- We are now going to investigate a class of models with quite different characteristics. They are used to model nonlinear dynamics for which the conditional excitation variance changes with time.
- One sub-class has been extensively used to model changes in volatility in financial time series.
- The other sub-class can be used to model natural phenomena with burst effects.

- Financial data (indices, stock values) generally present a high correlation at lag 1, i.e. values do not change much from one day to the next.
- This is why one generally works with the *returns* of a financial time series y_n, typically defined as:

$$x_n = \log \frac{y_n}{y_{n-1}} \quad \text{or} \quad x_n = \frac{y_n - y_{n-1}}{y_{n-1}}$$

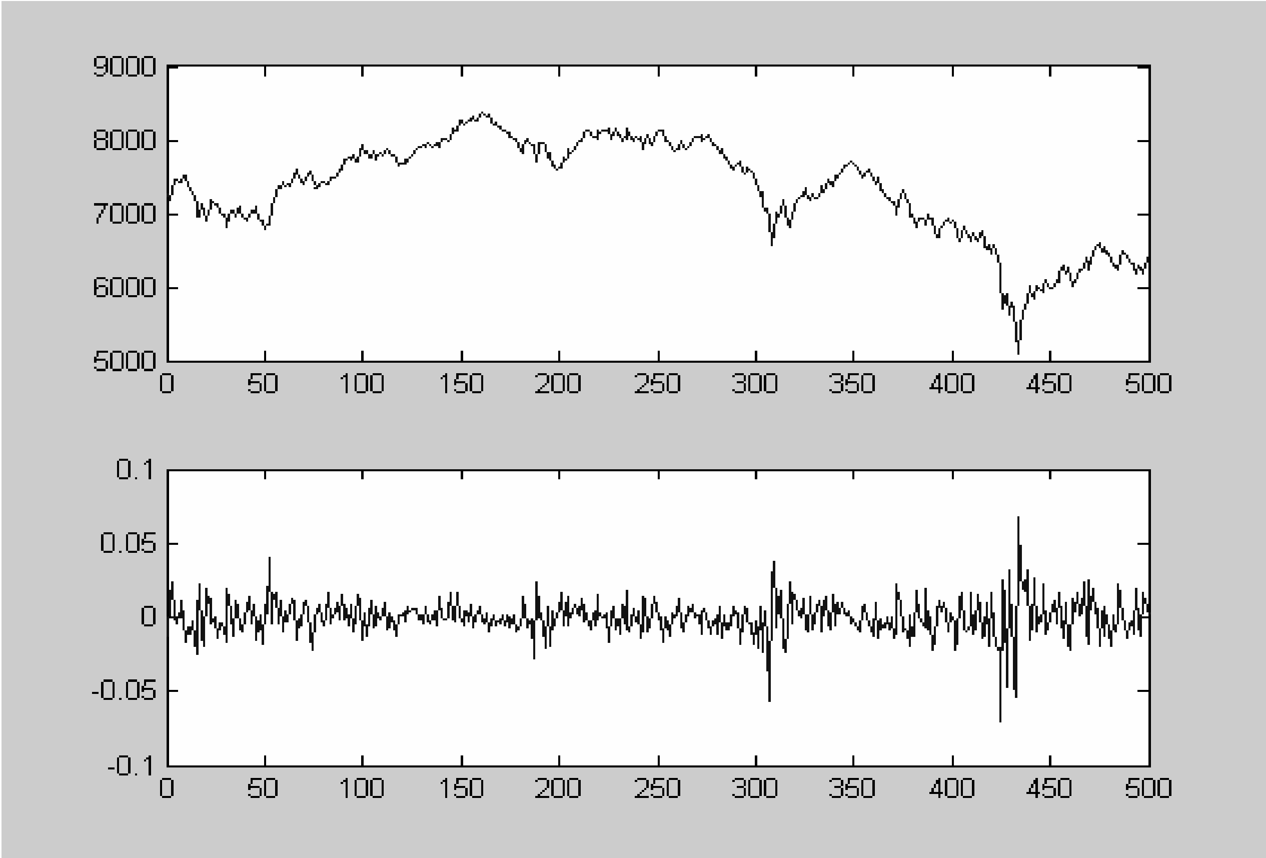
- Many financial returns share the following characteristics, which cannot be dealt with by linear models:
  - *Leptokurtosis*. Returns tend to have distributions that exhibit fat tails and excess peakedness near the man.
  - They tend to have a time-varying variance (*volatility*). The volatility often clusters, i.e. large returns (of either sign) are expected to follow large returns, and small returns (of either sign) to follow small returns.

– *Leverage effects*. The tendency fo volatility to rise more following a large price fall. It could be that traders react more to negative information than to positive one.
– *Long-range dependence*. The returns themselves usually show little correlation, while squared returns or absolute returns often show persistent autocorrelation.

