Forecasting Methods and Applications

Applied Data Analysis School

Lecture 5

Financial Market Volatility

November 2021

Cristina Amado
University of Minho

Plan of the second half of the course:

1. Time-dependent volatility and the ARCH model
2. The Generalized ARCH model
3. Families of univariate GARCH models
4. Nonlinear GARCH models [only for 5648]
5. Integrated and fractionally integrated GARCH and semi- and nonparametric ARCH models [only for 5648]
6. Intraday realized volatility models [only for 5648]
7. Multivariate GARCH models [only for 5648]

Lecture 5

Outline of the lecture:

- Stylized facts for financial returns
- Time-varying volatility
- Naive models of volatility
- Conditional heteroskedasticity
- The ARCH model

Reference list:

- Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of UK Inflation", *Econometrica* 50, 987 – 1007.
- Engle, R. F. and Patton, A. J. (2001), "What good is a volatility model?", *Quantitative Finance* 1, 237 – 245.
- Schwert, G. W. (1989), "Why Does Stock Market Volatility Change Over Time?" *The Journal of Finance* 44, 1115–1153.
- Teräsvirta, T. (2009), "An Introduction to Univariate GARCH Models", in T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), Handbook of Financial Time Series, New York: Springer, 17 – 42.
- Tsay, R. S. (2002), "Analysis of Financial Time Series", Wiley, Chapters 3.1 – 3.3.

Stylized facts for financial returns

1. Mean is constant, but volatility is time-varying

   ▷ There is strong serial correlation in squared or absolute returns. Serial independence between the squared values of the series is often rejected pointing towards the existence of non-linear relationships between subsequent observations.

2. No serial correlation in the returns

   ▷ Inverse relationship between volatility and serial correlation of stock returns. Return series usually show no serial correlation.

3. Volatility clustering in the return series

   ▷ 'Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes' (Mandelbrot, 1963).

Stylized facts for financial returns

4. Returns have a fat-tailed and skewed distribution. Financial returns are highly peaked (leptokurtic) and slightly asymmetric.

   ▷ Outliers: Large observations (particularly negative) in returns series.

   ▷ Asymmetry: Large negative returns are more frequent than large positive returns.

5. Leverage effects

   ▷ Changes in stock prices are often negatively correlated with changes in volatility.

6. Co-movements in volatility

   ▷ Evidence of common factors to explain volatility in multiple series.

Stylized facts for financial returns

Leverage effect

- If the price of the stock of a company declines, then the market capitalization of that company declines. (The market capitalization is the market price of that company's equities.) If the company's debt remains the same, then the market assessment of that company's debt-to-equity ratio has therefore increased. This makes the company riskier in the eye of the public. Therefore, the variance of the stocks returns of the company, which is a measure of the company's risk, increases.
- The result is a negative relation between returns and volatility. Thus, negative returns generate high volatility, while positive returns generate low volatility; see Fisher (1976).

Stylized facts for financial returns

- Consider the closing prices of the London Financial Times Stock Exchange 100 (FTSE100) and the Standard & Poor's 500 (S&P500).
- Data cover the period from 02-04-1984 until 20-11-2013.
- Transform the asset price into continuously compounded returns over the period $t-1$ to t:

$$y_t = \log(P_t) - \log(P_{t-1})$$

where P_t is the asset price at time t.

