Univariate non-linear time series models: ARCH, GARCH processes

3.1 Some 'stylized facts' of financial data

Returns typically have the following properties:

1. The returns' distribution has heavier tails than the normal distribution, i.e. returns of large absolute value happen much more often than implied by a normal distribution ('fat tails').
2. The returns' distribution is leptokurtic, i.e. $\kappa := \frac{E(X-E(X))^4}{\mathrm{Var}(X)^2} = E\left(\frac{X-EX}{\sqrt{\mathrm{Var}(X)}}\right)^4$ exceeds 3.
3. Returns are nearly uncorrelated, but there is significantly positive autocorrelation in absolute or squared returns.
4. Returns large in absolute value are often followed by returns of large absolute value, while absolutely small returns tend to be followed by small absolute values $\rightsquigarrow$ volatility clustering.

3.2 ARCH processes

3.2.1 ARCH(1) processes

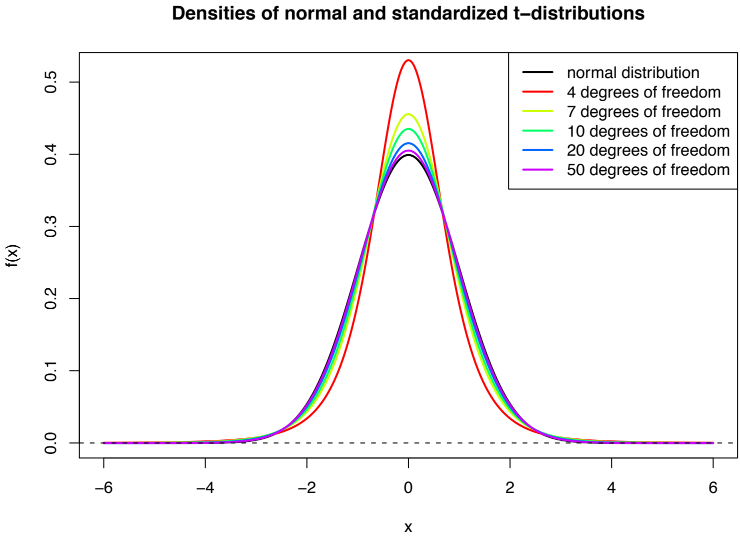
Definition (ARCH(1) process)

1. We call a stationary process X an autoregressive conditionally heteroscedastic process of order 1 or ARCH(1) process if there exist an independent white noise ε of unit variance $\sigma_\varepsilon^2 = 1$ and real numbers $\alpha_0 > 0, \alpha_1 \geq 0$ such that $X_t = \sigma_t \varepsilon_t$ with $\sigma_t^2 := \alpha_0 + \alpha_1 X_{t-1}^2$.
2. In this case, we call σ_t volatility and ε_t innovation (at time t).

