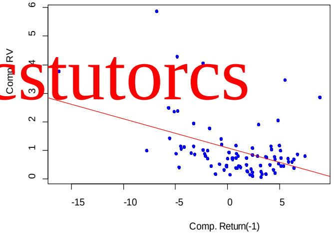
- Monthly realised variance:

  $RV$ =sample mean of squared daily returns in a month

- RV is negatively correlated to lagged monthly return.
  $\text{Cor}(RV, Return(-1)) = -0.419$.





## Motivation



- Importance of return volatility

  Asset pricing, risk management and portfolio selection

  Substantial dependence structure in volatility



- Clustering:
  - strong autocorrelations in squared returns,
  - large variations tend to be followed by large variations
- Asymmetry:
  - negative returns tend to cause more volatility than positives



- ARMA are unable to capture these features

  Conditional variance is constant in ARMA.

  Amend ARMA with a suitable conditional variance: ARCH and GARCH models.

# Volatility

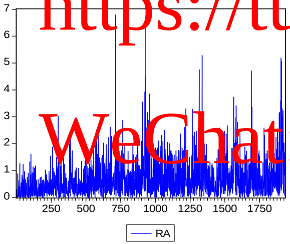


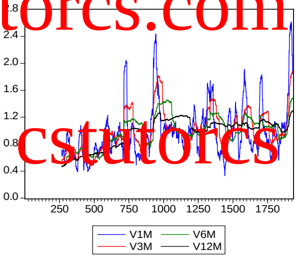
## Measures of return volatility (tendency of variation)

- Historical volatility: Sample variance or Stddev

eg. NYSE composite return: Sample Stddev





# Realized Volatility



Measures of return volatility

- **Realised volatility:** Realised variance = Sample mean of squared higher frequency returns

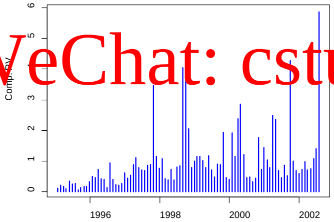
  (eg. daily RV = Sample mean of squared 5-min returns)



eg. NYSE composite return: Monthly realised variance

RV = Sample mean of squared daily returns in a month





# Realized Volatility



Measures of return volatility

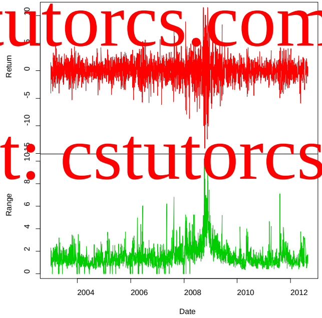
- **Range (high/low):**

  $100 \times$ ln(high/low) in a time interval (eg, a day)

eg. BHP
daily return and range







# Implied Volatility



**Implied volatility:**

standard deviation derived from options prices



- Option of an asset: the right to buy/sell the asset at a future time (maturity) at a fixed price (strike).
- Given the price of an option, maturity, strike and risk-free interest rate, the std deviation can be recovered from Black-Scholes formula, known as IV.
- IV represents market's opinions on the return's std deviation.



Black-Scholes formula:
price of an option $=$f(stdev, maturity,strike, $r_f-$rate)

# Implied Volatility



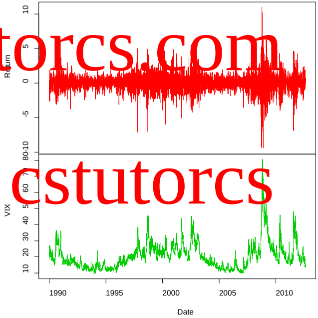
- **Implied volatility:**

eg. **VIX**: index of IVs of a set of options on the SP500 index

SP500 daily return & VIX







## Conditional Volatility



– Conditional variance of return

- $\sigma^2_{t+1|t} = \mathrm{Var}(r_{t+1}|\Omega_t)$ ,

  where $r_{t+1} = 100\ln(P_{t+1}/P_t)$ is the return and

  $\Omega_t$ is the information set at the end of period $t$.