Stylized facts for financial returns

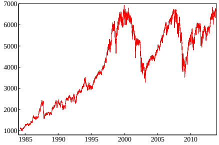


Figure 1: FTSE100

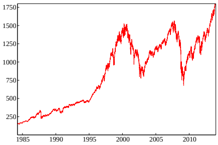


Figure 2: S&P500

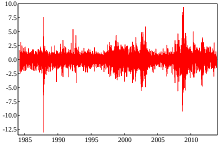


Figure 3: FTSE100 returns

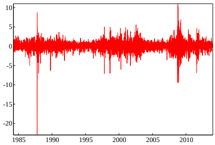


Figure 4: S&P500 returns

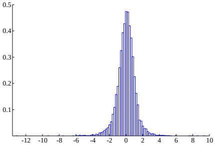


Figure 5: FTSE100 returns: histogram

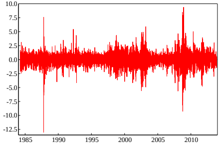


Figure 6: FTSE100 returns: ACF

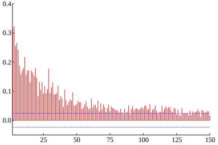


Figure 7: Sqrd. FTSE100 returns: ACF

Stylized facts for financial returns

Summary statistics for the return series

	Mean	Std.Dev.	Min.	Max.	Skewness	Ex.Kurtosis
FTSE100	0.024	1.1158	-13.029	9.384	-0.3777	8.3401
S&P500	0.032	1.1639	-22.900	10.957	-1.2771	27.996

Time-Dependent Volatility

Why do we need a volatility model?

- The main goal of a volatility model is to forecast volatility.
- A volatility model is used to forecast the absolute returns, predict quantiles or the entire density.
- These forecasts are used in risk management, derivative pricing and hedging, portfolio selection, options trading, and many other financial activities.
- In each case, it is the predictability of volatility that is required. For example:

  ▷ A risk manager must know today how likely his portfolio will decline in the future.

  ▷ To hedge a financial contract, the trader also needs to know how volatile is the forecast volatility.

  ▷ A portfolio manager may want to sell a stock or a portfolio before it becomes too volatile.

Time-Dependent Volatility

- In the early literature, asset returns were modelled as independent and identically distributed (i.i.d.) random process over time with zero mean and constant variance (Bachellier, 1900).
- Leptokurtosis led to a literature on modelling asset returns as i.i.d. random variables having some thick-tailed distribution (Mandelbrot, 1963).
- These models, although able to capture the leptokurtosis, could not account for the existence of non-linear temporal dependence such as volatility clustering.

Time-Dependent Volatility

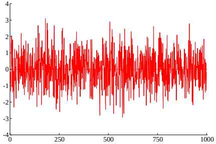


Figure 8: White Noise

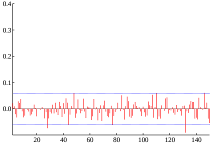


Figure 9: White Noise: ACF

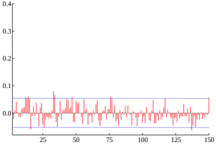


Figure 10: Sqrd. White Noise: ACF

Time-Dependent Volatility

Consider a simple model to replicate the time-varying volatility phenomenon.

- Let N_t be the number of news on day t.
- When a news arrives to the market, it is expected that the price of a given asset to be updated. This translates on a price change by a random quantity $u_{i,t}$, $i = 1, ..., N_t$.
- Then, if there is a (relevant) news on a given day t, we have:

$$\log P_t = \log P_{t-1} + \mu + u_{1,t}$$

- With two news the model becomes:

$$\log P_t = \log P_{t-1} + \mu + u_{1,t} + u_{2,t},$$

  and so on...

Time-Dependent Volatility

- The asset return is given by

$$y_t = \mu + \sum_{i=1}^{N_t} u_{i,t} \tag{1}$$

where $u_{i,t} \sim iid\ \mathcal{N}(0, \sigma^2)$ and independent of N_t. Thus,

$$\mathrm{Var}(y_t | N_t = n_t) = n_t \sigma^2.$$

- The larger the arrival of news n_t, the higher the volatility in the market.
- Refine the model assuming $N_t \sim P(\lambda_t)$, where $\lambda_t = N_{t-1} + 1$, i.e., the average number of news in t is a function of the news of the previous period.

