Économétrie non-linéaire

GARCH models

Gilles de Truchis - Elena Dumitrescu

Master EIPMC

September 2020

Plan

1 Stylized Facts

2 GARCH family of models

3 Multivariate GARCH models

Financial series

ARCH / GARCHmodels appeared in the context of the debate on the linear non-linear representation of stochastic temporal processes

Nonlinearity in variance

A major contribution of the ARCH literature is the finding that apparent changes in the volatility of economic time series may be predictable and result from a specific type of nonlinear dependence rather than exogenous structural change in variables (Berra et Higgins, 1993, page 315).

- specific representation of non-linearity
- simple modelling of uncertainty

Financial Series

Notations

S_t : asset (or portfolio) price at time t

p_t : asset (or portfolio) log-price at time t

r_t : the continuously compounded or log-return of a financial asset (or portfolio) at time t

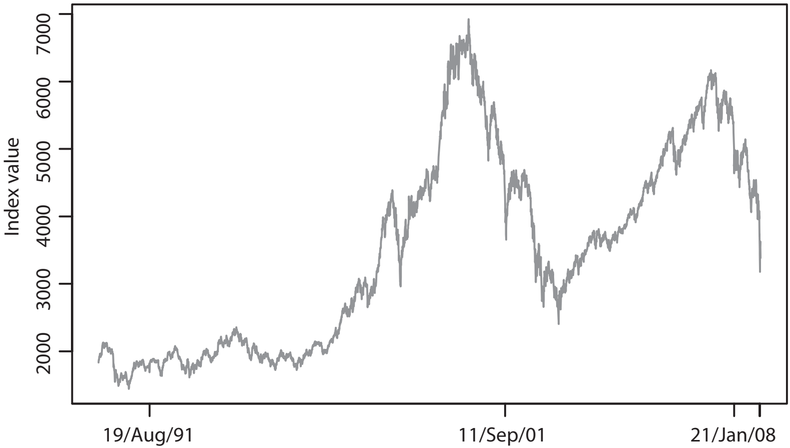
$r_t = p_t - p_{t-1}$

$r_t = log(1 + R_t)$ with $R_t = \frac{S_t - S_{t-1}}{S_{t-1}}$

- Their properties have been amply commented upon in the financial literature
- These stylized facts are mainly concerned with daily stock prices

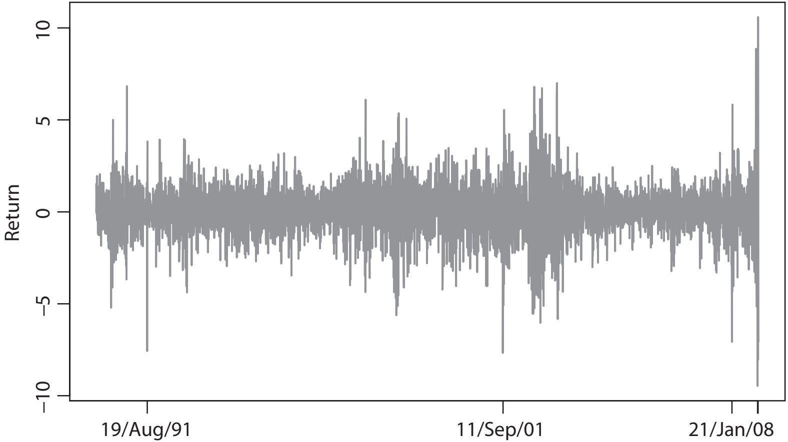
- Nonstationarity of price series
  - The stochastic process S_t is generally non-stationary in the sense of second-order stationarity

- Stationarity of return series
  - The stochastic process r_t is compatible with the second-order stationarity property



- Nonstationarity of price series
  - The stochastic process S_t is generally non-stationary in the sense of second-order stationarity

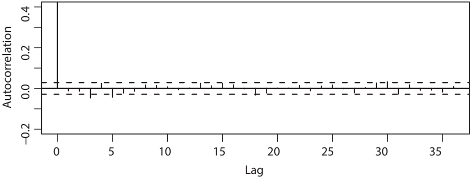
- Stationarity of return series
  - The stochastic process r_t is compatible with the second-order stationarity property



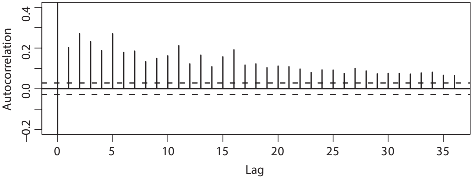
Autocorrelation

- Absence of autocorrelation for the price variations : (a)
  - The series of price variations generally displays small autocorrelations, making it close to a white noise (Efficient Market Hypothesis ou EMH)

- Autocorrelations of the squared price returns : (b)
  - Squared returns (r_t^2) or absolute returns ($|r_t|$) are generally strongly autocorrelated



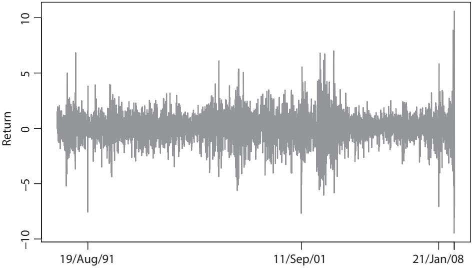
(a)



(b)

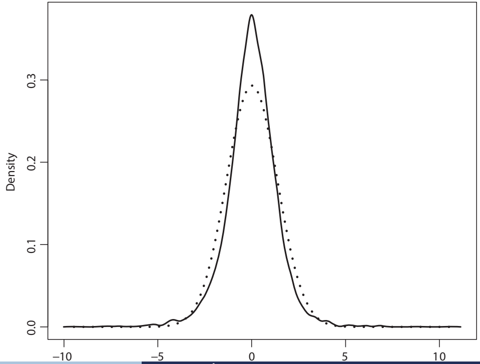
Volatility clustering

- Large absolute returns $|r_t|$ tend to appear in clusters
- Turbulent (high-volatility) sub-periods are followed by quiet (low-volatility) periods. These sub-periods are recurrent but do not appear in a periodic way (which might contradict the stationarity assumption)
- In other words, volatility clustering is not incompatible with a homoscedastic (i.e. with a constant variance) marginal distribution for the returns

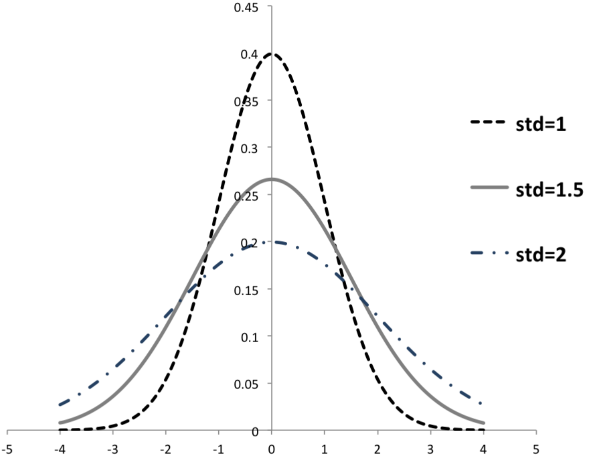


Fat-tailed distributions

- The empirical distribution of daily returns does not resemble a Gaussian one: classical tests typically lead to rejection of the normality assumption at any reasonable level
- The densities have fat tails and are sharply peaked at zero: they are called leptokurtic (check the coefficient of Kurtosis)
- When the time interval over which the returns are computed increases, leptokurticity tends to vanish and the empirical distributions get closer to a Gaussian (Aggregational Gaussianity property)



Fat tails vs different dispersion levels



Conditional fat tails

- Even after accounting for volatility clustering, (by using for example ARCH / GARCH models as we will see in the next section), the distribution of the residuals is leptokurtic
- Its kurtosis is however smaller than in the unconditional case (for the residuals of a simple ARMA type of model)

Leverage effects

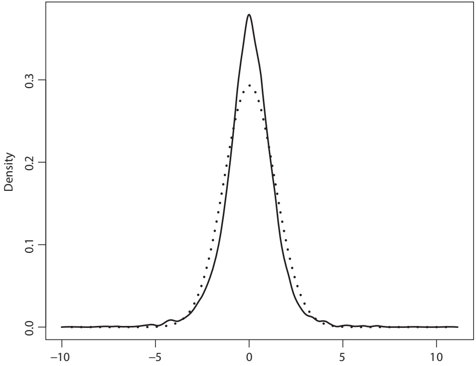
- Asymmetry in the response of volatility to positive and negative past returns, respectively
- A diminishing price generates an increase in volatility larger than a price increase of the same amount

h	1	2	3	4	5	6	7
$\hat{\rho}_r(h)$	−0.012	−0.014	−0.047	0.025	−0.043	−0.023	−0.014
$\hat{\rho}_{\lvert r\rvert}(h)$	0.175	0.229	0.235	0.200	0.218	0.212	0.203
$\hat{\rho}(r^+_{t-h}, \lvert r_t\rvert)$	0.038	0.059	0.051	0.055	0.059	0.109	0.061
$\hat{\rho}(-r^-_{t-h}, \lvert r_t\rvert)$	0.160	0.200	0.215	0.173	0.190	0.136	0.173

We use here the notation $r^+_t = \max(r_t, 0)$ and $r^-_t = \min(r_t, 0)$.

