2018, March 1st-13th **John Von Neumann Institute**

Master QCF

Financial Time Series, 20 hours+practical exercises.

Plus Corrections of exercises.

Forecasting discipline is an issue of Statistics. Indeed, the aim is to answer the following kind of problem: a system X is evolving in time, it is observed and one would like to predict the future. In our case, we are interested in financial data, for instance price processes modeling assets price. Anyway, practical observations tell us that the interesting matter is not the price process, but the RETURN processes, and mainly, their covariance function as a risk measure (cf. volatility in continuous time models). Actually, we do not try to model the price processes, but the risk concerning the returns, meaning we look for a model fitting the second order moments (meaning covariance function of return processes, which is more or less a risk measure). Generally, underlies a modeling problem: it is to find the mathematical "model" that realizes the better connection between a variable and the time.
The principle is to find a mathematical model fitting the covariance function (namely γ) as a function of time. Given the available observations, we try the "best" function f (the optimality criterion depending on the method) such as $\gamma \approx f(t)$ where t is time. Namely, we consider that the observations are a set $(X(t-i), i = 1, \cdots, n)$. Concerning Financial data, the more convenient models are ARCH and GARCH, due to stylised features.

Indeed, the financial data present some stylised facts:

- non stationarity of price series,
- absence of auto correlation for the price variations,
- unpredictability of returns,
- auto correlation of the squared price returns,
- volatility clustering $\Rightarrow$ prediction of squared returns,
- fat tailed distributions (leptokurticity),
- leverage effects,
- seasonality.

For all these facts, ARMA are not convenient for modeling price series. But, the correlation process, important since it is a measure of the risk, meaning more or less the volatility, could be modeled as a GARCH process. Anyway, we will start which basic definitions in ARMA area; as a first step, we will present processes ARMA which is a tool to model the covariance function $\gamma(k,j) = cov(X_k, X_j)$.

There is two parts: first one concerns ARMA processes, linear models:

- Processes **ARMA**: Box and Jenkins'methods, general features (sophisticated methods, where is exhibited a linear function of $X(t)$ and its past values $X(t-i)$, $i = 1, \cdots, n$).
- Delay operator, ARMA equations.
- Estimate of an ARMA process, covariance function.
- ARMA model identification, estimation of its parameters.

The second part concerns GARCH processes, convenient to model the correlation function, meaning the volatility in discrete time case. Processes **ARCH, GARCH** (cf. Gouriéroux, Zakoian) are similar to the previous one but the functions are no more linear. For instance $\forall t$, $X(t) = z_t \sqrt{\alpha_0 + \alpha_1 X(t-1)}$.

- Some non linear models
- Linear ARCH-GARCH models
- Identification
- Estimates and forecasting
- Tests based on the likelihood
- Some extensions
- Financial Applications.

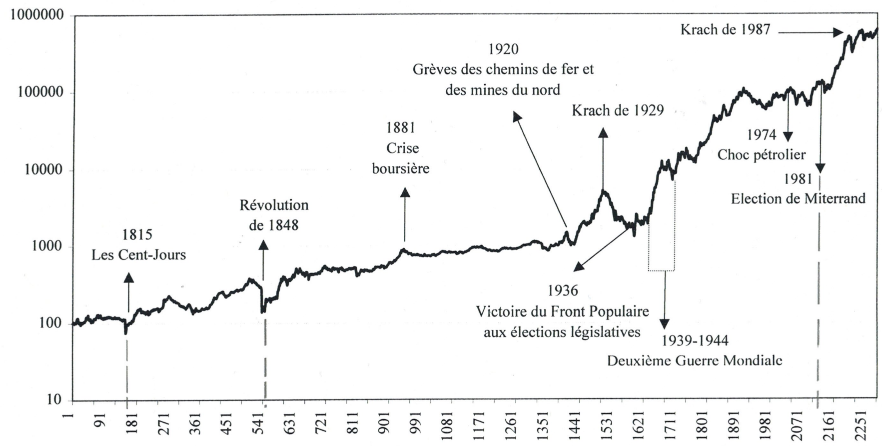
A selection criterion is obviously the quality of the forecast. We will propose statistical tests that allow to judge the goodness of fit (between the curve obtained and the observations). An empirical way could be added: to reserve some "witnesses spots" and to do the study, excluding them, and judging the error on witnesses.

For the concrete use of these methods it is recommended to use the free software "R":

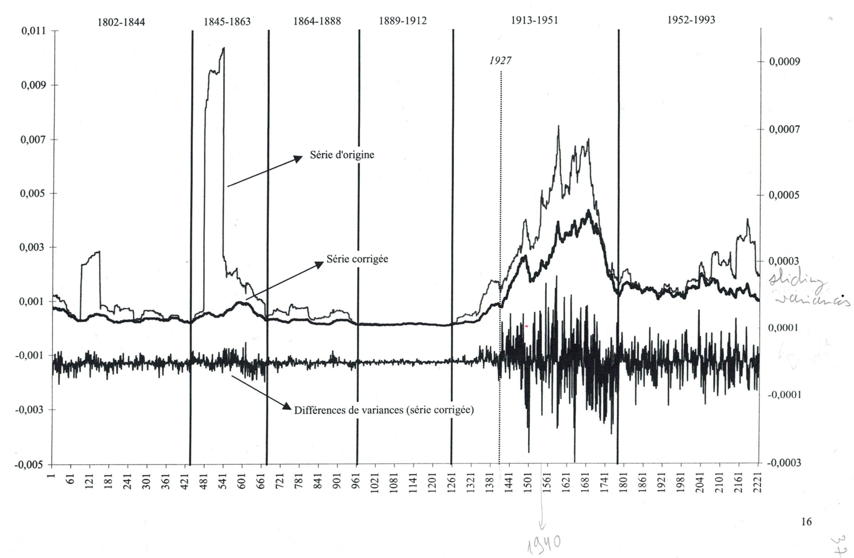
https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Some data can be found via Internet. For instance, historical prices, assets daily returns, on "yahoofinance" to get SP500, as a column Excel.

Séries mensuelles des cours des actions



Graphique 6
Variances glissantes (60 mois) et différences de variances



16

1 Box and Jenkins' methods, general features

Developed in the 70s, these are very powerful methods which make maximum use of the fact that the evolution of the studied time series is considered as **one** of the achievements of a stochastic process, endowed with a strong enough structure. Indeed, once highlighted the structure, this allows to predict more confidently the future series. The consideration is the need for a fairly long period of observations for the forecast being reliable. The authors recommend 5 **to** 6 **periods** in the case of periodic phenomena, and a minimum of 30 observations in other cases.

These methods work very well for short-term forecasts macroeconomic series, especially for the industrial production indexes. In Finance, this method does not concern the forecast of returns, but the one of volatility.

They are based on the assumption that each observation depends quite strongly on previous observations. Basically, this addiction to the past replaces multiplicity of observations (in Statistics) to estimate the settings by applying the law of large numbers. So are assumed strong enough assumptions, that the series is stationary, meaning the two first moments do not depend on time. If this is not the case, they must be done "stationary" by transformations (called filters) that remove trend and seasonality.

1.1 Definitions

Thus, we consider processes, random series, indexed in $\mathbb{Z}$ and taking their values in $\mathbb{C}$ (complex numbers) *but we restrict this course to real numbers* $\mathbb{R}$:

$$\forall n \in \mathbb{Z},\ X_n \text{ is a random variable } : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}).$$

We try to model the application $n \mapsto X_n$ with a trend part, a seasonal component, and measurement error.

Hypothesis: The observations x_n are the values of a centered, square-integrable, stationary, random process X_n, i.e. there exists a function γ on Z such that $\forall n$, $cov(X_n, X_{n-k}) = \gamma(k)$. Notice:

$$cov(X_n, X_{n-k}) = E[(X_n - E(X_n))(X_{n-k} - E(X_{n-k}))].$$

Remark 1.1 *Exercise 1: Actually for any $k \in \mathbb{Z}$, $\gamma(k) = \gamma(-k)$.*

Definition 1.2 *: Such a process is called a* **second order stationary time series***, S.T.S. for short.*

The function γ is called the **auto covariance function***.*

Moreover we define the **auto correlation function** $\rho : k \mapsto \rho(k) = \frac{\gamma(k)}{\gamma(0)}$.

There exists another notion: “strict stationarity” meaning the vectors $(X_1, \cdots, X_k)$ and $(X_{n+1}, \cdots, X_{n+k})$ have the same law, for any pair (k, n).

As for the covariance function γ, for any $k \in \mathbb{Z}$, $\rho(k) = \rho(-k)$ and we define the **correlogram**, graph of the application ρ, useful tool in analyzing the series as discussed later.

We also introduce:

Definition 1.3 *The* **partial auto correlation function**, *P.A.C.F., is defined on* $\mathbb{N}$ *as:*

$$r: \ \mathbb{N} \to \mathbb{R} \ ; \ r(p-n) = \ cor\ (X_n, X_p / X_{n+1}, \cdots, X_{p-1}), p > n,$$

meaning

$$r(p-n) = \frac{cov\ (X_n - X_n^*, X_p - X_p^*)}{\sqrt{Var\ (X_n - X_n^*)\ \ Var\ (X_p - X_p^*)}}$$

where X_j^* *is the orthogonal projection of* X_j *on the vector space* $S_{n,p}$ *generated by* $(X_{n+1}, \cdots, X_{p-1})$, *and completed by* $r(1) = \rho(1)$.

Exercise 2: this expression only depends on $(p-n)$.

Finally, we introduce the infinite dimensional matrix of variance-covariance process X.

Definition 1.4 *: The* **Toeplitz matrix** *is*

$$\Gamma, \ \gamma(i,j) = r(i-j), \ i, j \geq 1.$$

This is a symmetric matrix.

1.2 Examples of second order stationary times series, STS

First example of fundamental S.T.S. : the white noise.

Definition 1.5 *The (weak)* **white noise** *is a STS* (ε_k) *(with covariance function equal to* γ *with* $\gamma(k) = \sigma^2 \delta_{k,0}$.

If moreover there is independence between the random variables (ε_k), the white noise is said **strong**.
For example, this may be a Gaussian process with covariance matrix $\Gamma = \sigma^2 I_d$; in this case, there is in addition the orthogonality of the white noise components ε_n in L^2 and their independence, thanks to the Gaussian nature of the series.
A *strong* white noise is a white noise such that (ε_n) are i.i.d.

Remark 1.6 *We can show that the white noise covariance function checks the equality*

$$\gamma(n) = \frac{\sigma^2}{2\pi} \int_0^{2\pi} e^{in\lambda} d\lambda.$$

This "white noise process" is used to model the measurement error. If the series is not centered, the term is "colored noise".

Second example:

Definition 1.7 *A* **moving average** *is a STS as follows:*

$$X_n = \sum_{k \in Z} a_k \varepsilon_{n-k},$$

where the series $(a_k; k \in Z) \in l^2$ and ε is a white noise.
For short: M.A.= "moving average".

Proposition 1.8 *The covariance function of a moving average $X_n = \sum_{k \in Z} a_k \varepsilon_{n-k}$ is written as $\gamma(p) = \sum_{k \in Z} a_{p-k} a_{-k}$ $\forall p \in \mathbb{Z}$.*

Proof : We write X_n and X_{n-p} definition; firstly remark that these series are L^2 convergent using the hypothesis that the series $(a_k; k \in Z) \in l^2$. Secondly we compute their covariance, meaning the mean of the product since these random variables are centered:

$$E[X_n X_{n+p}] = \lim_{K \to \infty} \sum_{|k|<K} a_{p+k} a_k.$$

This limit exists since

$$\forall K > 0, \quad (\sum_{|k|<K} a_{p+k} a_k)^2 \leq \sum_{|k|<K} |a_{p+k}|^2 \sum_{|k|<K} |a_k|^2 < \infty.$$

This inequality is proved recursively: it is true for $K = 2$, and the property for $K-1$ implies it for K. •

Definition 1.9 *When there exists a finite number of non null coefficients a_k, i.e. $(a_0, \cdots, a_p)$, we say that X is a* **order p-moving average***, MA(p) for short.*

Third example: let ε be a white noise, and define the recursive series

$$X_n = \alpha X_{n-1} + \varepsilon_n.$$

Assuming that we know a particular element of the series, for instance X_0, assuming it is a centered random variable in L^2 we prove the following.

Proposition 1.10 *Let X be the process defined as*

$$X_n = \alpha X_{n-1} + \varepsilon_n, \; \forall n \in \mathbb{Z}, \; X_0 \in L^2, E[X_0] = 0.$$

Assuming $|\alpha| < 1$, and $E[X_n^2] \leq M^2$, $\forall n \in \mathbb{Z}^-$, then X is a STS.

Specifically, this is a moving average with coefficients $a_j = \alpha^j$, $j \geq 0$. Its covariance function is defined by $\gamma(k) = \frac{\alpha^k}{1-\alpha^2}$.

Proof = Exercise 3

Definition 1.11 *An order* 1 *auto regressive series* X *(*$AR(1)$ *for short) is a process depending only of the previous observation, step by step.*

At this point we can quote Francq and Zakoian [6] pp 7-11: Sections 1.3 Financial Series and Section 1.4 Random variance models which shows how ARMA processes are not appropriate to model Financial Series as it is written above in the introduction:

Indeed, once again, the financial data present some stylised facts:

- non stationarity of price series,
- absence of auto correlation for the price variations,
- unpredictability of returns,
- auto correlation of the squared price returns,
- volatility clustering $\Rightarrow$ prediction of squared returns,
- fat tailed distributions (leptokurticity),
- leverage effects,
- seasonality.

First Part: ARMA

2 Delay Operator, ARMA equations

In this subsection we consider that X is a STS. In AR(1) example, $X_n = aX_{n-1} + \varepsilon_n$ and $\forall(\varepsilon_n)$ (a given white noise) we get X_n as a function of X_{n-1}; more generally it is interesting to get formal this passage from $n-1$ to n. Firstly we have to define the spaces on which is defined this passage.

Definition 2.1 *The closed subspace generated by the set $\{X_p, p \in \mathbb{Z}, p \leq n\}$ in L^2 is denoted as H_n^X.*
This subspace of L^2, H_n^X, is named the **linear past** *of X.*
We note also:

$$H_{-\infty}^X = \cap_n H_n^X \ ; \ H_{+\infty}^X = \cup_n H_n^X = H^X.$$

$H_{-\infty}^X$ is named the asymptotic past, H^X the linear envelope.

These spaces are used to characterize two specific types of STS.
Following Francq and Zakoian [6] page 4, we consider $\varepsilon_{\mathbf{n}} := \mathbf{X_n} - \mathbf{P_{n-1}}(\mathbf{X_n})$, weak or strong white noise, where P_{n-1} is the L^2 orthogonal projector on H_{n-1}^X.

Definition 2.2 *When $H_{-\infty}^X = \{0\}$ the series is* **regular**.
When $H_{-\infty}^X = H^X$ the series is **singular**. *In this case, the linear pasts are constant and the "innovation" does not bring any information.*

A first example of regular STS is the white noise. Other examples in Section 4.4 (Exercise 6).
Actually because the process ε is non correlated, the vector space $H_n^\varepsilon = \mathbb{R}\varepsilon_n + H_{n-1}^\varepsilon$. So if $Y \in H_n^\varepsilon \cap H_{n-1}^\varepsilon$, firstly, $Y = a\varepsilon_n + P_{n-1}^\varepsilon(Y)$. But $Y \in H_{-\infty}^\varepsilon$ means that $Y \in H_{n-1}^\varepsilon$, so $a = 0$. And so on, $Y = 0$ and ε is a regular series.

Definition 2.3 *The operator $H^X =$ vect $\{X_n,\ n \in Z\}$ in L^2 which sends X_n to X_{n-1} is named the* **delay operator** *denoted S^X : $S^X(X_n) = X_{n-1}$.*

Proposition 2.4 *The operator S^X is the unique isometric from H^X to H^X which sends X_n to X_{n-1}. Moreover, $S^X(H^X) = H^X$.*

Proof : The operator S^X is defined on the $\{X_n, n \geq 0\}$ and is extended by linearity on any finite linear combinations of X_n. This is an isometric:

$$\begin{aligned}\| S^X(\sum_i a_i X_i) \|_2^2 &= \sum_{i,j} a_i a_j E[X_{i-1}X_{j-1}] \\ &= \sum_{i,j} a_i a_j \gamma(i-j) = \| \sum_i a_i X_i \|_2^2 .\end{aligned}$$

Thus we could extend this operator S^X by continuity on the whole H^X.
Uniqueness: it is a consequence of the fact that if T could be another solution, $T = S_X$ on any X_n, so on any finite linear combinations of X_n so by continuity on H^X.
Any element of H^X is a limit of finite linear combinations of X_n, image by S^X of finite linear combinations of X_n, so the equality $S^X(H^X) = H^X$. •

Theorem 2.5 *(WOLD): Any STS could be written as a unique sum of a regular and a singular parts:*

$$X = X^r + X^s$$

so that the spaces H^{X^r} and H^{X^s} are L^2 orthogonal.

Proof : Exercise 4

Proposition 2.6 *Both series X^r and X^s are too STS.*

Proof : Firstly by construction they are centered and in L^2.
Secondly we use the following:

Lemma 2.7 *For all $n \in Z$, $P_n^X \circ S^X = S^X \circ P_{n+1}^X$.*

Proof for all $p \in Z$, $P_n^X \circ S^X(X_p) = P_n^X(X_{p-1})$ is the unique vector in H_n^X such that $X_{p-1} - P_n^X(X_{p-1})$ is orthogonal to H_n^X. So we have to compute $\forall k \leq n$ the scalar product $\langle X_k, X_{p-1} - S^X \circ P_{n+1}^X(X_p)\rangle$. This scalar product is equal to:

$$\begin{aligned}\langle X_k, X_{p-1} - S^X \circ P_{n+1}^X(X_p)\rangle &= \gamma(k-p+1) - \langle S^X(X_{k+1}), S^X \circ P_{n+1}^X(X_p)\rangle \\ &= \gamma(k-p+1) - \langle X_{k+1}, P_{n+1}^X(X_p)\rangle\end{aligned}$$

since S^X is an isometry. Then we use $\forall k \leq n, X_{k+1} \in H_{n+1}^X$. Yields:

$$\langle X_k, X_{p-1} - S^X \circ P_{n+1}^X(X_p)\rangle = \gamma(k-p+1) - \langle X_{k+1}, P_{n+1}^X(X_p)\rangle = \gamma(k-p+1) - \langle X_{k+1}, X_p\rangle = 0.$$

•

We apply this lemma to the computation of the covariance function of the series X^s, whith $n \geq p$:

$$\begin{aligned}(X_n^s, X_p^s) = (P_{-\infty}^X(X_n), P_{-\infty}^X(X_p)) &= (S^X \circ P_{-\infty}^X(X_n), S^X \circ P_{-\infty}^X(X_p)) = \\ (P_{-\infty}^X \circ S^X(X_n), P_{-\infty}^X \circ S^X(X_p)) &= (P_{-\infty}^X(X_{n-1}), P_{-\infty}^X(X_{p-1}))\end{aligned}$$

which is exactly (X_{n-1}^s, X_{p-1}^s) by definition of X^s, step by step we go to

$$(X_n^s, X_p^s) = (X_{n-p}^s, X_0^s),$$

which only depends on the difference $n-p$; this proves the stationarity of the series (X^s).
Then, the part $X^r = X - X^s$ is too a STS: $X^r \in L^2$ with null expectation by linearity, and we easily check the stationarity of $E[(X_n^r, X_p^r)]$. More specifically we prove:

$$(X_n - X_n^s, X_p - X_p^s) = \gamma(n-p) - \gamma^s(n-p).$$

This shows the stationarity of X^r and the relation between the covariance functions $\gamma = \gamma^r + \gamma^s$. •

Remark 2.8 *When a STS is not singular, the strict inclusion* $\forall n$, $H^X_{n-1} \subset H^X_n$ *is satisfied. Indeed, if not, there exists* n *such that* $H^X_{n-1} = H^X_n$*, and with the lemma and the delay operator* S^X *we deduce that* $\forall n$, $H^X_{n-1} = H^X_n$*, so the series is singular.*

The following theorem provides a characterization of regular series.

Theorem 2.9 *A series* X *is regular if and only if there exists a sequence* (d_n) *in* $l^2(C)$ *and a white noise* ε *such that:*

$$X_n = \sum_{p \geq 0} d_p \varepsilon_{n-p}.$$

We could choose ε *so that the linear pasts of* X *and* ε *are identical; then this white noise and the associated sequence* (d_n) *are unique, except a possible multiplicative coefficient.*

Definition 2.10 *This white noise is named* **innovation white noise**.