- It should capture "clustering" or autocorrelations in squared returns, and facilitate predicting the return volatility

- Knowing it helps to
  - assess the risk of an asset via value-at-risk;
  - price options;
  - form mean-variance efficient portfolios.

# Conditional Volatility



## Exponentially weighted moving average (EWMA)

- The squared returns $\{r_t^2, r_{t-1}^2, \cdots, r_1^2\}$ carry info about the volatility as $E(r_t^2) \equiv$ variance.



- A weighted average of squared returns is an approximation to the conditional variance. Recent observations should weigh more.
- EWMA: for $0 < \lambda < 1$,

$$\sigma^2_{t+1|t} = (1-\lambda)(r_t^2 + \lambda r_{t-1}^2 + \lambda^2 r_{t-2}^2 + \cdots)$$



  - weights decay exponentially;
  - weights sum up to 1.
  - RiskMetrics recommend $\lambda = 0.94$

## EWMA



- EWMA: alternative formulation
  - $\sigma^2_{1|0} = r_1^2$
  - $\sigma^2_{t+1|t} = (1-\lambda)r_t^2 + \lambda\sigma^2_{t|t-1}$, for $t = 1,2,3,\ldots$



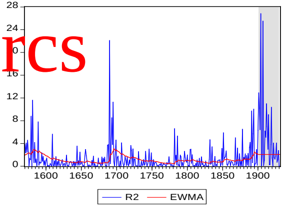
– Quick and easy;

– Can be used as 1-step ahead prediction.

eg. NYSE Composite return:

$\lambda = 0.94$





## ARCH (autoregressive conditional heteroskedasticity) Engle (1982) – Nobel price winner 1993



**Autoregressive conditional heteroscedasticity** (ARCH) models are a class of models where the conditional variance evolves according to an autoregressive process.



First define the conditional variance of the error term $u_t$ to be

$$\sigma_t^2 = \mathrm{var}\left(\mu_t \left| \mu_{t-1}, \mu_{t-2}, \ldots \right.\right) = E\left(\left(\mu_t - E\left(\mu_t\right)\right)^2 \left| \mu_{t-1}, \mu_{t-2}, \ldots \right.\right)$$

As it is usually assumed that $E\left(\mu_t\right) = 0$



$$\sigma_t^2 = \mathrm{var}\left(\mu_t \left| \mu_{t-1}, \mu_{t-2}, \ldots \right.\right) = E\left(\mu_t^2 \left| \mu_{t-1}, \mu_{t-2}, \ldots \right.\right) = E_{t-1}\left(\mu_t^2\right)$$

The ARCH(1) model assumes

$$\sigma_t^2 = E_{t-1}\left(\mu_t^2\right) = \alpha_0 + \alpha_1 \mu_{t-1}^2$$

**The conditional variance captures 'clustering': large past shock leads to large conditional variance.**

# ARCH (autoregressive conditional heteroskedasticity)



**Extensions**

- An ARCH($q$) model is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1\mu_{t-1}^2 + \alpha_2\mu_{t-2}^2 + \ldots + \alpha_q\mu_{t-q}^2$$



- Under ARCH, the conditional mean equation can take any form. An example of a full model would be

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \mu_t \qquad \mu_t \sim N\left(0, \sigma_t^2\right)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1\mu_{t-1}^2$$



**Alternative notation**

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \mu_t \\ \mu_t &= \nu_t\sigma_t \qquad \nu_t \sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1\mu_{t-1}^2 \end{aligned}$$

# Properties of ARCH(1)



- ARCH(1): $\mu_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$, $\Omega_{t-1} = \{y_{t-1}, \mu_{t-1}, y_{t-2}, \mu_{t-2} \cdots\}$ is the info set at the end of period $t-1$.