- Example: 500 daily values of SMI

time series

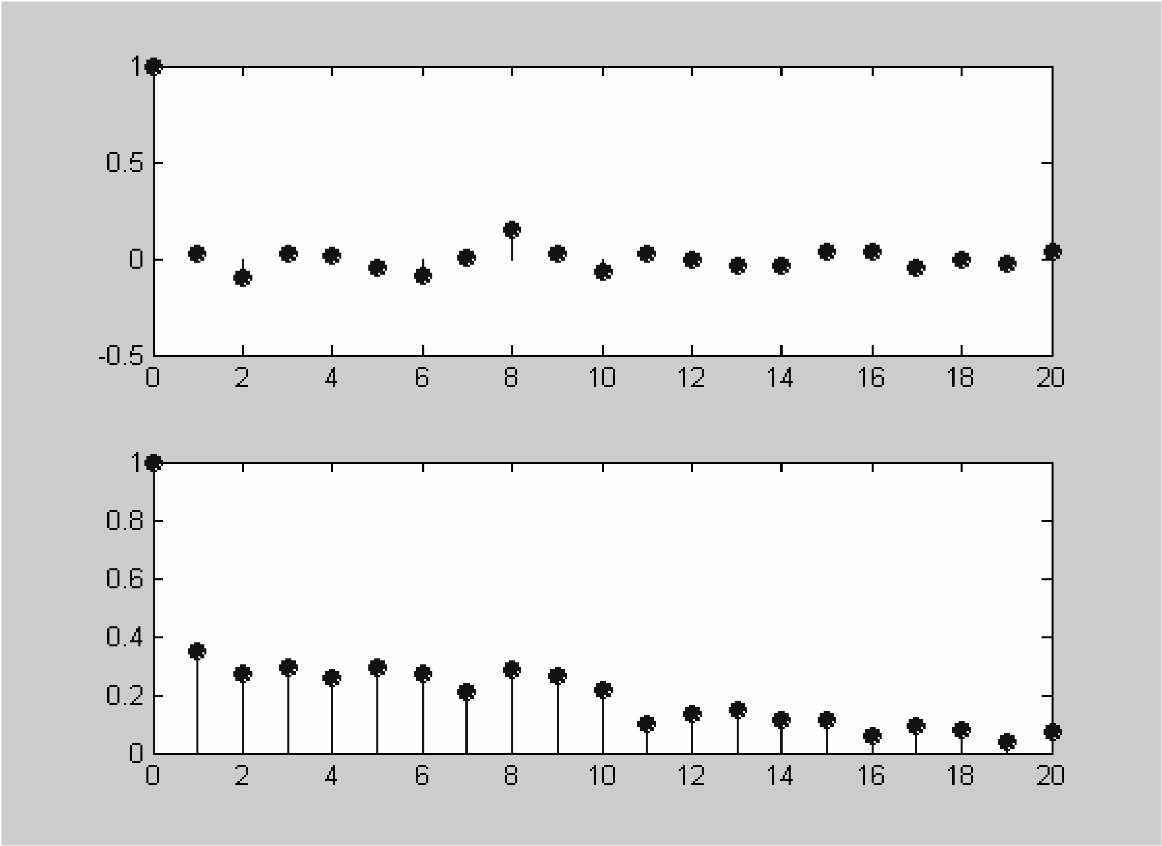
returns



- Normalized autocovariance estimates

returns

absolute returns



- We have seen that in the mean square sense one has:

$$x_n = \mathsf{E}[x_n|\mathbf{\Omega}_{n\text{-}1}] + \varepsilon_n$$

Where $\mathbf{\Omega}_{n\text{-}1}$ conveys all the information up to time n-1, and the innovation supposed to be white, conditionally and unconditionally homoscedastic, that is:

$$\mathsf{E}[\varepsilon_n^2 \mid \Omega_{n-1}] = \mathsf{E}[\varepsilon_n^2] = \sigma^2$$

- A convenient way to express conditional heteroscedasticity is:

$$\varepsilon_n = \omega_n \sqrt{h_n} \qquad \text{with} \quad h_n = \mathsf{E}[\varepsilon_n^2 \mid \Omega_{n-1}]$$

where ω_n is an i.i.d. Gaussian sequence with zero mean and unit variance, and h_n some function of $\mathbf{\Omega}_{n\text{-}1}$.

- In this way, the distribution of ε_n conditional on $\mathbf{\Omega}_{n\text{-}1}$ is Gaussian with zero mean and variance h_n.

- It is to be noted that the *unconditional* variance is still constant if $\mathsf{E}[h_n]$ is constant since:

$$\sigma^2 = \mathsf{E}[\varepsilon_n^2] = \mathsf{E}\left[\mathsf{E}[\varepsilon_n^2 \mid \Omega_{n-1}]\right] = \mathsf{E}[h_n]$$

- In the basic ARCH model introduced by Engle, is a linear function of the squares of the past innovations, which gives for an ARCH(1) model:

$$h_n = \alpha_0 + \alpha_1 \varepsilon_{n-1}^2$$

ARCH MODEL (4)

- Of course, one needs to have $h_n \geq 0$. this will be guaranteed if $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$.
- This ARCH(1) model can be recast as an AR(1) model:

$$\varepsilon_n^2 = \alpha_0 + \alpha_1 \varepsilon_{n-1}^2 + \nu_n \quad \text{with } \nu_n = \varepsilon_n^2 - h_n = h_n(\omega_n^2 - 1)$$

Note that $\mathrm{E}[\nu_n | \Omega_{n-1}] = 0$, since ω_n is not correlated with h_n and has unit variance.

- This AR model will be stable for $\alpha_1 < 1$, and the unconditional variance of ε_n is:

$$\sigma^2 = \mathrm{E}[\varepsilon_n^2] = \frac{\alpha_0}{1-\alpha_1}$$

- This AR model can thus be rewritten as:

$$\varepsilon_n^2 = \sigma^2 + \alpha_1(\varepsilon_{n-1}^2 - \sigma^2) + \nu_n$$

which indicates that squared innovations > σ^2 will have a tendency to perpetuate, i.e. the ARCH model indeed produces volatility clustering.

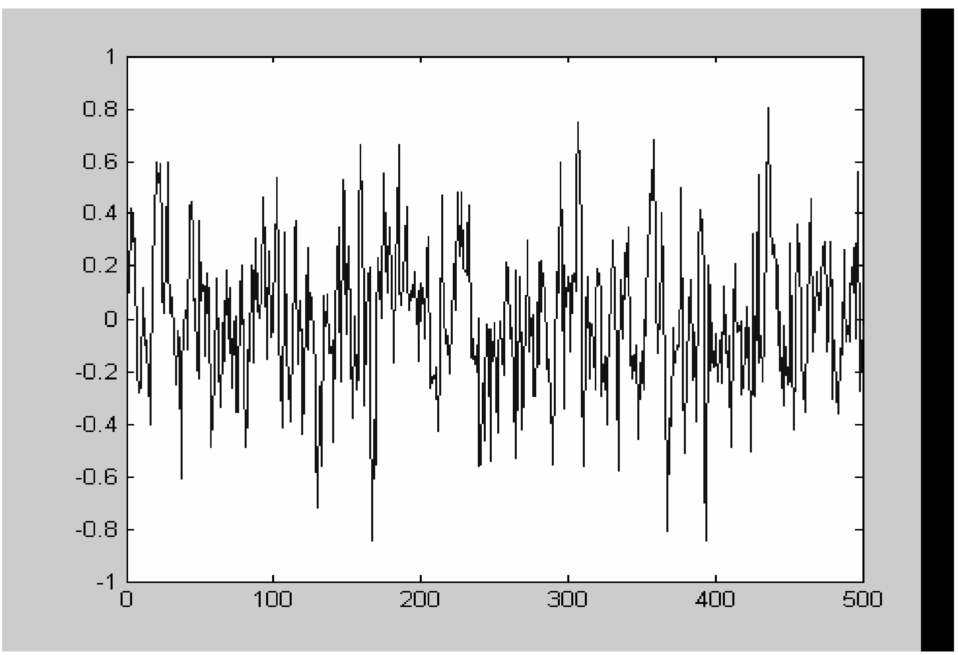
- Engle showed that the kurtosis of ε_n is given by:

$$\mathrm{K}_\varepsilon = \frac{\mathrm{E}[\varepsilon_n^4]}{\mathrm{E}[\varepsilon_n^2]^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2}$$

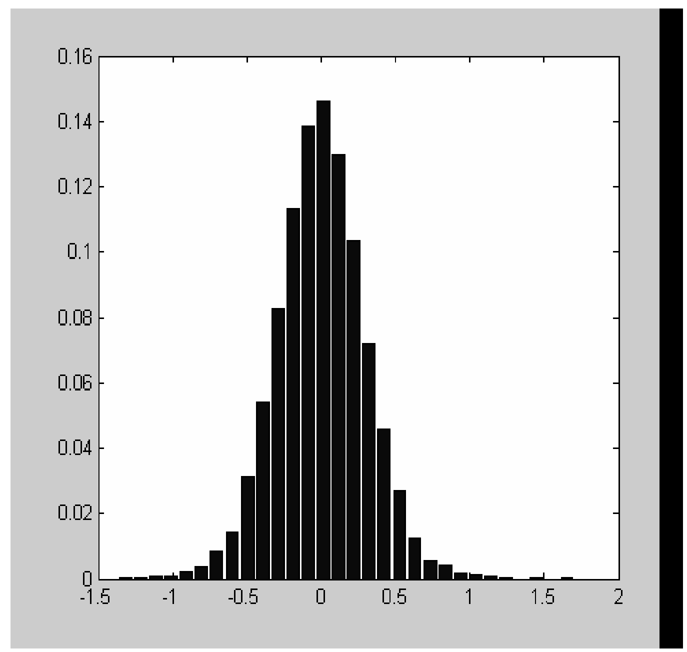
which makes sense if $3\alpha_1{}^2 < 1$. In that case, K_ε is larger than 3, which is the kurtosis for a Gaussian distribution. This property will reflect in the distribution of x_n.

- Example: $x_n = 0.5\, x_{n-1} + \varepsilon_n$, $\varepsilon_n \sim \mathrm{N}(0,h_n)$ with $h_n = 0.05 + 0.25\varepsilon_{n-1}{}^2$

x_n



histogram



- Of course it is possible to define an ARCH(q) model with:

$$h_n = \alpha_0 + \alpha_1 \varepsilon_{n-1}^2 + \alpha_2 \varepsilon_{n-2}^2 + \cdots + \alpha_q \varepsilon_{n-q}^2$$

in which $\alpha_i \geq 0$ for all i ensures positiveness of h_n.

- Having $q > 1$ permits to model better the persistent autocorrelation between the squared innovations ε_n^2. However, for financial time series, q must be very large.

- This is why Bollerslev proposed an enhanced model, the generalized ARCH (GARCH) one. The basic idea is to make the evolution on h_n autoregressive. This gives for a GARCH(1,1) model:

$$h_n = \alpha_0 + \alpha_1 \varepsilon_{n-1}^2 + \beta_1 h_{n-1}$$

- Although it is possible to consider higher order GARCH(p,q) models, this simple representation has been found adequate in many financial applications.

- Conditions $\alpha_{1,2} \geq 0$ and $\beta_1 \geq 0$ guarantee $h_n \geq 0$.
- Condition $\alpha_1 + \beta_1 < 1$ guarantees the model to be stable.
- The unconditional variance and kurtosis take the values:

$$\sigma^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1} \qquad \mathrm{K}_\varepsilon = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-2\alpha_1\beta_1-\beta_1^2}$$

- It may also be shown that the autocorrelation of the squared innovations ε_n^2 decrease like $(\alpha_1 + \beta_1)^k$. Thus, if $\alpha_1 + \beta_1$ is close to 1, this decrease is slow.
- Many extensions to the GARCH model have been proposed. For instance the GJR-GARCH model introduced by Glosten, Jagannathan and Runkle:

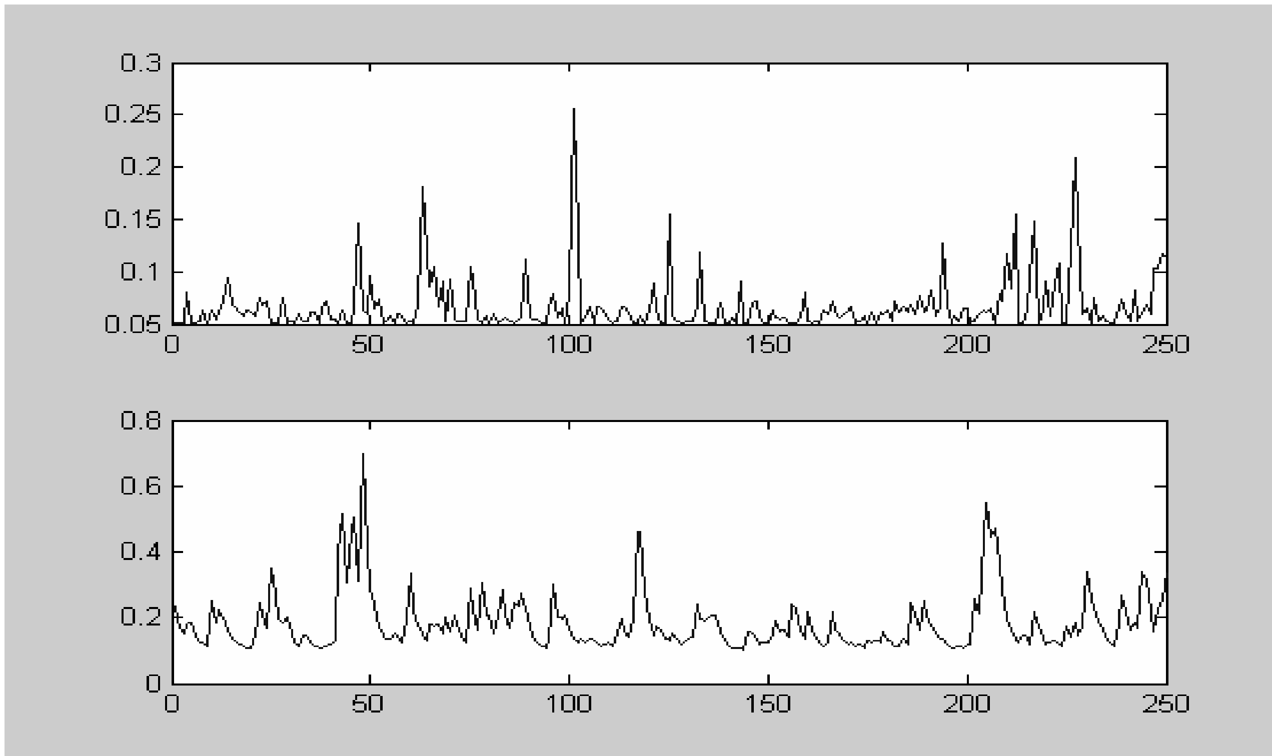
$$h_n = \alpha_0 + \alpha_1 \varepsilon_{n-1}^2 (1 - I[\varepsilon_{n-1} > 0]) + \gamma_1 \varepsilon_{n-1}^2 (1 - I[\varepsilon_{n-1} < 0] + \beta_1 h_{n-1}$$

with $I[.]$ the indicator function. It is a threshold model which allows for asymmetric effects in the volatility.

- Example: $x_n = 0.5\, x_{n\text{-}1} + \varepsilon_n, \quad \varepsilon_n \sim \mathrm{N}(0,h_n)$ with $h_n = 0.05 + 0.5\, h_{n\text{-}1} + 0.25\varepsilon_{n\text{-}1}^{\ 2}$

h_n for ARCH(1)

h_n for GARCH(1,1)



- The principle is simple. The innovations are estimated using some model:

$$\hat{\varepsilon}_n = x_n - G(\boldsymbol{x}_{n-1})$$

- And a test of hpyothesis is performed on the regression:

$$\hat{\varepsilon}_n^2 = a_0 + a_1\hat{\varepsilon}_{n-1}^2 + \cdots + a_m\hat{\varepsilon}_{n-m}^2 + u_n$$

to detect correlation in the squared estimates.

- Due to the specific nature of the innovation ε_n, it is not possible to apply least squares estimation, and one must resort to a maximum likelihood approach.
- If one hypothesizes a NAR general structure:

$$x_n = G(\boldsymbol{x}_{n-1};\phi) + \varepsilon_n$$

with ϕ the parameter vector defining $G(.)$, the total parameter vector is $\theta = [\phi\ \delta]$, δ being the parameters describing the conditional variance.

- Let us define $f(.)$ the probability density of the "core" innovations ω_n, which are i.i.d. with unit variance. Then the probability density of ε_n is:

$$\frac{1}{\sqrt{h_n}} f\left(\frac{\varepsilon_n}{\sqrt{h_n}}\right)$$

- So the log-likelihood for the nth observation will be:

$$l_n(\theta) = \ln f\left(\frac{\varepsilon_n}{\sqrt{h_n}}\right) - \ln\frac{1}{\sqrt{h_n}} = \ln f\left(\frac{x_n - G(\boldsymbol{x}_{n-1})}{\sqrt{h_n}}\right) - \ln\frac{1}{\sqrt{h_n}}$$

- So, if $f(.)$ is the Gaussian density:

$$l_n(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln h_n - \frac{\varepsilon_n^2}{2h_n}$$

- Then the maximum likelihood of θ is obtained by maximizing:

$$\sum_{n=1}^{N} l_n(\theta)$$

i.e. the sum of the log likelihoods on all available samples. It is a nonlinear problem that must be solved using iterative optimization techniques.

- We will examine briefly a different type of ARCH models, called state-dependent variance (SVR) models, defined by:

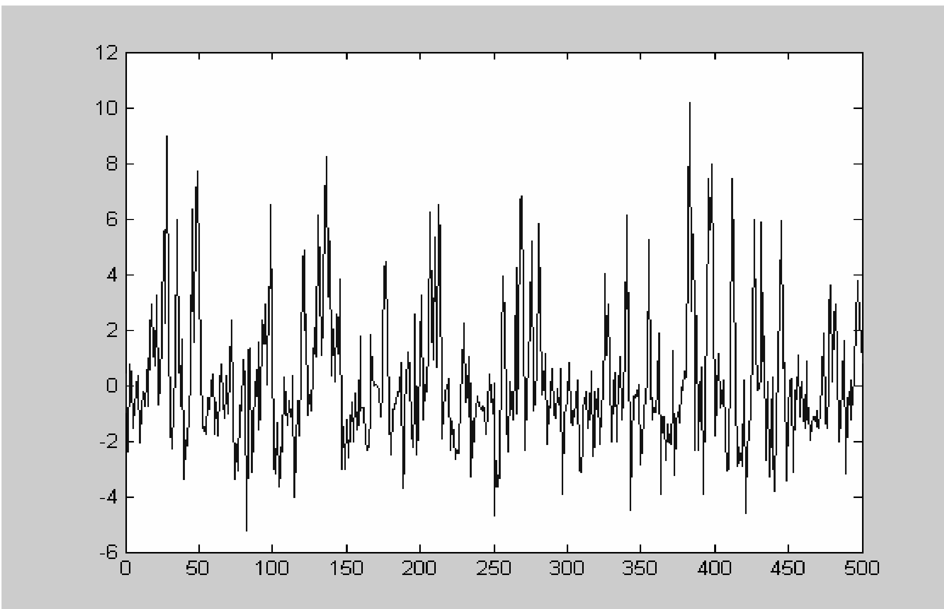
$$x_n = g(x_{n-1},\ldots, x_{n-p}) + s(x_{n-1},\ldots, x_{n-p})\varepsilon_n$$