The interest of such series lies in the following corollary: the projection on the past is so extremely simple.

Corollary 2.11 *Let* X *be a regular series and* ε *its innovation white noise; for all* $m \leq n$,

$$P^X_m(X_n) = \sum_{p \geq n-m} d_p \varepsilon_{n-p}.$$

Proof of the theorem: By definition $X_n \in H^\varepsilon_n$, so $H^X_n \subset H^\varepsilon_n$, $\cap_n H^X_n \subset \cap H^\varepsilon_n = \{0\}$ since ε is regular, and X is regular.

Conversely, let X be a regular series. Let the process $v_n = X_n - P^X_{n-1}(X_n)$; this is a STS since we could compute its covariance function:

$$\forall n, \quad \| v_n \| = \| S^X(X_{n+1}) - P^X_{n-1} \circ S^X(X_{n+1}) \| = \| X_{n+1} - P^X_n(X_{n+1}) \| = \| v_{n+1} \|$$

denoted $\sigma^2 = \gamma(0)$. By definition, $v_n \in H^X_n$ and is orthogonal to H^X_{n-1} so to the previous v_i: so it is a STS, and more specifically a white noise denoted $a\varepsilon_n$.

By definition, $X_n = a\varepsilon_n + P^X_{n-1}(X_n)$, $\varepsilon_n \in H^X_n$ and is orthogonal to H^X_{n-1}, thus H^X_n is the direct sum $\mathbb{R}\varepsilon_n \oplus H^X_{n-1}$. By induction we get that H^X_n is the direct sum $\oplus_{0 \leq i \leq j} \mathbb{R}\varepsilon_{n-i} \oplus H^X_{n-j-1}$. On this direct sum we get the decomposition

$$X_n = \sum_{0 \leq i \leq j} a_i \varepsilon_{n-i} + P^X_{n-j-1}(X_n)$$

Since X is a regular series, $\lim_{j \to \infty} P^X_{n-j-1}(X_n) = 0$ and X is equal to $\sum_{0 \leq i} a_i \varepsilon_{n-i}$, which is the expected form.

As a consequence, $X_n \in H^\varepsilon_n$ and since previously we knew that, $\varepsilon_n \in H^X_n$, these two spaces are identical.

Uniqueness: we assume that there exists a pair (ε', d'), (white noise, $l^2(C)$ element), solution of the problem, so satisfying

$$\forall n, P_n^{\varepsilon'} = P_n^X = P_n^{\varepsilon} \text{ and } X_n = \sum_{0\le i} d_i \varepsilon_{n-i} = \sum_{0\le i} d'_i \varepsilon'_{n-i}.$$

On both hands of this equality we apply the operator P_{n-1}^X, we get :

$$P_{n-1}^{\varepsilon}(X_n) = \sum_{1\le i} d_i \varepsilon_{n-i}; P_{n-1}^{\varepsilon'}(X_n) = \sum_{1\le i} d'_i \varepsilon'_{n-i}.$$

But $P_n^{\varepsilon'} = P_n^{\varepsilon}$ so the difference is null and $\forall n$, $d'_0 \varepsilon'_n = d_0 \varepsilon_n$ meaning the uniqueness except a possible multiplicative coefficient. •

The proof of the corollary is obvious since the operators P_m^X and P_m^{ε} are the same, as are the corresponding spaces H_m^X and H_m^{ε}.

Remark 2.12 *The identity between these two families of spaces is interpreted as follows: Linear pasts of X and ε coincide. If X is observed up to time $n-1$, the additional information provided by really new X_n is represented by $a\varepsilon_n = X_n - P_{n-1}^X(X_n)$, the 'innovation' as we called it previously.*

More generally, we will now study the class of STS, solution of " ARMA " equations, written using the delay operator S^X.

Definition 2.13 *Let X be a STS and ε a white noise, P and Q two polynomials with complex coefficients. We say that X* **is solution of ARMA(P,Q) equation** *if this process satisfies for any n in Z :*

$$P(S^X)(X_n) = Q(S^{\varepsilon})(\varepsilon_n), \tag{1}$$

meaning there exist complex coefficients $(a_0, \cdots, a_p)$ and $(b_0, \cdots, b_q)$ such that $\forall n \in \mathbb{Z}$:

$$\sum_{i=0}^{p} a_i X_{n-i} = \sum_{i=0}^{q} b_i \varepsilon_{n-i}. \tag{2}$$

In case of $p = 0$, X is **MA(q)** ; in case of $q = 0$, X is **AR(p)**. In the general case we say that X is **ARMA(p,q)**.
Such an equation could be solved, either to get X function of process ε or the converse so that we could "forecast" X_n based solely on its past. Roughly speaking, this consists in a "reverse" of operators $P(S^X)$ and $= Q(S^{\varepsilon})$. This is out of our agenda, but the following Section 2.1 is an important result which will be useful in the second part of this course.

2.1 ARMA Equation: resolution

Let $A_P(X) = A_Q(\varepsilon)$ an ARMA equation.

Theorem 2.14 *(Fejer-Riesz) Let P et Q be non nul polynomials with no common roots, those of P have modulus $\neq 1$. Then the ARMA equation is solvable as soon as the modulus of P roots are > 1 and those of $Q \geq 1$.*

Definition 2.15 *This ARMA representation of X is called* **canonical Fejer-Riesz canonical representation**.

3 Estimate of an ARMA Process covariance function

We come back to the observation of a STS, supposed to be stationary, non necessarily centered:

$$X_1, \cdots, X_N,$$

The first step is to estimate $E(X)$ and covariance function γ_X.

According to standard probability or statistics lecture notes in case of sampling, $E(X)$ is estimated by Cesàro mean, that is justified by the large numbers law (cf. [4]). But the required assumptions are either the independence of the observations or the martingale property for the process. Neither of these assumptions is checked in the case of STS. Nevertheless, with similar proofs to those seen during Probability course, we get same type results. This is what will be used to justify an approximate of mean, covariance function.

Insert work with R: 'plotobs(X)' to draw the series graph; mean(X); acf(X) to get correlogram, variogram, partial correlogram...see TD-TP Agnes Lagnoux.

3.1 Large numbers law

Lemma 3.1 *Let $X_1, \cdots, X_n,\ n \in N$ be a sequence of random variables with mean m. We put $S_n := \sum_{i=1}^n X_i$ and assume:*

$$\exists M > 0,\ Var(X_n) \leq M^2,\ Var(S_n) \leq nM^2,\ \forall n \geq 1.$$

Then $\frac{1}{n}S_n \to m$ in L^2 and almost surely, when n goes to infinity.

Proof :
(i) $Var(\frac{1}{n}S_n) = E[\frac{1}{n}S_n - m]^2$ since by hypothesis $E(S_n) = nm$. But $Var(\frac{1}{n}S_n) = \frac{1}{n^2}Var(S_n) \leq \frac{1}{n}M^2 \to 0$ when n goes to infinity, so the convergence in L^2.
(ii) Let $Z_k = \sup\{|\frac{1}{n}S_n - m|,\ n \in [k^2, (k+1)^2[\}$. We put $Y_j := X_j - m$ so:

$$\frac{1}{n}S_n - m = \frac{1}{n}S_{k^2} + \frac{1}{n}(X_{k^2+1} + \cdots + X_n - nm) = \frac{1}{n}(S_{k^2} - k^2 m + Y_{k^2+1} + \cdots + Y_n).$$

Then we deduce the bound:

$$Z_k \leq \frac{1}{k^2}(|S_{k^2} - k^2 m| + |Y_{k^2+1}| + \cdots + |Y_{(k+1)^2-1}|)$$

so the L^2 norm satisfies:

$$\|Z_k\|_2 \leq \frac{1}{k^2}(\|S_{k^2} - k^2 m\|_2 + \|Y_{k^2+1}\|_2 + \cdots + \|Y_{(k+1)^2-1}\|_2).$$

By hypothesis the first term is bounded by Mk, and any following terms $(k+1)^2-1-k^2 = 2k$ are equal to the X standard deviation bounded by M:

$$\|Z_k\|_2 \leq \frac{1}{k^2}(Mk + 2kM) = 3M/k.$$

Thus the series $E(\sum_k Z_k^2) = \sum_k E(Z_k^2) \leq \sum_k 9M^2/k^2$ is convergent, proving that Z_k converges almost surely, when k goes to infinity, exactly meaning $\frac{1}{n}S_n - m$ converges almost surely to zero, meaning $\frac{1}{n}S_n$ converges almost surely to m when n goes to infinity. •

We apply this lemma to a STS: since $Var(X_n) = \gamma^X(0)$ the first hypothesis is satisfied. The second hypothesis concerns

$$Var(S_n) = Var(\sum_{i=1}^{n} X_i) = \sum_{1 \leq i,j \leq n} \gamma^X(i-j) = n\gamma^X(0) + 2(n-1)\gamma^X(1) + \cdots + 2\gamma^X(n-1)$$

the order of which not necessarily being nM.
But for instance a MA(q) process satisfies this hypothesis since in this case there exists a finite number of non null $\gamma^X(i)$, $\gamma(k) = 0$ for all $k > q$:

$$Var(S_n) \leq n(\gamma(0) + \cdots + \gamma(q)).$$

Exercise 7: under the assumption of the lemma above, in case of an AR(1), $X_n = aX_{n-1} + \varepsilon_n$ prove that the covariance is $\gamma^X(k) = \frac{a^k}{1-a^2}$.

3.2 Covariance function estimate, *acf, pacf*

Let k be fixed in $\mathbb{N}$ (if $k < 0$, $\gamma(k) = \gamma(-k)$). Using the large numbers law (or rather Lemma 3.1), if the series $Y : n \to X_n X_{n+k}$ has "good" properties, a $\gamma^X(k)$ reasonable estimate is:

$$\bar{\gamma}_n(k) = \frac{1}{n}\sum_{j=1}^{n} X_j X_{j+k}.$$

For that remark that we need observations at least from time 1 to $n+k$.
If we have only n observations, we propose:

$$\gamma_n^*(k) = \frac{1}{n}\sum_{j=1}^{n-k} X_j X_{j+k}.$$

Both estimates have the following properties:

(i) **Bias**

$$E[\bar{\gamma}_n(k)] = \gamma(k),$$

meaning this estimate has a null bias $\forall n$.

$$E[\gamma_n^*(k)] = \frac{n-k}{n}\gamma(k) \to \gamma(k),$$

this estimate bias is asymptotically null.
Exercise 8: Applying Lemma 3.1, compute the biais of these both estimates.

(ii) **Convergence and quadratic error**: here we need more hypotheses. To apply Lemma 3.1, $E(X_n X_{n+k}) = \gamma(k)$ but we also need the existence of a constant M such that $Var(X_n X_{n+k}) \leq M^2$ and $Var(\sum_{i=1}^n X_i X_{i+k}) \leq nM^2$ meaning we would need at least $X \in L^4$ and $\sup_n E(X_n^4) \leq M^2$. Now to check the second hypothesis:

$$\sum_{1\leq i,j\leq n} E[X_i X_{i+k} X_j X_{j+k}] - n^2\gamma^2(k) \leq nM^2$$

we could (for instance) assume that the series law is Gauss.

Anyway, since it is a stylized fact that price processes are not Gaussian, we can not use such an hypothesis.
But even if we can not assume normal law, we nevertheless get:

Proposition 3.2 *(cf. [1], p. 104) Let X be a STS in L^4 such that $\sup_n E(X_n^4) \leq M^2$ and*

$$\lim_{|n-m|\to\infty} [E[X_n X_{n+k} X_m X_{m+k}] - \gamma^2(k)] = 0.$$

Then $\bar{\gamma}_n(k) \to \gamma(k)$ in L^2.

Proof : Exercise 9

The following is to skip for a first lecture.

Definition 3.3 *A sequence of real random variable $(X_n, n \in Z)$ is said to be* **Gauss** *when any real linear combination of X_i follows a Gauss law.*

Remark that in such a case the vector space H^X contains only Gauss variables since the Gauss laws are preserved under L^2 convergence.

In this case we get the following for $\gamma^*(k)$.

Proposition 3.4 *Let X be a Gaussian STS with spectral density = restriction to Π of a continuous function on $\mathbb{R}$ with period 2π, then $\forall k$: $\gamma_n^*(k)$ goes to $\gamma(k)$ almost surely and in L^2.*

Admit the proof.

•

Theorem 3.5 *Let X be a Gaussian centered STS with covariance function γ such that:*

$$\sum_{k=1}^{\infty} k|\gamma(k)| < \infty.$$

Then $\forall$ fixed $k \in \mathbb{N}$ the vector

$$\sqrt{n}(\gamma_n^*(0) - \gamma(0), \cdots, \gamma_n^*(k) - \gamma(k))$$

weakly converges to the centered Gauss law with covariance matrix Γ:

$$\Gamma_{ij} = \sum_{k \in Z} \gamma(k)\gamma(k+i+j) + \gamma(k)\gamma(k+i-j). \tag{3}$$

The proof is easy but tedious. Look at the details for instance in [1] pages 111 et sq.

Remark 3.6 *As a consequence of this theorem we could notice:*

$$\mathbb{P}\{\sqrt{n}|\gamma_n^*(k) - \gamma(k)| \leq \alpha\} \to \int_{-\alpha}^{+\alpha} f^\Gamma(x)dx.$$

So we could get a confidence interval for the parameter $\gamma(k)$, with confidence level ε deduced from α ($\varepsilon = \int_{-\alpha}^{+\alpha} f^\Gamma(x)dx$):

$$\gamma(k) \in]\gamma_n^*(k) - \frac{\alpha}{\sqrt{n}}, \gamma_n^*(k) + \frac{\alpha}{\sqrt{n}}[.$$

(iii) **Comparison** between $\bar{\gamma}$ and γ^*: In the case where $\sup_n E(X_n^4) \leq M^2$ when $n \to \infty$, k being fixed, we get, Exercise 9:

$$\|\bar{\gamma}_n(k) - \gamma_n^*(k)\|_2 \leq \frac{k}{n}M \to 0.$$

Routines R: acf, pacf, to give an example.

4 ARMA model Identification, estimation of its parameters

Cf. Chapter 5.2 [6].
We assume that the changes in the time series (differentiation, seasonal fitting) have been made so that we have an effective centered STS, and that the obtained series is real, with **a rational spectrum** meaning that there exists p and $q \in \mathbb{N}$, polynomials P degree p and Q degree q, a white noise ε such that the series X is solution to the ARMA equation $A_P X = A_Q \varepsilon$.

The aim is to find p, q, P, Q to identify the model. We have n observations of X and we suppose that the covariance function γ is known, actually estimated according to the method provided in Section 3.2.
R command: arima, monmodele= ; X= ; with model parameters, simulation of processes, plotobs(X) ; mean(X) ; acf(X) which gives correlogram, variogram; pacf(X), etc.

4.1 Estimation of P coefficients

Hypothesis : we suppose that p, q are known in $\mathbb{N}$ and function γ is known, we put $a_0 = 1$.
(p, q) is minimal, meaning there does not exists polynomials P' and Q' with smaller degrees than p, q in the ARMA equation.
We detail the ARMA equation $A_P X = A_Q \varepsilon$:

$$\sum_0^p a_i X_{n-i} = \sum_0^q b_l \varepsilon_{n-l}.$$

We do the scalar product in L^2 of this equality with X_{n-m} for any $m \geq q+1$ so that X_{n-m} is orthogonal to $(A_Q\varepsilon)_n$. For any $m \geq q+1$:

$$\sum_0^p a_i \gamma(m-i) = 0,$$

let the set of linear equations the solution of which being the vector a in R^p:

$$\sum_1^p a_i \gamma(m-i) = -\gamma(m), \ \forall m \geq q+1.$$

With $m = q+1, \cdots, q+p$, we get a system of equations named **Yule-Walker equations**; we denote R_{pq} the matrix of this system of p equations and p unknown variables:

$$R_{pq} = \begin{vmatrix} \gamma(q) & \cdots & \gamma(q+1-p) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \gamma(q-1+p) & \cdots & \gamma(q) \end{vmatrix}$$

and Γ_{q+1}^{q+p} the vector with coordinates $\gamma(m), \ m = q+1, \cdots, q+p$.

Proposition 4.1 *If X is an ARMA(p,q) process, (p, q) being minimal, the matrix R_{pq} is invertible and the coefficients of the polynomial P are the coordinates of the vector*

$$a = -R_{pq}^{-1}\Gamma_{q+1}^{q+p}.$$

Proof : We assume that $\det R_{pq} = 0$, meaning there exists p coefficients α_i (at least one is non null) such that :

$$\sum_{i=0}^{p-1} \alpha_i \gamma(q+j-i) = 0, \ \forall j = 0, \cdots, p-1.$$

On the other hand, for $j = p$, using Yule-Walker equations, we replace $\gamma(q+p-i)$:

$$\sum_{i=0}^{p-1} \alpha_i \gamma(q+p-i) = -\sum_{i=0}^{p-1} \alpha_i \sum_1^p a_j \gamma(q+p-i-j) = -\sum_{j=1}^{p} a_j \sum_{i=0}^{p-1} \alpha_i \gamma(q+p-i-j)$$

which is a sum of null terms for $p-j=p-1,\cdots,0$ since $\det R_{pq}=0$. By induction, step by step, we get for $j\geq 0$:

$$\sum_{i=0}^{p-1}\alpha_i\gamma(q+j-i)=0.$$

This exactly reflects the fact that $\forall j\geq 0$:

$$E[\sum_{i=0}^{p-1}\alpha_i X_{n-i}X_{n-j-q}]=0,$$

meaning $\forall n\geq 0, \sum_{i=0}^{p-1}\alpha_i X_{n-i}$ is orthogonal to $H^X_{n-q}=H^\varepsilon_{n-q}$ and we compute its coordinates in $(H^\varepsilon_{n-q})^\perp$:

$$\langle\sum_{i=0}^{p-1}\alpha_i X_{n-i},\varepsilon_{n-q+l}\rangle=\sum_{i=0}^{p-1}\alpha_i\langle X_{n-i},\varepsilon_{n-q+l}\rangle$$

for $l=1,\cdots,q$ and equal to 0 for $l>q$. Moreover using stationarity hypothesis $\langle X_{n-i},\varepsilon_{n-q+l}\rangle$ does not depend on n: since the white noise ε is the innovation white noise X is expressed as a function of ε and this scalar product is stationary.
Denoting γ_l the coordinate of $\sum_{i=0}^{p-1}\alpha_i X_{n-i}$ on ε_{n-q+l}:

$$\sum_{i=0}^{p-1}\alpha_i X_{n-i}=\sum_{l=1}^{q}\gamma_l\varepsilon_{n-q+l},$$

which is an ARMA(p-1,q-1) relation and contradicts the hypothesis that the pair (p,q) is 'minimal'. •

4.2 Estimation of Q coefficients

This is a much more difficult problem and we will only give a weak approach! We assume P is known (we estimated it in previous subsection), q and γ are also known. We put

$$Y_n=\sum_0^p a_k X_{n-k}.$$

We will only put the problem, then its resolution states on numerical analysis. The existence of solutions is proved, but not the uniqueness. The Y covariance function is computed as a function of the (b_i) using that $Y=A_Q\varepsilon$:

$$\begin{aligned}
\gamma^Y(0) &= \sum_0^q b_k^2\\
\gamma^Y(1) &= \sum_1^q b_k b_{k-1}\\
\gamma^Y(j) &= \sum_j^q b_k b_{k-j}\\
\gamma^Y(q) &= b_q b_0
\end{aligned}$$

We look for a solution b such that the corresponding polynomial Q admits only zeros with modulus ≥ 1.
Exercise 10: solve this system for $q = 1, 2$.
For $q = 1$, $b_i^2, i = 0, 1$ are $\frac{1}{2}\left(\gamma(0) \pm \sqrt{\gamma(0)^2 - 4\gamma(1)^2}\right)$ so we need $\gamma(0) \geq 2\gamma(1)$.
For $q = 2$..... awful computations !