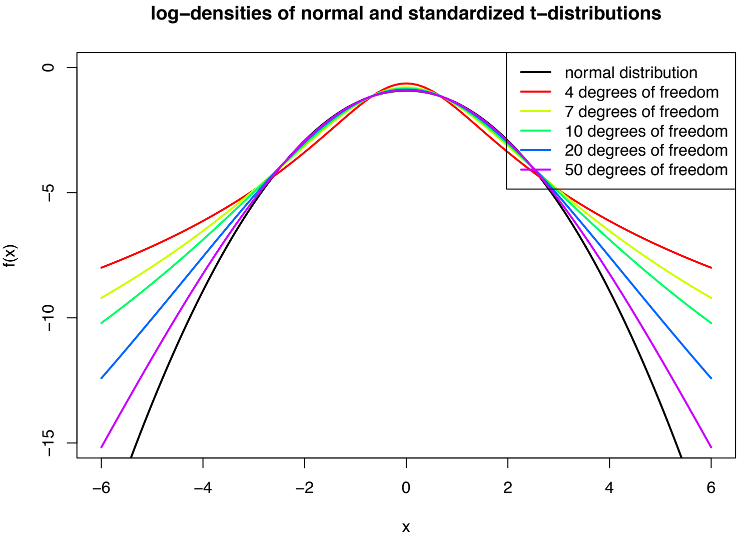
Remarks I

1. Often, the innovations' distribution is also specified: the easiest case is to assume normal innovations, ε is then a Gaussian white noise. Sometimes, however, one needs ε_t to be heavy-tailed: in this case, one often uses t_ν, the t-distribution with $\nu > 2$ degrees of freedom: $\varepsilon_t \sim \sqrt{\frac{\nu-2}{\nu}} t_\nu$.
2. A necessary and essentially also sufficient condition for the existence of an ARCH(1) process is $E \ln(\alpha_1 \varepsilon_1^2) < 0$.
3. Because the variance of X is finite only if $\alpha_1 < 1$, we usually restrict ourselves to $\alpha_1 < 1$.
4. An ARCH(1) process X is a white noise with variance $\sigma_X^2 = \frac{\alpha_0}{1-\alpha_1}$.

Density plots for t-distributions I



Density plots for t-distributions II



Remarks II

5. For an ARCH(1) process X, we have:
   - $E(X_t|X_{t-1}, X_{t-2}, \ldots) = 0$,
   - $E(X_t^2|X_{t-1}, X_{t-2}, \ldots) = \sigma_t^2$
   - for $\mathrm{Var}(X_t|X_{t-1}, X_{t-2}, \ldots)$, defined as $E((X_t - E(X_t|X_{t-1}, X_{t-2}, \ldots))^2|X_{t-1}, X_{t-2}, \ldots)$: $\mathrm{Var}(X_t|X_{t-1}, X_{t-2}, \ldots) = \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$, which explains the name ARCH.
6. Under suitable assumptions, the square X^2 of an ARCH(1) process an AR(1) process with mean $\frac{\alpha_0}{1-\alpha_1}$.
7. An ARCH(1) process X with normal innovations ε_t has finite fourth moments if and only if $\alpha_1 < \frac{1}{3}\sqrt{3}$. In this case, its kurtosis κ_X equals $\kappa_X = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2}$.

Remarks III

8. ARCH(1) processes have heavy tails, even if the innovations do not.
9. ARCH(1) processes display volatility clustering.
10. For $0 \leq \alpha_1 < 1$, we have $\sigma_t^2 = \alpha_0(1 + \sum_{k=1}^{\infty} \alpha_1^k \prod_{j=1}^{k} \varepsilon_{t-j}^2)$ and
$$X_t = \sigma_t \varepsilon_t = \varepsilon_t \sqrt{\alpha_0(1 + \sum_{k=1}^{\infty} \alpha_1^k \prod_{j=1}^{k} \varepsilon_{t-j}^2)}.$$

Estimating ARCH(1) models

1. One possibility: making use of the fact that the squared process X_t^2 is an AR(1) process with mean $\frac{\alpha_0}{1-\alpha_1}$:
   $X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + Z_t$ with the white noise $Z_t := X_t^2 - \sigma_t^2$.
   The problem is then reduced to the estimation of an AR model which for instance can be done using the Yule-Walker estimator. Under mild assumptions, this approach delivers consistent and asymptotically normal estimators which however are not efficient. Nonetheless, these estimators may be useful as start values for an efficient estimator that has to be computed iteratively.
2. Under certain conditions, asymptotically normal and efficient estimators are given by maximum likelihood.

Computing the ML estimator I

The likelihood function $L(\alpha_0, \alpha_1 | x_1, \ldots, x_n)$ given realisations $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$ can be written as a product of the unconditional density of X_1 and the conditional densities of X_j given $X_1, \ldots, X_{j-1}$:

$$L(\alpha_0, \alpha_1 | x_1, \ldots, x_n) = f_{X_1}(x_1 | \alpha_0, \alpha_1) \prod_{j=2}^{n} f_{X_j | X_1 = x_1, \ldots, X_{j-1} = x_{j-1}}(x_j | \alpha_0, \alpha_1)$$

As $X_j = \sigma_j \varepsilon_j$ and $\sigma_j^2 = \alpha_0 + \alpha_1 X_{j-1}^2$, we have:

$$f_{X_j | X_1 = x_1, \ldots, X_{j-1} = x_{j-1}}(x_j | \alpha_0, \alpha_1) = \frac{1}{\sigma_j} f_\varepsilon\left(\frac{x_j}{\sigma_j}\right),$$

with $\sigma_j := \sqrt{\alpha_0 + \alpha_1 x_{j-1}^2}$ and f_ε the density of the innovations ε.