Models for Volatility

- There is an enormous amount of research on volatility models, but it continues to be an important line of research for both practitioners and academics.
- Problems of specifying a volatility model: σ_t^2 is unobserved; σ_t^2 is non-negative.
- Types of volatility models:

1. Historical volatility
2. Implied volatility
3. Exponential weighted moving average models
4. Autoregressive volatility models

Naive Models of Volatility

- The easiest way to capture volatility clustering is by letting tomorrow's variance be the simple average of the most recent p observations:

$$\sigma^2_{t+1} = \frac{1}{p}\sum_{i=1}^{p} y^2_{t+1-i} \tag{2}$$

- This is known as the **moving average** variance estimator. We move the window one unit, add a new observation and drop the last one, such that the size of the window remains constant, i.e., equal to p.

- Notes:

  $\triangleright$ Only the latest p observations are used. Since σ^2_{t+1} depends on lagged information, it can be thought of as the prediction (made in t) of the volatility in $t+1$.

  $\triangleright$ This method can produce quite abrupt changes in the estimate.

Naive Models of Volatility

In general, for a point in time t :

$$\sigma_t^2 = \frac{1}{p}\sum_{i=1}^{p} y_{t-i}^2 \qquad (3)$$

Here, we assume $\mathrm{E}(y_t|\mathcal{F}_{t-1}) = 0$. By moving the window, the estimator becomes a moving average.

Changing slightly the notation and re-writing:

$$h_t = \sum_{i=1}^{p} \frac{1}{p} \varepsilon_{t-i}^2$$

where $\varepsilon_t = y_t$.

- This is an estimator of the variance where past squared forecasting errors have the same weight equal to $1/p$.

Naive Models of Volatility

Remarks:

- What is the best choice for p?
  $\triangleright$ A high p leads to an excessively smoothly evolving σ_{t+1}, whereas a low p leads to an excessively irregular pattern of σ_{t+1}.
- The estimator puts equal weights on the past p observations (from $t-1$ until $t-p$) for computing the volatility.
- The estimator implies that the observations outside the period from $t-1$ until $t-p$ have weight equal to zero. This means that an extreme observation inside (outside) the sample window will overestimate (underestimate) the estimated volatility.
- The autocorrelation function of the squared returns suggests that a more gradual decline is needed in the effect of past returns on today's variance.

Naive Models of Volatility

- However, we might think that the recent past is more relevant than the far past.
- An alternative is to consider decreasing weights.
- The decay of the weights quantifies the degree of memory of the process:

  ▷ The weights measure the influence of variations of the past prices on today's volatility.

  ▷ Low (high) weights indicate fast (slow) decay.

Naive Models of Volatility

- An alternative to estimator (3) is the **Exponential Weighted Moving Average** (EWMA, proposed by J.P. Morgan), where weights decrease exponentially as we move backward in time:

$$\begin{aligned} \sigma_t^2 &= (1-\lambda)\sum_{i=1}^{\infty}\lambda^{i-1}y_{t-i}^2, \qquad 0<\lambda<1 \\ &= (1-\lambda)(y_{t-1}^2+\lambda y_{t-2}^2+\lambda^2 y_{t-3}^2+\ldots) \end{aligned} \tag{4}$$

  where λ is the decay factor: it determines the relative weights and the effective amount of data used in estimating volatility.

- The weights $\omega_i = (1-\lambda)\lambda^{i-1}$ sum up to 1:

$$\sum_{i=1}^{\infty}\omega_i = (1-\lambda)(1+\lambda+\lambda^2+\ldots) = 1.$$

Naive Models of Volatility

- From (4), we have

$$\sigma_{t-1}^2 = (1-\lambda)(y_{t-2}^2 + \lambda y_{t-3}^2 + \lambda^2 y_{t-4}^2 + \ldots) \tag{5}$$

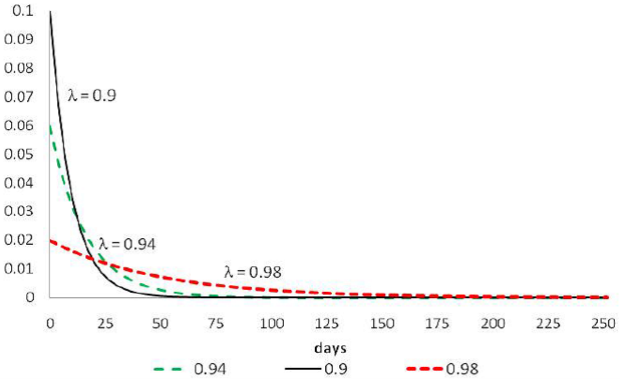
- Using (5), we can write (4) as:

$$\sigma_t^2 = (1-\lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2 \tag{6}$$

  that is, today's volatility can be estimated as the average of yesterday's variance and yesterday's squared returns.