But the aim is to find the polynomial Q and there is another method, easier but using complex numbers and what is called "spectral density". Since Y is MA(q) process, its spectral density is known to be

$$f(\lambda) = \frac{1}{2\pi}\sum_{-q}^{+q} \gamma^Y(k)e^{-ik\lambda} = \frac{1}{2\pi}|Q(e^{-i\lambda})|^2$$

where you only have to know that $z = e^{-i\lambda}$ is 2 dimensional, $(\cos(\lambda), -\sin(\lambda))$, and satisfies $1/z = (\cos(\lambda), \sin(\lambda)) = e^{i\lambda}$. So we have to deal with: meaning

$$Q(z)Q(1/z) = \gamma^Y(0) + \sum_{1}^{+q} \gamma^Y(k)(z^k + z^{-k})$$

With the change of variable $Z = z + 1/z$ we compute $z^k + z^{-k}$ as polynomial of Z, for instance:

$$z^2 + z^{-2} = Z^2 - 2.$$

Thus $Q(z)Q(1/z)$ could be written as a polynomial $U(Z)$ the zero of which Z_j are linked to those of Q by the relation $Z_j = z_j + 1/z_j$.

Practically, once found U and its zeros, we deduce those of Q, chosen with modulus ≥ 1. The coefficients b are got from the expansion of $\Pi_j(z - z_j)$.

Routines R: for instance for ARMA(2,1) needs arima commands:

```
arima(x, order = c(2,0,1)),
    seasonal = list(order = c(2,0,1), period = NA),
    xreg = NULL, include.mean = TRUE,
    transform.pars = TRUE,
    fixed = NULL, init = NULL,
    method = c("CSS-ML", "ML", "CSS"), n.cond,
    SSinit = c("Gardner1980", "Rossignol2011"),
    optim.method = "BFGS",
    optim.control = list(), kappa = 1e6)
X.ord=c(2,9,1)
X.arima=arima(X,ord=X.ord)
```

Remark: CSS= Conditional Square Sum.

4.3 Characterization of parameters p and q

Definition 4.2 *A rational spectrum ARMA process is said to be with* **minimal type** (p, q) *when in the "canonical Fejer-Riesz relation", the degrees of P and Q are exactly p and q.*
More concretely: (p, q) is minimal when there does not exist polynomials P' and Q' with smaller degrees than p, q in the ARMA equation.

Consequence: if an ARMA(p',q') process is minimal type (p, q), necessarily $p' \geq p, q' \geq q$.

Theorem 4.3 *A regular STS X is minimal type $(0, q)$ if and only if*

$$\gamma(m) = 0,\ \forall |m| \geq q + 1 \ et\ \gamma(q) \neq 0.$$

Proof Exercise 11.

Definition 4.4 *Let (p, q) be a pair of positive numbers. We say that a real sequence $r_n,\ n \in Z$* **satisfies a** (p, q) **induction** *if there exists coefficients $(\alpha_0, \cdots, \alpha_p)$ with $\alpha_0 = 1, \alpha_p \neq 0$, such that $\sum_0^p \alpha_j r_{m-j} = 0,\ \forall m \geq q + 1$.*

The induction is **minimal** (p, q) *if any pair (p', q') satisfying the property above are such that $p' \geq p, q' \geq q$.*

As we saw that in Subsection 4.1, the sequence $\gamma(n)$ of an ARMA(p,q) satisfies a minimal (p, q) induction. With the γ (or at least their estimates), we can find p and q highlighting the minimal induction. A priori it is not so obvious but this property is equivalent to others properties which are easier to check numerically.

Lemma 4.5 *Let a sequence $(x_m,\ m \in Z)$ and the matrix $R_{s,t}$ with (i, j) coefficient equal to x_{i-j}, i and j going from 1 to s. If $r_{s,0} \neq 0$, the following are equivalent:*

(i) The sequence $(x_m,\ m \in Z)$ satisfies a minimal induction (p, q) relation;

(ii) among the determinants $r_{s,t}$, we have $r_{s,t} \neq 0$ while $s \leq p$ or $t \leq q$, and $r_{s,t} = 0$ if $s \geq p + 1$ and $t \geq q + 1$.

(iii) $r_{p+1,q} \neq 0$ and $r_{p,q+1} \neq 0$ and $r_{p+1,j} = 0$ if $j \geq q + 1$.

(iv) $r_{p+1,q} \neq 0$ and $r_{p,q+1} \neq 0$ and $r_{i,q+1} = 0$ si $i \geq p + 1$.

Here $r_{p,q}$ will denote the determinant of the matrix $R_{p,q}$ defined in Section 4.1.

Remark 4.6 *In case of ARMA process, $R_{s,0}$ is the variance matrix of the vector $(X_1, \cdots, X_s)$. The lemma hypothesis corresponds to the case where the series X is non singular.*

So this hypothesis is not too strong:
Exercise, if X is non singular, prove that $r_{s,0} \neq 0$, meaning: prove that $r_{s,0} = 0$ implies X is singular.

The lemma proof is tedious, for a complete proof, look at [1], pp. 137-138.

Proposition 4.7 *Let X be a rational spectrum STS. It is minimal type ARMA (p, q) if and only if the covariance function satisfies a minimal (p, q) induction relation.*
In this case the induction relation is the one which provides the coefficients (a_i) of the polynomial P:

$$\gamma(m) + a_1\gamma(m-1) + \cdots + a_p\gamma(m-p) = 0, \ \forall m \geq q.$$

Definition 4.8 *The* **order** s **partial auto correlation** *of X, denoted as $\Phi(s)$, is the last coordinate of the vector $-R_{s,0}^{-1}\Gamma_1^s$.*

Previously it was denoted r (Definition 1.3)

$$r(p-n) = \frac{\text{cov }(X_n - X_n^*, X_p - X_p^*)}{\sqrt{\text{Var }(X_n - X_n^*) \text{ Var }(X_p - X_p^*)}}.$$

Proposition 4.9 *Let a rational spectrum non singular real STS X. It is an AR(p) process if and only if $\Phi(s) = 0, \ \forall s \geq p+1$ and $\Phi(p) \neq 0$.*

Proof : Exercise 12 proves the necessary condition.
Conversely, to prove the sufficient condition, we use Lemma 4.5. We consider the Cramer system:

$$R_{s,0}\alpha = -\Gamma_1^s.$$

We noticed that, for a non singular series, $r_{s,0} = det \ R_{s,0} \neq 0$. By performing the Cramer resolution, the last coordinate of α is:

$$-\frac{det \ R'_{s,0}}{r_{s,0}}$$

where $R'_{s,0}$ is the matrix $R_{s,0}$ with the last column replaced by Γ_1^s. Using a sequence of s permutations, we see that $R'_{s,0}$ is actually $R_{s,1}$, and the last coordinate of α is: $(-1)^s\frac{r_{s,1}}{r_{s,0}}$.
We then can express the hypothesis

$$\Phi(s) = 0, \ \forall s \geq p+1 \text{ and } \Phi(p) \neq 0$$

as $r_{s,1} = 0 \ \forall s \geq p+1$ and $r_{p,1} \neq 0$, meaning the property (iv) in Lemma 4.5 when $q = 0$ which is a characterization of an AR(p). •

4.4 Exercises

Below ε is a white noise with variance $= 1$ on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

1. Let a STS X, the covariance function is $\gamma(k) = E[X_n X_{n-k}]$. Prove that for any $k \in \mathbb{Z}$, $\gamma(k) = \gamma(-k)$ and

$$|\gamma(n)| \leq \gamma(0) \ ; \ \gamma(-n) = \gamma(n)$$

and $\forall k$, $\forall k$-uplet $(n_1, \cdots, n_k)$ and $\forall k$-uplet of real numbers $(c_1, \cdots, c_k)$, we get:

$$\sum_{i,j} c_i c_j \gamma(n_i - n_j) \geq 0$$

meaning that the Toeplitz matrix T is positive, meaning for all non equal to 0 vector $c \in \mathbb{R}^k$, $\tilde{c}Tc > 0$.

Indication: to study the variance of the random variable $\sum_i c_i X_{n_i}$.

2. Prove that the partial auto correlation function, P.A.C.F., defined on $\mathbb{N}$ as:

$$r: \ \mathbb{N} \to \mathbb{R} \ ; \ \ \text{cor } (X_n, X_p / X_{n+1}, \cdots, X_{p-1}), p > n,$$

meaning

$$\frac{\text{cov } (X_n - X_n^*, X_p - X_p^*)}{\sqrt{\text{Var } (X_n - X_n^*) \text{ Var } (X_p - X_p^*)}}$$

where X_j^* is the orthogonal projection of X_j on the vector space generated by $(X_{n+1}, \cdots, X_{p-1})$, depends only on the lag $p - n$.

3. Prove that an AR(1) is a STS, proof of Proposition 1.10.
Prove that a moving average $\sum_{k \in \mathbb{Z}} a_k \varepsilon_{n-k}$ is a STS. Compute its covariance function with the coefficients a_i.

4. Wold's Theorem: Prove that the regular and singular parts of a STS are still STS.

$$X_n^s = P_{-\infty}^X(X_n) \ ; \ X_n^r = X_n - P_{-\infty}^X(X_n).$$

5. *MidTerm Test 2016*: Let X be a STS such that $\forall n \in \mathbb{Z}$, $X_n = \varepsilon_n + \alpha \varepsilon_{n-1}, \ |\alpha| = 1$.
(i) Prove that for all n, $H_n^X \subset H_n^\varepsilon$.
(ii) Prove that $\forall n \in Z$ and for all $p \geq 1$, there exists a constant β_p such that: $\varepsilon_n + \beta_p \varepsilon_{n-p} \in H_n^X$.
(iii) Prove that for all $n \in Z$ and for all $p \geq 1$,

$$\| \varepsilon_n - P_n^X(\varepsilon_n) \|_2^2 \leq \| P_{n-p}^\varepsilon(\varepsilon_n - P_n^X(\varepsilon_n)) \|_2 \ .$$

Indication: inside the squared norm, keep one of the factors and decompose the other using $\varepsilon_n + \beta_p \varepsilon_{n-p}$.
(iv) Deduce that ε is an innovation white noise for X.

6. Look at the regularity of the STS:
$X_n = g(n) X_0$ where g is an application from $\mathbb{Z}$ to $\mathbb{R}$ such that X is a STS;
White noise
Moving average $\sum_{k \in \mathbb{N}} a_k \varepsilon_{n-k}$;
Moving average $\sum_{k \in \mathbb{Z}} a_k \varepsilon_{n-k}$;
AR(1).

7. Under the assumption of Lemma 3.1, in case of an AR(1), $X_n = aX_{n-1} + \varepsilon_n$ prove that the covariance is $\gamma^X(k) = \frac{a^k}{1-a^2}$.

8. Applying Lemma 3.1, compute the biais of the estimates $\bar{\gamma}_n(k) = \frac{1}{n}\sum_{j=1}^n X_j X_{j+k}$, $\gamma_n^*(k) = \frac{1}{n}\sum_{j=1}^{n-k} X_j X_{j+k}$.

9. Assuming the existence of moments in L^4 ($\sup_n E[X_n^4] \leq M$),
(i) study the quadratic convergence, $\lim_{n\to\infty} E[(\bar{\gamma}_n(k) - \gamma(k))^2]$.
(ii) Moreover, bound the norm $\|\bar{\gamma}_n(k) - \gamma_n^*(k)\|_2$.

10. Let X be an ARMA(p,1) $\sum_0^p a_k X_{n-k} = b_0\varepsilon_n + b_1\varepsilon_{n-1}$. Propose estimates of b_0, b_1.

11. Prove Theorem 4.3

12. Prove necessary condition of Proposition 4.9.

13. Let X be a STS defined as $X_n = \varepsilon_n + 0.8\varepsilon_{n-1} - 0.2\varepsilon_{n-2}$.
(i) What is the regularity of this STS ? Compute its covariance function.
(ii) Prove that $\varepsilon_n + \varepsilon_{n-1} \in H_n^X$. Then use the scheme of exercise 5 above to prove that ε is an innovation white noise for X.

14. II. of 2016 Midterm test. Let be ε an ARCH(1): there exists a white noise η, strong or weak, there exists α_0 and α_1 positive constants, and a process h such that $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$(also denoted σ_t^2) such that $\varepsilon_t = \sigma_t\eta_t$.
(i) Prove that (ε_t^2) is an AR(1): $\varepsilon_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \mu_t$ where $\mu_t = \varepsilon_t^2 - h_t$ is the **innovation process**, meaning the supplementary information given by ε_t after time $t-1$.
(ii) Prove $E(\mu_t) = 0$.
(iii) We now assume $\alpha_1 < 1$, then ε satisfies the following $E[\varepsilon_t/\mathcal{F}_{t-1}] = 0$; $Var(\varepsilon_t) = \frac{\alpha_0}{1-\alpha_1}$,
(iv) (ε_t^2) is a stationary process,
(v) and the conditional variance of ε given the past $\mathcal{F}_{t-h}, h > 0$, is non constant in time:

$$Var(\varepsilon_t/\mathcal{F}_{t-h}) = \alpha_1^h\varepsilon_{t-h}^2 + \alpha_0\frac{1-\alpha_1^h}{1-\alpha_1}.$$

(vi) Study the case $\alpha_1 \geq 1$.

Second Part: Processes ARCH, GARCH

5 Some non linear models

We now look for models taking in account the stylized facts of financial series. The standard ARMA can not do that. Remember that our aim is to model financial series such that a forecasting could be efficient: perhaps a linear combination of past values could be a forecast. But the linearity does not allow to take in account asymmetry, leptokurticity, for instance. Actually, given the past at time t, meaning the σ-algebra $\mathcal{F}_{t-1} := \sigma(X_{t-i}, i \geq 0)$, the better approximation (in L^2 sense) is the conditional expectation with respect to the $\sigma-$algebra $\mathcal{F}^X_{t-1}$ generated by the past of process X, $\{X_{t-i}, i \geq 1\}$:

$$\hat{X}_t = E[X_t/\mathcal{F}^X_{t-1}].$$

Thinking of the first part, there exists a white noise ε generating a sub filtration $\mathcal{F} \subset \mathcal{F}^X$.

Recall that a process η is a **strong** *white noise if it is a series of centered i.i.d. random variables, $Var(\eta_n) = 1$, and* **a weak** *white noise if the linear projection of η_t on the past $\{\eta_{t-i}, i \geq 1\}$ is 0, but there is no more the independence, the η_i are only non correlated.*

Thus, Campbell, Lo, McKinlay propose

$$X_t = g(\eta_{t-1}, \eta_{t-2}, ...) + \eta_t h(\eta_{t-1}, \eta_{t-2}, ...).$$

the first term is the conditional expectation, the second term is the forecasting error. So there could be two types of non linear models, according to g or h could be non linear. Anyway, there is two different approaches.

- Non linear extensions of ARMA processes, which take in account the asymmetric features.
- ENGLE (1982): Autoregressive Conditional heteroscedasticity (=ARCH), with estimates of the variance.

Francq and Zakoian [6]; see also BOLLERSEV (1986) [3].

Definition of **heteroscedasticity**: the covariance function is not stationary; Here we deal with "Conditional Heteroscedasticity" (so CH in the acronym GARCH, G being for "general"): the conditional covariance is non stationary.

Below, η is a white noise process, meaning (η_t) are centered independent identically distributed with variance $= 1$. Some authors consider $\sigma^2_\eta \neq 1$. Actually, non correlated η could be a sufficient condition.

5.1 First approach

Here are some examples, but it is not an exhaustive description.
(i) GRANGER-ANDERSEN (1978):

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \eta_{t-j} + \sum_{i,j \geq 0} \lambda_{i,j} X_{t-i} \eta_{t-j}.$$

(ii) EXPAR

$$X_t = \mu + \sum_{i=1}^{p} [\alpha_i + \beta_i \exp(-\gamma X_{t-i}^2] + \eta_t.$$

(iii) Markov switching models, threshold auto regressive (TONG, 1978) meaning

$$X_t = \Phi_1(S^X)(X)\mathbf{1}_{\{X_{t-d}>\gamma\}} + \Phi_2(S^X)(X)\mathbf{1}_{\{X_{t-d}\leq\gamma\}} + \eta_t,$$

where Φ_i are polynomial functions, S^X is the delay operator, γ is the threshold, meaning according to this threshold there is a switching.
This method allows to take in account the asymmetry feature.

5.2 Second approach, Engle

The aim is to model the conditional variance of the price variation:

$$\varepsilon_t = \log(1 + r_t) \; ; \quad \text{where } r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \; ; \; \sigma_t^2 = Var(\varepsilon_t / \varepsilon_{t-i}, i > 0).$$

Then we look at ARCH/GARCH: Let us suppose that $X = \log p$ is an AR(1) process:

$$X_t = \theta X_{t-1} + \varepsilon_t.$$

$\mathcal{F}_{t-1}$ is the σ-algebra generated by $(\varepsilon_{t-i}, i > 0)$, sub σ-algebra of the one generated by the prices observations $(X_{t-i} = \log p_{t-i}, i > 0)$.

Then the conditional expectation $E[X_t/\mathcal{F}_{t-1}] = \theta X_{t-1}$. On the one hand $\varepsilon_t = X_t - \theta X_{t-1}$. On the other hand, recursively we get $E(X_t) = \theta E(X_{t-1}) = \theta^t x_0$ which goes to 0 when $t \to \infty$ if $|\theta| < 1$. Similarly, the conditional variance $E[(X_t - \theta X_{t-1})^2/\mathcal{F}_{t-1}] = E(\varepsilon^2) = \sigma_\varepsilon^2$ does not depend on time and (once again recursively) the variance is $E(X_t^2) = \frac{\sigma_\varepsilon^2}{1-\theta^2}$, with such models we can not forecast the changes in forecasting errors
So the estimates of the variance are constant and do not highlight any evolution in the time....Thus it could be better to model the covariance function as a process.

ENGLE's hypothesis is: ARCH model for the volatility process. Such models can take in account the non stationarity of the variance along the time. There exists two types of such models:

- Linear ARCH with quadratic specification of conditional variance:
  ARCH(q), GARCH(q), IGARCH (p,q);

- Non linear ARCH with asymmetric specification of variance: EGARCH(p,q), TARCH(q), APAGARCH....(cf. below Chapter 10)

We think of random variance model, ARMA are not convenient to model these facts, such as the property of conditional heteroscedasticity: $Var(\varepsilon_t/\varepsilon_{t-i}, i > 0)$ is non constant, below we will put $\varepsilon_t = \sigma_t \eta_t$, where $\sigma_t > 0$ is $\mathcal{F}_{t-1}$ measurable, and η is a white noise, η_t being independent on $\mathcal{F}_{t-1}$. ε is the current price variation, $\varepsilon_t := X_t - E[X_t/\mathcal{F}_{t-1}]$, $E[\varepsilon_t/\mathcal{F}_{t-1}] = 0$, $E[\varepsilon_t^2/\mathcal{F}_{t-1}] = \sigma_t^2$, σ represents the volatility process, ε is a weak white noise with kurtosis, cf. [6] (1.7): $\frac{E(\varepsilon_t^4)}{(E(\varepsilon_t^2)^2} = \kappa_\eta[1 + \frac{Var(\sigma^2)}{(E(\sigma_t^2))^2}]$, where κ_η is η kurtosis coefficient.

Look at some models [6] p. 11 with random variance models, $\varepsilon_t = \sigma_t \eta_t$ where $\sigma_t > 0$ is $\mathcal{F}_{t-1}$ measurable, η white noise independent of $\mathcal{F}_{t-1}$:

- Conditionally heteroscedastic (=GARCH) process, where the filtration is induced by the past of the process (ε_t) and the volatility at time t, is a function of $(\varepsilon_i, i < t)$. In the standard case, the volatility at time t is a LINEAR function of $(\varepsilon_i, i < t)$.
- Stochastic volatility processes: $\log \sigma_t = \omega + \varphi \log \sigma_{t-1} + v_t$ where v is a strong white noise independent of η. We can say that actually, $\log \sigma$ is an AR(1) process.
- Switching regime models: $\sigma_t = \sigma(\Delta_t, \mathcal{F}_{t-1})$ where Δ_t represents a 'regime', unobservable process, independent of η. Conditionally to Δ_t, σ_t could be a GARCH. The process Δ is for instance a finite-state Markov chain (cf. Markov-switching models).

6 Linear ARCH-GARCH models

6.1 ARCH(1)

The first model is Engel's one:

$$\varepsilon_t = \eta_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2},$$

where η is a weak white noise (centered, $E(\eta_t/\mathcal{F}_{t-1}) = 0$, and variance 1) and $\alpha_i, i = 0, 1$ are positive real parameters. It is equivalent to write $\varepsilon_t = \eta_t \sqrt{h_t}$ with $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$.