Computing the ML estimator II

Substituting this into the likelihood function, we can simplify:

$$L(\alpha_0, \alpha_1 | x_1, \ldots, x_n) = f_{X_1}(x_1 | \alpha_0, \alpha_1) \prod_{j=2}^{n} \left(\frac{1}{\sigma_j} f_\varepsilon\left(\frac{x_j}{\sigma_j}\right)\right)$$

The log-likelihood function is then

$$l(\alpha_0, \alpha_1 | x_1, \ldots, x_n) = \ln f_{X_1}(x_1 | \alpha_0, \alpha_1) + \sum_{j=2}^{n} \left(\ln f_\varepsilon\left(\frac{x_j}{\sigma_j}\right) - \ln \sigma_j\right) .$$

Even for normal innovations ε, the distribution of X_t is not known in exact form such that the term $\ln f_{X_1}(x_1 | \alpha_0, \alpha_1)$ can not be given explicitly. For large n, this term is typically small as compared to the rest, which is the reason why it is simply neglected.

Computing the ML estimator III

This approach is called conditional ML, maximising

$$l^c(\alpha_0, \alpha_1 | x_1, \dots, x_n) = \sum_{j=2}^{n} (\ln f_\varepsilon(\frac{x_j}{\sigma_j}) - \ln \sigma_j) \ .$$

After specifying the innovations' density f_ε, the actual calculation is done iteratively.

- If the model is correctly specified, that is if the (unknown) true distribution equals the specified one, then the ML estimator is asymptotically efficient.
- If, however, the innovations' distribution is misspecified, then the ML estimator typically is no longer consistent, though asymptotically normal.

Computing the ML estimator IV

If, however, the standard normal distribution is used as the innovations' distribution, then the estimator will be consistent and asymptotically normal, even if the innovations' true distribution is not normal. The corresponding estimator is then called quasi maximum likelihood or QML estimator. It is however less efficient than the ML estimator computed for the correctly specified innovations' distribution.

Remark

To guarantee the existence of higher moments, α_1 may not be chosen too large. For instance, the existence of fourth moments, together with normal innovations, implies $\alpha_1 < \frac{1}{3}\sqrt{3}$. Under this condition, on the other hand, the acf of the squared process X_t^2 will decay rapidly, whereas financial data typically exhibit an only slowly decaying acf of squared returns. A first way out of this dilemma is given by ARCH(q) processes.

3.2.2 ARCH(q) processes

Definition (ARCH(q) process)

1. We call a stationary process X an autoregressive conditionally heteroscedastic process of order q or ARCH(q) process if there exist an independent white noise ε of unit variance $\sigma_\varepsilon^2 = 1$ and real numbers $\alpha_0 > 0, \alpha_1 \geq 0, \dots, \alpha_q \geq 0$ such that $X_t = \sigma_t \varepsilon_t$ with $\sigma_t^2 := \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_q X_{t-q}^2$.
2. In this case, we call σ_t volatility and ε_t innovation (at time t).

Remarks I

1. Regarding the innovations' distribution, the same holds true as for ARCH(1) processes.
2. ARCH processes were introduced in 1982 by Robert F. Engle who was awarded with the Nobel prize in economics in 2003.
3. In 1992, Bougerol and Picard established both necessary and sufficient conditions on the innovations' distribution and the ARCH parameters for the existence of ARCH processes. For these, we refer to the literature.
4. As the variance of X is only finite if $\alpha_1 + \ldots + \alpha_q < 1$, we usually consider only this case.

Remarks II

5. An ARCH(q) process X is a white noise with variance $\sigma_X^2 = \frac{\alpha_0}{1-\alpha_1-\ldots-\alpha_q}$.
6. For an ARCH(q) process X, we have: the conditional variance of X_t given the history of X is $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_q X_{t-q}^2$, explaining the name ARCH(q).
7. Under suitable conditions, the square X^2 of an ARCH(q) process is an AR(q) process with mean $\frac{\alpha_0}{1-\alpha_1-\ldots-\alpha_q}$:
   $X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_q X_{t-q}^2 + Z_t$ with the white noise $Z_t := X_t^2 - \sigma_t^2$.