with $s(.) \geq 0$. That is, the conditional variance of the innovations is now a function of the past signal samples. With this model class, it is even easier to generate time series with non Gaussian properties.

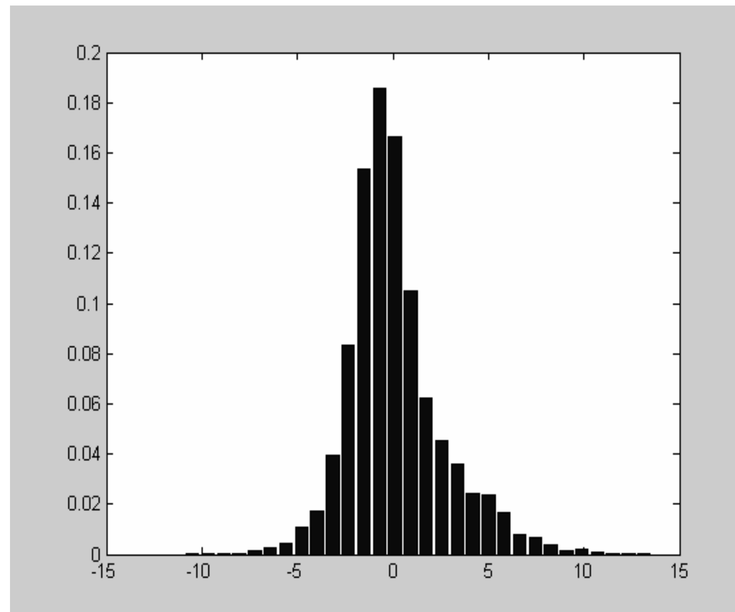
- Example: $x_n = 0.5x_{n-1} + [2+\text{sign}(x_{n-1})]\varepsilon_n$,

$\varepsilon_n \sim N(0,h_n)$

x_n



histogram



- Since $\Omega_{n\text{-}1}$ conveys *all* the information up to time n-1, all that has been said before on the conditional heteroscedasticity and unconditional homoscedasticity of GARCH models remains true for SDV ones.
- However, the stability of SDV models has to be assessed globally. We are going to derive sufficient conditions on $g(.)$ and $s(.)$ for geometrical ergodicity.

- An SDV model can be seen as a Markov one by introducing the state vector $\boldsymbol{x}_n = [x_n, \ldots, x_{n-p+1}]^\mathsf{T}$. This gives:

$$\boldsymbol{x}_n = G(\boldsymbol{x}_{n-1}) + s(\boldsymbol{x}_{n-1})\varepsilon_n \mathbf{e}_1$$

with $G(\boldsymbol{x}_{n-1}) = g(\boldsymbol{x}_{n-1})\mathbf{e}_1 + \mathbf{I}_1\boldsymbol{x}_{n-1}$. The vector $\mathbf{e}_1$ is p-dimensional with a first element equal to one and all others null. The $p\times p$ matrix $\mathbf{I}_1$ has ones on its first lower diagonal and zero elements everywhere else.

- Let us recall that two sufficient conditions for geometrical ergodicity are:

$$\mathrm{E}\left(\|\boldsymbol{x}_n\| - \|\boldsymbol{x}_{n\text{-}1}\| \,\middle|\, \boldsymbol{x}_{n\text{-}1} = \boldsymbol{x}\right) \leq \gamma, \ \ \|\boldsymbol{x}\| \leq \alpha \qquad (1)$$

$$\mathrm{E}\left(r\|\boldsymbol{x}_n\| - \|\boldsymbol{x}_{n\text{-}1}\| \,\middle|\, \boldsymbol{x}_{n\text{-}1} = \boldsymbol{x}\right) \leq -\beta, \ \ \|\boldsymbol{x}\| > \alpha \qquad (2)$$

with r a constant > 1, constants α, β, and $\gamma > 0$. the symbol ||.|| denotes *any* norm. For condition (1):

$$\mathrm{E}(\|\boldsymbol{x}_n\| - \|\boldsymbol{x}_{n-1}\| \,|\, \boldsymbol{x}_{n-1} = \boldsymbol{x}) = \mathrm{E}(\|G(\boldsymbol{x}) + s(\boldsymbol{x})\varepsilon_n \mathbf{e}_1\| - \|\boldsymbol{x}\|)$$

- One has: $\|G(\boldsymbol{x})+s(\boldsymbol{x})\varepsilon_n\mathbf{e}_1\| \leq \|G(\boldsymbol{x})\|+s(\boldsymbol{x})\,|\varepsilon_n|$
- Thus:

$$\mathrm{E}\left(\|G(\boldsymbol{x})+s(\boldsymbol{x})\varepsilon_n\mathbf{e}_1\|-\|\boldsymbol{x}\|\right)\leq \mathrm{E}\left(\|G(\boldsymbol{x})\|+s(\boldsymbol{x})\,|\varepsilon_n|-\|\boldsymbol{x}\|\right)$$

$$=\left(\|G(\boldsymbol{x})\|+s(\boldsymbol{x})\mathrm{E}[\,|\varepsilon_n|\,]-\|\boldsymbol{x}\|\right)$$

- Let us assume that $\mathrm{E}[|\varepsilon_n|]$ is finite (which is the case if ε_n is Gaussian). Then if $\|\boldsymbol{x}\| \leq \alpha$ and $g(.)$, $s(.)$, are continuous, it is clear that this last quantity is bounded, and condition (1) is satisfied.