Definition 6.1 *A process ε is said to be an ARCH(1) if there exists a (strong or weak) white noise η (satisfying also $E[\eta_t/\mathcal{F}_{t-1}] = 0$) and a process h such that $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ (usually denoted as σ_t^2) and*

$$\varepsilon_t = \sigma_t \eta_t, \ \alpha_i > 0, i = 0, 1.$$

Actually we remark that (ε_t^2) is an AR(1):

$$\sigma_t^2 = h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \Leftrightarrow \varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \mu_t$$

where $\mu_t := \varepsilon_t^2 - h_t$ is the **innovation process**, meaning the supplementary information given by ε_t after time $t-1$.
Remark that $E(\mu_t) = E(\sigma_t^2 \eta_t^2 - \sigma_t^2) = 0$ **and** $E(\mu_t/\mathcal{F}_{t-1}) = \sigma_t^2[E(\eta_t^2 - 1)/\mathcal{F}_{t-1}] = 0$.

The following proposition proves that in case of $\alpha_1 < 1$, $(\varepsilon_t^2)_t$ is a stationary process:

Proposition 6.2 *Assume that $\alpha_1 < 1$, and let ε be an ARCH(1) process satisfying: there exist K and M such that $\sup_{t\leq K} E(\varepsilon_t^2) \leq M$. It satisfies the following*

$$E[\varepsilon_t/\mathcal{F}_{t-1}] = 0 \ ; \ Var(\varepsilon_t) = \frac{\alpha_0}{1-\alpha_1}.$$

There exists a similar result for GARCH(1,1) if $\alpha + \beta < 1$.
Proof Exercise 15

Such properties allow us to consider such process as an error model, for instance the variation, the volatility.

Proposition 6.3 *Assume $\alpha_1 < 1$. The conditional variance of ε given the past $\mathcal{F}_{t-h}, h > 0$, is non constant in time:*

$$Var(\varepsilon_t/\mathcal{F}_{t-h}) = \alpha_1^h \varepsilon_{t-h}^2 + \alpha_0 \frac{1-\alpha_1^h}{1-\alpha_1}.$$

It is the main feature of such models: ε is similar to a homoscedastic white noise but its conditional variance is not stationary.
Proof Exercise 16 (i)

Proposition 6.4 *The conditional covariance is null*

$$cov(\varepsilon_t, \varepsilon_{t+k}/\mathcal{F}_{t-h}) = 0 \ \forall h \geq 1, \ k \geq 1.$$

Proof Exercise 16 (ii)

Proposition 6.5 *We assume that there exist K and M such that $\sup_{t\leq K} E(\varepsilon_t^2) \leq M$.*
(i) Assume $\alpha_1 > 0$ and almost surely $\alpha_0 + \mu_t > 0$. Then the process (ε_t^2) defined by the ARCH(1) Definition 6.1 is positive. So we have some conditions on the support of random variable μ_t.
(ii) $\alpha_0 > 0$ and $0 < \alpha_1 < 1 \Leftrightarrow$ the variance of random variable ε_t exists.

Proof : (i) Definition 6.1 is $\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \mu_t$, the assumptions imply $\varepsilon_t > 0$ almost everywhere.
(ii) the proof of $\Rightarrow$ is Proposition 6.2.
Conversely under the assumtion $\varepsilon_t \in L^2$, and the fact that μ_t are centered, once again we get the recursive formula

$$E[\varepsilon_t^2] = \alpha_0 \frac{1-\alpha_1^h}{1-\alpha_1} + \alpha_1^h E[\varepsilon_{t-h}^2].$$

Using that for h great enough $E[\varepsilon_{t-h}^2] \leq M$, this series is converging if and only if $|\alpha_1| < 1$. Since the result is positive, the limit $\frac{\alpha_0}{1-\alpha_1} > 0$, so $\alpha_0 > 0$. •

Under the hypothesis "the process ε belongs to L^4", we get (cf. Berra and Higgins, 1993):

Proposition 6.6 *(i) Assume* $\forall t,\ \varepsilon_t \in L^4$ *and* $E[\eta_t^4] = 3$. *Then*
$E[\varepsilon_t^4/\mathcal{F}_{t-1}] = 3(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^2$.
(ii) If there exist K *and* M *such that* $\sup_{t \leq K} E(\sigma_t^4) \leq M$ *and* $3\alpha_1^2 < 1$,

$$E[\varepsilon_t^4] = 3(\alpha_0^2 + 2\alpha_0\alpha_1 E(\varepsilon_{t-1}^2) + \alpha_1^2 E[\varepsilon_{t-1}^4] = \frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)}.$$

(iii) Thus we get the kurtosis

$$kurtosis = \frac{E[\varepsilon_t^4]}{(E[\varepsilon_t^2])^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} > 3.$$

Remark that the kurtosis is greater than the Gaussian law kurtosis. So this model can take in account this leptokurtic feature of the observed data.
Proof Exercise 17.

6.2 Models with ARCH(q) errors

Definition 6.7 *A process* ε *is said to be an ARCH(q) if there exist a white noise* η *and a process* h *such that* $h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$, $\alpha_0 > 0$, $\alpha_i \geq 0$ *for* $i \geq 1$, *and*

$$\varepsilon_t = \eta_t \sqrt{h_t}.$$

Similarly to ARCH(1), we have results for ARCH(q) models, at least $E[\varepsilon_t/\mathcal{F}_{t-1}] = 0$ and $Var[\varepsilon_t/\mathcal{F}_{t-1}] = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$.

Thus we now consider a financial series AR(1): (X_t) and its residual, a "weak white noise", meaning $\varepsilon_t := X_t - E[X_t/\mathcal{F}_{t-1}]$ and $X_t = \mu + \rho X_{t-1} + \varepsilon_t,\ |\rho| < 1$.

But actually ε is an ARCH(q) process; there exists a weak white noise η so that

$$\varepsilon_t = \eta_t \sqrt{h_t},\ h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2.$$

Thus, for instance if $q = 1$, the process ε satisfies Propositions 6.2, 6.3 , 6.4, 6.5.
Exercise 18 : In case of a linear model, we can prove recursively

$$\forall h > 0,\ E[X_t/\mathcal{F}_{t-h}] = \mu + \rho E[X_{t-1}/\mathcal{F}_{t-h}] = \mu \left(\frac{1-\rho^h}{1-\rho}\right) + \rho^h X_{t-h}.$$

Then we have some properties for initial process X:

Proposition 6.8 *Let an AR(1) X with error ε being an ARCH(1). Then*

$$Var(X_t/\mathcal{F}_{t-h}) = \left(\frac{\alpha_0}{1-\alpha_1}\right)\left[\left(\frac{1-\rho^{2h}}{1-\rho^2}\right) - \alpha_1\left(\frac{\alpha_1^h-\rho^{2h}}{\alpha_1-\rho^2}\right)\right] + \alpha_1\left(\frac{\alpha_1^h-\rho^{2h}}{\alpha_1-\rho^2}\right)\varepsilon_{t-h}^2.$$

Proof Exercise 19

As a corollary: $Var(X_t/\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$. This means that the forecasting error admits a non constant variance: the confident interval size depends on time, via the values of ε_{t-h}.

6.3 GARCH(p,q)

[[6] 2.1 pp. 19 et sq.]
Once again we consider the white noise of a financial time series, meaning $\varepsilon_t := \log p_t - \log p_{t-1}$. But we now assume that ε is a GARCH(p,q) process.

It could happen that q could be too large, and in such a case, following Box and Jenkins, we would apply the "parsimony" principle to the process h.
Anyway, the practitioners usually only consider GARCH(1,1) even if it is not always convenient, so we have to be careful, cf. [6] page 205, ZakTab205.pdf; very often the hypothesis "the model is ARCH(5)" is accepted at level 0.05 and 0.01. The model is GARCH(1,1) is accepted only for DJU, Nasdaq.

Table 8.4 Portmanteau test p-values for adequacy of the ARCH(5) and GARCH(1,1) models for daily returns of stock market indices, based on m squared residual autocovariances. p-values less than 5% are in bold, those less than 1% are underlined.

	m											
	1	2	3	4	5	6	7	8	9	10	11	12
Portmanteau tests for adequacy of the ARCH(5)												
CAC	0.194	**0.010**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DAX	0.506	0.157	0.140	**0.049**	**0.044**	0.061	0.080	0.119	0.140	0.196	0.185	0.237
DJA	0.441	0.34	0.139	0.002	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DJI	0.451	0.374	**0.015**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DJT	0.255	0.514	0.356	**0.044**	**0.025**	**0.013**	**0.020**	**0.023**	**0.000**	**0.000**	**0.000**	**0.000**
DJU	0.477	0.341	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
FTSE	0.139	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Nasdaq	0.025	**0.031**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Nikkei	**0.004**	**0.000**	**0.001**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SMI	0.502	0.692	0.407	0.370	0.211	0.264	0.351	0.374	0.463	0.533	0.623	0.700
S&P 500	0.647	0.540	**0.012**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Portmanteau tests for adequacy of the GARCH(1,1)												
CAC	0.312	0.379	0.523	0.229	0.301	0.396	0.495	0.578	0.672	0.660	0.704	0.743
DAX	0.302	0.583	0.574	0.704	0.823	0.901	0.938	0.968	0.983	0.989	0.994	0.995
DJA	0.376	0.424	0.634	0.740	0.837	0.908	0.838	0.886	0.909	0.916	0.938	0.959
DJI	0.202	0.208	0.363	0.505	0.632	0.742	0.770	0.812	0.871	0.729	0.748	0.811
DJT	0.750	0.100	0.203	0.276	0.398	0.518	0.635	0.721	0.804	0.834	0.885	0.925
DJU	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
FTSE	0.733	0.940	0.934	0.980	0.919	0.964	0.328	0.424	0.465	0.448	0.083	0.108
Nasdaq	0.523	**0.024**	0.061	**0.019**	**0.001**	**0.001**	**0.002**	**0.001**	**0.002**	**0.001**	**0.001**	**0.002**
Nikkei	0.049	0.146	0.246	0.386	0.356	0.475	0.567	0.624	0.703	0.775	0.718	0.764
SMI	0.586	0.758	0.908	0.959	0.986	0.995	0.996	0.999	0.999	0.999	0.999	0.999
S&P 500	0.598	0.364	0.528	0.643	0.673	0.394	0.512	0.535	0.639	0.432	0.496	0.594

Figure 1: Portmanteau test.

Definition 6.9 *A process ε is a GARCH(p,q) if there exist a weak white noise η, integers p and q, real numbers $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, so that*

$$\varepsilon_t = \eta_t \sigma_t, \;\; \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

$$\varepsilon_t \in L^2, \;\; E[\varepsilon_t / \mathcal{F}_{t-1}^{\varepsilon}] = 0,$$

with sufficient conditions to insure $\sigma_t^2 \geq 0$.

Recall: η is a white noise, $E(\eta_t/\mathcal{F}_{t-1}) = 0, E(\eta_t^2) = 1$.
We recall $\mu_t := \varepsilon_t^2 - \sigma_t^2$ the **innovation process** for ε^2 , which is an uncorrelated process; it is also defined as $\mu_t = \varepsilon_t^2 - Var(\varepsilon_t/\mathcal{F}_{t-1})$.
Remark that $E(\varepsilon_t/\mathcal{F}_{t-1}) = 0$ since η_t is centered and independent on $\mathcal{F}_{t-1}$.
So $Var(\varepsilon_t/\mathcal{F}_{t-1}) = E(\varepsilon_t^2/\mathcal{F}_{t-1}) = \sigma_t^2$ since $Var(\eta_t) = 1$.

Exercise 20: Prove that the process (ε_t^2) is an ARMA$(\sup(p,q), p)$ process, more specifically:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{p \vee q} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + \mu_t - \sum_{j=1}^{p} \beta_j \mu_{t-j}.$$

This is named an “ARMA representation”.

Thus it could be an idea to apply usual ARMA methods to the process ε^2 to identify p, q, α_i, β_j.... but actually it is not really convenient, because the lack of strict stationarity.

Example 1: GARCH(1,1): $\varepsilon_t = \eta_t \sigma_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ induces an ARMA(1,1):

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + \mu_t - \beta_1 \mu_{t-1}.$$

Proof : $\varepsilon_t^2 = \mu_t + \sigma_t^2 = \mu_t + \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \mu_t + \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1(\varepsilon_{t-1}^2 - \mu_{t-1})$.
If $\alpha_1 + \beta_1 < 1$, and if $\exists M, \exists N, \sup_{t \leq N} E[\varepsilon_t^2] \leq M$, by induction we get

$$Var(\varepsilon_t) = E[\varepsilon_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

Here we once again have a necessary condition, $\alpha_1 + \beta_1 < 1$, for the existence of $Var(\varepsilon_t)$ and the stationarity of the process (ε_t^2).

More generally we now look for a link between kurtosis and conditional heteroscedasticity, using $\sigma^2 = \varepsilon^2 - \mu$:

Proposition 6.10 *Let $\varepsilon_t = \eta_t \sigma_t$ with η a Gaussian white noise (so $E(\eta_t^4) = 3$) and $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$. Then $E[\varepsilon_t^4] \geq 3 \left(E[\varepsilon_t^2]\right)^2$ and*

$$Kurtosis = 3 + 3 \frac{Var[E(\varepsilon_t^2/\mathcal{F}_{t-1})]}{\left(E[\varepsilon_t^2]\right)^2}.$$

Proof cf. Proposition 6.6, case ARCH(1).
$E[\varepsilon_t^4] = 3E[\sigma_t^4]$; $E[\varepsilon_t^2] = E[\sigma_t^2]$; so the Kurtosis$= \frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} = 3\frac{E[\sigma_t^4]}{E[\sigma_t^2])^2}$.

On the other hand: $E(\varepsilon_t^2/\mathcal{F}_{t-1}) = \sigma_t^2$, so $Var[E(\varepsilon_t^2/\mathcal{F}_{t-1})] = E(\sigma_t^4) - (E(\sigma_t^2))^2$, and $E(\sigma_t^4) = (E(\sigma_t^2))^2 + Var[E(\varepsilon_t^2/\mathcal{F}_{t-1})]$.

So we get the result putting these moments inside the ratio Kurtosis expression. •

Example 2, [6] p. 45: GARCH(1,1) with η a Gaussian white noise, $\varepsilon_t = \eta_t\sigma_t$, $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$.
Exercise 21: Prove that $\varepsilon_t \in L^4$ only if $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1$. In this case

$$kurtosis = 3\frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}.$$

Stylized facts:
- process ε^2 is correlated, ε is not; e.g. $\varepsilon_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\mu_{t-1}$, $\mu_{t-1} = \varepsilon_{t-1}^2 - \sigma_{t-1}^2$. We assume $\varepsilon_t \in L^4$, then we can look at $cor(\varepsilon_t^2, \varepsilon_{t-l}^2)$ (cf. [6] remark 2.1 page 20).

Definition 6.11 *A process ε is a STRONG GARCH(p,q) if there exist a strong white noise η, p and q, $\alpha_0 > 0$, $\alpha_i, i = 1...p$, $\beta_j, j = 1....q$ so that*

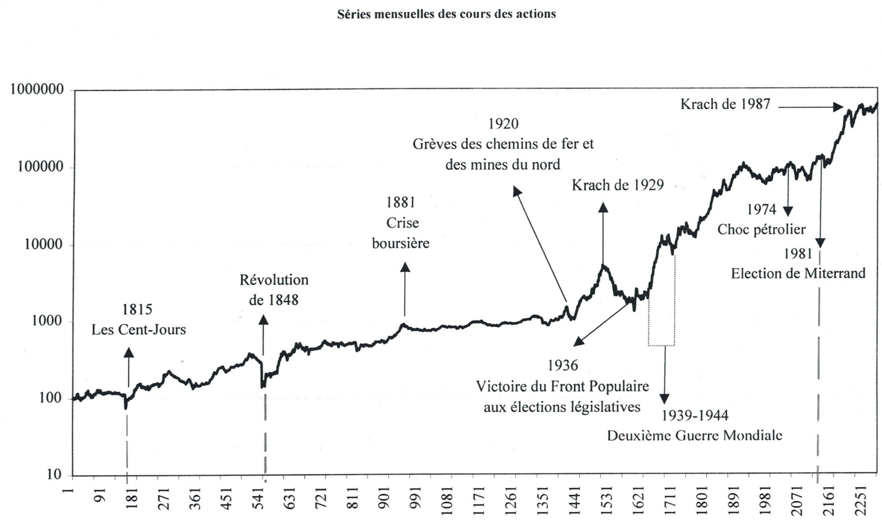
$$\begin{aligned} \varepsilon_t &= \sigma_t\eta_t,\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j\sigma_{t-j}^2, \\ \varepsilon_t &\in L^2,\ E[\varepsilon_t/\mathcal{F}_{t-1}^\varepsilon] = 0, \end{aligned} \tag{4}$$

with sufficient conditions to insure $\sigma_t^2 \geq 0$.

Remark that using $\varepsilon_{t-i}^2 = \sigma_{t-i}^2\eta_{t-j}^2$, $\sigma_t^2 = \alpha_0 + \sum_{i=1}^q(\alpha_i\eta_{t-i}^2 + \beta_i)\sigma_{t-i}^2$, meaning σ^2 is an $AR(p \vee q)$ with **random** coefficients.

Properties of simulated paths:
look at fig 2.1 [6] pp 21 et sq. and real data "Bourse de Paris"

showing the volatility clustering property, succession of large magnitudes of $|\varepsilon_t|$ then low magnitudes of $|\varepsilon_t|$. Large absolute values are not uniformly distributed but tend to cluster.



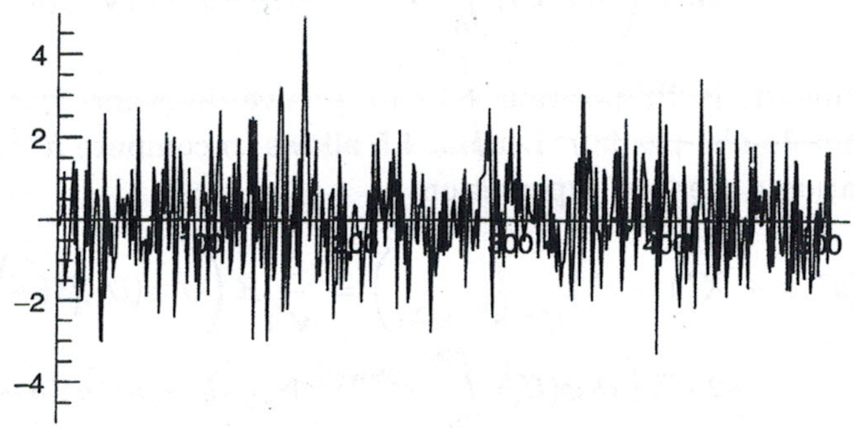


Figure 2.1 Simulation of size 500 of the ARCH(1) process with $\omega = 1$, $\alpha = 0.5$ and $\eta_t \sim \mathcal{N}(0, 1)$.

Stationarity study: The aim is to look at the stationary solutions to Equation in Definition 4, similarly to what we did in ARMA study. As an example, look at GARCH(1,1): let a strong white noise η and

$$\varepsilon_t = \sigma_t \eta_t, \ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

and denote the polynomial function $a(z) = \alpha_1 z^2 + \beta_1$. With this notation we get

$$h_t = \sigma_t^2 = \alpha_0 + (\alpha_1 \eta_t + \beta_1)\sigma_{t-1}^2 = \alpha_0 + a(\eta_t)\sigma_{t-1}^2,$$

so recursively we get

$$h_t = \alpha_0(1 + \sum_{i\geq 1} a(\eta_{t-1})....a(\eta_{t-i})).$$

The following proposition proves the strict stationarity of any strong GARCH(1,1).