Remarks III

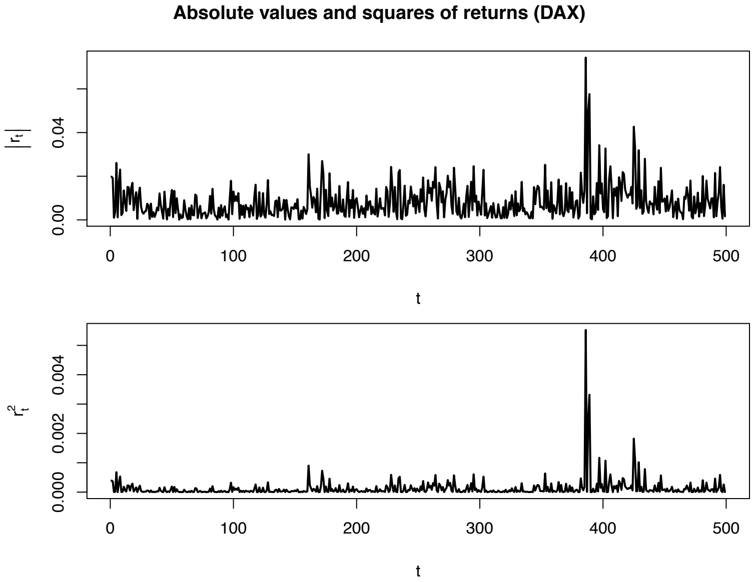
8. For the fourth moment of an ARCH(q) process X with normal innovations ε_t to be finite, we must have $\alpha_1^2 + \ldots + \alpha_q^2 < \frac{1}{3}$.
9. ARCH(q) processes have heavy tails, even if the innovations do not.
10. ARCH(q) processes display volatility clustering.

Testing for 'ARCH effects'

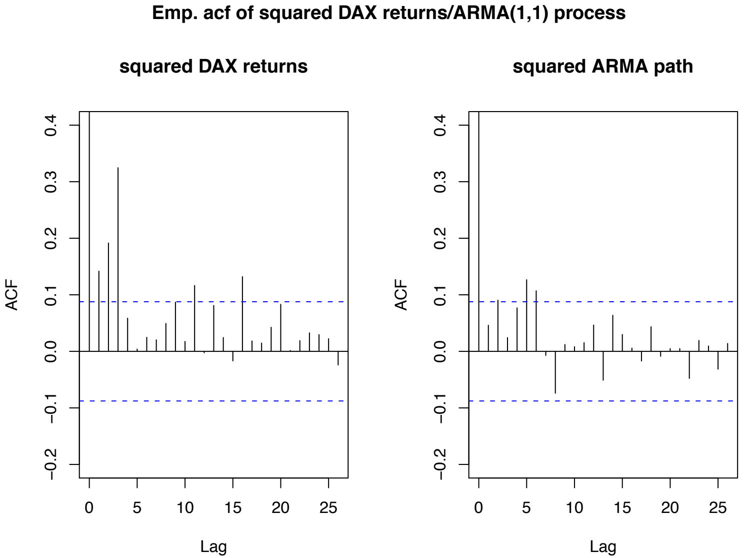
Adressing the question whether a given white noise displays 'ARCH effects':
if we want to test the null hypothesis of an independent white noise against the alternative of 'ARCH effects' being present, there are several asymptotically equivalent alternatives:

1. Portmanteau tests applied to the squared time series
2. regressing the squares on a constant and lagged squares of the time series: here we may either make use of the usual F-statistic or of the approximately χ^2-distributed product nR^2, where R^2 is the coefficient of determination.

Absolute values/squares of DAX returns



ACF of squared DAX returns



Estimation of ARCH(q) models I

The estimation of ARCH(q) models poses essentially the same problems as that of ARCH(1) models, the only differences being the larger number of parameters and the fact that the conditional log-likelihood now differs from the exact log-likelihood by q summands.