Leverage effects vs Gain/Loss Asymmetry

- Gain/Loss Asymmetry : The distribution of prices may be asymmetric, there are more drops than increases (check the coefficient of Skewness)



Seasonality

- Calendar effects: the day of the week, the proximity of holidays, among other seasonalities, may have significant effects on returns
- Following a period of market closure, volatility tends to increase, reflecting the information cumulated during this break
- The seasonal effect is also very present for intraday series (beyond the scope of this course)

Table: January effect

Average return (monthly %)		
Period	January	Other months
1904-1928	1.3	0.44
1929-1940	6.63	-0.6
1940-1974	3.91	0.7
1904-1974	3.84	0.42

Seasonality

- Calendar effects: the day of the week, the proximity of holidays, among other seasonalities, may have significant effects on returns
- Following a period of market closure, volatility tends to increase, reflecting the information cumulated during this break
- The seasonal effect is also very present for intraday series (beyond the scope of this course)

Table: Week-end effect

		Monday	Tuesday	Wednesday	Thursday	Friday
French (1980)	1953-1977	-0.17	0.02	0.1	0.04	0.09
Gibbons and Hess (1981)	1962-1978	-0.13	0	0.1	0.03	0.08

Summary

- Any satisfactory statistical model for daily returns must be able to capture these main stylized facts, mainly leptokurticity, the unpredictability of returns, and the existence of positive autocorrelations in the squared and absolute returns
- Classical formulations (such as ARMA models) centered on the second-order structure are inappropriate
- There is evidence of conditional heterokedasticity (time-varying volatility):

$$\mathbb{V}(r_t | r_{t-1}, r_{t-2}, ...) \not\equiv const$$

- Conditional heteroscedasticity is perfectly compatible with stationarity (in the strict and secondorder senses), just as the existence of a nonconstant conditional mean is compatible with stationarity

Plan

1. Stylized Facts
2. GARCH family of models
3. Multivariate GARCH models

Modelling Approaches

Objective : account for the very specific nature of financial series (price variations or log-returns, interest rates, etc.)

Recall the traditional forecast analysis (cf. Box et Jenkins)

Exemple:
Stationary AR (1): $r_t = \theta r_{t-1} + \varepsilon_t$, with ε_t i.i.d $N(0, \sigma_\varepsilon^2)$

$\mathbb{E}(r_{t+1}) = 0$

$\mathbb{E}(r_{t+1} | r_t, r_{t-1}, ...) = \theta r_t$

provides a model specification for the conditional mean

Modelling Approaches

Engle (1982)'s idea : account for other conditional moments of the return processus

But, for an AR(1) process

$\mathbb{E}(r_{t+1}^2) = \sigma_\varepsilon^2/(1-\theta^2)$

$\mathbb{E}(r_{t+1}^2 | r_t, r_{t-1}, ...) = \sigma_\varepsilon^2$

are constants

Such models are unable to measure changes in forecast error variance although we want them to be impacted by their past evolution

Modelling Approaches

Solution

- Models that capture time-varying volatility are written in the multiplicative form
$$r_t = \sigma_t z_t$$
- where (z_t) and (σ_t) are real processes such that:
  - σ_t is measurable with respect to a σ-field, denoted $\mathcal{I}_{t-1}$;
  - (z_t) is a weak white noise process with unit variance, z_t being independent of $\mathcal{I}_{t-1}$ and $\sigma(r_u; u < t)$;
  - $\sigma_t > 0$
- This formulation implies that the sign of the current price variation (that is, the sign of r_t) is that of z_t, and is independent of past price variations
- Most importantly, if the first two conditional moments of r_t exist, they are given by $\mathbb{E}(r_t|\mathcal{I}_{t-1}) = 0, \quad \mathbb{E}(r_t^2|\mathcal{I}_{t-1}) = \sigma_t^2$
- The random variable σ_t is called the volatility of r_t

Modelling Approaches

- As $Cov(r_t, r_{t-h}) = \mathbb{E}(z_t)\mathbb{E}(\sigma_t r_{t-h}) = 0$, (r_t) is a weak white noise.
- The series of squares, on the other hand, generally have nonzero autocovariances: (r_t) is thus not a strong white noise
- The kurtosis coefficient of r_t, if it exists, is related to that of z_t, denoted k_z by $\frac{\mathbb{E}(r_t^4)}{\{\mathbb{E}(r_t^2)\}^2} = k_\eta[1 + \frac{Var(\sigma^2)}{\{\mathbb{E}(\sigma^2)^2\}}]$
- Hence, the leptokurticity of financial time series can be taken into account in two different ways:
  - either by using a leptokurtic distribution for the weak white noise sequence (z_t),
  - or by specifying a process (σ_t^2) with a great variability

Modelling Approaches

Different classes of models can be distinguished depending on the specification adopted for σ_t

- Conditionally heteroscedastic (or GARCH-type) processes
  - Here $\mathcal{I}_{t-1} = \sigma(r_s; s < t)$ is the σ-field generated by the past of r_t
  - The volatility is here a deterministic function of the past of r_t
  - Processes of this class differ by the choice of a specification for this function
  - The standard GARCH models are characterized by a volatility specified as a linear function of the past values of r_t^2
- Stochastic volatility processes
  - Here $\mathcal{I}_{t-1}$ is a σ-field generated by $v_t, v_{t-1}, \cdots$, where $(v_t\ )$ is a strong white noise and is independent of $(z_t\ )$
  - volatility is a latent process
  - a popular specification is the one where the process $\log \sigma_t$ follows an AR(1)

Conditionally heteroscedastic processes

- In these models, the key concept is the conditional variance, that is, the variance conditional on the past
- We can reproduce the autocorrelation empirically seen in conditional volatility by using the information in the previous value(s) of the squared returns

  $\Rightarrow$ in an ARCH(q) specification, perturbations follow an autoregressive process of order q

  $\Rightarrow$ ARCH(q) are autoregressive conditionally heteroskedastic models

  $\mathbb{V}(r_t) = const$

  $\mathbb{V}(r_t|\mathcal{I}_{t-1}) = f(r_{t-1}, r_{t-2}, \cdots ; \theta)$

ARCH test

Usual (Ljung-Box) autocorrelation test on squared returns

- H0: $\rho_1 = \rho_2 = \ldots = \rho_K = 0$

$$Q_{LB}(K) = T(T+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{T-k}\xrightarrow[T\to\infty]{d}\chi^2(K),$$

where $\hat{\rho}_k$ is the empirical autocorrelation

ARCH-LM test

- Auxiliary regression

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1\hat{\varepsilon}_{t-1}^2 + \cdots + \phi_p\hat{\varepsilon}_{t-p}^2 + \eta_t$$

- H0': $\phi_1 = \cdots = \phi_p = 0$
- Test-statistic: LM(p)$= T \times R^2 \xrightarrow[T\to\infty]{d} \chi^2(p)$

Family of ARCH-type models

- linear models
(ARCH(q), GARCH(p, q) et IGARCH(p, q))

- non-linear models (i.e. asymmetric models)
(EGARCH(p, q), GJRGARCH(p,q), TGARCH(p, q)...)

ARCH models Engle (1982)

Definition

r_t follows an ARCH(1) if

$r_t = z_t\sqrt{(\sigma_t^2)}$, $\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2$

z_t - strong white noise

σ_t^2 - deterministic and positive process conditionally on the σ-field

$$\begin{aligned}\mathbb{V}(r_t|\mathcal{I}_{t-1}) &= \mathbb{V}(z_t\sqrt{(\sigma_t^2)}|\mathcal{I}_{t-1})\\ &= \sigma_t^2\mathbb{V}(z_t|\mathcal{I}_{t-1})\\ &= \sigma_t^2,\ \mathbb{V}(z_t|\mathcal{I}_{t-1}) \text{ normalized to } 1\end{aligned}$$

$\Rightarrow \sigma_t^2$ is the conditional variance of r_t

Modèles ARCH

Moments of ARCH(1) process

i)

$$\begin{aligned}\mathbb{E}(r_t|\mathcal{I}_{t-1}) &= \mathbb{E}(z_t\sigma_t|\mathcal{I}_{t-1}) \\ &= \sigma_t\mathbb{E}(z_t|\mathcal{I}_{t-1}) = 0 \text{ if } z_t \text{ is weak white noise}\end{aligned}$$

$$\mathbb{E}(r_t) = \mathbb{E}(\mathbb{E}(r_t)|\mathcal{I}_{t-1}) = 0$$

ii)

$$\begin{aligned}\mathbb{V}(r_t|\mathcal{I}_{t-1}) &= \mathbb{V}(z_t\sigma_t|\mathcal{I}_{t-1}) \\ &= \sigma_t^2\mathbb{V}(z_t|\mathcal{I}_{t-1}) \\ &= \sigma_t^2\mathbb{V}(z_t) \\ &= \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2.\end{aligned}$$

$$\mathbb{V}(r_t) = \mathbb{E}\big((r_t - \mathbb{E}(r_t))^2\big) = \mathbb{E}(r_t^2)$$

From the autoregressive structure of r_t^2 in ARCH models under stationnarity hypothesis ($\mathbb{E}(r_t^2) = \alpha_0 + \alpha_1\mathbb{E}(r_t^2)$) we have that

$$\mathbb{E}(r_t^2) = \mathbb{V}(r_t) = \frac{\alpha_0}{1-\alpha_1}$$

ARCH(q) Models

Definition

r_t follows an ARCH(q) Process if

$r_t = z_t \sigma_t$

with $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2$

and where z_t is a weak white noise such that $\mathbb{E}(z_t) = 0$ *et* $\mathbb{E}(z_t^2) = \sigma_z^2$.