- Model (6) is also known as the **RiskMetrics** model.

Naive Models of Volatility



Conditional Heteroskedasticity

Equation (1) suggests the specification:

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t \xi_t \tag{7}$$

- Here the conditional variance of ε_t is not constant.

Assume the following:

- H1: $\xi_t \sim iid(0, 1)$
- H2: ξ_t is independent of ε_{t-k}, $k \in \mathbb{N}$
- H3: σ_t is $\mathcal{F}_{t-1}-$ mensurable, where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$, i.e., σ_t^2 only depends of observable variables in moment $t-1$.

Conditional Heteroskedasticity

Multiplicative stochastic processes of the form

$$\varepsilon_t = \sigma_t \xi_t \tag{8}$$

are heteroskedastic processes as the variance is time-varying.

Heteroskedastic processes:

1. Traditional case: Assume that $\sigma_t = \alpha z_{t-1}$, where $z_{t-1} > 0$ is an exogenous variable.
2. Conditional heteroskedastic case: To acommodate volatility clustering, we need a model allowing that when ε_{t-1}^2 is large (small), on average, then ε_t^2 will also be of large (small) size. For example, in the simplest case we could have

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2, \quad \omega > 0, \alpha_1 \geq 0. \tag{9}$$

Here volatility σ_t^2 is assumed to be a random variable function of random past squared shocks determined by the model.

Conditional Heteroskedasticity

Assume (8) and assumptions H1-H3. In addition, assume H4: $\xi_t \sim \mathcal{N}(0,1)$.

Then,

$$\begin{aligned} \mathrm{E}(\varepsilon_t) &= 0 \\ \mathrm{Var}(\varepsilon_t) &= \mathrm{E}(\varepsilon_t^2) = \mathrm{E}(\sigma_t^2) \\ \mathrm{E}(\varepsilon_t^3) &= 0 \Longrightarrow \mathit{Skewness}_\varepsilon = 0 \end{aligned}$$

and since

$$\mathit{Kurtosis}_\varepsilon = \frac{\mathrm{E}(\varepsilon_t^4)}{\mathrm{E}(\varepsilon_t^2)^2} > 3$$

the marginal distribution of ε_t is leptokurtic.

Remark: The random variable ξ_t does not have to be normal. It can have a fat-tailed distribution, such as the Student's-t distribution.

Autoregressive Conditional Heteroskedasticity

The autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) was the first model to provide a systematic framework for modelling volatility.

Main idea: To assess the Friedman (1977) hypothesis that the unpredictability of inflation was a primary cause of business cycles. The uncertainty due to this unpredictability would affect the investor's behaviour and could change over time.

Note: First application to macro data (wage-price equation), not to financial data.

Autoregressive Conditional Heteroskedasticity

Model: Let the returns be

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mathsf{E}(y_t|\mathcal{F}_{t-1}) \\ \varepsilon_t &= \sigma_t \xi_t \end{aligned}$$

and assume H1-H4. Then,

$$\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0 \quad \text{and} \quad \mathsf{Var}(\varepsilon_t|\mathcal{F}_{t-1}) = \sigma_t^2.$$

The following conditional variance defines an ARCH model of order q, i.e., ARCH(q):

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2$$

Necessary and sufficient conditions for the positivity of the conditional variance: $\omega > 0$, $\alpha_j \geq 0$, $i = 1, \ldots, q-1$, and $\alpha_q > 0$.