  $\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2, \ \alpha_0 > 0, \ 0 \leq \alpha_1 < 1$

  - Its conditional variance is time varying: Var$(\mu_t|\Omega_{t-1}) = \sigma_t^2$, CI$(95\%) =?$
  - It is WN:(Use LIE) $E(\mu_t) = 0$, Var$(\mu_t) = \frac{\alpha_0}{1-\alpha_1}$, Cov$(\mu_t, \mu_{t-j}) = 0$

    But it is NOT independent WN or iid WN. Why?

# Proof of properties



### Definition (Law of Iterated Expectations)

For a random variable $Y$ and information sets $\Omega_1$ and $\Omega_2$, the the LIE states that

$$E\left(Y|\Omega_1\right) = E\left(E\left(Y|\Omega_2\right)|\Omega_1\right),$$



where information set $\Omega_1$ is included in information set $\Omega_2$.
Example: $E\left(Y_t|\Omega_{t-2}\right) = E\left(E\left(Y_t|\Omega_{t-1}\right)|\Omega_{t-2}\right)$
Special Case: If $\Omega_1$ is empty set, then $E\left(Y\right) = E\left(E\left(Y|\Omega_2\right)\right)$.

$\mu_t = \nu_t\sigma_t = \nu_t\sqrt{\alpha_0 + \alpha_1\mu_{t-1}^2}$, where $\nu_t$ is $N(0,1)$



1. Unconditional Expectation of $\mu_t$. We have that $\mu_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$:

$$\begin{aligned} E\left(\mu_t\right) &= E\left[E\left[\mu_t|\Omega_{t-1}\right]\right] && (1) \\ E\left[\mu_t|\Omega_{t-1}\right] &= 0 && (2) \\ E\left(\mu_t\right) &= 0. && (3) \end{aligned}$$

## Proof of properties



$\mu_t = \nu_t \sigma_t = \nu_t \sqrt{\alpha_0 + \alpha_1 \mu_{t-1}^2}$, where $\nu_t$ is $N(0,1)$

2. Unconditional variance of $\mu_t$. We have that



$$
\begin{aligned}
E\left(\mu_t^2\right) &= E\left[E\left[\mu_t^2 | \Omega_{t-1}\right]\right] && (4)\\
&= E\left[E\left[\nu_t^2\left(\alpha_0 + \alpha_1 \mu_{t-1}^2\right) | \Omega_{t-1}\right]\right] && (5)\\
&= E\left[\left(\alpha_0 + \alpha_1 \mu_{t-1}^2\right) E\left[\nu_t^2 | \Omega_{t-1}\right]\right] && (6)\\
&= E\left[\alpha_0 + \alpha_1 \mu_{t-1}^2\right] = E\left[\alpha_0 + \alpha_1 E\left[\mu_{t-1}^2 | \Omega_{t-2}\right]\right] && (7)\\
&= \alpha_0 + \alpha_1 E\left[\alpha_0 + \alpha_1 \mu_{t-2}^2\right] && (8)\\
&= \cdots = \alpha_0\left(1 + \alpha_1 + \alpha_1^2 + \cdots + \alpha_1^{t-1}\right) + \alpha_1^t E\left[\mu_0^2\right] && (9)
\end{aligned}
$$



As $t \to \infty$, the unconditional variance converges if $\alpha_1 < 1$ to:
$E\left(\mu_t^2\right) = \frac{\alpha_0}{1-\alpha_1}$. $\longrightarrow$ Unconditionally, the process $\mu_t$ is **homoskedastic.**

# Properties of ARCH(1)



- It can be alternatively expressed as: $\mu_t = \sigma_t v_t$, $v_t \sim iidN(0, 1)$, where $v_t = \mu_t/\sigma_t$ is the **standardised shock.**
- When model is correct, $v_t^2$ should have no autocorrelation
- The unconditional distribution of $\mu_t$ is NOT normal, with heavy tails (kurtosis $> 3$).