• This model fulfils the martingale difference and time-varying conditional variance properties

$\mathbb{E}(r_t | r_{t-1}) = 0$ and $\mathbb{V}(r_t | r_{t-1}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2$

ARCH-errors model

Linear autoregressive model $Y_t = \mathbb{E}(Y_t|Y_{t-1}) + \varepsilon_t$

with ε_t a weak white noise

$$\mathbb{E}(\varepsilon) = 0 \text{ and } \mathbb{E}(\varepsilon_t\varepsilon_s) = 0 \text{ if } s \neq t,$$

satisfying the martingale difference hypothesis

$$\mathbb{E}(\varepsilon_t|\varepsilon_{t-1}) = 0.$$

We assume that the residuals have an ARCH(q) representation: $\varepsilon_t = z_t\sigma_t$

with $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$

where z_t is a strong white noise

ARCH-errors model

- **AR(1) - ARCH(1) Exemple**

$$Y_t = \mu + \rho Y_{t-1} + \varepsilon_t,\ \varepsilon_t = z_t \sigma_t$$

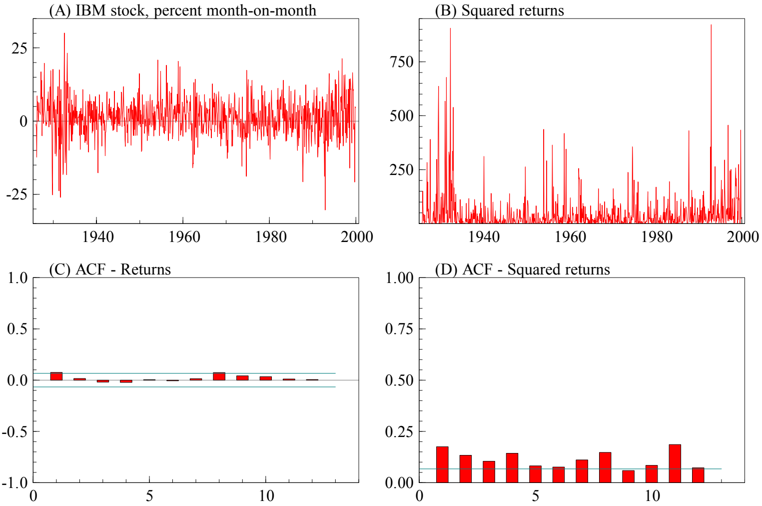
with $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ and $|\rho| < 1$

The model describes the evolution of both the conditional mean and the conditional variance of Y_t through time

- ε_t : résiduals
- z_t : standardized residuals

Residuals ε_t satisfy the properties of an ARCH process:
martingale difference; time-varying conditional variance; zero conditional auto-covariances; leptokurtic distribution

IBM example


(A) IBM stock, percent month-on-month
25
0
-25
1940
1960
1980
2000
(B) Squared returns
750
500
250
1940
1960
1980
2000
(C) ACF - Returns
1.0
0.5
0.0
-0.5
-1.0
0
5
10
(D) ACF - Squared returns
1.00
0.75
0.50
0.25
0.00
0
5
10

IBM example

Coeff.	Estimate	Std. Error	t-stat
ω	0.2605	0.0155	16.785
α_1	0.0366	0.0099	3.700
α_2	0.0809	0.0123	6.575
α_3	0.0657	0.0118	5.585
α_4	0.0866	0.0133	6.525
α_5	0.1035	0.0140	7.420
α_6	0.0746	0.0125	5.943
α_7	0.0780	0.0130	6.002
α_8	0.0892	0.0135	6.452
α_9	0.0875	0.0134	6.530
α_{10}	0.0789	0.0130	6.074

Non-negativity constraints and Limits of ARCH

- Non negativity constraints
  - For a ARCH(1) model , $\alpha_0 \geq 0; \alpha_1 \geq 0$
  - For an ARCH(q) model, $\alpha_i \geq 0, \ \forall i = 0, 1, \ldots, q$

- Limitations of ARCH(q) models
  - q, number of lags of the squared residuals, is potentially very large
  - Non negativity constraints might be violated

Generalised ARCH(1,1) or GARCH(1,1) Models

- Due to Bollerslev (1986)
- The conditional variance of the error term (or one-period ahead estimate of the variance of the error term) depends on own past values and on past values of the squared residuals

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

  with $\alpha_0 > 0, \alpha_1 > 0$ and $\beta > 0$
- GARCH(1,1) model containing 3 parameters is a very parsimonious infinite ARCH model
- The GARCH forecast variance is a weighted average of three different variance forecasts. One is a constant variance that corresponds to the long run average. The second is the forecast that was made in previous period. The third is the new information that was not available when the previous forecast was made
- This could be viewed as a variance forecast based on one period of information. The weights on these three forecasts determine how fast the variance changes with new information and how fast it reverts to its long run mean

GARCH(p,q) Models

Definition

A process ε_t satisfies a GARCH(p, q) representation if

$$\varepsilon_t = z_t\sigma_t,$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i\sigma_{t-i}^2,$$

where z_t is a weak white noise

and where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., q$ and $\beta_i \geq 0$, $i = 1, ..., p$

- The conditional variance of the error term depends on own p past values and on q past values of the squared residuals
- But in general a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data

GARCH Models

Conditional Moments

$\mathbb{E}(\varepsilon_t|\varepsilon_{t-1}) = 0$

$\mathbb{V}(\varepsilon_t|\varepsilon_{t-1}) = \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,$

Unconditional Variance

$$\mathbb{V}(r_t) = \mathbb{E}(r_t - \mathbb{E}(r_t)^2) = \mathbb{E}(r_t^2)$$

$$\mathbb{V}(r_t) = \mathbb{E}(r_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)}$$

GARCH Models

- Under additional assumptions (implying the second-order stationarity of ε_t^2), we can state that if ε_t is GARCH(p, q), then ε_t^2 is an ARMA(p, q) process

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 + v_t - \sum_{i=1}^{p} \beta_i v_{t-i}$$

  where $v_t = \varepsilon_t^2 - \sigma_t^2$ are the innovations of the process

- GARCH processes are hence able to capture the characteristic feature of financial series is that squared returns are autocorrelated
- The sum $\alpha+\beta$ is referred to as the persistence of the conditional variance process

GARCH Models

- Contrary to standard time series models (ARMA), the GARCH structure allows the magnitude of the noise ε_t^2 to be a function of its past values.
- Thus, periods with high volatility level (corresponding to large values of ε_{t-i}^2) will be followed by periods where the fluctuations have a smaller amplitude.

GARCH Stationarity

Definition

A process ε_t satisfies a GARCH(p, q) representation if

$$\varepsilon_t = z_t \sigma_t,$$
$$\sigma_t^2 = \alpha_0 + \textstyle\sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,$$

where z_t is a weak white noise and where $\alpha_0 > 0$, $\alpha_i \geq 0$,$i = 1, ..., q$ and $\beta_i \geq 0$, $i = 1, ..., p$ is **asymptotically second-order stationary** *if and only if*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \leq 1$$

GARCH

Drost & Nijman, 1993 define 3 types of GARCH

- The strong GARCH where z_t is a Strong White Noise
- The semi-strong GARCH where z_t is a Weak White Noise
- The weak GARCH where only projections of the conditional variance are considered

Examples of GARCH(1,1)

	S&P500		DAX	
	statistic	std error	statistic	std error
Daily returns				
α_0	0.0074	0.0012	0.0248	0.0031
α_1	$\boxed{0.0513}$	0.0039	0.0910	0.0065
β_1	$\boxed{0.9422}$	0.0042	0.8954	0.0069
weekly returns				
α_0	0.0829	0.0292	0.2369	0.0634
α_1	0.1015	0.0165	0.1091	0.0165
β_1	0.8872	0.0174	0.8642	0.0195
Monthly returns				
α_0	0.6531	0.4497	3.4344	1.8789
α_1	0.1297	0.0419	0.1276	0.0487
β_1	0.8444	0.0505	0.7837	0.0817

i) $\alpha_1+\beta_1 \sim\leq 1$; ii) $\beta_1 >> \alpha_1$; iii) $\hat{\beta}_1 > 0.9$ and $\hat{\alpha}_1 < 0.1$ for daily data

IGARCH(p,q) (Engle and Bollerslev, 1987)

- When
$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$$
the model is called an integrated GARCH(p, q) or IGARCH(p, q) model (see Engle and Bollerslev, 1986)
- There is a unit root in the autoregressive part of the ARMA representation of ε_t^2 representation
- Returns are strictly stationary with an infinite variance

Maximum Likelihood (ML)