Additionally, q might have to be determined by the data. To this end, there exist several approaches:

1. As the squared process is an AR(q) process, q might be chosen by looking at the squared process' pacf: one chooses the largest q for which the estimated pacf of the squared process is significantly different from 0.
2. q might also be estimated by using model selection criteria.

Estimation of ARCH(q) models II

For financial returns, the estimation of ARCH(q) models often leads to quite large values of q which is undesirable for the following reasons:

1. The number of parameters to be estimated is quite large which is numerically demanding.
2. Frequently, parameter estimates are obtained that are not in accord with the conditions for stationarity or existence of moments.
3. A too large number of parameters might lead to the phenomenon of 'overfitting'.

To get rid of these problems, one may use the GARCH model (<u>G</u>eneralized ARCH) which was introduced in 1986 by Bollerslev and also - under a different name - by Taylor.

3.3 GARCH processes

3.3.1 GARCH$(1,1)$ processes

Definition (GARCH$(1,1)$ process)

1. We call a stationary process X a generalized autoregressive conditionally heteroscedastic process of order $(1,1)$ or GARCH$(1,1)$ process if there exist an independent white noise ε of unit variance $\sigma_\varepsilon^2 = 1$ and real numbers $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ such that there exists a stochastic process σ_t with $X_t = \sigma_t \varepsilon_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.
2. In this case, we call σ_t volatility and ε_t innovation (at time t).

Remarks I

1. Regarding the innovations' distribution, the same holds true as for ARCH(1) and ARCH(q) processes.
2. A necessary and essentially also sufficient condition for the existence of a GARCH$(1,1)$ process is given by $E \ln(\alpha_1 \varepsilon_1^2 + \beta_1) < 0$.
3. As the variance of X is finite if and only if $\alpha_1 + \beta_1 < 1$, one usually considers only the case $\alpha_1 + \beta_1 < 1$.
4. A GARCH$(1,1)$ process X is a white noise with variance $\sigma_X^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$.

Remarks II

5. For a GARCH$(1,1)$ process X, we have $E(X_t|X_{t-1}, X_{t-2}, \dots) = 0$ and $\mathrm{Var}(X_t|X_{t-1}, X_{t-2}, \dots) = \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.
6. If X has finite fourth moments, then the square X^2 is an ARMA$(1,1)$ process with mean $\frac{\alpha_0}{1-\alpha_1-\beta_1}$: $X_t^2 = \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 + Z_t - \beta_1 Z_{t-1}$ with the white noise $Z_t := X_t^2 - \sigma_t^2$.
7. For a GARCH$(1,1)$ process X with normal innovations, the fourth moment is finite if and only if $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. In this case, the kurtosis κ_X of X equals $\kappa_X = 3\frac{1-(\alpha_1+\beta_1)^2}{1-3\alpha_1^2-2\alpha_1\beta_1-\beta_1^2}$.
8. GARCH$(1,1)$ processes have heavy tails, even if the innovations do not.

Remarks III

9. GARCH$(1, 1)$ processes display volatility clustering.
10. GARCH$(1, 1)$ models are usually estimated by ML or QML. In contrast to the estimation of ARCH models, we have additionally the difficulty that σ_t^2 depends on σ_{t-1}^2, necessitating start values for σ_0^2 (and X_0^2). Mostly, one chooses for these the variance of the data which is an estimator of the unconditional mean $EX_t^2 = E\sigma_t^2$.
11. GARCH$(1, 1)$ models are widely used in practice, as on the one hand, there are only a few parameters to be estimated, and on the other hand, they typically deliver a good fit to the data. However, there exists also a generalization of GARCH$(1, 1)$ processes, namely GARCH(p, q) processes.

3.3.2 GARCH(p, q) processes

Definition (GARCH(p, q) process)

1. We call a stationary process X a generalized autoregressive conditionally heteroscedastic process of order $(1, 1)$ or GARCH$(1, 1)$ process if there exist an independent white noise ε of unit variance $\sigma_\varepsilon^2 = 1$ and real numbers $\alpha_0 > 0, \alpha_1 \geq 0, \ldots, \alpha_q \geq 0$, $\beta_1 \geq 0, \ldots, \beta_p \geq 0$ such that there exists a stochastic process σ_t with $X_t = \sigma_t \varepsilon_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_q X_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2$.
2. In this case, we call σ_t volatility and ε_t innovation (at time t).