Theorem 6.12 *([6] pp 24-25 and Cor p. 26) Assume* $\gamma = E[\log(a(\eta)] < 0$, *where* $a(x) = \alpha_1 x + \beta_1$, *then the series*

$$h_t = \alpha_0(1 + \sum_{i\geq 1} a(\eta_{t-1})....a(\eta_{t-i})),$$

converges almost surely and $\varepsilon_t = \eta_t\sqrt{h_t}$ *is the unique strictly stationary solution of the system (4), case* $p = q = 1$.
Moreover the process ε *is* $\mathcal{F}^\eta$ *adapted and ergodic.*
If $\gamma \geq 0$ *and* $\alpha_0 > 0$, *there exists no strictly stationary solution of (4).*

Proof : only an idea....
(i) Using large numbers theorem, remark that $\frac{1}{n}\ln[a(\eta_{t-1})....a(\eta_{t-i})] = \frac{1}{n}\sum_{l=1}^{i}\ln a(\eta_{t-l}) \to \gamma < 0$, so the order of $a(\eta_{t-1})....a(\eta_{t-i})$ is about $e^{n\gamma} = (e^\gamma)^n$ which is a convergent series when $\gamma < 0$.
Then starting with the definition of h and a we get

$$h_t = \alpha_0(1 + \sum_{i\geq 1}(\alpha_1\eta_{t-1}^2 + \beta)....(\alpha_1\eta_{t-i}^2 + \beta)),$$

so

$$E(\varepsilon_t^2) = E(h_t) = \alpha_0(1 + \sum_{i\geq 1} E[(\alpha_1\eta_{t-1}^2 + \beta_1)....(\alpha_1\eta_{t-i}^2 + \beta_1)].$$

Since the process η is non correlated and $E(\eta_j^2) = 1$

$$E(\varepsilon_t^2) = E(h_t) = \alpha_0(1 + \sum_{i\geq 1}(\alpha_1 + \beta_1)^i)$$

which does not depend on t and exists as soon as $\alpha_1 + \beta_1 < 1$.
(ii) By definition the process ε is $\mathcal{F}^\eta$-adapted. Ergodic definition: $\forall k \in \mathbb{N}$, $\forall B$ Borel set in $\mathbb{R}^k$ $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{k}\mathbf{1}_B(Z_t, ..., Z_{t+k-1}) = \mathbb{P}\{(Z_t, ..., Z_{t+k-1}) \in B\}$.
(iii) If $\gamma \geq 0$, the series in (i) is not converging, no solution. •

Corollary 6.13 *When* $t \to \infty : \gamma > 0 \Rightarrow \sigma_t^2 \to_{a.s.} \infty$ *and*

$$\gamma > 0 \textit{ and } E[|\log\eta_t^2|] < \infty \Rightarrow \varepsilon_t^2 \to_{a.s.} \infty.$$

Second-order stationarity of solutions to Equation (4) th 2.2 page 27 [6]:

Theorem 6.14 • *Let $\alpha_0 > 0$, if $\alpha_1 + \beta_1 \geq 1$, there does not exist a solution to Equation (4).*
• *If $\alpha_1 + \beta_1 < 1$, the process defined by*

$$\varepsilon_t = \eta_t \sigma_t;\ \sigma_t^2 = \alpha_0 + \sum_{i \geq 1} \Pi_{j=1}^i a(\eta_{t-j})),\ a(x) = \alpha_0 x^2 + \beta_1,$$

is a second order stationary solution to Equation (4).
• *More precisely, ε is a weak white noise. Moreover there exists no other adapted second-order stationary solution.*

Proof • [6] Th 2.2 page 27, case $p = q = 1$: Suppose there exists a solution to (4) then $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = E(\sigma_t^2) = \alpha_0 + (\alpha_1 + \beta_1) E(\sigma_{t-1}^2)$. Recursively we get $E(\sigma_t^2) = \frac{\alpha_0}{1-(\alpha+\beta)}$ if and only if $\alpha_1 + \beta_1 < 1$, so there is no solution in case of $\alpha_1 + \beta_1 \geq 1$.

• If $\alpha_1 + \beta_1 < 1$, look at

$$\varepsilon_t = \eta_t \sigma_t;\ \sigma_t^2 = \sum_{i \geq 1} \Pi_{j=1}^i a(\eta_{t-j})),\ a(x) = \alpha_0 x^2 + \beta_1.$$

As previously, η is a white noise process, independent on the process σ. Thus $E(\varepsilon_t^2) = E(\eta_t^2) E(\sigma_t^2) = E(\sigma_t^2)$ which is computable via the recursive equation

$$E(\sigma_t^2) = \alpha_0 + \sum_{i \geq 1} E[a(\eta_t)]^i$$

This sum is finite if and only if $E[a(\eta_t)] < 1$ which is equivalent to $\alpha_1 + \beta_1 < 1$.
Denote that the assumptions implies that $\varepsilon_{t-1}^2 = \eta_{t-1}^2 \sigma_{t-1}^2$, so on the one hand,

$$\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + \alpha_1 \eta_{t-1}^2 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + a(\eta_{t-1})$$

which on the other hand can be identified to σ_t^2 meaning that actually the process defined in the theorem is actually solution to Equation (4).

• The process ε is a weak white noise: it is centered as η is, and uncorrelated for the same reason, since σ and η are uncorrelated.
Under this assumption, we see above that necessarily any solution of Equation (4) satisfies the definition of the theorem with such a definition of function a. ([6] chap 2 pp 28 et sq.)

7 Identification [6] Chap. 5

Let $(p_1, \cdots, p_n)$ be observed prices of a centered stationary process, deduced from a financial time series, $X = \log p$. The log-price variation,$X_t - X_{t-1}$, should coincide with its innovation process ε: $\forall t$, $\varepsilon_t = \log \frac{p_t}{p_{t-1}} = X_t - X_{t-1}$ where p is the financial time series. Notice this series ε is dependent though uncorrelated:

X stationary so ε_t is centered; concerning the ε covariance function: $cov(\varepsilon_t, \varepsilon_{t-s}) = E[\eta_t\sigma_t\eta_{t-s}\sigma_{t-s}] = E[E[\eta_t/\mathcal{F}_{t-1}]\sigma_t\eta_{t-s}\sigma_{t-s}] = 0$. Cf. [6] page 93 line -1

We have to identify the model GARCH(p,q), meaning identify the orders p, q and the coefficients:

$$\begin{aligned} \varepsilon_t &= \sigma_t\eta_t,\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \\ \varepsilon_t &\in L^2,\ E[\varepsilon_t/\mathcal{F}_{t-1}^{\varepsilon}] = 0, \alpha_0 > 0,\ \alpha_i \geq 0, i = 1...p,\ \beta_j \geq 0, j = 1....q. \end{aligned}$$

7.1 Autocorrelation check for white noise

Recall that the theoretical covariances $E[\varepsilon_n\varepsilon_{n+k}] = 0\ \forall k \neq 0$. They can be estimated by **SACV** (S for sample):

$$\hat{\gamma}(k) = \hat{\gamma}(-k) = n^{-1} \sum_{i=1,n-k} \varepsilon_i \varepsilon_{i+k}$$

and the autocorrelation function, **SACR** by

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}.$$

We now need to test the null hypothesis: $\gamma = 0$. It is done by the following theorem which is similar to a central limit theorem, so provides confident intervals:

Theorem 7.1 *Let the GARCH process defined above satisfying $\varepsilon_n \in L^4$ and the symmetric covariance $m \times m$ matrix Γ_m defined by $\Gamma_m(i,j) = E[\varepsilon_{n+1-i}^2 \varepsilon_{n+1-j}^2]$. Then when $n \to \infty$, the distribution of the m vector SACV $\sqrt{n}\hat{\gamma}_m = \sqrt{n}(\hat{\gamma}(1), ..., \hat{\gamma}(m))$ goes to a centered vector Gaussian law with variance matrix Γ_m.*
Let $\Gamma_\rho(m) := \frac{1}{[E(\varepsilon_n^2)]^2}\Gamma_m$, then the distribution of the m vector SACR $\sqrt{n}\hat{\rho}_m = \sqrt{n}(\hat{\rho}(1), ..., \hat{\rho}(m))$ goes to a centered vector Gaussian law with variance matrix $\Gamma_\rho(m)$.

Look for R code to draw a given number of $\hat{\rho}(j)$ and their confident intervals, [6] page 96.

Then there exists "Portmanteau" tests for checking that the data is a strong realization of a strong white noise, it involves the statistic

$$Q_m^{LB} := n(n+2) \sum_{i=1,m} \hat{\rho}^2(i)/(n-i).$$

Under null hypothesis: "ε is a weak white noise", Q_m^{LB} asymptotic distribution is χ_m^2. So the null hypothesis is rejected as soon as $Q_m^{LB} \geq (1-\alpha)$ quantile of χ_m^2.
But there exists a more robust statistic using estimated covariance and correlation matrices

$$\widehat{\Gamma}_m := \left[\widehat{\Gamma}_m(i,j) := \frac{1}{n} \sum_{k=1,n} \varepsilon_k^2 \varepsilon_{k-i}\varepsilon_{k-j}\right],\ \widehat{\Gamma}_\rho(m) := \frac{1}{[E(\varepsilon_n^2)]^2}\widehat{\Gamma}_m.$$

Theorem 7.2 *Let the GARCH process defined above satisfying $\varepsilon_n \in L^4$ and the symmetric covariance $m \times m$ matrix Γ_m defined by $\Gamma_m(i,j) = E[\varepsilon^2_{n+1-i}\varepsilon^2_{n+1-j}]$. The Portmanteau statistic $Q_m = n\hat{\rho}'_m \widehat{\Gamma}_\rho(m)^{-1}\hat{\rho}_m$ has an asymptotic χ^2_m distribution.*

Let $r(k) = Corr(\varepsilon_t, \varepsilon_{t-k}/\mathcal{F}_{t-1})$ named partial auto correlation, and its estimate $\hat{r}(k)$, the sample partial auto correlation **SPAC**. It could be easily computed with Durbin's algorithm [6] p. 355. It satisfies the convergence in law $\sqrt{n}\hat{r}(k) \to \mathcal{N}(0,1)$, $\forall k > p$ for an AR(p), but be careful: with such too narrow confident intervals, we could be wrong rejecting the null hypothesis.

Think that tests based on SPAC could be more powerful than the ones based on SACR (cf. [6] pp 97-99).

7.2 Identifying the ARMA orders of an ARMA-GARCH

Let ARMA-GARCH model, ARMA(P,Q), GARCH(p,q):

$$X_t - \sum_{i=1,P} a_i X_{t-i} = \varepsilon_t - \sum_{i=1,Q} b_i \varepsilon_{t-i}$$

where ε is a GARCH (weak) white noise as defined above. The first task is to identify the orders P and Q. Recall that $\rho_X(k) = 0\ \forall k > Q$, and $r_X(k) = 0\ \forall k > P$:
let us refer to the first part, explicitly Chapter 4 Section 4.3.
From now on we assume that the law of η_t is symmetric.
We will identify (P,Q) using the "corner method": look at the $(j \times j)$ Toeplitz matrix $D(i,j)$ with $D(i,j)_{k,l} := \rho_X(i-1+k, i+1-l)$ and $\Delta(i,j)$ its determinant. Since $\rho_X(h) = \sum_{i=1,p} a_i \rho_X(h_i) = 0$ for all $h > Q$, $\Delta(i,j) = 0$ as soon as $i > P, j > Q$. Thus we look for P and Q such that $(P+1, Q+1)$ is a corner of 0 in the table of $\Delta(i,j)$.
Note that this is automatically done by R routines.

7.3 Identifying the GARCH orders of an ARMA-GARCH

In this case we could use the same methods for the ARMA(p,q) process $(\varepsilon^2_t)_t$: corner method, cf. [6] Section 5.3.1. and above in these notes Lemma 4.5.

We could also look at the estimates of SACV and SPACV....

But *a priori* the most used pair is $p = q = 1$..... even if it is not the most convenient model, cf. Tables in [6].

7.4 Lagrange multiplier test for conditional homoscedasticity

[6] pp 111-116. The purpose is to test the absence of "GARCH effect", meaning a null hypothesis $H_0 : \alpha_{01} = \cdots = \alpha_{0q} = 0$. We introduce "Lagrange multiplier" the statistic

$$LM_n := \frac{1}{n}\frac{\partial}{\partial \alpha'} l_n(\hat{\theta}^c)\hat{I}^{22}\frac{\partial}{\partial \alpha} l_n(\hat{\theta}^c) = \frac{1}{n}\sum_{h=1}^{q}\left[\frac{1}{\hat{\kappa}_\eta - 1}\sum_{t=1}^{n}(\frac{\varepsilon^2_t}{\hat{\omega}^c} - 1)\frac{\varepsilon^2_{t-h}}{\hat{\omega}^c}\right]^2,$$

to complete after Sections 8.3 and 8.4 defining the notations: $\hat{\theta}^c$, $\hat{\kappa}_\eta$, and the function l_n.

8 Estimates and forecasting

Francq and Zakoian [6] Chapters 6,7
We here suppose that p and q are known and we present two methods to estimate: ordinary least squares (OLS) or (quasi) Maximal Likelihood (MLE).

The aim is to estimate the unknown parameter $\theta = (\alpha_i,\ i = 1, ...q;\ \beta_j, j = 1, ..., p)$ for a GARCH(p,q) process.

OLS is a useful method but there is two drawbacks:
- OLS estimate is not efficient, (less good than MLE estimate...)
- Hypothesis L^8 is needed for better properties.

An improvement is provided with "feasible generalized least squares" (FGLS).

8.1 OLS to estimate ARCH(q) models

Recall the model:

$$\begin{aligned} \varepsilon_t &= \sigma_t\eta_t,\ \sigma_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_{i0}\varepsilon_{t-i}^2,\ \omega_0 > 0, \\ \varepsilon_t &\in L^2,\ E[\varepsilon_t/\mathcal{F}_{t-1}^\varepsilon] = 0 \end{aligned} \tag{5}$$

where (η_t) are independent identically distributed, centered and $E(\eta_t^2) = 1$ meaning η is a strong white noise.
The parameter to be estimated is $\theta_0 = (\omega_0, \alpha_{i0}) \in \mathbb{R}^{q+1}$, the observations are $(\varepsilon_t, t = 1, ..., n)$, for instance $\varepsilon_t = \log(1 + r_t) = \log p_t - \log p_{t-1}$ where r_t is the return, we denote the vector of observations as $Z_{t-1} := (1, \varepsilon_{t-1}^2, \cdots, \varepsilon_{t-q}^2)$. We get the scalar product in $\mathbb{R}^{q+1}$:

$$\varepsilon_t^2 = \langle Z_{t-1}, \theta_0\rangle + \mu_t$$

i.e. $Y = X\theta_0 + U$ in $\mathbb{R}^n$, where $U = (\mu_t = \sigma_t^2(\eta_t^2 - 1),\ t = 1, ...n)$, X is the matrix with 1 in the first colum, elsewhere the X elements are $\varepsilon_{i-j}^2, i = 0, ..., n-1;\ j = 0, ..., q-1$.

The consistent OLS estimate is defined as

$$\hat{\theta}(n) := \arg\min \|Y - X\theta\|^2 = (\tilde{X}X)^{-1}\tilde{X}Y. \tag{6}$$

Under assumptions: $\varepsilon \in L^4$ and satisfies (5) above, the following is an estimate of $\sigma_0^2 = Var(\mu_t) = E[(\varepsilon_t^2 - \langle Z_{t-1}, \theta_0\rangle)^2]$:

$$\hat{\sigma}^2(n) := \frac{1}{n-q-1}\|Y - X\hat{\theta}(n)\|^2 = \frac{1}{n-q-1}\sum_{i=1}^{n}\left(\varepsilon_t^2 - \hat{\omega} - \sum_{i=1}^{q}\hat{\alpha}_i\varepsilon_{t-i}^2\right)^2.$$

Theorem 8.1 *(i) If ε is a strictly stationary non anticipative ($\mathcal{F}^\eta$ adapted) solution to system (5) with $\omega_0 > 0$, $\varepsilon_t \in L^4$, $\mathbb{P}\{\eta_t^2 = 1\} < 1$ then almost surely the sequence of estimates $\hat{\theta}(n) \to \theta_0$ and $\hat{\sigma}^2(n) \to \sigma^2$.*
(ii) If moreover $\varepsilon_t \in L^8$ then there is a central limit Theorem, convergence in distribution:

$$\sqrt{n}(\hat{\theta}(n) - \theta_0) \to_{\mathcal{L}} \mathcal{N}\left(0, (E(\eta^4) - 1)A^{-1}BA\right)$$

where A is the matrix $E[Z_{t-1}\tilde{Z}_{t-1}]$ and $B = E[\sigma_t^4 Z_{t-1}\tilde{Z}_{t-1}]$.

These matrices are "information matrices" and can be approximated by

$$\hat{A} := \frac{1}{n}\sum_{t=1}^{n} Z_{t-1}\tilde{Z}_{t-1};\ \hat{B} := \frac{1}{n}\sum_{t=1}^{n} \hat{\sigma}_t^4 Z_{t-1}\tilde{Z}_{t-1}.$$

8.2 FGLS to estimate ARCH(q) models

Remember that the error vector is $\mu_t = \sigma_t^2(\eta_t^2 - 1) = \varepsilon_t^2 - \sigma_t^2$, $E(\mu_t) = 0$, the observations being the ε_t (and recall that $\eta_t \in L^2$ for all t). The $\mathcal{F}_{t-1}$ conditional variance is $Var(\mu_t/\mathcal{F}_{t-1}) = Var(\eta_t^2)\sigma_t^4$.
Let $\theta = (\omega, \alpha) \in \mathbb{R}^{q+1}$, we denote the application $\sigma_t^2 : \theta \to \omega + \langle\alpha, \varepsilon_t\rangle$, $\sigma_t^4 : \theta \to (\omega + \langle\alpha, \varepsilon_t\rangle)^2$, and $\widehat{\Omega} := Diag(\sigma_1^{-4}(\hat{\theta}(n)), ..., \sigma_n^{-4}(\hat{\theta}(n)))$ where $\hat{\theta}(n)$ is the OLS estimate given in Equation (6).

Theorem 8.2 *Under the assumptions of Theorem 8.1 (i), the FGLS estimator defined as*

$$\tilde{\theta}_n := (\tilde{X}\widehat{\Omega}X)^{-1}\tilde{X}\widehat{\Omega}Y$$

almost surely goes to θ_0 and

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_{\mathcal{L}} \mathcal{N}\left(0, Var(\eta_t^2)J^{-1}\right)$$

where $J = E\left(\sigma_t^{-4} Z_{t-1}\tilde{Z}_{t-1}\right)$ is positive definite.

Remember that $(Z_t, t \in N)$ are the observations.
We skip the too long proof, cf. [6] Th 6.3 pp. 132-134. but here are some elements:

- J is a positive definite matrix.
- $\tilde{\theta}_n - \theta_0 = \left(\sum_{t=1,n} \sigma_t^{-4}(\hat{\theta}(n))Z_{t-1}Z'_{t-1}\right)^{-1}\left(\sum_{t=1,n} \sigma_t^{-4}(\hat{\theta}(n))Z_{t-1}\mu_t\right)$.
- Taylor expansion of the above expression.
- Bound
$$\frac{\|\frac{2}{n}\sum_{t=1,n} \sigma_t^{-6}(\theta^*)Z_{t-1}\mu_t \times Z'_{t-1}(\tilde{\theta}_n - \theta_0)\|}{\|\hat{\theta}(n) - \theta_0\|}$$
$(\theta^* \in (\theta_0, \hat{\theta}(n)))$ to prove the almost sure convergence.

- Then apply central limit theorem.

By the way, remark that such an estimator $\tilde{\theta}_n$ is the orthogonal projection of Y under the norm $\|X\|^2 := \tilde{X}\widehat{\Omega}X$.
Remark that we need $n >> q$: q parameters are to be estimated using n observations.

8.3 Constrained OLS to estimate ARCH(q) models

A problem could occur: the estimate of one component θ^i could be non positive.... So it is convenient to add this constraint and we turn to the constrained optimization problem:

$$\hat{\theta}(n)^c := \arg \min_{\theta \in \mathbb{R}_+^{q+1}} \|Y - X\theta\|^2.$$

Since $\theta \to \|Y - X\theta\|^2$ is a convex application and $\|Y - X\theta\| \geq \|X\theta\| - \|Y\| \to \infty$ when $\|\theta\| \to \infty$ in $\mathbb{R}_+^{q+1}$, $\hat{\theta}(n)^c$ exists.

We get the following properties, cf. Theorems 6.5, 6.6, 6.7 [6].

Theorem 8.3 *(i) If* $rank(X) = q + 1$, $\hat{\theta}(n)^c = \hat{\theta}(n) \Leftrightarrow \hat{\theta}(n) \in \mathbb{R}_+^{q+1}$,
(ii) If $rank(X) = q + 1$, $\hat{\theta}(n)^c = \arg\min_{\theta \in \mathbb{R}_+^{q+1}} (\hat{\theta}(n) - \theta)\tilde{X}X(\hat{\theta}(n) - \theta)$.
(iii) Under the assumptions of Theorem 8.1 (i), $\hat{\theta}(n)^c \to \theta_0$ *almost surely.*

8.4 Quasi-maximal Likelihood (QML)

QML method needs stronger assumptions than the previous method. It provides consistent and asymptotically normal estimators in case of strictly stationary GARCH processes.