- Method used to estimate parameters of ARCH and GARCH models; idea is to choose the parameters that maximize the chance (likelihood) of the data occurring
- Easy to implement once the density function of z_t is specified
- Let us call θ the vector of the parameters to be estimated.
- If z_t are assumed to be normally distributed, then the log likelihood function for a sample of T observations is:

$$\ell(\varepsilon_1, \varepsilon_2, ..., \varepsilon_T; \theta) = \sum_{t=1}^{T} log f\left(\varepsilon_t | \mathcal{I}_{t-1}\right)$$

$$= -\frac{T}{2} log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log\left(\sigma_t^2(\theta)\right) - \frac{1}{2}\sum_{t=1}^{T} \frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)},$$

where $\frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)} = z_t^2$

Maximum Likelihood (ML)

$$z_t^2 = z_t^2(\theta) = \frac{r_t - \mathbb{E}(r_t|\mathcal{I}_{t-1})}{\sigma_t^2}$$

and

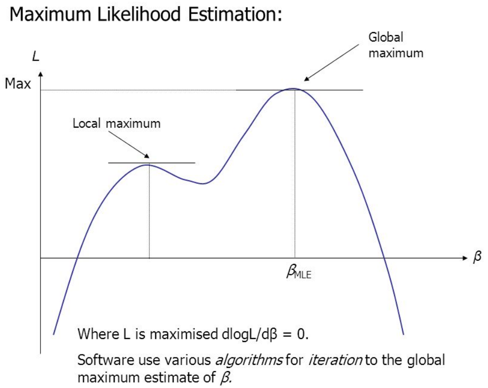
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

- Note that σ_t^2 is not observed for $t = 0, -1, \ldots, -p+1$
- To initialize the process, the unobserved squared residuals are
  1. set to their sample mean;
  2. set to the unconditional variance;
  3. obtained using a pre-sample;
  4. or considered as additional parameters to be estimated
- Under the normality hypothesis, this estimator

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_T^{-1}(\theta_0))$$

Maximum Likelihood (ML)

- The estimator of θ does not have a closed-form formula
- numerical optimization methods are used
- We need
  1. Initial condition
  2. Moving rule
  3. Stopping rule



Maximum Likelihood (ML) Example

Observation	*Residuals* u_i	$u_i \times u_i$	*GARCH(1,1) variance,* σ_t^2	$-ln(\sigma_t^2) - \frac{u_i^2}{\sigma_t^2}$
30/04/2004	-0.01195	0.00014	0.00014	7.85
31/05/2004	-0.01082	0.00012	0.00014	8.03
30/06/2004	-0.00015	0.00000	0.00014	8.91
31/07/2004	0.00719	0.00005	0.00010	8.70
31/08/2004	-0.00272	0.00001	0.00009	9.26
30/09/2004	0.01046	0.00011	0.00007	7.99
...	...	...	...	...
30/11/2013	-0.00038	0.00000	0.00004	10.18
31/12/2013	0.00254	0.00001	0.00003	10.11
31/01/2014	0.00594	0.00004	0.00003	9.25
28/02/2014	0.00367	0.00001	0.00004	9.79
31/03/2014	-0.00150	0.00000	0.00004	10.09

α_0	α_1	β
0.00	0.28	0.66

LLF to maximize 987.93

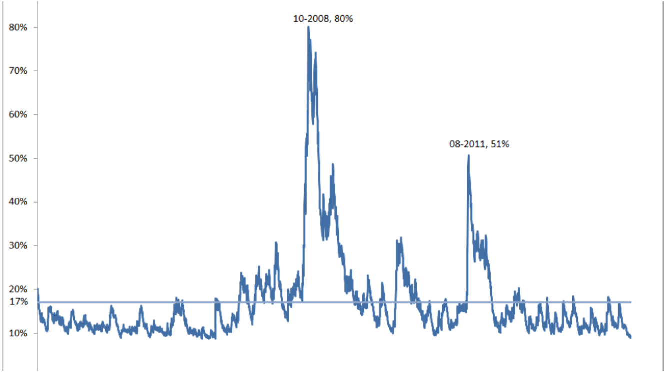
$$LLF = \sum_{i=1}^{m} \left[-ln(\sigma_t^2) - \frac{u_i^2}{\sigma_t^2} \right]$$

Maximum Likelihood (ML) Example 2

- Annualized GARCH(1,1) volatility fitted to daily US market returns

$$\sigma_t^2 = \underset{(4.46)}{0.0} + \underset{(8.59)}{0.09}\varepsilon_{t-1}^2 + \underset{(80.19)}{0.9}\sigma_{t-1}^2$$

- Volatility is
  - Time-varying
  - Persistent
  - Mean reverting



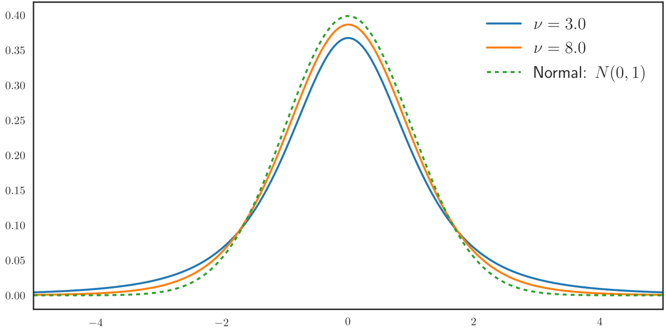
GARCH Extensions

- GARCH models are doing better than ARCH models (e.g., persistence in volatility) but there are still some issues (e.g., tails are not enough fat, etc).
- Improvements in various directions:
  - Non-normality of the conditional distribution:
    e.g. GARCH-t model
  - Asymmetric GARCH models :
    e.g. Exponential GARCH model (EGARCH), Threshold GARCH model (TGARCH), GJR model
  - Trade-off mean vs variance : e.g. GARCH-in-mean model

GARCH Extensions

In case of non-normality:

- Student distribution
- Skewed Student distribution
- Generalized error distribution (GED)



GARCH Extensions

Student distribution

If z_t follos a Student distribution with v degrees of fredom, where $v \in \mathbb{R}$ satsfies $v > 2$, then the log-likelihood associated with an observation and the parameter set θ is given by :

$$\begin{aligned}\ell(\theta, \varepsilon_t) = & \log\big[\Gamma\big(\frac{v+1}{2}\big)\big] - \log\big[\Gamma\big(\frac{v}{2}\big)\big] \\ & - 0.5\big[\log[\pi(v-2)] + \log(\sigma_t^2) + (1-v)\log\big(1+\frac{z_t^2}{v-2}\big)\big]\end{aligned}$$

with $\Gamma(.)$ the Gamma function and where $z_t = \frac{\varepsilon_t - \mathrm{E}\varepsilon_t}{\sigma_t}$

GARCH Extensions

Asymmetric GARCH

The GJR-GARCH(1,1) Model

- Due to Glosten, Jaganathan and Runkle, 1993

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{t-1}$$

where $I_{t-1} = \begin{cases} 1 \text{ if } \varepsilon_{t-1} < 0 \\ 0 \text{ otherwise} \end{cases}$

- For a leverage effect, we would see $\gamma > 0$
- We require $\alpha_1 + \gamma \geq 0$ and $\alpha_1 \geq 0$ for non-negativity
- We require $\alpha_1 + 0.5\gamma + \beta < 1$ for stationarity

GARCH Extensions

Asymmetric GARCH

The GJR-GARCH(1,1) Model

- Using monthly S&P 500 returns, December 1979- June 1998
- Estimating a GJR model, we obtain the following results

$$r_t = \underset{(3.198)}{0.172}$$

$$\sigma_t^2 = \underset{(16.372)}{1.243} + \underset{(0.437)}{0.015}\varepsilon_{t-1}^2 + \underset{(14.999)}{0.498}\,\sigma_{t-1}^2 + \underset{(5.772)}{0.604}\varepsilon_{t-1}^2 I_{t-1}$$

Extension 4. GARCH Asymétriques

TGARCH(1,1), (Zakoian, 1994)

A process ε_t satisfies a TGARCH(1,1) representation if and only if

$$\varepsilon_t = z_t \sigma_t$$

$$\sigma_t = \alpha_0 + \alpha_{pos} \mathrm{I}_{\varepsilon_{t-i} \geq 0} \varepsilon_{t-1} - \alpha_{neg} \mathrm{I}_{\varepsilon_{t-i} < 0} \varepsilon_{t-1} + \beta_1 \sigma_{t-1}$$

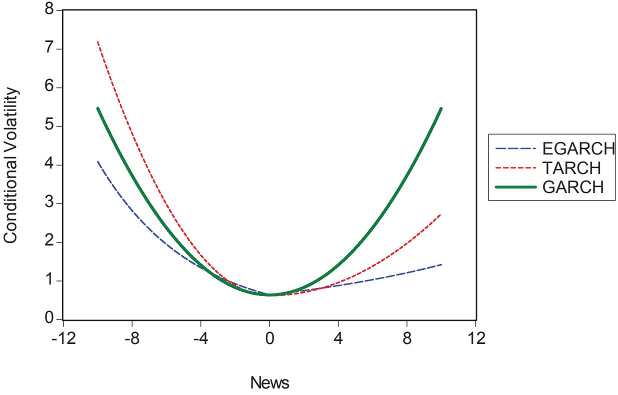
where the standardized residual z_t is a weak white noise and $\mathrm{I}_{\varepsilon_{t-1}<0}$ is an indication function such that $\mathrm{I}_{\varepsilon_{t-i}<0} = 1$ si $\varepsilon_{t-i} < 0$ et $\mathrm{I}_{\varepsilon_{t-i}<0} = 0$ sinon