Remarks I

1. Regarding the innovations' distribution, the same holds true as for ARCH(1) and ARCH(q) processes.
2. For GARCH(p, q) processes, one tacitly assumes that at least one of the coefficients $\alpha_1, \ldots, \alpha_q$ is positive, because otherwise the volatility proces σ^2 does not depend on X^2.
3. In 1992, Bougerol and Picard derived necessary and sufficient conditions for the existence of GARCH(p, q) processes.
4. As the variance of X is finite if and only if $\alpha_1 + \ldots + \alpha_q + \beta_1 + \ldots + \beta_p < 1$, we usually restrict ourselves to the case $\alpha_1 + \ldots + \alpha_q + \beta_1 + \ldots + \beta_p < 1$.

Remarks II

5. In the literature, the roles of p and q are sometimes reversed such that occasionally a GARCH(p, q) process is to be understood as a GARCH(q, p) process according to the above definition.
6. In 1992, Nelson and Cao showed that the non-negativity of the coefficients $\alpha_1, \ldots, \alpha_q$ and $\beta_1, \ldots, \beta_p$ is only sufficient for positive σ_t^2 but not necessary. They were able to develop necessary and sufficient conditions for positive σ_t^2 for GARCH$(1, q)$- and GARCH$(2, q)$ processes.
7. A GARCH(p, q) process X is a white noise with variance $\sigma_X^2 = \frac{\alpha_0}{1-\alpha_1-\ldots-\alpha_q-\beta_1-\ldots-\beta_p}$.

Remarks III

8. For a GARCH(p, q) process X, we have $E(X_t|X_{t-1}, X_{t-2}, \ldots) = 0$ and $\mathrm{Var}(X_t|X_{t-1}, X_{t-2}, \ldots) = \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_q X_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2$.
9. If X has finite fourth moments, then the squared process X^2 is an ARMA$(\max(p, q), p)$ process with mean $\frac{\alpha_0}{1-\alpha_1-\ldots-\alpha_q-\beta_1-\ldots-\beta_p}$: $X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_q X_{t-q}^2 + \beta_1 X_{t-1}^2 + \ldots + \beta_p X_{t-p}^2$ $+ Z_t - \beta_1 Z_{t-1} - \ldots - \beta_p Z_{t-p}$ with the white noise $Z_t := X_t^2 - \sigma_t^2$.
10. Necessary and sufficient conditions for the existence of moments of X were derived by Ling and McAleer in 2002.
11. GARCH(p, q) processes have heavy tails, even if the innovations do not.

Remarks IV

12. GARCH(p, q) processes display volatility clustering.
13. The estimation of GARCH(p, q) models is similar to that of GARCH($1, 1$) models: they are usually estimated using ML or QML, we again have the need for start values for both the process itself and the volatility process: $X_0^2, \dots, X_{1-q}^2$ and $\sigma_0^2, \dots, \sigma_{1-p}^2$ have to be chosen. Again, the usual approach is to that these equal to the sample variance of the data.
14. In practice, the by far most prominent model is the GARCH($1, 1$) model, as the following results for Google searches for GARCH(p, q) reveal (Year 2008; in brackets: years 2007, 2006, 2005):

p/q	1	2	3
1	ca. 149000 (125000, 106000, 2590)	5580 (585, 241, 74)	271 (139, 34, 14)
2	1170 (404, 235, 85)	1160 (427, 247, 55)	71 (41, 20, 8)
3	95 (41, 25, 3)	153 (30, 19, 9)	118 (52, 34, 5)

GARCH$(1,1)$ estimation (short example)

```
> DAXgarch <- garch(dDAXlp,order=c(1,1),trace=FALSE)
> DAXgarch

Call:
garch(x = dDAXlp, order = c(1, 1), trace = FALSE)

Coefficient(s):
       a0          a1          b1  
6.153e-06  1.308e-01  8.257e-01  

> Box.test(residuals(DAXgarch)^2,lag=5)

        Box-Pierce test

data:  residuals(DAXgarch)^2
X-squared = 4.4986, df = 5, p-value = 0.4801
```

ACF of (squared) GARCH(1,1) residuals