- For condition (2), the same development leads to:

$$\mathsf{E}(r\,\|G(\mathbf{x})+s(\mathbf{x})\varepsilon_n\mathbf{e}_1\|-\|\mathbf{x}\|)\le\mathsf{E}(r\,\|G(\mathbf{x})\|+rs(\mathbf{x})\,|\varepsilon_n|-\|\mathbf{x}\|)$$

$$=\|\mathbf{x}\|\left(r\frac{\|G(\mathbf{x})\|}{\|\mathbf{x}\|}+cr\frac{s(\mathbf{x})}{\|\mathbf{x}\|}-1\right)$$

with $c = \mathsf{E}[|\varepsilon_n|]$ finite.

- Now it is possible to find some α such that this quantity can be made negative for $\|\mathbf{x}\| > \alpha$, regardless of c, by assuring that:

$$\lim_{\|\mathbf{x}\|\to\infty}\frac{\|G(\mathbf{x})\|}{\|\mathbf{x}\|}<1 \quad\text{and}\quad \lim_{\|\mathbf{x}\|\to\infty}\frac{s(\mathbf{x})}{\|\mathbf{x}\|}=0$$

- In the case the AR part is linear, i.e.:

$$g(x_{n-1},\dots, x_{n-p}) = a_1 x_{n-1} + a_2 x_{n-2} + \dots + a_p x_{n-p}$$

then in the Markov representation:

$$G(\boldsymbol{x}_{n-1}) = \mathbf{A}\boldsymbol{x}_{n-1} \quad \text{with} \quad \mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ \mathbf{0} & & 1 & 0 \end{bmatrix}$$

- The condition on stability translates into eigenvalues of A having modulus < 1, which is equivalent to the poles being in the unit circle.

- When $p = 1$, it is actually possible to derive an approximation of the probability density function of x_n, when the innovation is unit-variance Gaussian.
- First, one notice that with:

$$x_n = g(x_{n\text{-}1}) + s(x_{n\text{-}1})\varepsilon_n$$

the conditional probability $P(x_n|x_{n\text{-}1}=y)$ is given by:

$$P(x_n \mid x_{n-1} = y) = \frac{1}{s(y)\sqrt{2\pi}} \exp\left[-\frac{[x_n - g(y)]^2}{2s^2(y)}\right]$$

PDF OF SDV(1) MODEL (2)

- Since $P(x) = \int P(x|y)P(y)\,dy$, the equilibrium pdf must be a solution (invariant) of the so-called *master equation*:

$$P(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{P(y)}{s(y)} \exp\left[-\frac{[x - g(y)]^2}{2s^2(y)} \right] dy$$

- This is of course impossible to solve exactly in most situations.

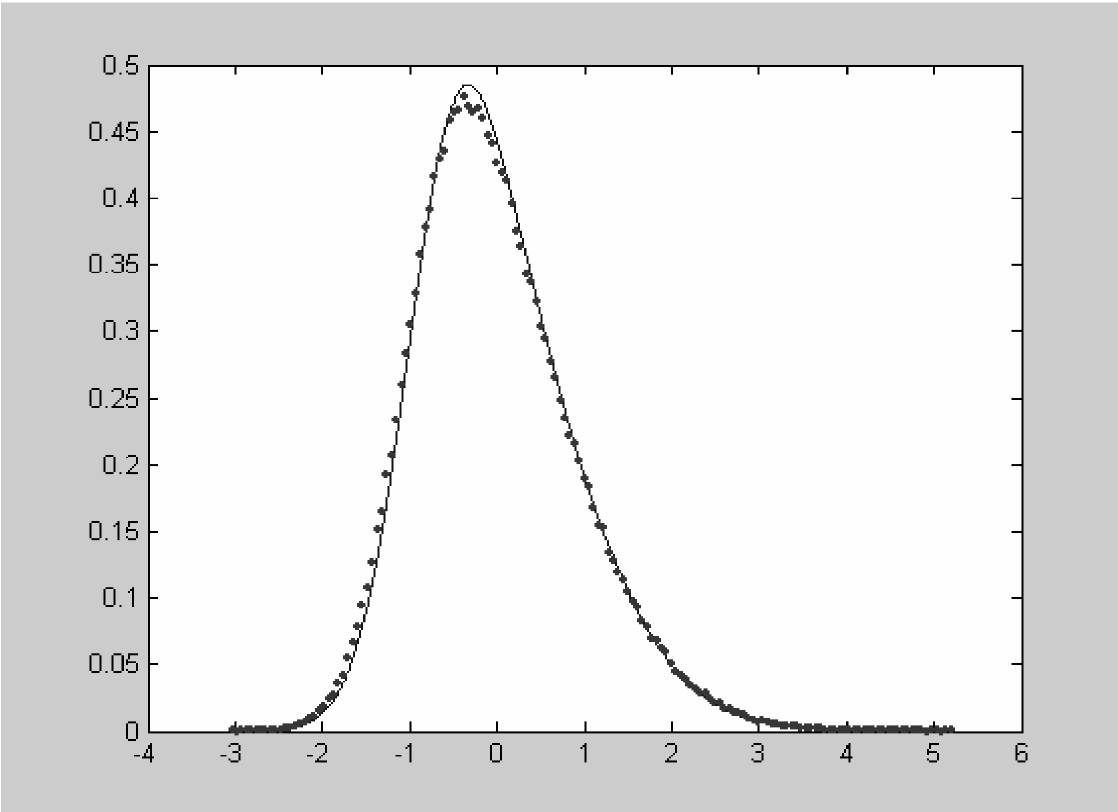
- But one may use an *ansatz* (i.e. well chosen representation) to obtain an approximate solution. In this case it is:

$$P(x) = Z(x)\exp[-\Phi(x)] \quad \text{with } \Phi(x) = \int^{x} \frac{\psi(u)}{s^2(u)}\,du$$