• The asymmetric dynamics is specified for the squared root and not for the conditional variance

News Impact Curves

Asymmetric GARCH

- The news impact curve plots the next period volatility (σ_t) that would arise from various positive and negative values of ε_{t-1}, given an estimated model
- News Impact Curves for S&P 500 Returns using Coefficients from GARCH and GJR Model Estimates



News Impact Curves For Daily Volatility of the Dow Jones Industrial Average (DJIA) 1915-2001

EGARCH

Asymmetric GARCH

- Suggested by Nelson (1991)
- The variance equation is given by

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} a_i z_{t-i} + \sum_{i=1}^{q} b_i(|z_{t-i}| - \mathrm{E}[|z_{t-i}|]) + \sum_{i=1}^{p} \beta_i log(\sigma_{t-i}^2)$$

- Advantages of the model
  - Since we model the $\log \sigma_t^2$, then even if the parameters are negative, σ_t^2 will be positive
  - one identifies a sign effect $a_i z_{t-i}$ and a magnitude effect $b_i(|z_{t-i}| - \mathrm{E}[|z_{t-i}|])$
  - $\mathrm{E}[|z_{t-i}|]$ depends on the distribution of z_t

$$\mathrm{E}[|z_t|] = \sqrt{\frac{2}{\pi}} \text{ Loi Gaussienne}$$

$$\mathrm{E}[|z_t|] = 2\frac{\gamma(\frac{v}{2})\sqrt{v-2}}{\sqrt{\pi}(v-1)\Gamma(\frac{v}{2})} \quad \text{Loi de Student (v)}$$

GARCH-in-mean or GARCH-M (Engle, 1987)

- Asset pricing models suppose that higher risks should be rewarded by higher returns
- The GARCH-in-mean model lets the mean of an asset's returns to be determined by its lagged conditional volatility

$$r_t = \mu + \delta G(\sigma_t^2) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta \sigma_{t-1}^2,$$

  with $G(\sigma_t^2)$ a linear, log-linear or square-root function
- The parameter δ can be interpreted as the price of risk and can thus be assumed to be positive
- Hence, if $\delta > 0$, increases in risk (given by increases in conditional volatility) lead to higher mean returns

Plan

1 Stylized Facts

2 GARCH family of models

3 Multivariate GARCH models

Multivariate GARCH models

- While the volatility of univariate series has been the focus of the previous chapters, modeling the comovements of several series is of great practical importance
- The standard linear modeling of real time series has a natural multivariate extension through the framework of the vector ARMA (VARMA) models
- Similarly, here we introduce the concept of multivariate GARCH model
- Essential for asset pricing and risk management crucially depend on the conditional covariance structure of the assets of a portfolio

Multivariate GARCH models

Let us denote by r_t a column vector of k asset returns and the vector of their conditional expectations by μ_t

- ... returns's equation implies a **conditional covariance** matrix H_t:

$$r_t - \mu_t = \varepsilon_t = H_t^{1/2} z_t, \quad H_t^{1/2}(H_t^{1/2})' = H_t$$

  - ε_t is a vector, not a scalar as previously
  - where H_t is a matrix $k \times k$ with elements h_{ijt}
  - and z_t is i.i.d Gaussian such that $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t z_t') = I$ with I a $k \times k$ identity matrix

- The **conditional covariance** matrix H_t takes the form

$$H_t = f(H_{t-1}, H_{t-2}, \cdots, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots)$$

- If $H_t^{1/2}$ exists, H_t is positive definite

  $\Rightarrow$ the transformation $f(.)$ ought to insure that H_t is symmetric and positive definite

- But $f(H_{t-1}, H_{t-2}, \cdots, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots)$ is complex...

Multivariate GARCH models

Choosing a specification for H_t is obviously more delicate than in the univariate framework because:

(i) H_t should be (almost surely) symmetric, and positive definite for all t

(ii) the specification should be simple enough to be amenable to probabilistic study (existence of solutions, stationarity, ...), while being of sufficient generality

(iii) the specification should be parsimonious enough to enable feasible estimation

(iv) but, the model should not be too simple to be able to capture the - possibly sophisticated - dynamics in the covariance structure

Multivariate GARCH models

- Moreover, it may be useful to have the so-called **stability by aggregation property**
- If $\varepsilon_t = H_t^{1/2} z_t$ is satisfied, the process $(\tilde{\varepsilon}_t)$ defined by $\tilde{\varepsilon}_t = P\varepsilon_t$, where P is an invertible square matrix, is such that

$$\mathbb{E}(\tilde{\varepsilon}_t|\tilde{\varepsilon}_u, u < t) = 0, \quad \mathbb{V}(\tilde{\varepsilon}_t|\tilde{\varepsilon}_u, u < t) = \tilde{H}_t = PH_tP$$

- The stability by aggregation of a class of specifications for H_t requires that the conditional variance matrices $\tilde{H}_t$ belong to the same class for any choice of P
- Relevance: if the components of the vector ε_t are asset returns, $\tilde{\varepsilon}_t$ is a vector of portfolios of the same assets, each of its components consisting of amounts (coefficients of the corresponding row of P) of the initial assets

Multivariate GARCH models

- Generally z_t is assumed to follow the multivariate Gaussian distribution, $z_t \sim N(0, I)$, since it provides the basis of QML estimation as in the univariate case
- Another choice of density for z_t is the multivariate t
- Multivariate skewed distributions can also be used (e.g. the skewed-t of Bauwens and Laurent, 2005)
- As in the univariate case, distributions with fat-tails and skewness are usually better fitting data than the Gaussian

Multivariate GARCH models

- Unlike the ARMA models, however, the GARCH model specification does not suggest a natural extension to the multivariate framework
- Indeed, the (conditional) expectation of a vector of size k is a vector of size k, but the (conditional) variance is a $k \times k$ matrix
- Important milestones are
  - the BEKK model of Engle and Kroner (1995)
  - the constant conditional correlation (CCC) model of Bollerslev (1990)
  - the dynamic correlation model (DCC) of Engle (2002a)
  - the time-varying correlation (TVC) model of Tse and Tsui (2002)
- Earlier models had too many parameters to be useful for modeling more than two asset returns jointly (e.g. VEC model)

VEC models

- Take the case of a bivariate GARCH(p,q) ($k = 2$):

$$H_t = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{pmatrix} \Rightarrow h_t = vec(H_t) = \begin{pmatrix} h_{11t} \\ h_{12t} \\ h_{21t} \\ h_{22t} \end{pmatrix}$$

- Note: The operator $vec(.)$ consists in vectorizing a matrix by stacking the columns of the matrix on top of one another
- Using this operator, Engle et Kroner (1995) propose the VEC model:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i vec(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{i=1}^{p} \beta_i h_{t-i}$$

  with ω a $k \times 1$ vector, and α_i and β_i $k \times k$ matrices
- **Problem:** the model is big and some equations are redundant

  e.g. $h_{12t} = h_{21t}$ as H_t is a covariance matrix

  - it will not in general produce positive definite covariance matrices H_t

VECH models

- Apply a $vech(.)$ operator now to the previous $GARCH(p,q)$:

$$H_t = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{pmatrix} \Rightarrow h_t = vech(H_t) = \begin{pmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{pmatrix}$$

- Note: The operator $vech(.)$ consists in vectorizing a matrix by stacking the columns of the lower triangular part of its argument square matrix
- One obtains the VECH model where $h_t = vech(H_t)$

Definition

The process ε_t is said to admit a VEC-GARCH(p,q) representation (relative to the i.i.d sequence z_t) if it satisfies

$$\varepsilon_t = H_t^{1/2} z_t, \text{ where } H_t \text{ is positive definite such that}$$

$$vech(H_t) = \omega + \sum_{i=1}^{q} A^{(i)} vech(\varepsilon_{t-1}\varepsilon'_{t-1}) + \sum_{j=1}^{p} B^{(j)} vech(H_{t-j}),$$

where ω is a vector of size $\{k(k+1)/2\} \times 1$, and the $A^{(i)}$ and $B^{(j)}$ are matrices of dimension $k(k+1)/2 \times k(k+1)/2$.