8.4.1 Conditional QL

We observe $(\varepsilon_1, ...\varepsilon_n)$. We suppose that: p and q are known and the model is

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t,\ \sigma_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \\ \varepsilon_t &\in L^2,\ E[\varepsilon_t / \mathcal{F}_{t-1}^{\varepsilon}] = 0, \end{aligned}$$

η being independent identically distributed, centered, with variance equal to 1.

The parameter to be estimated is $\theta = (\omega, \alpha, \beta) \in \Theta \subset \mathbb{R}_*^+ \times (\mathbb{R}^+)^{p+q}$. We suppose that conditionally to initial values $(\varepsilon_0, ..., \varepsilon_{1-q}, \sigma_0^2, ..., \sigma_{1-p}^2)$, the law of the vector ε is Gaussian, meaning the likelihood

$$L_n(\theta) = \Pi_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\theta)}} \exp(-\frac{\varepsilon_t^2}{2\tilde{\sigma}_t^2(\theta)})$$

where recursively

$$\tilde{\sigma}_t^2(\theta) := \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2(\theta),$$

the choice of the initial values could be $\varepsilon_0 = ... = \varepsilon_{1-q} = \sigma_0^2 = ... = \sigma_{1-p}^2 =$ the common value $\frac{\omega}{1-\sum\alpha-\sum\beta}$. But, if this choice gives negative values for the parameters, a better choice could be ω or ε_1^2: ε_1^2 is observed, and ω is a component of the parameter θ to be estimated.

Then the QMLE is

$$\hat{\theta}(n) := \arg\max_{\theta\in\Theta}(\theta \to L_n(\theta)).$$

Actually, we look for the arg min of the application

$$\theta \to l_n(\theta) := -2\log L_n(\theta) = \sum_{t=1}^{n} [\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\theta)} + \log(\tilde{\sigma}_t^2(\theta))],$$

$$\hat{\theta}(n) := \arg\min_{\theta\in\Theta}(\theta \to l_n(\theta)).$$

Anyway, there does not exist an explicit expression for this estimator, but it could be exhibited via numerical procedures.

We admit the following theorem.

Theorem 8.4 *The QML estimator $\hat{\theta}(n)$ is solution to the system in $\mathbb{R}_*^+ \times (\mathbb{R}^+)^{p+q}$*

$$\sum_{t=1}^{n} \frac{\varepsilon_t^2 - \tilde{\sigma}_t^2(\theta)}{\tilde{\sigma}_t^4(\theta)} \nabla_\theta \tilde{\sigma}_t^2(\theta) = 0$$

with initial values ω or ε_1^2.
(i) Assume the true parameter $\theta_0 \in \Theta$, compact subset of $\mathbb{R}_*^+ \times (\mathbb{R}^+)^{p+q}$, plus some technical assumptions (cf. [6] p.144 and A.i below), then almost surely $\hat{\theta}(n) \to_{n\to\infty} \theta_0$.
(ii) If moreover $\theta_0 \in \Theta^o$ (means $\theta_0 \notin \partial\Theta$) and $\kappa_\eta = E(\eta_t^4) < \infty$, then we get a CLT:

$$\sqrt{n}(\hat{\theta}(n) - \theta_0) \to_{\mathcal{L}} \mathcal{N}(0, (E(\eta_t^4) - 1)J^{-1})$$

where the matrix J is defined by $J_{i,j} = E_{\theta_0}[\frac{\partial^2 l_n(\theta_0)}{\partial_i \partial_j}]$

Assumptions:

A.1: $\theta_0 \in \Theta, \Theta$ is compact.

A.2: $\gamma(A_0) < 0$ $\gamma(A_0)$ *being deduced from some matrices*, $\forall\theta \in \Theta : \sum_{j=1}^{p} \beta_j < 1$.

A.3: η_t^2 law is non degenerate, $E(\eta_t^2) = 1$, *for instance* $\mathbb{P}(\eta_t^2 = 1) < 1$, *cf. Theorem 8.1.*

A.4: If $p > 0$, $\sum_{i=1}^{q} \alpha_{i0} z^i$ and $1 - \sum_{j=1}^{q} \beta_{j0} z^j$ polynomials have no common roots, $\sum_{i=1}^{q} \alpha_{i0} \neq 0$, $\alpha_{q0} + \beta_{p0} \neq 0$.

Exercise 22: Give the log likelihood of a process ARCH(1) $\varepsilon_t = (\sqrt{\omega + \alpha\varepsilon_{t-1}^2})\eta_t$. Stationarity imposes $\alpha < 1$. Recall (η_t) are iid, standard Gaussian law. The conditional law of ε_t given $\mathcal{F}_{t-1}$ is $\mathcal{N}(0, \omega + \alpha\varepsilon_{t-1}^2)$. So the density is

$$\frac{1}{\sqrt{2\pi(\omega + \alpha\varepsilon_{t-1}^2)}} \exp[-\frac{y^2}{2(\omega + \alpha\varepsilon_{t-1}^2)}]$$

Considering that the observations are $\varepsilon_0, ...\varepsilon_n$ the likelihood is

$$\Pi_{t=1}^n \frac{1}{\sqrt{2\pi(\omega + \alpha\varepsilon_{t-1}^2)}} \exp[-\frac{\varepsilon_t^2}{2(\omega + \alpha\varepsilon_{t-1}^2)}]$$

and the loglikelihood is $l_n(\theta) = \sum_{t=1}^n [\frac{\varepsilon_t^2}{\omega+\alpha\varepsilon_{t-1}^2} + \ln(\omega + \alpha\varepsilon_{t-1}^2)]$. The MLE has to minimize $\theta = (\omega, \alpha) \to \sum_{t=1}^n [\frac{\varepsilon_t^2}{\omega+\alpha\varepsilon_{t-1}^2} + \ln(\omega + \alpha\varepsilon_{t-1}^2)]$. Remark that for each term in the above sum the application $x \to \frac{\varepsilon_t^2}{x} + \ln x$ is not convex but its derivative is negative then positive. So we can argue that $\theta \to \ln_n(\theta)$ could have a minimum. Anyway, such a minimum does not exist in a close form but it could be provided with numerical procedures, as a solution of the system

$$\partial_\omega l_n(\theta) = \sum_{t=1}^n \left(\frac{-\varepsilon_t^2}{(\omega+\alpha\varepsilon_{t-1}^2)^2} + \frac{1}{\omega+\alpha\varepsilon_{t-1}^2} \right) = 0$$
$$\partial_\alpha l_n(\theta) = \sum_{t=1}^n \varepsilon_{t-1}^2 \left(\frac{-\varepsilon_t^2}{(\omega+\alpha\varepsilon_{t-1}^2)^2} + \frac{1}{\omega+\alpha\varepsilon_{t-1}^2} \right) = 0$$

8.4.2 Estimation of ARMA-GARCH models by QML

We now look at ARMA(P,Q)-GARCH(p,q) model: The process X is observed (for instance log of prices):

$$X_t - c_0 = \sum_{i=1}^{P} a_{0i}(X_{t-i} - c_i) + e_t - \sum_{j=1}^{Q} b_{0j} e_{t-j},$$

but the noise is not directly observed and is modeled as a GARCH(p,q) process .

$$e_t = \eta_t \sigma_t, \;\; \sigma_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} e_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} \sigma_{t-j}^2.$$

In this case the parameter to be estimated is $\phi = (c, a, b, \alpha, \beta)$. With QML method we get estimators which are consistent and asymptotically Gaussian. (cf. [6] page 150, 7.21)

Routines in R to estimate these parameters by MLE:
garchFitControl, MLE is the by default method; on another hand, method=CSS-ML, or ML or CSS...
ML means Maximal Likelihood, and CSS means Contributed Squared Sum, it minimizes the sum of squared residuals.

8.5 Forecast and confident intervals

With the identified model we now can predict the future behavior of the time series. With the central limit theorems, (cf. Theorems 8.1 (ii), 8.2 ,8.4 (ii)) any estimated parameters actually are estimated via a confident interval, so the forecasting is an interval for any time.

R routines: look at fGarchUse.pdf, predict(....) ; value....

9 Tests based on the likelihood

Cf. Francq and Zakoian [6] Chapter 8, pp. 185-206.
The asymptotic normality of QML estimators allows to test the estimated model. But be careful: when some true coefficients are null, in this case, asymptotic normality fails.... indeed, for a $\theta_{0i} = 0$, we should have $\sqrt{n}\hat{\theta}_i(n) = \sqrt{n}(\hat{\theta}_i(n) - \theta_{0i}) \geq 0$ almost surely, impossible for a Gaussian random variable !

9.1 Test of second order stationarity assumption

In such a test, the null hypothesis H_0 is a necessary condition for stationarity, look at Example 1 in Section 6.3: $\alpha + \beta < 1$. If not, the ε_t^2 is not integrable, the stationarity fails.

We mean to test

$$H_0 : \sum_i \alpha_{0i} + \sum_j \beta_{0j} < 1 \text{ against } H_1 : \sum_i \alpha_{0i} + \sum_j \beta_{0j} \geq 1$$

since the assumption H_0 is necessary for the series belonging to L^2. Let us consider a vector c with all components equal to 1, H_0 is exactly $\langle c, \theta_0 \rangle < 1$. Under convenient hypotheses we deduce from the asymptotic normality of $\hat{\theta}(n)$ the convergence in distribution

$$\sqrt{n}(\langle c, \hat{\theta}(n) \rangle - \langle c, \theta_0 \rangle) \to_{\mathcal{L}} \mathcal{N}(0, (E(\eta_t^4) - 1)c' J^{-1} c)$$

recalling $J_{i,j} = E_{\theta_0}[\frac{\partial^2 l_n(\theta_0)}{\partial_i \partial_j}]$. We have to replace unknown parameters $E(\eta_t^4)$ and J by their estimates: $\hat{\kappa}_\eta$ and $\hat{J}$:

$$\hat{\kappa}_\eta := \frac{1}{n}\sum_{t=1}^n \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\hat{\theta}(n))}; \ \hat{J} := \frac{1}{n}\sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}(n))}(\nabla_\theta \tilde{\sigma}^2)(\hat{\theta}(n))(\nabla_\theta \tilde{\sigma}^2)'(\hat{\theta}(n)).$$

Proposition 9.1 *Under assumptions in Theorem 8.4 (ii), a test of assumption H_0 at level α is defined by the critical region*

$$\{T_n := \sqrt{n}\frac{\langle c, \hat{\theta}(n) \rangle - 1}{\sqrt{(\hat{\kappa}_\eta - 1)c' \hat{J}^{-1} c}} > \Phi^{-1}(1 - \alpha)\}$$

where Φ is the normal Gaussian law distribution function.

9.2 Case of $\theta_0 \in \partial\Theta$

This subsection is to skip in a first reading, cf. [6] Section 8.2 page 187.

Recall $\tilde{l}_t(\theta) = \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} + \log\tilde{\sigma}_t^2$, $\tilde{I}_n(\theta) = \frac{1}{n}\sum_t \tilde{l}_t(\theta)$ and the Hessian matrix $J_n (= D_\theta^2 \tilde{I}_n)(\theta_0)$. We define the normalized score vector

$$Z_n := -J_n^{-1}\sqrt{n}(\nabla_\theta I_n)(\theta_0).$$

Theorem 9.2 *(Theorem 8.1 [6])*
Under convenient hypotheses, the asymptotic distribution of $\sqrt{n}(\hat{\theta}(n) - \theta_0)$ is the one of the statistic

$$\lambda^\Lambda := \arg\min_{\lambda\in\Lambda}(\lambda - Z)' J(\lambda - Z)$$

where J law is $\mathcal{N}(0, (\kappa_\eta - 1)J^{-1}$.

9.3 Portmanteau tests, [6] p. 205

The "residuals methods" mean that we extract the residuals, which could be a white noise, so we have to check the correlogram of these residuals.
Concerning the ARMA models we test the significance of the residual correlations. For GARCH models, we look at the square residual auto covariances

$$\hat{r}(h) := \frac{1}{n}\sum_{t=1+|h|}(\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-|h|}^2 - 1), \text{ where } \hat{\eta}_t^2 := \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\theta)}.$$

We recall the following estimates:

$$\hat{\kappa}_\eta := \frac{1}{n}\sum_t \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\hat{\theta}(n))};\ \hat{J} := \frac{1}{n}\sum_t \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}(n))}(\nabla_\theta\tilde{\sigma}^2)(\hat{\theta}(n))(\nabla_\theta\tilde{\sigma}^2)'(\hat{\theta}(n)),$$

and the $m \times (p+q+1)$ matrix $\hat{C}_m$ defined as:

$$\hat{C}_m(i,j) = -\frac{1}{n}\sum_t(\hat{\eta}_{t-i}^2 - 1)\frac{1}{\tilde{\sigma}_t^2(\hat{\theta}(n))}(\partial_{\theta_j}\tilde{\sigma}^2)(\hat{\theta}(n)).$$

Theorem 9.3 *Let the matrix $\hat{D} := (\hat{\kappa}_\eta - 1)^2 I_m - (\hat{\kappa}_\eta - 1)\hat{C}_m\hat{J}^{-1}\hat{C}_m'$.*
Under assumptions in Theorem 8.4 (ii), $n\hat{r}_n'\hat{D}^{-1}\hat{r}_n \to_{\mathcal{L}} \chi_m^2$.
Thus we reject the GARCH(p,q) model when

$$n\hat{r}_n'\hat{D}^{-1}\hat{r}_n > \chi_m^2(1-\alpha).$$

Remark that the GARCH(1,1) model is too often assumed by the practitioners as [6] presents page 205. Here they show that CAC or S&P are not GARCH(1,1) but could be better modeled as ARCH(5).

Table 8.5 p-values for tests of the null of a GARCH(1,1) model against the GARCH(1,2), GARCH(1,3), GARCH(1,4) and GARCH(2,1) alternatives, for returns of stock market indices and exchange rates. p-values less than 5% are in bold, those less than 1% are underlined.

	Alternative											
	GARCH(1,2)			GARCH(1,3)			GARCH(1,4)			GARCH(2,1)		
	W_n	R_n	L_n	W_n	R_n	L_n	W_n	R_n	L_n	W_n	R_n	L_n
Daily returns of indices												
CAC	**0.007**	**0.033**	**0.013**	**0.005**	**0.000**	**0.001**	**0.024**	0.188	**0.040**	0.500	0.280	0.500
DAX	**0.002**	**0.001**	**0.003**	**0.001**	**0.000**	**0.000**	**0.001**	0.162	**0.014**	0.350	**0.031**	0.143
DJA	0.158	0.337	0.166	0.259	0.285	0.269	0.081	0.134	0.064	0.500	0.189	0.500
DJI	**0.044**	0.100	**0.049**	0.088	0.071	0.094	0.107	0.143	0.114	0.500	**0.012**	0.500
DJT	0.469	0.942	0.470	0.648	**0.009**	0.648	0.519	0.116	0.517	0.369	0.261	0.262
DJU	0.500	**0.000**	0.500	0.643	**0.000**	0.643	0.725	**0.001**	0.725	**0.017**	**0.000**	**0.005**
FTSE	0.080	0.122	0.071	0.093	0.223	0.083	0.213	0.423	0.205	0.458	0.843	0.442
Nasdaq	0.469	0.922	0.468	0.579	0.983	0.578	0.683	0.995	0.702	0.500	0.928	0.500
Nikkei	**0.004**	**0.002**	**0.004**	**0.042**	0.332	0.081	0.052	0.526	0.108	0.238	**0.000**	**0.027**
SMI	0.224	0.530	0.245	0.058	0.202	0.063	0.086	0.431	0.108	0.500	0.932	0.500
SP 500	0.053	0.079	**0.047**	0.089	**0.035**	0.078	0.055	0.052	**0.043**	0.500	**0.045**	0.500
Weekly returns of indices												
CAC	**0.017**	0.143	**0.049**	**0.028**	0.272	0.068	0.061	0.478	0.142	0.500	0.575	0.500
DAX	0.154	**0.000**	**0.004**	0.674	0.798	0.674	0.667	0.892	0.661	**0.043**	**0.000**	**0.000**
DJA	0.194	**0.001**	0.052	0.692	0.607	0.692	0.679	0.899	0.597	**0.003**	**0.000**	**0.000**
DJI	0.173	**0.000**	**0.030**	0.682	0.482	0.682	0.788	0.358	0.788	**0.000**	**0.000**	**0.000**
DJT	0.428	0.623	0.385	0.628	0.456	0.628	0.693	0.552	0.693	**0.002**	**0.000**	**0.004**
DJU	0.500	0.747	0.500	0.646	**0.011**	0.646	0.747	**0.038**	0.747	0.071	**0.003**	**0.017**
FTSE	0.188	0.484	0.222	0.183	0.534	0.214	0.242	0.472	0.272	0.500	0.532	0.500
Nasdaq	0.441	0.905	0.448	0.387	0.868	0.412	0.199	0.927	0.266	0.069	0.961	0.344
Nikkei	0.500	0.140	0.500	0.310	0.154	0.260	0.330	0.316	0.462	**0.030**	0.138	0.053
SMI	0.500	0.720	0.500	0.217	0.144	0.150	0.796	0.754	0.796	0.314	0.769	0.360
SP 500	0.117	**0.000**	**0.001**	0.659	0.114	0.659	0.724	0.051	0.724	**0.000**	**0.000**	**0.000**
Daily exchange rates												
\$/€	0.452	0.904	0.452	0.194	0.423	0.181	0.066	**0.000**	**0.015**	0.500	**0.002**	0.500
¥/€	**0.037**	**0.000**	**0.002**	0.616	0.090	0.618	0.304	**0.000**	0.227	0.136	**0.000**	**0.000**
£/€	0.439	0.879	0.440	0.471	0.905	0.464	0.677	0.981	0.677	0.258	0.493	0.248
CHF/€	0.141	**0.000**	**0.012**	0.641	0.152	0.641	0.520	0.154	0.562	**0.012**	**0.000**	**0.000**
C\$/€	0.500	0.268	0.500	0.631	0.714	0.631	**0.032**	**0.000**	**0.002**	**0.045**	**0.045**	**0.029**

Look at the tables from [6], Table page 205 H_0 is rejected when the p-value, $p = \mathbb{P}_{H_0}[\chi^2_m \leq n\hat{r}'_n \hat{D}^{-1}\hat{r}_n$ computed], is more than α.
Table page 206 is testing GARCH(1,1) against others GARCH and use statistics W_n, R_n, L_n, defined in [6] section 8.3.1. The table provides the corresponding p-values.

Use in R routines: residuals-methods.

GED, see https://en.wikipedia.org/wiki/Generalized_normal_distribution, means 'General Exponential Distribution'.

10 Some extensions, [6] Chapter 10.

Above, we modeled the conditional variances σ_t as linear functions of the squared past innovations $\varepsilon_{t-h}, h > 0$. But from an empirical point of view, there exist important drawbacks: actually in the previous models, we do not take in account the sign of innovations. However, the conditional asymmetry is a stylized fact: the volatility μ increase due to a price decrease is stronger than the one resulting from a price increase of the same magnitude.
Think of $p_t \downarrow$ *yields* $(\Delta\sigma_t)^-$ *and* $p_t \uparrow$ *yields* $(\Delta\sigma_t)^+$, *then* $(\Delta\sigma_t)^- \geq (\Delta\sigma_t)^+$.