Autoregressive Conditional Heteroskedasticity

Consider now the ARCH(1) model:

$$
\begin{aligned}
& y_t = \varepsilon_t, \qquad \varepsilon_t = \sigma_t \xi_t, \\
& \varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2), \\
& \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2
\end{aligned}
$$

For simplicity, write $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) \equiv \mathsf{E}_{t-1}(\varepsilon_t) = 0$ and $\mathsf{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) \equiv \mathsf{E}_{t-1}(\varepsilon_t^2) = \sigma_t^2$.

Remarks:

- There is no source of randomness independent of ξ_t in the variance equation.
- At time t, the variable ε_{t-1} is observable!
- Because ε_{t-1} is observed at time t, given ω and α_1, the variance σ_t^2 is observed.

Autoregressive Conditional Heteroskedasticity

$$\mathsf{E}(\varepsilon_t) = \mathsf{E}(\sigma_t \xi_t) = \mathsf{E}(\mathsf{E}_{t-1}(\sigma_t \xi_t)) = \mathsf{E}(\sigma_t \mathsf{E}_{t-1}(\xi_t)) = 0$$

$$\begin{aligned} \mathsf{E}(\varepsilon_t^2) &= \mathsf{E}(\sigma_t^2 \xi_t^2) = \mathsf{E}(\mathsf{E}_{t-1}(\sigma_t^2 \xi_t^2)) \\ &= \mathsf{E}(\mathsf{E}_{t-1}(\sigma_t^2)\mathsf{E}_{t-1}(\xi_t^2)) \\ &= \mathsf{E}(\mathsf{E}_{t-1}(\sigma_t^2)1) = \mathsf{E}(\sigma_t^2) \end{aligned} \tag{10}$$

Notes:

- There is a random term ε_{t-1}^2 entering σ_t^2, but this term is known at time t.
- Thus, conditional on information available at time $t-1$, and given the parameters ω and α_1, the variance σ_t^2 is not a random variable anymore.

Autoregressive Conditional Heteroskedasticity

Key observations:

- Remember $\mathrm{E}(\varepsilon_t^2) = \mathrm{E}(\sigma_t^2)$ and $\mathrm{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2$.
- Standardized residuals $\varepsilon_t / \sqrt{\sigma_t^2}$ do not display time-varying variance.
- Squared standardized residuals $\varepsilon_t^2 / \sigma_t^2$ have mean equal to 1, but centralized standardized residuals $(\varepsilon_t^2 / \sigma_t^2 - 1)$ have mean equal to 0.

Properties of ARCH models:

- Unconditional mean of ε_t :

$$\mathsf{E}(\varepsilon_t) = \mathsf{E}(\mathsf{E}_{t-1}(\sigma_t \xi_t)) = \mathsf{E}(\sigma_t \mathsf{E}_{t-1}(\xi_t)) = 0 \tag{11}$$

- Unconditional variance of ε_t :

$$\mathsf{Var}(\varepsilon_t) = \mathsf{E}(\varepsilon_t^2) = \mathsf{E}(\mathsf{E}_{t-1}(\varepsilon_t^2)) = \mathsf{E}(\omega + \alpha_1 \varepsilon_{t-1}^2) = \omega + \alpha_1 \mathsf{E}(\varepsilon_{t-1}^2) \tag{12}$$

  Because ε_t is a stationary process with $\mathsf{E}(\varepsilon_t) = 0$ and $\mathsf{Var}(\varepsilon_t) = \mathsf{Var}(\varepsilon_{t-1}) = \mathsf{E}(\varepsilon_{t-1}^2)$, we have

$$\mathsf{Var}(\varepsilon_t) = \frac{\omega}{1 - \alpha_1}$$

  from (12).

- Since ε_t is a martingale difference sequence (MDS), it is an uncorrelated process: $\mathsf{E}(\varepsilon_t \varepsilon_{t-j}) = 0$, for $j = 1, 2, \ldots$
- Since $\mathsf{Var}(\varepsilon_t)$ must be positive we need $0 \leq \alpha_1 < 1$.