VEC models

- In particular, for a VECH-GARCH(1,1)

$$h_t = \begin{pmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} \varepsilon^2_{1,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon^2_{2,t-1} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{pmatrix}$$

- every conditional covariance is a function of lagged conditional variances as well as lagged cross-products of all components
- More parsimonious model than the VEC-GARCH
- But the VECH-GARCH still implies a big number of coefficients
- **Problem:** VEC and VECH are not able to generally insure that H_t is positive definite

Diagonal VECH-GARCH models

- To further simplify the model and its estimation, one may assume that volatilities and covariances depend only on their past values (Bollerslev, Engle, and Wooldridge, 1988)

  $\Rightarrow$ Non-diagonal coefficients of A_i and B_i are null

- E.g. VECH-GARCH(1,1)

$$h_t = \begin{pmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{pmatrix}$$

- More parsimonious than the VEC-GARCH

- The VEC-GARCH is stable by aggregation

- In this case it is also possible to obtain conditions for positive definiteness of H_t for all t

BEKK model

- Developed by Baba, Engle, Kraft and Kroner, in a preliminary version of Engle and Kroner (1995)

Definition

Let (z_t) denote an i.i.d. sequence with common distribution. The process (ε_t) is called a BEKK-GARCH(p, q), with respect to the sequence (z_t), if it satisfies

$$\varepsilon_t = H_t^{1/2} z_t$$

$$H_t = C'C + \sum_{n=1}^{N} \sum_{i=1}^{q} A_{in} \varepsilon_{t-i} \varepsilon'_{t-i} A'_{in} + \sum_{n=1}^{N} \sum_{i=1}^{p} B'_{in} H_{t-i} B'_{in}$$

with A_{in}, B_{in}, $n \in \{1, \ldots, N\}$, and C matrices of dimension $k \times k$

- Each BEKK model implies a unique VECH model, while the converse implication is not true
- The BEKK class contains the diagonal models by choosing diagonal matrices A_{ik} and B_{jk}

BEKK model

Definition

H_t **is positive definite** *if matrices* H_{t-i}*,* $i = 1, \ldots, p$*, are almost surely positive definite and*

$$ker\{C\} \bigcap_{j=1}^{p} \bigcap_{n=1}^{N} ker\{B_{jn}\} = \{0\}$$

- This is a weak condition, requiring only that C and B_{jn} are full rank (e.g. triangular C with positive diagonal elements)
- an identifiability restriction is needed, $H_{jj,t}$ being invariant to a change of sign of the j-th row of any matrix A_i
- BEKK-GARCH(1,1) in the bivariate case (k=2) with $N = 1$

$$H_t = C'C + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}' \begin{pmatrix} h_{11,t-1}^2 & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1}^2 \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$$

Stationarity of the BEKK model

Definition

Let C be an upper triangular $n \times n$ matrix and A_{in}, B_{in} be $n \times n$ parameter matrices. Let z_t be an i.i.d. process with mean zero and unit variance. Hence z_t is independent of $\mathcal{I}_{t-1}$, and $cov(z_t|\mathcal{I}_{t-1}) = cov(z_t) = I$.

There exists a covariance stationary BEKK process ε_t, such that $\varepsilon_t = H_t^{1/2} z_t$, where $H_t = cov(\varepsilon_t|\mathcal{I}_{t-1})$ and $\mathcal{I}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ if and only if all the eigenvalues of $\sum_{i=1}^{q} A_{in} \otimes A_{in} + \sum_{n=1}^{N} \sum_{i=1}^{p} B_{in} \otimes B_{in}$ are less than one in modulus.

General BEKK model

- A general BEKK- GARCH(1,1) representation requires $N > 1$
- But the number of parameters increases with N
- Need to find the minimal N which eliminates restrictions defined by the need of positive definiteness and identifiability
- Engle et Kroner (1995) give two such conditions:
  - Denote $w = k(k+1)/2$. Then N should be big enough such that the total number of elements in the matrices is at least w^2
  - Let $a_{i,j,n}$ be the i,j-th element of a matrix $A_{1,n}$. Then, there should exist a matrix $A_{1,n}$ which contains the pair $a_{il,n}$, $a_{jm,n}$ or the pair $a_{jl,n}$, $a_{im,n}$ for all i, j, l, m between 1 and n
- Similar restriction are needed for matrices B_{1n}

General BEKK model

Example

- For $k = 2$, the following matrices satisfy the two aforementioned conditions

$$A_{11} = \begin{pmatrix} a_{11,1} & a_{12,1} \\ 0 & a_{22,1} \end{pmatrix}; \quad A_{12} = \begin{pmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & 0 \end{pmatrix}; \quad A_{13} = \begin{pmatrix} a_{11,3} & 0 \\ a_{21,3} & a_{22,3} \end{pmatrix}$$

- The coefficients of a BEKK representation are difficult to interpret (highly artificial constraints on the volatilities and covolatilities)

Stability of the BEKK model by aggregation

Definition

Let (ε_t) be a BEKK-GARCH (p, q) process. Then, for any invertible $m \times m$ matrix P, the process $\tilde{\varepsilon}_t = P\varepsilon_t$ is a BEKK- GARCH(p, q) process.

Proof.

Letting $\tilde{H}_t = PH_tP'$, $\tilde{G} = PC'CP'$, $\tilde{A}_{in} = PA_{in}P^{-1}$ and $\tilde{B}_{in} = PB_{in}P^{-1}$, we get

$$\tilde{\varepsilon}_t = \tilde{H}_t^{1/2} z_t$$

$$\tilde{H}_t = \tilde{G} + \sum_{n=1}^{N} \sum_{i=1}^{q} \tilde{A}_{in}\varepsilon_{t-i}\varepsilon'_{t-i}\tilde{A}'_{in} + \sum_{n=1}^{N} \sum_{i=1}^{p} \tilde{B}'_{in}H_{t-i}\tilde{B}'_{in},$$

and $\tilde{G}$ being a positive definite matrix, the result is proved. $\square$

Estimation of the BEKK model

- Under the assumption that z_t are i.i.d. conditionally on initial values, the quasi log-likelihood function of the BEKK model is given by

$$L_n(\theta) = L_n(\theta; \varepsilon_1, \ldots, \varepsilon_n) = \sum_{t=1}^{n} -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|H_t| - \frac{1}{2}\varepsilon_t' H_t^{-1}\varepsilon_t,$$

where

$$\varepsilon_t = H_t^{1/2} z_t$$

$$H_t = C'C + \sum_{n=1}^{N}\sum_{i=1}^{q} A_{in}\varepsilon_{t-i}\varepsilon_{t-i}' A_{in}' + \sum_{n=1}^{N}\sum_{i=1}^{p} B_{in}' H_{t-i} B_{in}'$$

Estimation of the BEKK model

- Comte and Lieberman (2003) provide conditions for strong consistency and asymptotic normality of the quasi maximum likelihood estimator
  - Strong consistency
    $$\hat{\theta}_n \to \theta_0 \textit{ almost surely when } n \to \infty$$
  - Asymptotic normality
    $$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1} I J^{-1}),$$
    where J is a positive definite matrix and I is a positive semi-definite matrix, defined by
    $$I = \mathbb{E}\Big(\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\Big), \quad J = \mathbb{E}\Big(\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\Big)$$

Conditional correlations

- Multivariate GARCH models allow one to compute conditional **variances and covariances**

- Conditional correlations can hence be reconstructed

$$\rho_{ij,t} = \frac{h_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t}}}$$

with $\{i, j\} = 1, \cdots, k$ and $i \neq j$

Constant Conditional Correlations models

- Suppose that for a multivariate GARCH process of the form $\varepsilon_t = H_t^{1/2}\tilde{z}_t$ all the past information on ε_{it}, involving all the variables $\varepsilon_{j,t-i}$, is summarized in the conditional variance
- The standardized innovations $z_{it} = h_{ii,t}^{-1/2}\varepsilon_{it}$ are sequences of i.i.d (0,1) variables generally correlated
- Denote the covariance matrix $R = \mathbb{V}(z_t) = (\rho_{i,j})$, with $z_t = (z_{1t}, \ldots, z_{kt})$
- In CCC models the conditional covariances $h_{ij,t}$ are obtained as $h_{ij,t} = \rho_{i,j}\sqrt{(h_{ii,t}h_{jj,t})}$ for $i \neq j$ and they are time varying although the correlations are constant
- In matrix notations,

$$H_t = D_t R D_t = \rho_{i,j}\sqrt{(h_{ii,t}h_{jj,t})}$$

with D_t a $k \times k$ diagonal matrix with $\sqrt{h_{11,t}}, \cdots, \sqrt{h_{kk,t}}$ on its main diagonal

$\Rightarrow$ H_t is positive-definite if $h_{ii,t}$ is positive for all i and R_t is positive-definite

Constant Conditional Correlations models

Definition

Let $\tilde{z}_t$ be a sequence of i.i.d. variables. A process ε_t is called CCC-GARCH(p, q) if it satisfies

$$\begin{aligned}
\varepsilon_t &= H_t^{1/2}\tilde{z}_t \\
H_t &= D_t R D_t \\
\boldsymbol{h}_t &= \boldsymbol{\omega} + \sum_{s=1}^{q} \boldsymbol{A}_s \boldsymbol{\varepsilon}_{t-s} + \sum_{v=1}^{p} \boldsymbol{B}_v \boldsymbol{h}_{t-v},
\end{aligned}$$

where $R = cov(z_t z_t')$ is a correlation matrix, $D_t = diag(\sqrt{\boldsymbol{h}_t})$, $\boldsymbol{h}_t$ is the vector of k conditional variances with elements $(h_{ii,t})$, $\boldsymbol{\varepsilon}_t$ is the vector of k squared innovations (non-standardized), ω is a $m \times 1$ vector with positive coefficients, $\boldsymbol{A}_s$ and $\boldsymbol{B}_v$ are $k \times k$ matrices with nonnegative coefficients