If the law is symmetric, $cov(\sigma_t, \varepsilon_{t-h}) = 0,\ \forall h > 0$, which is equivalent to $cov(\varepsilon_t^+, \varepsilon_{t-h}) = cov(\varepsilon_t^-, \varepsilon_{t-h}) = 0,\ \forall h > 0$. This is an hypothesis which could be often rejected, look at table below ([6] p. 246) where the correlations $\rho((\varepsilon_t, \varepsilon_{t-h}),\ \rho(\varepsilon_t^+, \varepsilon_{t-h})$ and $\rho(\varepsilon_t^-, \varepsilon_{t-h})$ are not equal to zero altogether. We can observe $cov(\sigma_t, \varepsilon_{t-h}) < 0$, meaning a *leverage effect*: the volatility increases dramatically after bad news, but increases moderately after good news.

h	1	2	3	4	5	10	20	40
$\rho(\varepsilon_t, \varepsilon_{t-h})$:	0.030	0.005	−0.032	0.028	−0.046*	0.016	0.003	−0.019
$\rho(\lvert\varepsilon_t\rvert, \lvert\varepsilon_{t-h}\rvert)$:	0.090*	0.100*	0.118*	0.099*	0.086*	0.118*	0.055*	0.032
$\rho(\varepsilon_t^+, \varepsilon_{t-h})$:	0.011	−0.094*	−0.148*	−0.018	−0.127*	−0.039*	−0.026	−0.064*

Table 1: Empirical autocorrelations CAC40 series, period 1988-1998

10.1 Exponential GARCH model: EGARCH

Definition 10.1 *Let η a strong white noise. Then (ε_t) is called an EGARCH process if it satisfies*

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t, \\ \log \sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i g(\eta_{t-i}) + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2, \end{aligned} \tag{7}$$

where

$$g(\eta_{t-i}) = \theta\eta_{t-i} + \zeta(|\eta_{t-i}| - E(|\eta_{t-i}|), \tag{8}$$

and $\omega, \beta, \theta, \zeta \in \mathbb{R}$.

Exercise 23: Remark that the volatility σ has a multiplicative dynamics. The log allows the coefficients to be negative or positive. Actually $\varepsilon_t = \sigma_t\eta_t$ and the dynamics of σ is

$$\sigma_t^2 = \exp \circ \ln \sigma_t^2 = e^{\omega}\Pi_{i=1}^q \exp(\alpha_i g(\eta_{t-i}))\Pi_{j=1}^p (\sigma_{t-j})^{2\beta_j}.$$

However we would like that the innovations of large modulus should increase the volatility. Thus we add some constraints on the coefficients, for instance

$$-\zeta < \theta < \zeta, \;\; \alpha_i \geq 0, \;\; \beta_j \geq 0.$$

The coefficient θ reflects the asymmetry property: look at the model $\log \sigma_t^2 = \omega + \theta\eta_{t-1}$ with $\theta < 0$. ([6] page 247, 3):
in this case $\sigma_t^2 = e^{\omega}e^{\theta\eta_{t-1}}$ and $\sigma_t^2 - e^{\omega} = e^{\omega}(e^{\theta\eta_{t-1}} - 1)$.
If $\eta_{t-1} < 0$, the variation $\sigma_t^2 - e^{\omega} = e^{\omega}(e^{\theta\eta_{t-1}} - 1)$ is less than the one when $\eta_{t-1} > 0$, because of the asymetry of the function exp, $e^{-a} < e^a$ when $a > 0$.

The specification $g(\eta_{t-i}) = \theta\eta_{t-i} + \zeta(|\eta_{t-i}| - E(|\eta_{t-i}|)$ allows for sign and modulus effects:
- sign effect with $\theta\eta_{t-i}$,
- modulus effect with $\zeta(|\eta_{t-i}| - E(|\eta_{t-i}|)$.

So we also could take θ depending on the lag:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i(\theta_i\eta_{t-i} + |\eta_{t-i}| - E(|\eta_{t-i}|) + \sum_{j=1}^p \beta_j \log \sigma_{t-j}^2.$$

Theorem 10.2 *Assume that $g(\eta_t)$ is not almost surely equal to zero and that the polynomials $\alpha(z) = \sum_{i=1}^q \alpha_i z^i$ and $\beta(z) = 1 - \sum_{j=1}^p \beta_j z^j$ have no common root, $\alpha(z)$ being non identically null. Then the system (7) admits a strictly stationary and $\mathcal{F}$-adapted solution if and only if the roots z_i of polynomial β are outside the unit circle (meaning in $\mathbb{R}$, $|z_i| > 1$ for any i). This solution satisfies $E[(\log \varepsilon_t^2)^2] < \infty$ as soon as $E[(\log \eta_t^2)^2] < \infty$ and $E[g^2(\eta_t)] < \infty$,*

Proof exercise 24: Prove the theorem in case of $p = 1$, $|\beta| < 1$, $E[(\log \eta_t^2)^2] < \infty$ and $G = E[g^2(\eta_t)] < \infty$.

(i) We here have to use Theorem 2.14 in Section 2.1: $\omega + \sum_{i=1}^q \alpha_i g(\eta_{t-i}) + \beta \log \sigma_{t-1}^2$ could be seen as an AR(1,q) equation

$$\log \sigma_t^2 - \beta \log \sigma_{t-1}^2 = \omega + \sum_{i=1}^q \alpha_i g(\eta_{t-i})$$

with the polynomials $Q(z) = \sum_{i=1}^{q} z^i$ and $P(z) = 1 - \beta_1 z$. Theorem in Section 2.1 asks P and Q have no common root and $\beta < 1$.
(ii) Since $\varepsilon_t^2 = \eta_t^2 \sigma_t^2$, $\log \varepsilon_t^2 = \log \eta_t^2 + \log \sigma_t^2$, and $E[(\log \eta_t^2)^2] < \infty$ we have to prove $E[(\log \sigma_t^2)^2] < \infty$.
(iii) The asumption $E[g^2(\eta_t)] < \infty$ and the fact that (η_t) are uniformly distributed implies

$$E[(\sum_{i=1}^{q} \alpha_i g(\eta_{t-i}))^2] = \sum_{i=1}^{q} \alpha_i^2 E[(g(\eta)^2] < \infty.$$

(iv) Since $|\beta| < 1$, admitting $\log \sigma_{t-k}^2$ almost surely bounded in the past, the induction

$$\log \sigma_t^2 = \beta \log \sigma_{t-1}^2 + \omega + \sum_{i=1}^{q} \alpha_i g(\eta_{t-i})$$

is solved as

$$\log \sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{i=1}^{q} \alpha_i \sum_{k \geq 0} \beta^k g(\eta_{t-i-k}).$$

Thus the L^2 norm of $\log \sigma_t^2$ exists as soon as $G = E[(g(\eta))^2] < \infty$ and

$$E[(\sum_{i=1}^{q} \alpha_i \sum_{k \geq 0} \beta^k g(\eta_{t-i-k}))^2] = G \sum_{j \geq 1} \beta^{2j} (\sum_{i=1}^{q} \alpha_i \beta^{-i})^2].$$

The following theorem is to skip.

Theorem 10.3 *Let m be an integer and suppose*

$$\mu_{2m} := E[\eta_t^{2m}] < \infty; \;\; \Pi_{i=1}^{\infty} E[\exp(|m\lambda_i g(\eta_t)|)] < \infty$$

then $E[\varepsilon_t^{2m}] = \mu_{2m} e^{m\omega^*} \Pi_{i=1}^{\infty} g_\eta(m\lambda_i)$ where λ_i are defined as the coefficient of the development of $\frac{\alpha(z)}{\beta(z)} = \sum_i \lambda_i z^i$, $\omega^* = \omega/\beta(1)$, $g_\eta(x) = E[\exp(xg(\eta_t))]$.

10.2 Threshold GARCH model: TGARCH

To take in account asymmetry, we specify the conditional variance of positive and negative part of ε_t: $\varepsilon_t^+ = \sup(\varepsilon_t, 0)$, $\varepsilon_t^- = \sup(-\varepsilon_t, 0)$, $\varepsilon_t = \varepsilon_t^+ - \varepsilon_t^-$, $|\varepsilon_t| = \varepsilon_t^+ + \varepsilon_t^-$.

Definition 10.4 *Let η be a strong white noise. Then (ε_t) is called an TGARCH if it satisfies*

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t, \\ \sigma_t &= \omega + \sum_{i=1}^{q} (\alpha_{i+} \varepsilon_{t-i}^+ + \alpha_{i-} \varepsilon_{t-i}^-) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}, \end{aligned} \tag{9}$$

where $\omega, \beta_j, \alpha_{i\pm} \in \mathbb{R}$.

Remark that if $\omega > 0$ and $\alpha_{i\pm}, \beta_j \geq 0$, then $\sigma_t > 0$. In this case $\sigma_t = \sqrt{E[\varepsilon_t^2/\mathcal{F}_{t-1}]}$ is the conditional standard deviation. Such a model allows the lags i to have an influence on the past, so the asymmetry is taken into account.

Figure 10.1 in [6] page 251 stress the difference between GARCH and TGARCH. GARCH is symmetric: $\varepsilon_t = \sqrt{1+0.38\varepsilon_{t-1}^2}\eta_t$ and TGARCH is asymmetric: $\varepsilon_t = (1 - 0.5\varepsilon_{t-1}^+ - 0.2\varepsilon_{t-1}^-)\eta_t$.

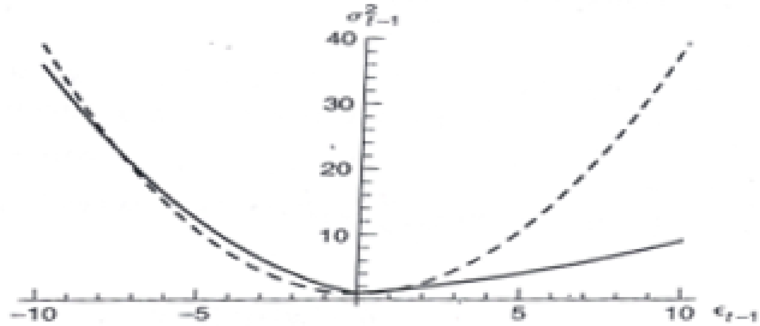


Figure 10.1 News impact curves for the ARCH(1) model, $\epsilon_t = \sqrt{1+0.38\epsilon_{t-1}^2}\eta_t$ (dashed line), and TARCH(1) model, $\epsilon_t = (1 - 0.5\epsilon_{t-1}^- + 0.2\epsilon_{t-1}^+)\eta_t$ (solid line).

Under the constraints $\omega > 0$ and $\alpha, \beta \geq 0$, since $\sigma > 0$, then $\varepsilon_t^\pm = \sigma_t \eta_t^\pm$, and the conditional standard deviation is $\sigma_t = \omega + \sum_{i=1}^{\max(p,q)} a_i(\eta_{t-i})\sigma_{t-i}$, with $a_i(z) = \alpha_{i+}z^+ + \alpha_{i-}z^- + \beta_i$.

In case of TGARCH(1,1) model, if $E[\log(\alpha_{1+}\eta^+ + \alpha_{1-}\eta^- + \beta_1)] < 0$, then the model is stationary (cf. Theorem 6.12).

Finally, we look at the moments of σ_t to go to the kurtosis. We can prove that the mth moment exists if and only if $E[a^m(\eta_t)] < 1$.
Exercise 25: TGARCH(1,1) model, $\varepsilon_t = \eta_t\sigma_t$, $\sigma_t = \omega + a(\eta_{t-1})\sigma_{t-1}$, with $a(z) = \alpha_+ z^+ + \alpha_- z^- + \beta$, assume $E[a^m(\eta_t)] < 1$.
(i) Prove that the assumption $\|a(\eta)\|_2 < 1$ implies the condition $E[\ln a(\eta)] < 0$ in case of $\beta = 1$ and symmetrical law for the η_t, ;
(ii) Compute the moments of η_t to provide skewness and kurtosis $\kappa_\varepsilon = 3\frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2}$. Cf. [6] p. 252.

(0) Case $\beta_1 = 0$: $a(\eta) = \alpha_+\eta^+ + \alpha_-\eta^- = \alpha_+|\eta|\mathbf{1}_{\eta>0} + \alpha_-|\eta|\mathbf{1}_{\eta<0} = |\eta|(\alpha_+\mathbf{1}_{\eta>0} + \alpha_-\mathbf{1}_{\eta<0})$ so looking at the two disjoint sets according to the sign of η:

$$\log a(\eta) = \log|\eta| + \log(\alpha_+)\mathbf{1}_{\eta>0} + \log\alpha_-\mathbf{1}_{\eta<0}.$$

Using the symmetry of η law

$$E[\log a(\eta)] = E[\log|\eta|] + \frac{1}{2}(\log(\alpha_+) + \log\alpha_-) = E[\log|\eta|] + \log\sqrt{\alpha_+\alpha_-}$$

or $E[2\log a(\eta)]) = E[\log|\eta|^2] + \log(\alpha_+\alpha_-)$ meaning

$$\exp E[2\log a(\eta)]) = (\alpha_+\alpha_-)\exp E[\log|\eta|^2].$$

We now use $(\alpha_+\alpha_-) \leq \frac{1}{2}(\alpha_+^2 + \alpha_-^2)$ and Minkowski inequality:
$\exp E[\log|\eta|^2] \leq E[\exp\log|\eta|^2] = E[|\eta|^2]$, so

$$\exp E[2\log a(\eta)]) \leq \frac{1}{2}(\alpha_+^2 + \alpha_-^2)E[|\eta|^2] = E[(a(\eta)^2],$$

and we conclude that $E[(a(\eta)^2] < 1$ yields $\exp E[2\log a(\eta)] < 1$ and $E[2\log a(\eta)] < 0$.

(i) We first control the mth moment of ε_t by the bound $\|\varepsilon_t\|_m = \|\eta\|_m\|\sigma_t\|_m$ since η_t are iid and η_t is independant on σ_t.

(ii)From $\|\sigma_t\|_m \leq |\omega| + \|a(\eta)\|_m\|\sigma_{t-1}\|_m$, we recursively deduce

$$\|\sigma_t\|_m \leq |\omega|\frac{1-\|a(\eta)\|_m^k}{1-\|a(\eta)\|_m} + \|a(\eta)\|_m^k\|\sigma_{t-k}\|_m.$$

If moreover $\sup_{t\leq -N}\|\sigma_{t-k}\|_m$ are uniformly bounded, the following are too and $\|\sigma_t\|_m < \infty$.

(iii) Now compute these moments: $E[\sigma_t] = \omega + E[a(\eta_{t-1})]E[\sigma_{t-1}] = \omega + \|a(\eta)\|_1 E[\sigma_{t-1}]$. Under the previous assumptions we solve this induction:

$$E[\sigma_t] = \frac{\omega}{1-\|a(\eta)\|_1}.$$

Now let $\sigma_t^2 = \omega^2 + 2\omega a(\eta_{t-1})\sigma_{t-1} + a^2(\eta_{t-1})\sigma_{t-1}^2$, so

$$E[\sigma_t^2] = \omega^2 + 2\omega\|a(\eta)\|_1 E[\sigma_{t-1}] + \|a(\eta)\|_2^2 E[\sigma_{t-1}^2]$$

$$= \omega^2 + 2\omega\|a(\eta)\|_1\frac{\omega}{1-\|a(\eta)\|_1} + \|a(\eta)\|_2^2 E[\sigma_{t-1}^2] = \omega^2\frac{1+\|a(\eta)\|_1}{1-\|a(\eta)\|_1} + \|a(\eta)\|_2^2 E[\sigma_{t-1}^2],$$

we solve this induction: $E[\sigma_t^2] = \frac{\omega^2(1+\|a(\eta)\|_1)}{(1-\|a(\eta)\|_1)((1-\|a(\eta)\|_2^2)}$.
(iii) Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean: $E\left[(\frac{\sigma_t-\mu}{\sigma})^3\right]$.

Now let $\sigma_t^3 = \omega^3 + 3\omega^2 a(\eta_{t-1})\sigma_{t-1} + 3\omega a^2(\eta_{t-1})\sigma_{t-1}^2 + a^3(\eta_{t-1})\sigma_{t-1}^3$, after once again a standard induction:

$$E[\sigma_t^3] = \frac{\omega^3(1+2\|a(\eta)\|_1 + 2\|a(\eta)\|_2^2 + \|a(\eta)\|_1\|a(\eta)\|_2^2)}{(1-\|a(\eta)\|_1)((1-\|a(\eta)\|_2^2)(1-\|a(\eta)\|_3^3)}.$$

(iv) Fourth moment: after tedious computations, to be checked !! Denote $a_i = \|a(\eta)\|_i^i,\ i=1,2,3,4$.

$$E[\sigma_t^4] = \omega^4\frac{1+3a_1+5a_2+3a_3+3a_1a_2+5a_1a_3+3a_2a_3+a_1a_2a_3}{(1-a_1)((1-a_2)(1-a_3)(1-a_4)}.$$

10.3 Asymmetric Power GARCH model, APAGARCH, APARCH

The following gathers GARCH, TGARCH, Log-GARCH.

Definition 10.5 *Let η a strong white noise. Then (ε_t) is called an APAGARCH process if it satisfies*

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t, \\ \sigma_t^\delta &= \omega + \sum_{i=1}^{q} (\alpha_i(|\varepsilon_{t-i}| - \zeta_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta, \end{aligned} \tag{10}$$

where $\omega > 0, \delta > 0,$ and $\beta, \alpha \in \mathbb{R}^+$.

• We recover GARCH model with $\delta = 2, \zeta = 0$.
• The case $\delta = 1$ is the TGARCH model.
• Using $\log \sigma_t = \lim_{\delta \to 0} \frac{\sigma_t^\delta - 1}{\delta}$, the log-GARCH model can be interpreted as the limit of APAGARCH when δ goes to 0.

The role of parameter ζ in ARCH(1) model can be seen in the following :

$$\sigma_t^2 = \omega + \alpha_1(1 - \zeta_1)^2 \varepsilon_{t-1}^2 \text{ when } \ \varepsilon_{t-1} \geq 0,$$

$$\sigma_t^2 = \omega + \alpha_1(1 + \zeta_1)^2 \varepsilon_{t-1}^2 \text{ when } \ \varepsilon_{t-1} < 0.$$

Thus the choice of $\zeta_1 > 0$ ensures that a negative innovation ε has more impact on the current volatility σ_t^2 than positive ones of the same magnitude.

Stationarity : we can write

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p \vee q} a_i(\eta_{t-i}) \sigma_{t-i}^\delta,$$

with $a_i(z) = \alpha_i(|z| - \zeta z)^\delta + \beta_i$. Remind Theorem 6.12: Such a process is stationary if and only if

$$E[\log(\alpha_1(|\eta_t| - \zeta_1 \eta_t) + \beta_1)^\delta] < 0.$$

Exercise 26: In the case $\beta_1 = 0$ and when the law of η_t is symmetric, express this condition, cf. [6] (10.24) page 257.
$\sigma_t^\delta = \omega + \alpha(|\varepsilon_{t-1}| - \zeta \varepsilon_{t-1})^\delta = \omega + a(\eta)\sigma_{t-1}^\delta$.

$$\log a(\eta) = \log \alpha + \log |\eta|^\delta + \mathbf{1}_{\eta > 0} \log(1 - \zeta)^\delta + \mathbf{1}_{\eta < 0} \log(1 + \zeta)^\delta.$$

We use the symetry of η law

$$E[\log a(\eta] = \log \alpha + E[\log |\eta|^\delta] + \frac{1}{2}(\log(1-\zeta)^\delta + \log(1+\zeta)^\delta) = \log \alpha + E[\log |\eta|^\delta] + \frac{1}{2} \log(1-\zeta^2)^\delta.$$

The condition $E[\log a(\eta] < 0$ is equivalent to

$$E[\log |\eta|^\delta] < -\log(\alpha(1 - \zeta^2)^{\frac{\delta}{2}}).$$

Remark that $\exp E[\log |\eta|^\delta] \leq E[\exp \log |\eta|^\delta] = E[|\eta|^\delta]$, so a sufficient condition for $E[\log a(\eta] < 0$ is

$$E[|\eta|^\delta] < \frac{1}{\alpha(1-\zeta^2)^{\frac{\delta}{2}}}.$$

•

11 Financial Applications, [6] Chapter 12, pp. 311-326.

11.1 Relation between GARCH and continuous time models

Consider a Wiener filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with a Brownian motion W. Recall a stochastic differential equation (SDE) for $X_t = \log p_t$, p being a price process.

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t,\ X_0 = x_0.$$

Under convenient hypotheses on the volatility σ and the drift μ (Lipschitz property, sub linear increase) there exists a unique strong solution. The concrete interpretation of these parameters μ and σ is:

$$\mu(x) = \lim_{h\to 0} h^{-1}E[X_{t+h} - X_t/X_t = x];\ \sigma(x)\sigma(x)' = \lim_{h\to 0} h^{-1}Var[X_{t+h} - X_t/X_t = x].$$

- To this diffusion X is associated an infinitesimal operator

$$L = \mu\partial_x + \frac{1}{2}\sigma^2\partial^2_{xx}.$$

Using Itô's formula we get for any C^2_b function φ

$$\varphi(X_t) - \varphi(x_0) - \int_0^t L\varphi(X_s)ds \text{ is a martingale.}$$

Such diffusion could have a stationary distribution. For instance look at

$$dX_t = (\omega + \mu X_t)dt + \sigma X_t dW_t,\ X_0 = x_0.$$

Actually it can be proved that, for any t, the distribution density of X_t law is

$$\frac{1}{\Gamma(\zeta)}\left(\frac{2\omega}{\sigma^2}\right)^\zeta \exp(-\frac{2\omega}{x\sigma^2})x^{-1-\zeta} \text{ where } \zeta = 1 - \frac{2\mu}{\sigma^2},$$

and the distribution of $1/X_t$ is the law $\Gamma(\frac{2\omega}{\sigma^2}, \zeta)$, [6] page 313.