- And after some (lengthy) developments one gets, with *C* a normalizing constant:

$$P(x) = \frac{C}{s^2(x)}\exp\left[2\int^{x} \frac{g(u)-u}{s^2(u)}\,du\right]$$

- Example $g(x) = 0.9x$ $g(x) = 0.4 + 0.1 \tan^{-1}(x)$

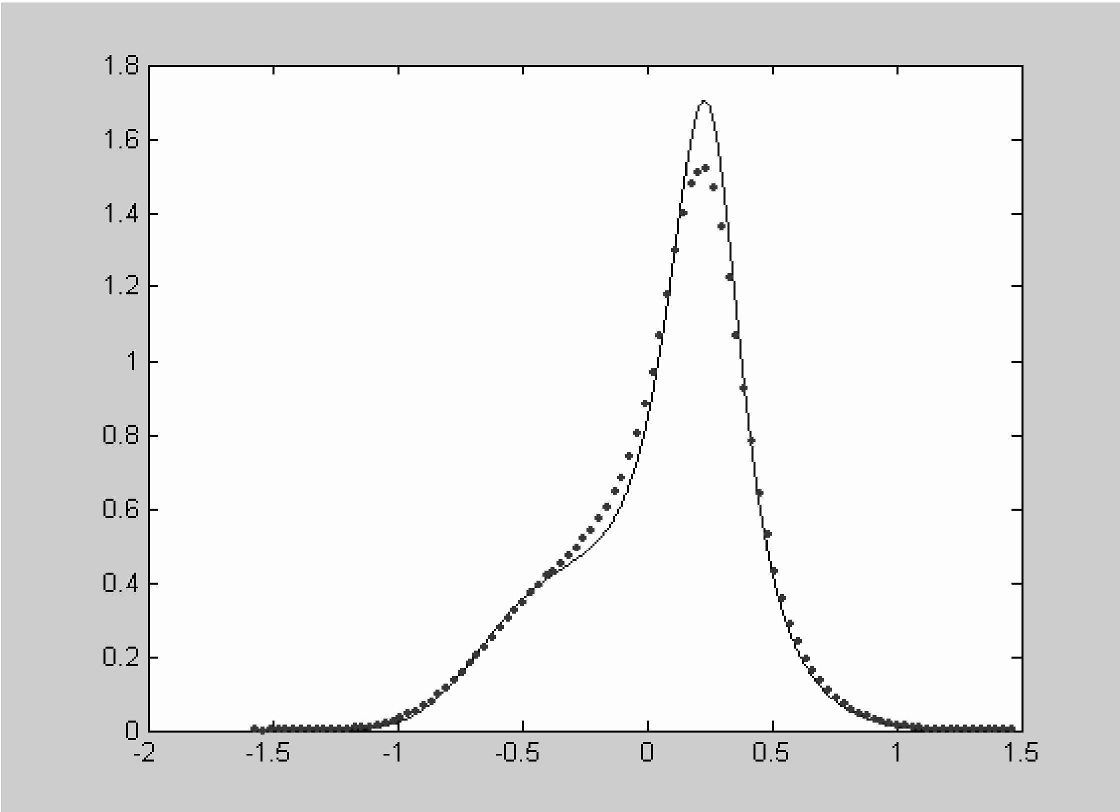


—— approximation

• histogram-based pdf estimate

PDF OF SDV(1) MODEL (5)

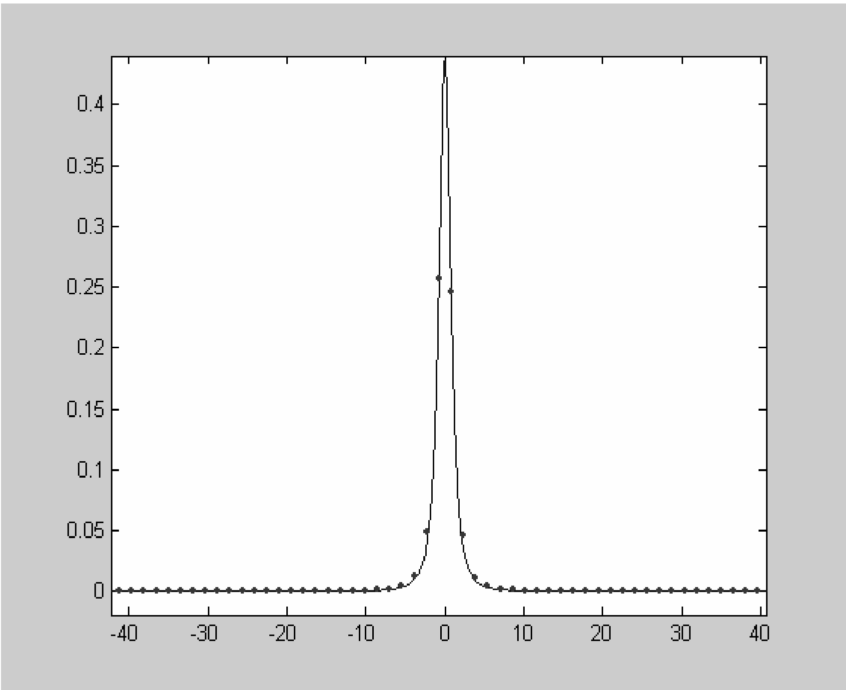
- Example $g(x) = \tan^{-1}(x)$ $g(x) = 0.2\text{-}0.1\exp[-10(x-0.25)^2]$