Properties of ARCH models:

- If $\varepsilon_t \sim \mathrm{ARCH}(1)$, then ε_t^2 has a stationary $\mathrm{AR}(1)$ representation. Let

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2$$

  then

$$\begin{aligned} \sigma_t^2 + \varepsilon_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \varepsilon_t^2 \\ \varepsilon_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \varepsilon_t^2 - \sigma_t^2 \\ \varepsilon_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + v_t, \end{aligned} \qquad (13)$$

  where $v_t = \varepsilon_t^2 - \sigma_t^2$.
- Since

$$\mathrm{E}(v_t|\mathcal{F}_{t-1}) = \mathrm{E}(\varepsilon_t^2 - \sigma_t^2|\mathcal{F}_{t-1}) = \mathrm{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) - \sigma_t^2 = 0,$$

  then $v_t \sim$ MDS, and therefore an uncorrelated process.
- Note: If $\varepsilon_t \sim \mathrm{ARCH}(q)$, then ε_t^2 has a stationary $\mathrm{AR}(q)$ representation.

Properties of ARCH models:

- ε_t^2 exhibits volatility mean reversion. Assume that $\varepsilon_t \sim \text{ARCH}(1)$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \tag{14}$$

- From (10) we have

$$\mathsf{E}(\varepsilon_t^2) = \mathsf{E}(\sigma_t^2) = \overline{\sigma}^2 = \frac{\omega}{1-\alpha_1} \Longrightarrow \overline{\sigma}^2(1-\alpha_1) = \omega \tag{15}$$

Using (13) and (15) we get

$$\varepsilon_t^2 - \overline{\sigma}^2 = \alpha_1(\varepsilon_{t-1}^2 - \overline{\sigma}^2) + v_t.$$

Thus,

$$\mathsf{E}(\varepsilon_{t+k}^2 | \mathcal{F}_{t-1}) - \overline{\sigma}^2 = \alpha_1^k(\varepsilon_{t-1}^2 - \overline{\sigma}^2) \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty.$$

Properties of ARCH models:

- For the ARCH(1) model we have:

$$\mathrm{E}(\varepsilon_t^4) = 3\mathrm{E}(\sigma_t^4)$$
$$\mathrm{E}(\sigma_t^4) = \frac{\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

Therefore, the ARCH(1) process has finite fourth moment (vol of vol) iff $1 - 3\alpha_1^2 > 0$, i.e., $0 \leq \alpha_1^2 < 1/3$.

- The unconditional kurtosis of ε_t is

$$k_\varepsilon = \frac{\mathrm{E}(\varepsilon_t^4)}{(\mathrm{Var}(\varepsilon_t^2))^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} = 3 + \frac{6\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Autoregressive Conditional Heteroskedasticity

Testing against ARCH effects:

1. Fit the conditional mean to remove the linear dependence in the data. A simple AR model may be needed for some daily return series. Assuming the conditional mean equal to zero may be an acceptable approximation for daily data

2. Test the residuals for a higher-order dependence.

   ▷ Two tests are available for testing against conditional heteroskedasticity: McLeod and Li (1983) (Ljung-Box statistics of ε_t^2) and the Engle (1982) (Lagrange multiplier test).

Autoregressive Conditional Heteroskedasticity

- Engle (1982) test:

1: Calculate $\widehat{\varepsilon}_t = y_t - \widehat{\mu}_t$

2: Regress $\widehat{\varepsilon}_t^2$ on q lags of itself

$$\widehat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \widehat{\varepsilon}_{t-1}^2 + \alpha_2 \widehat{\varepsilon}_{t-2}^2 + \ldots + \alpha_p \widehat{\varepsilon}_{t-q}^2 + u_t \tag{16}$$

3: Under $\mathrm{H}_0 : \alpha_1 = \ldots = \alpha_q = 0$, the test statistic

$$LM = TR^2 \overset{asy}{\sim} \chi^2(q)$$

where R^2 is computed from regression (16) and T is the number of observations.

4: If the null hypothesis is rejected, then there are ARCH effects.