- Note that $\varepsilon_t = D_t z_t$, where $z_t = R^{1/2}\tilde{z}_t$ is a centered vector with covariance matrix R such that $\varepsilon_{i,t} = h_{ii,t}^{1/2} z_{i,t}$
- Note that $h_{ii,t}$ may depend on the past of all the components of ε_t

Strict stationarity of the CCC model

Definition

The CCC-GARCH(p,q) model admits a second-order stationary solution if the vector of parameters is such that the roots of the polynomial $\det(I - \sum_{i=1}^{s}(A_i + B_i)\lambda)$ *with* $s = sup(p, q)$, *are outside the unit cercle. This solution is unique and ergodic.*

Strict stationarity of the CCC model

Definition

A necessary and sufficient condition for the existence of a strictly stationary and nonanticipative solution process for the CCC model is $\gamma < 0$, where γ is the top Lyapunov exponent of the sequence $\{D_t, t \in \mathcal{Z}\}$ (see Aue, Hormann, Horvath, and Reimherr, 2009). This stationary and nonanticipative solution, when $\gamma < 0$, is unique and ergodic, with

$$D_t = \begin{pmatrix} \Omega_t A_1 & \Omega_t A_2 & \dots & \Omega_t A_q & \Omega_t B_1 & \Omega_t B_2 & \dots & \Omega_t B_p \\ I & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & I & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & I & \dots & 0 & 0 & \dots & 0 \\ A_1 & A_2 & \dots & A_q & B_1 & B_2 & \dots & B_p \\ 0 & 0 & \dots & 0 & I & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & I & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & I & 0 \end{pmatrix},$$

and $\Omega_t = diag(z_{it}^2)$ matrix.

Estimation of CCC models

- by the quasi-maximum likelihood method
- Overall there are $k + k^2(p+q) + k(k-1)/2$ parameters to estimate $\theta = (\boldsymbol{\omega}', \alpha', \beta', \rho')'$
- Let $(\varepsilon_1, \dots, \varepsilon_n)$ be a sample of length n of the unique nonanticipative and strictly stationary variable ε_t of the CCC model
- Conditionally on nonnegative initial values $\varepsilon_0, \dots, \varepsilon_{1-q}, h_0, \dots, h_{1-p}$, the Gaussian quasi-likelihood is written as

$$L_n(\theta) = L_n(\theta; \varepsilon_1, \dots, \varepsilon_n) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{k/2}|H_t|^{1/2}} \exp\Big(-\frac{1}{2}\varepsilon_t' H_t^{-1} \varepsilon_t \Big),$$

where H_t are recursively defined, for $t \geq 1$, by

$$H_t = D_t R D_t, \quad D_t = \{\mathrm{diag}(\boldsymbol{h_t})\}^{1/2} \tag{1}$$

$$\boldsymbol{h_t} = \boldsymbol{h_t}(\theta) = \boldsymbol{\omega} + \sum_{s=1}^{q} \boldsymbol{A}_s \boldsymbol{\varepsilon}_{t-s} + \sum_{v=1}^{p} \boldsymbol{B}_v \boldsymbol{h}_{t-v} \tag{2}$$

Estimation of CCC models

- Under the assumption that each conditional variance is specified as a function of its own lags and the i^{th} element of ε_t (denoted by ε_{it}), for example, by a GARCH(1,1) equation, an important simplification is obtained in QML estimation
- This assumption splits the log-likelihood function into two parts

$$
\begin{aligned}
l_n(\theta) = \log L_n(\theta) &= -\frac{1}{2}\sum_{t=1}^{n}(2\log|D_t| + \log|R| + z_t' R z_t) \\
&= -\frac{1}{2}\sum_{t=1}^{n}(2\log|D_t| + z_t' z_t) \\
&\quad -\frac{1}{2}\sum_{t=1}^{n}(\log|R| + z_t' R z_t - z_t' z_t)
\end{aligned}
$$

- The parameters of the conditional variances appear only in D_t (first term), while the parameters of the conditional correlation matrix R_t appear only in the second term

Estimation of CCC models

- So the estimation can be performed in two steps
  - Estimate univariate GARCH models for each asset $i = 1, \ldots, k$ and construct standardized residuals

$$z_t = D_t^{-1} \varepsilon_t$$

  - In a second step, estimate the correlation model (i.e. the constant conditional correlations) based on

$$\mathbb{E}(z_t z_t') = D_t^{-1} H_t D_t^{-1} = R,$$

    where R is symmetric and positive definite

- Remark: The separate estimation of each conditional variance model and of the correlation model is the key to enable estimation of MGARCH models of conditional correlations when k is large, where large means more than, say, 5
- Remark 2: The price to pay for this is the impossibility of including spillover terms in the conditional variance equations, i.e. terms involving $\varepsilon_{t-1,j}$ or $h_{t-1,j}$ for $j \neq i$

Estimation of CCC models

- A QMLE of θ is defined as a measurable solution $\hat{\theta}_n = \arg\underset{\theta\in\Theta}{max} L_n(\theta)$
- Under several assumption the following asymptotic properties of the QMLE estimator can be established (Francq and Zakoïan, 2010)
  - Strong consistency

$$\hat{\theta}_n \to \theta_0 \; almost \; surely \; when \; n \to \infty$$

  - Asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1} I J^{-1}),$$

    where J is a positive definite matrix and I is a positive semi-definite matrix, defined by

$$I = \mathbb{E}\Big(\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'}\Big), \quad J = \mathbb{E}\Big(\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\Big)$$

CCC models

- The hypothesis of CCCs is not tenable except for specific cases and short periods
- Several tests of the null hypothesis of constant correlations exist: see Longin and Solnik (1995), Tse (2000), Engle and Sheppard (2001), Bera and Kim (2002a), and Silvennoinen and Terasvirta (2005). The tests differ because of the specification of the alternative hypothesis
- Indeed, many empirical work show that the matrix R is time-varying

$$H_t = D_t R_t D_t,$$

with R_t measurable with respect to the past variables $\{\varepsilon_u, u < t\}$

- Dynamic conditional correlations GARCH (DCC-GARCH) of Engle et Sheppard (2001) is the most well known multivariate approach introducing dynamics for the conditional correlation
- For reasons of parsimony, it seems reasonable to choose diagonal matrices $\boldsymbol{A}_s$ and $\boldsymbol{B}_v$ as discussed on slide 80 regarding the definition of CCC models (on slide 76), corresponding to univariate GARCH models for each component

DCC models

- Dynamic conditional correlations GARCH models are an extension of CCC-GARCH, obtained by introducing a dynamic for the conditional correlation (Engle 2002)

Definition

The DCC process is a martingale difference sequence ε_t relative to a given filtration $\mathcal{I}_t$, whose conditional covariance matrix $H_t = cov(\varepsilon_t|\mathcal{I}_{t-1})$ satisfies

$$H_t = D_t R_t D_t$$

where $D_t = diag(h_{11,t}^{1/2} \dots h_{kk,t}^{1/2})$ and R_t is a $k \times k$ **time varying** *correlation matrix of z_t.*
Besides, $h_{ii,t}$ is defined as univariate GARCH(p, q) model where the usual restrictions for non-negativity and stationarity are imposed.

- The univariate GARCH models can have different orders
- The number of parameters to be estimated is quite large when k is large (e.g. equal to $(k+1)(k+4)/2$ in bivariate case for a DCC(1,1))

DCC models

Different DCC(1,1) models are obtained depending on the specification of R_t

- Simple GARCH-like formulation

$$R_t = \theta_0 R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1},$$

  with R a constant correlation matrix, and Ψ_{t-1} the empirical correlation matrix of $z_{t-1}, \dots, z_{t-M}$

- $$R_t = \text{diag}(Q_t)^{-1/2} Q_t \, \text{diag}(Q_t)^{-1/2},$$

  where

$$Q_t = \theta_0 \bar{Q} + \theta_1 z_{t-1} z'_{t-1} + \theta_2 Q_{t-1}$$

  with $\theta_1 > 0$, $\theta_2 > 0$, $\theta_1 + \theta_2 < 1$, $\theta_0 = 1 - \theta_1 - \theta_2$, and $\bar{Q} = cov(z_t z'_t)$

- One can test the assumption of constant conditional covariance matrix through the restriction $\theta_2 = \theta_3 = 0$
- Both ensure that H_t is positive definite if R_t is positive definite with elements in the unit cercle. For this, Q_t and its initial value have to be positive definite.

General DCC (M,N)

$$\varepsilon_t = H_t^{1/2} \tilde{z}_t$$
$$H_t = D_t R_t D_t$$

where

$$D_t = diag(h_{11,t}^{1/2} \dots h_{kk,t}^{1/2}) \tag{3}$$
$$R_t = \mathrm{diag}(Q_t)^{-1/2} Q_t \, \mathrm{diag}(Q_t)^{-1/2}, \tag{4}$$
$$Q_t = (1 - \sum_{i=1}^{M} \theta_{1i} - \sum_{j=1}^{N} \theta_{2j}) \bar{Q} + \sum_{i=1}^{M} \theta_{1i} z_{t-i} z'_{t-i} + \sum_{j=1}^{N} \theta_{2j} Q_{t-j} \tag{5}$$

- Less parsimonious than DCC(1,1)

Estimation of the DCC model

Suppose that the process z_t is multivariate Gaussian distributed such that $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t z_t') = I$.