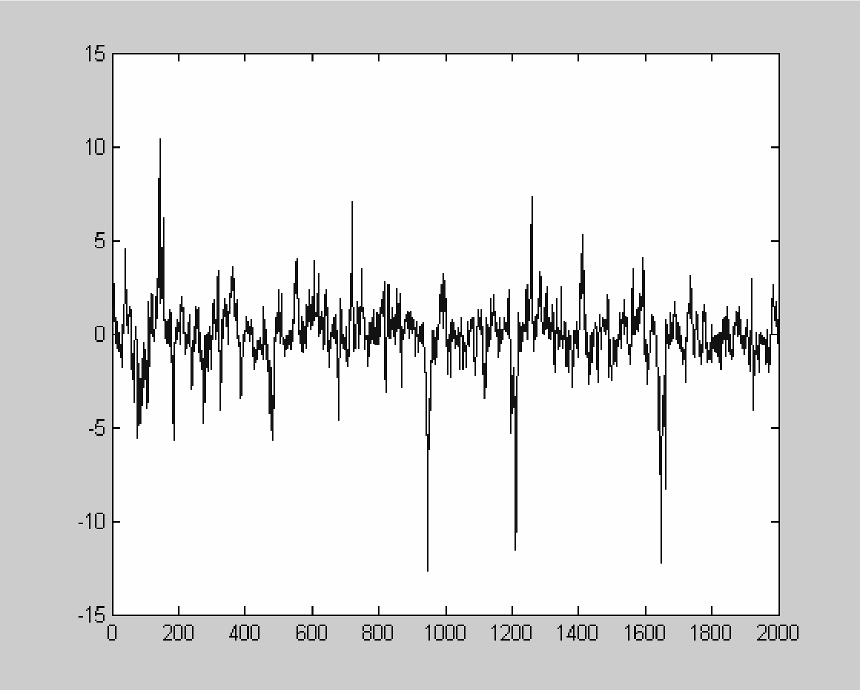


—— approximation

• histogram-based pdf estimate

- Example $g(x) = 0.9x$ $g(x) = 0.5(0.7x^2+1)^{1/2}$





- Due to the conditional heteroscedasticity, it is necessary as for the ARCH models to use a maximum likelihood approach to estimate $g(.)$ and $s(.)$.
- A parametric description of must be provided, which assures $s > 0$, and obeys the condition for stability. Also, it is preferable that $s(.)$ has the universal approximation capability. For these reasons, an RBF network may be a good choice.

- Since optimization algorithms have to be used, it is advised to start from a good initial solution.
- Such a good initial condition is obvious the least squares one. One first estimate $g(.)$ to compute the residuals:

$$\eta_n = x_n - \hat{g}(x_{n-1}, \cdots, x_{n-p})$$

- If the ε_n are supposed Gaussian, then the joint density of the residuals is:

$$P(\eta_1,\cdots,\eta_N;\theta)=\prod_{i=1}^{N}\frac{1}{s(\boldsymbol{x}_{i-1};\theta)\sqrt{2\pi}}\exp\left[-\frac{\eta_i^2}{2s^2(\boldsymbol{x}_{i-1};\theta)}\right]$$

- Maximizing log[P(.)] amounts to make the residuals homoscedastic, i.e. to define a function $s(\boldsymbol{x}_{n\text{-}1})$ such that the $\{\eta_n/s(\boldsymbol{x}_{n\text{-}1}))$ have a Gaussian probability density with unit variance [1].

- A non parametric approach has also been described in [2] in the case $s(.)$ depends only on the sample x_{n-1}. The interval $J = [a\ b]$ containing all these samples is divided into K equal sub-intervals $J_k = [t_k\ t_{k+1}]$ with $t_1 = a$ and $t_{K+1} = b$.
- Let us define the random sets:

$$V(k) = \{n, 2 \leq n \leq N, x_{n-1} \in J_k\}$$

And by $|V(k)|$ their cardinal.

- Once the estimates of the innovations $\{\varepsilon_n\}$ have been computed on the available samples $n = 1, \ldots, N$, an estimate of $s(.)$ in the interval J_k is obtained by:

$$\hat{s}_k^2 = \frac{1}{|V(k)|} \sum_{n \in V(k)} \hat{\varepsilon}_n^2$$

- The estimate is naturally extended outside J by using the first and last estimates.

- Note the following feature of an SDV model with linear AR part ($k > 0$):

$$x_n = a_1 x_{n\text{-}1} + \ldots + a_p x_{n\text{-}p} + s(\boldsymbol{x}_{n\text{-}1})\varepsilon_n$$

$$x_n x_{n\text{-}k} = a_1 x_{n\text{-}1} x_{n\text{-}k} + \ldots + a_p x_{n\text{-}p} x_{n\text{-}k} + s(\boldsymbol{x}_{n\text{-}1}) x_{n\text{-}k} \varepsilon_n$$

- If the model is ergodic (thus stationary)

$$\mathrm{E}[x_n x_{n\text{-}k}] = R_{xx}(k) = a_1 R_{xx}(k\text{-}1) \ldots + a_p R_{xx}(k\text{-}p)$$
$$+ \mathrm{E}[s(\boldsymbol{x}_{n\text{-}1}) x_{n\text{-}k} \varepsilon_n]$$

- But ε_n is independent both from the state vector $\boldsymbol{x}_{n\text{-}1}$ (and thus $s(\boldsymbol{x}_{n\text{-}1})$) and from $x_{n\text{-}k}$. This gives:

$$\mathrm{E}[s(\boldsymbol{x}_{n\text{-}1})x_{n\text{-}k}\varepsilon_n] = \mathrm{E}[s(\boldsymbol{x}_{n\text{-}1})x_{n\text{-}k}].\mathrm{E}[\varepsilon_n] = 0$$

- So

$$R_{xx}(k) = a_1 R_{xx}(k\text{-}1) \ldots + a_p R_{xx}(k\text{-}p)$$

For $k \neq 0$. The output of the SDV model follows the same Yule-Walker equations as the linear AR model.

1. P. H. Franses and D. van Dijk, *Non-Linear Time Series Models in Empirical Finance*, Cambridge Univ. Press, 2000.
2. J. Fan and Q. Yao, *Nonlinear Time Series – Nonparametric and Parametric Methods*, Springer, NY, 2003.
3. J.-M. Vesin, "A nonlinear autoregressive signal model with state-dependent gain," *Signal Processing*, vol. 26, 1992, pp. 37-48.
4. J. Diebold, "Testing the functions defining a nonlinear autoregressive time series," *Stoch. Proc. Appl.*, vol. 36, 1990, pp. 85-106.