# MLE of ARCH(1)



- An example: $AR(1) - ARCH(1)$

$$y_t = c + \phi_1 y_{t-1} + \mu_t, \mu_t|\Omega_{t-1} \sim N(0, \sigma_t^2), \tag{10}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2, \tag{11}$$

$$\alpha_0 > 0, 0 \leq \alpha_1 < 1. \tag{12}$$



- Likelihood of $\{y_1, y_2, \cdots, y_{T-1}, y_T\}$:

$$L(\Theta) = f(y_T|\Omega_{T-1}) f(y_{T-1}|\Omega_{T-2}) \cdots f(y_2|\Omega_1) f(y_1)$$

$$f(\mu_t|\Omega_{t-1}) = (2\pi\sigma_t^2)^{-1/2} \exp\{-\frac{(y_t - c - \phi_1 y_{t-1})^2}{2\sigma_t^2}\}. \tag{13}$$



- ML Estimator maximises the Log likelihood function

$$\ln L(\Theta) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\left[\ln(\sigma_t^2) + \frac{(y_t - c - \phi_1 y_{t-1})^2}{\sigma_t^2}\right].$$

# MLE of ARCH(1)



- ML estimators are generally consistent with an asymptotic normal distribution.



- The above holds even when the conditional normality $\mu_t|\omega_{t-1} \sim N(0, \sigma_t^2)$ is **incorrectly** assumed, as long as the conditional mean and conditional variance are correctly specified.



- With robust quasi ML standard errors, inference is standard.

# Example



eg: NYSE composite return: AR(1)-ARCH(5)



Dependent Variable: RC
Method: ML - ARCH (Marquardt) - Normal distribution

Sample (adjusted): 3 1931
Included observations: 1929 after adjustments
Convergence achieved after 17 iterations
Bollerslev-Wooldridge robust standard errors & covariance
Variance backcast: OFF
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*RESID(-2)^2 + C(6)*RESID(-3)^2 + C(7)*RESID(-4)^2 + C(8)*RESID(-5)^2

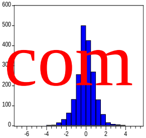
| | Coefficient | Std. Error | z-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.073023 | 0.020198 | 3.615442 | 0.0003 |
| AR(1) | 0.109876 | 0.026312 | 4.175893 | 0.0000 |
| | | Variance Equation | | |
| C | [illegible] | [illegible] | [illegible] | 0.0000 |
| RESID(-1)^2 | [illegible] | [illegible] | [illegible] | [illegible] |
| RESID(-2)^2 | [illegible] | [illegible] | [illegible] | [illegible] |
| RESID(-3)^2 | [illegible] | [illegible] | [illegible] | [illegible] |
| RESID(-4)^2 | [illegible] | [illegible] | [illegible] | [illegible] |
| RESID(-5)^2 | 0.059128 | 0.023724 | 2.492331 | 0.0127 |

| | | | |
|---|---|---|---|
| R-squared | 0.001836 | Mean dependent var | 0.035168 |
| Adjusted R-squared | -0.001801 | S.D. dependent var | 1.006452 |
| S.E. of regression | 1.007357 | Akaike info criterion | 2.664282 |
| Sum squared resid | 1949.371 | Schwarz criterion | 2.687361 |
| Log likelihood | -2561.700 | F-statistic | 0.504880 |
| Durbin-Watson stat | 2.069097 | Prob(F-statistic) | 0.831451 |
| Inverted AR Roots | .11 | | |

**Correlogram of E2**

Sample: 1 1931
Included observations: 1929

| | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | 0.216 | 0.216 | 90.395 | 0.000 |
| 2 | 0.165 | 0.124 | 142.97 | 0.000 |
| 3 | 0.187 | 0.137 | 210.51 | 0.000 |
| 4 | 0.112 | [illegible] | [illegible] | 0.000 |
| 5 | [illegible] | 0.116 | [illegible] | 0.000 |
| 6 | [illegible] | 0.038 | [illegible] | 0.000 |
| 7 | [illegible] | 0.062 | 354.28 | 0.000 |
| 8 | 0.115 | [illegible] | [illegible] | 0.000 |
| 9 | 0.100 | 0.028 | 399.24 | 0.000 |
| 10 | 0.088 | 0.010 | 414.33 | 0.000 |
| 11 | 0.104 | 0.041 | 435.21 | 0.000 |
| 12 | 0.105 | 0.035 | 456.75 | 0.000 |
| 13 | 0.061 | -0.012 | 464.07 | 0.000 |
| 14 | 0.050 | -0.015 | 468.87 | 0.000 |
| 15 | 0.046 | -0.007 | 472.97 | 0.000 |
| 16 | 0.098 | 0.059 | 491.72 | 0.000 |