- The DCC model can be estimated by a two-step procedure as the conditional variance $H_t = D_t R_t D_t$ can be divided into volatility part and correlation part (Engle 2002)
- The method is thought to produce consistent but not efficient estimators
- The log-likelihood takes the form of

$$
\begin{aligned}
l_n(\theta) &= -\frac{1}{2}\sum_{t=1}^{n}\left(\log(|H_t|) + \varepsilon_t^{'} H_t^{-1}\varepsilon_t\right) \\
&= -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|R_t|) + \varepsilon_t^{'} D_t^{-1} R_t^{-1} D_t^{-1}\varepsilon_t\right)
\end{aligned}
$$

Estimation of the DCC model

- In the first step the likelihood involves replacing R_t with the identity matrix I

$$l_{1,n}(\theta_a) = -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|I|) + \varepsilon_t^{'} D_t^{-1} I^{-1} D_t^{-1}\varepsilon_t\right)$$
$$= -\frac{1}{2}\sum_{i=1}^{k}\sum_{t=1}^{n}\left(\log(h_{ii,t}) + \frac{\varepsilon_{ii,t}^2}{h_{ii,t}}\right),$$

  where θ_a corresponds to the vector of parameters of the univariate GARCH model for all returns series

- Once θ_a is estimated, $h_{ii,t}$ is estimated such that z_t and $\bar{Q}$ can be estimated as well

- In the second step, $\theta_b = (\theta_1, \theta_2)$ is estimated, given the estimated parameters from step one

$$l_{2,n}(\theta_b|\hat{\theta}_a) = -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|R_t|) + \varepsilon_t^{'} D_t^{-1} R_t^{-1} D_t^{-1}\varepsilon_t\right)$$
$$= -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|R_t|) + z_t' R_t^{-1} z_t\right)$$

Estimation of the DCC model

- Asymptotic properties of the two-step estimation procedure have been studied in Engle and Sheppard (2001)
- However, Aielli (2009) showed that the estimation of Q by $\hat{R}$ is **inconsistent** since

$$\mathbb{E}(z_t z_t) = \mathbb{E}(\mathbb{E}(z_t z_t' | \mathcal{I}_{t-1})) = \mathbb{E}(R_t) \neq \mathbb{E}(Q_t)$$

- The consistent DCC (cDCC) relies on a consistent specification of Q_t

$$Q_t = (1-\theta_1-\theta_2)\bar{Q} + \theta_1 \operatorname{diag}(Q_{t-1}^{1/2}) z_{t-1} z_{t-1}' \operatorname{diag}(Q_{t-1}^{1/2}) + \theta_2 Q_{t-1},$$

such that $\bar{Q}$ is the unconditional covariance matrix of $\operatorname{diag}(Q_{t-1}^{1/2}) z_t$

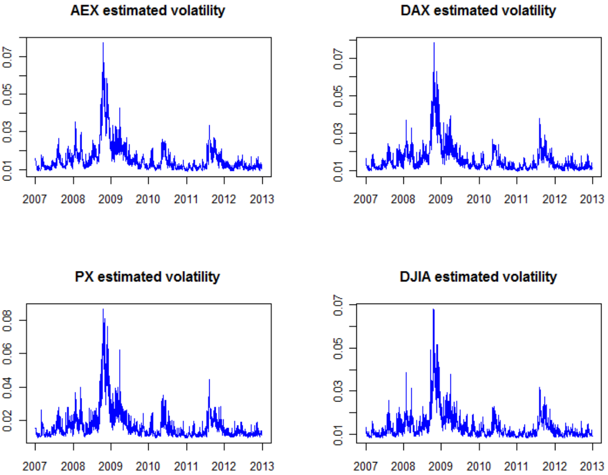
Multivariate GARCH models: example

- Returns on 4 stock market indices: AEX, DAX, PX and DJIA from January 2007 to December 2012

Unconditional Correlation coefficients of the returns series

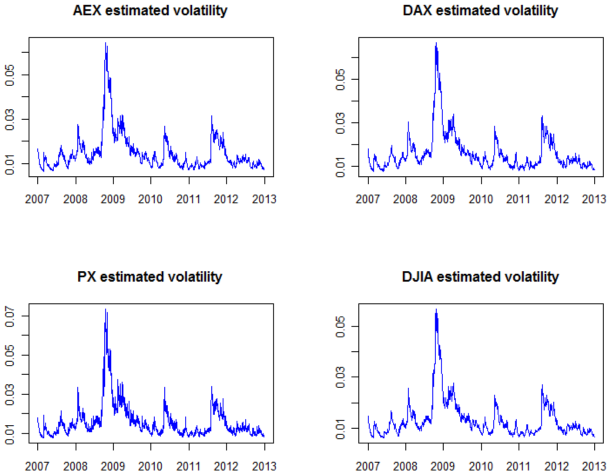
	AEX	DAX	PX	DJIA
AEX	1			
DAX	0.8568444	1		
P	0.5330840	0.4924072	1	
DJIA	0.5630591	0.6086716	0.3260289	1

Multivariate GARCH models: example



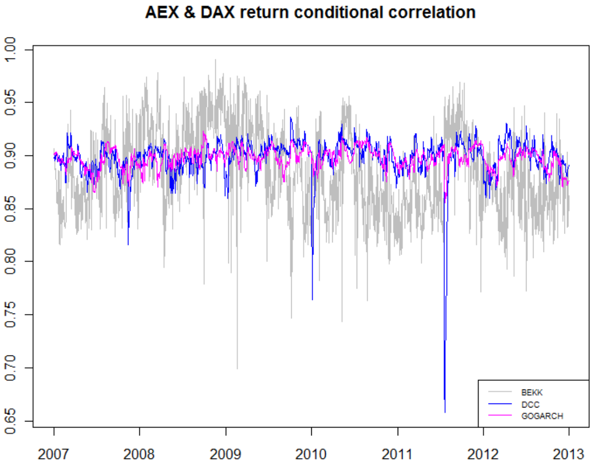
- BEKK model

Multivariate GARCH models: example

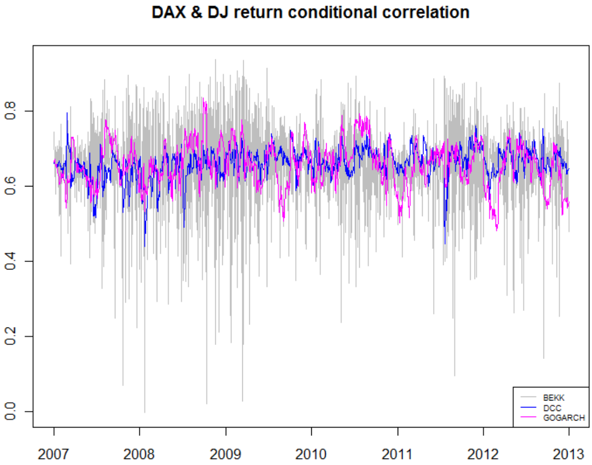


- DCC model: smoother volatilities

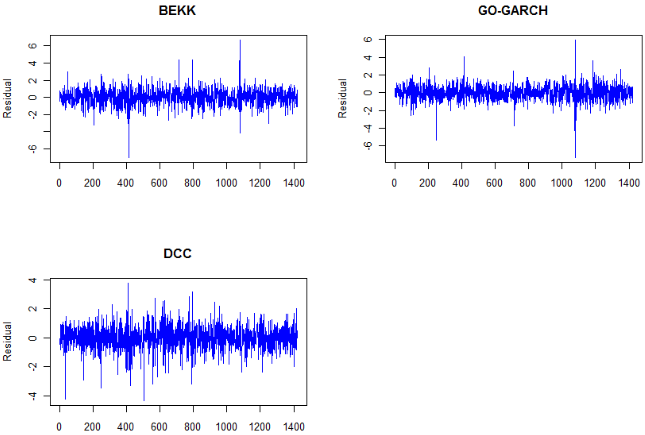
Multivariate GARCH models: example



Multivariate GARCH models: example

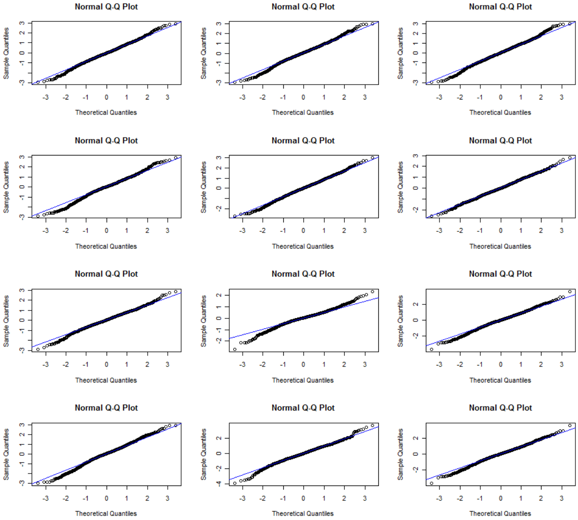


Multivariate GARCH models: example



- Residuals for AEX (similar for the other three series)

Multivariate GARCH models: example



- Residual QQ plots (for each of the four series AEX, DAX, PX, DJIA using BEKK, GO-GARCH and DCC model)