Series: E
Sample 1 1931
Observations 1929

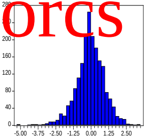
| | |
|---|---|
| Mean | -0.033709 |
| Median | -0.037415 |
| Maximum | 5.392314 |
| Minimum | -6.773783 |
| Std. Dev. | 1.004962 |
| Skewness | -0.198864 |
| Kurtosis | 7.131158 |
| Jarque-Bera | 1384.431 |
| Probability | 0.000000 |

**Correlogram of V2**



Sample: 1 1931
Included observations: [illegible]

| | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | [illegible] | [illegible] | [illegible] | [illegible] |
| 2 | -0.013 | -0.013 | 0.3671 | 0.832 |
| 3 | -0.012 | -0.012 | 0.6386 | 0.888 |
| 4 | -0.044 | -0.044 | 4.3525 | 0.360 |
| 5 | -0.033 | -0.033 | 6.4623 | 0.264 |
| 6 | 0.036 | 0.035 | 8.9570 | 0.176 |
| 7 | 0.025 | 0.023 | 10.143 | 0.181 |
| 8 | 0.028 | 0.026 | 11.669 | 0.167 |
| 9 | 0.020 | 0.019 | 12.458 | 0.189 |
| 10 | 0.030 | 0.033 | 14.199 | 0.164 |
| 11 | 0.025 | 0.030 | 15.392 | 0.165 |
| 12 | 0.045 | 0.049 | 19.325 | 0.081 |
| 13 | 0.010 | 0.014 | 19.523 | 0.108 |
| 14 | 0.015 | 0.019 | 19.955 | 0.132 |
| 15 | 0.001 | 0.004 | 19.958 | 0.174 |
| 16 | 0.023 | 0.026 | 21.019 | 0.178 |




Series: V
Sample 1 1931
Observations 1929

| | |
|---|---|
| Mean | -0.048030 |
| Median | -0.043219 |
| Maximum | 3.427925 |
| Minimum | -5.188789 |
| Std. Dev. | 0.999109 |
| Skewness | -0.413633 |
| Kurtosis | 4.558816 |
| Jarque-Bera | 250.3099 |
| Probability | 0.000000 |

## Example



eg. NYSE composite return: AR(1)-ARCH(5)

- Squared residuals (E2) of AR(1) have strong autocorrelation.
  Squared standardised residuals (V2) are not autocorrelated
- Residuals (E) of AR(1) have larger kurtosis.
  Standardised residuals (V) larger negative skewness.
- Normality is rejected for both E and V.



Two essential checks for the 'adequacy' of a model

- Adequate mean equation: E (residuals) has no autocorrelation;
- Adequate variance equation: V2 has no autocorrelation

# Comments and limitations of ARCH



## Advantages of ARCH

- It is able to capture 'clustering' in return series or the autocorrelation in squared returns.
- It facilitates volatility forecasting.
- It explains, partially, non-normality in return series.



## Limitations of ARCH

- In ARCH(q), the $q$ may be selected by AIC, SIC or LR test. The correct value of q might be very large. The model might not be parsimonious. (eg. ARCH(1) would not work for the composite return)
- The conditional variance $\sigma_t^2$ cannot be negative: Requires non-negativity constraints on the coefficients. Sufficient (but not necessary) condition is: $\alpha_i \geq 0$ for all $i = 0, 1, 2, \cdots q$. Especially for large values of $q$ this might be violated