- A second point concerns the simulation of these diffusion trajectories (paths). We can proceed to the *Euler discretization*:

$$\tilde{X}^h_{(n+1)h} = \tilde{X}^h_{nh} + h\mu(X^h_{nh}) + \sigma(X^h_{nh})(W_{(n+1)h} - W_{nh}),\ \tilde{X}^h_0 = x_0.$$

Actually, we consider the increment $W_{(n+1)h} - W_{nh}$ as $\sqrt{h}\varepsilon_{n+1}$ where ε is a Gaussian white noise. It can be proved that the process $\tilde{X}_n$ converges in distribution to X when $h \to 0$. Black-Schole's model admits an exact simulation. Indeed

$$\log X_t = (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$$

so

$$\log X_{t+h} = \log X_t + (\mu - \frac{1}{2}\sigma^2)h + \sigma\sqrt{h}\varepsilon_t.$$

• The third point considers the GARCH models as approximation of a diffusion(cf. [6] page 315). Let μ_t be the conditional mean of the returns

$$\mu_t = \xi + \lambda\sigma_t, \ \lambda > 0,$$

where λ and μ are some parameters and (σ_t) is the volatility process. In GARCH model, (σ_t^2) is an ARCH. More generally, let η be a white noise and look at the system [6] 12.16

$$\begin{aligned} X_t &= X_{t-1} + f(\sigma_t) + \sigma_t\eta_t, \\ g(\sigma_t) &= \omega + a(\eta_{t-1})g(\sigma_{t-1}). \end{aligned} \tag{11}$$

With $g(z) = z^2$ and $a(z) = \alpha z^2 + \beta$ in (11) we get a GARCH(1,1) model.

Recall the results on stationarity (Theorem 6.12): $E[\log a(\eta_t)] < 0 \Rightarrow X_t - X_{t-1}$ exists as an adapted and strictly stationary process. Thus the approximation is

$$\begin{aligned} \tilde{X}_{nh} &= \tilde{X}_{(n-1)h} + f(\tilde{\sigma}_{nh}) + \sqrt{h}\tilde{\sigma}_{nh}\eta_n^h, \\ g(\tilde{\sigma}_{(n+1)h}) &= \omega_h + a_h(\eta_n^h)g(\tilde{\sigma}_{nh}). \end{aligned} \tag{12}$$

The following theorem is th 12.2 in [6].

Theorem 11.1 *Assume $\exists\delta > 0, \ \omega_h, \ \eta_n^h, \ a_h, \ 0 < \rho^2 \le \zeta$, such that when $h \to 0$*
(13)
$$h^{-1}\omega_h \to \omega; \ h^{-1}(1 - E[a_h(\eta_n^h)]) \to \delta; \ h^{-1}Var[a_h(\eta_n^h)] \to \zeta; \ h^{-1/2}Cov(a_h(\eta_n^h), \eta_n^h) \to \rho,$$

$$\limsup_{h\to 0} h^{-1-\delta}E\left[(a_h(\eta_n^h) - 1)^{2(1+\delta)}\right] < +\infty,$$

then when $h \to 0$, the system (12) converges to

$$\begin{aligned} dX_t &= f(\sigma_t)dt + \sigma_t dW_t^1, \\ dg(\sigma_t) &= (\omega - \delta g(\sigma_t))dt + g(\sigma_t)(\rho dW_t^1 + \sqrt{\zeta - \rho^2}dW_t^2. \end{aligned} \tag{14}$$

Exercise 27: in the following example, check the above assumptions (13).
Example: $\omega_h = h\omega$, $a_h(z) = 1 - h\delta + \sqrt{h}(\rho z + \sqrt{\zeta - \rho^2}\eta')$,

η' being independent of η; η and $\eta' \in L^{2(1+\delta)}$.

11.2 Option pricing

• The aim is to "price" the derivatives, for instance option (call and put option), at the "maturity" T. We denote K the 'strike' (or 'exercise', cf. Stochastic calculus applied to Finance): considering the assets price process (S_t), we look for the price of $(S_T - K)^+$ or $(K - S_T)^+$. In Black-Scholes model the dynamics is

$$S_t = \mu S_t dt + \sigma S_t dW_t, \ \log S_t = \log S_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t.$$

The discretized version is

$$\log S_t - \log S_{t-1} = \mu - \frac{1}{2}\sigma^2 + \sigma \varepsilon_t$$

where ε is a Gaussian white noise.
But actually, σ is not constant (cf. estimated volatility with data from "Bourse de Paris").
The price is given by Feymann-Kac formula:

$$C(S,t) = e^{-r(T-t)} E_Q[g(S_T)/\mathcal{F}_t] \tag{15}$$

where Q is a risk neutral probability measure. So is deduced the famous Black-Scholes' formula (cf. Lecture Notes Stochastic Calculus applied to Finance). In particular, the partial derivative $\partial_S C(S_t,t)$ provides the hedging portfolio (named "delta"). We want to extend such a scheme to more general diffusions:

$$dS_t = \mu(t,S_t)dt + \sigma(t,S_t)dW_t.$$

• Actually the drift μ has no influence on the hedge price, the important point is the volatility σ. We first consider the "historic" volatility based on the observations, for instance $(S_0, S_1, \cdots, S_n)$:

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=0}^{n-1}(S_i - S_{i-1})^2$$

is an estimate of σ^2 if this one is constant on the observed period.

An alternative solution is the "implied volatility": actually the formula (15) depends on σ (as a constant). The application $\sigma \to C(S,\sigma,t)$ is increasing and the prices (S_t) are observed. This application is not analytically invertible. But with numerical computations, we can extract σ from the observations.
Exercise 28: prove that $\sigma \to C(S,\sigma,t)$ is increasing (cf. Jeanblanc-Yor) so at least numerically invertible.
This is the "implied volatility", and in the practical cases, it is not a constant.

• We now turn to the option pricing when the underlying volatility is a GARCH process. We observe the process Z (which as usually generates the filtration $\mathcal{F}$) up to the maturity T. Suppose that at the time $t < T$ there exists a price $C_t(Z,g,T)$ for the payoff $g(Z_T)$. It can be proved that there exists a random variable $M_{t,T} > 0$, $\mathcal{F}_T$-measurable, which is called *stochastic discount coefficient* (SDF) and such that

$$C_t(Z,g,T) = E[g(Z_T)M_{t,T}/\mathcal{F}_t].$$

suite A RETRAVAILLER

This applied to the "zero-coupon bond" of maturity T (meaning $g = 1$) yields $B(t,T) = E[M_{t,T}/\mathcal{F}_t]$. So $C_t(Z, g, T) = E[g(Z_T)M_{t,T}/\mathcal{F}_t]$ can be written as

$$C_t(Z, g, T) = B(t,T)E[g(Z_T)\frac{M_{t,T}}{B(t,T)}/\mathcal{F}_t].$$

• So we can introduce the "forward risk neutral probability"

$$\pi_{t,T} := M_{t,T}\mathbb{P},$$

we deduce (using the probability change formula)

$$E_{\pi_{t,T}}[g(Z_T)/\mathcal{F}_t] = \frac{E[g(Z_T)M_{t,T}/\mathcal{F}_t]}{E[M_{t,T}/\mathcal{F}_t]},$$

and $C_t(Z, g, T) = B(t,T)E_{\pi_{t,T}}[g(Z_T)/\mathcal{F}_t]$.

• Suppose that actually (in discrete time) $M_{t,T} = \Pi_{i=t}^{T-1} M_{i,i+1}$. Notice the constraints

(16) $$B(t, t+1) = E[M_{t,t+1}/\mathcal{F}_t],\ S_t = E[S_{t+1}M_{t,t+1}/\mathcal{F}_t],$$

this is (12.41) in [6] which actually is the martingale property for the discounted price process, which means that $S_t M_{t,t} = E_{\pi_{t,t+1}}[S_{t+1}M_{t,t+1}/\mathcal{F}_t]$.
So recursively

$$E[g(Z_T)M_{t,T}/\mathcal{F}_t] = B(t, t+1)E[g(Z_T)\Pi_{i=t+1}^{T-1}B(i, i+1)\Pi_{i=t}^{T-1}\frac{M_{i,i+1}}{B(i, i+1)}/\mathcal{F}_t].$$

We introduce a "risk neutral probability measure"

$$\pi^*_{t,T} := E[\Pi_{i=t}^{T-1}\frac{M_{i,i+1}}{B(i, i+1)}/\mathcal{F}_t] \times \mathbb{P}$$

such that

$$E[g(Z_T)M_{t,T}/\mathcal{F}_t] = B(t, t+1)E^*[g(Z_T)\Pi_{i=t+1}^{T-1}B(i, i+1)/\mathcal{F}_t].$$

• Pricing formulas, two exercises

Exercise 29: Look at Black-Scholes model, $Z_t = \log S_t - \log S_{t-1} = \mu - \frac{1}{2}\sigma^2 + \sigma\varepsilon_t$, one step SDF is defined as $B(t, t+1) = e^{-r}$, $M_{t,t+1} = \exp(a + bZ_{t+1})$.
With the constraint $B(t, t+1) = E[M_{t,t+1}/\mathcal{F}_t]$ (16) we get $e^{-r} = E[\exp(a + bZ_{t+1}/\mathcal{F}_t]$, and $S_t = E[S_{t+1}M_{t,t+1}/\mathcal{F}_t]$ means $1 = E[e^{a+(b+1)Z_{t+1}}/\mathcal{F}_t]$.

Use that the law of Z_{t+1} given $\mathcal{F}_t$ is the Gaussian law $(\mu - \frac{1}{2}\sigma^2, \sigma^2)$
to prove the existence of a and b. Then define the risk neutral probability with its characteristic function $E_\pi(e^{uZ_{t+1}}]$
$E(e^{aX}) = e^{ma+\frac{1}{2}a^2\sigma^2}$ if X Gaussian (m, σ^2)

Exercise 30: GARCH-type model: $Z_t = \log S_t - \log S_{t-1} = \mu_t + \varepsilon_t$, $\varepsilon_t = \sigma_t \eta_t$ where η is a white noise. Suppose that the filtrations generated by ε, Z, η are the same. Once again $B(t, t+1) = e^{-r}$, and the SDF $M_{t,t+1} = \exp(a_t + b_t \eta_{t+1})$, where the processes a and b are $\mathcal{F}$-adapted. The constraints (16) lead to $a_t = -r - \frac{1}{2} b_t^2$, $b_t \sigma_{t+1} = r - \mu_{t+1} - \frac{1}{2}\sigma_{t+1}^2$. The risk neutral probability measure $\pi_{t,t+1}$ will be defined through its characteristic function. Under the probability measure $\pi_{t,t+1}$,

$$Z_t = r - \frac{1}{2}\sigma^2 + \varepsilon_t^*,\ \varepsilon_t^* = \sigma_t \eta_t^*$$

where $\eta^* = \frac{1}{\sigma_t}(Z_t - r + \frac{1}{2}\sigma^2)$ is a white noise under the risk neutral probability measure. We can check that (12.49) [6]:

$$C_t(Z, g, T) = e^{-r(T-t)} E^{\pi_{t,T}}[g(Z_T)/\mathcal{F}_t^S] = E[g(Z_T)\Pi_{i=t}^{T-1} M_{i,i+1}/\mathcal{F}_t^S].$$

• Under the hypothesis that a and b do not depend on t, we consider a *numerical pricing of Option Prices* since a closed expression is not available. (cf. Example 12.5 [6] p. 325). We consider the GARCH model:

$$\begin{aligned} \log(S_t/S_{t-1}) &= r + \lambda\sigma_t - \frac{1}{2}\sigma_t^2 + \varepsilon_t, \\ \varepsilon_t &= \sigma_t \eta_t\ ;\ \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \end{aligned}$$

Such a pricing is done using independent simulations of the η paths, at the step i

$$S_T^i = S_t^i \exp[(T-t)r - \frac{1}{2}\sum_{s=t+1}^{T}(\sigma_s^i)^2 + \sum_{i=t+1}^{T}\sigma_s^i \eta_s^i,$$

where $(\sigma_s^i)^2 = \hat{\omega} + (\sigma_{s-1}^i)^2(\hat{\alpha}\eta_{s-1}^i + \hat{\beta})$ is computed recursively.

11.3 Value at risk and other risk measures

Market risk is the risk of change in the value of a financial position ; *Credit risk* is the risk of not receiving repayments on outstanding loans (borrower default); *Model risk* can be defined as the risk due to the use of a mis-specified model; etc. There is also operational risk, liquidity risk... These risks increased in the two last decades. So Basel Committee on Banking Supervision set new rules against these risks, meaning Basel I, Basel II, for instance look at https://en.wikipedia.org/wiki/Basel_II

11.3.1 Value at risk, VaR

This one is the most widely used risk measure in financial institutions. Look at
http://www.gloriamundi.org/

Definition 11.2 *$VaR(\alpha)$ is the value such that the portfolio value V satisfies*

$$\mathbb{P}\{V_t - V_{t+h} > VaR(\alpha)/\mathcal{F}_t\} < \alpha.$$

This means that the loss $V_t - V_{t+h}$ has to be less than $VaR(\alpha)$ with probability greater than $1-\alpha$. $VaR(\alpha)$ is the $(1-\alpha)$-quantile of the conditional loss distribution.

We consider the value V of a portfolio on d assets, and the loss between time t and $t+h$:

$$V_t = \sum_{i=1}^{d} a_i S_t^i,\ \ L_{t,t+h} = -\sum_{i=1}^{d} a_i S_t^i (e^{r_{t,t+h}} - 1) = V_t - V_{t+h}.$$

We can prove that

$$VaR_{t,h}(\alpha) = \inf\{x \in \mathbb{R},\ \mathbb{P}\{L_{t,t+h} \le x/\mathcal{F}_t\} \ge \alpha\}.$$

Recall $r^i_{t,t+h} := \log S^i_{t+h} - \log S^i_t$.

11.3.2 Other risk measures

• Previously, the variance (volatility) is only used to measure the risk. Anyway, this one hides the sign of the variations.
• In insurance is used *expected shortfall*, $ES(\alpha)$:

Definition 11.3

$$ES_{t,h}(\alpha) = E_t\left[L_{t,t+h}/\{L_{t,t+h} > VaR_{t,h}(\alpha)\}\right]$$

where E_t means $E[./\mathcal{F}_t]$.

When $L^+_{t,t+h} \in L^1$, and admits a distribution absolutely continuous,

$$E_t[L_{t,t+h}\mathbf{1}_{L_{t,t+h} > VaR_{t,h}(\alpha)}\}] = E_t[L_{t,t+h}/L_{t,t+h} > VaR_{t,h}(\alpha)] \times \mathbb{P}_t\{L_{t,t+h} > VaR_{t,h}(\alpha)\}$$

By definition of the VaR, actually $\mathbb{P}_t\{L_{t,t+h} > VaR_{t,h}(\alpha)\} = \alpha$ so

$$ES_{t,h}(\alpha) = \frac{1}{\alpha} E_t[L_{t,t+h}\mathbf{1}_{L_{t,t+h} > VaR_{t,h}(\alpha)}\}].$$

Exercise 31: prove that in this last case $ES_{t,h}(\alpha) = \frac{1}{\alpha}\int_0^\alpha VaR_{t,h}(u)du$, Exercise 12.16 in [6].
Plus: example 12.11 page 332.

Let X be a Gaussian random variable (m, σ^2): by definition,

$$1 - \alpha = \mathbb{P}\{X \le VaR(\alpha)\} = \mathbb{P}\{m + \sigma X_0 \le VaR(\alpha)\} = \Phi(\frac{VaR(\alpha) - m}{\sigma})$$

Thus, by monotonicity, $\Phi^{-1}(1-\alpha) = \frac{VaR(\alpha)-m}{\sigma}$ and $VaR(\alpha) = m + \sigma\Phi^{-1}(1-\alpha)$.

We now look at

$$\frac{1}{\alpha}\int_0^\alpha VaR_{t,h}(u)du = \frac{1}{\alpha}\int_0^\alpha (m+\sigma\Phi^{-1}(1-u))du = m+\frac{\sigma}{\alpha}\int_0^\alpha \Phi^{-1}(1-u)du = m+\frac{\sigma}{\alpha}\int_{1-\alpha}^1 \Phi^{-1}(v)dv$$

We now operate the change of variable $v = \Phi(y)$ so

$$\int_{1-\alpha}^1 \Phi^{-1}(v)dv = \int_{\Phi^{-1}(1-\alpha)}^\infty y\phi(y)dy$$

- *Distortion risk measure* (DRM) when $L_{t,t+h}$ admits a density law strictly positive f and note the distribution function F, $F' = f$.

We remark that $ES_{t,h}(\alpha) = \frac{1}{\alpha}\int_0^1 F^{-1}(1-u)\mathbf{1}_{[0,\alpha]}(u)du$, (using $\frac{1}{\alpha}E_t[L_{t,t+h}\mathbf{1}_{L_{t,t+h}>VaR_{t,h}(\alpha)}\}] = \frac{1}{\alpha}\int_{VaR_{t,h}(\alpha)}^\infty lf(l)dl$ and the change of variable $l = F^{-1}(1-u)$) so more generally, we introduce for any distribution G on $[0,1]$

$$r(F;G) = \int_0^1 F^{-1}(1-u)\mathbf{1}_{[0,\alpha]}(u)dG(u).$$

We then recover the previous risk measures:

- $Var(\alpha)$ with G is Dirac in α,
- $ES(\alpha)$ with G uniform law on $[0,\alpha]$.

We skip *Coherent risk measures* page 333, def 12.2.

11.3.3 Estimation methods

We here only present the so called *GARCH-based estimation*: we observe K returns (for instance $K = 250$): $r_{t+h-i} = \log p_{t+h-i} - \log p_{t-i},\ i = h, h+K-1$, $\Delta P_t = \log p_t - \log p_{t-1}$. Consider the example 12.9 [6] page 330: $h = d = a = 1$.

$$(\Delta P_t)^2 = (\omega + \alpha_1(\Delta P_{t-1})^2)U_t^2$$

U_t standard Gaussian law. It is exactly an ARCH(1): $\varepsilon_t = \Delta P_t$, $U_t = \eta_t$, $\sigma_t = \sqrt{\omega + \alpha_1(\varepsilon_{t-1})^2}$. Then the conditional law of the loss $L_{t,t+1}$ is $\mathcal{N}(0, \omega + \alpha_1(\Delta P_t)^2)$. Therefore

$$VaR_{t,1}(\alpha) = \sqrt{\omega + \alpha_1(\Delta P_t)^2}\Phi^{-1}(1-\alpha).$$

It is more problematic when $h > 1$... Then we work with $r_t = \Delta_1 \ln p_t$ assumed to be stationary, then we get $VaR_{t,1}(\alpha) = (1 - e^{q_t(1,\alpha)})p_t$ where $q_t(1,\alpha)$ is the α quantile of the conditional law of r_{t+1}. This one, $q_t(1,\alpha)$, can be estimated by

$$\hat{\sigma}_{t+1}\hat{F}^{-1}(\alpha)$$

where $\hat{\sigma}_t^2$ is the conditional variance estimated by a GARCH model, and F^{-1} is an estimate of the distribution of the normalized residuals.

The suggested steps are the following

- Fit a model for instance on a GARCH(1,1) on the n observations and deduce an estimate of $\hat{\sigma}_t^2, t = 1, \cdots, n+1$.
- Simulate a large number N of $\varepsilon_{n+1}, \cdots, \varepsilon_{n+h}$:
  - simulate the values of the iid (according to the distribution function $\hat{F}$) $\eta_{n+1}^i, \cdots, \eta_{n+h}^i$,
  - set $\sigma_{n+1}^i = \hat{\sigma}_{n+1}$ and $\varepsilon_{n+1}^i = \sigma_{n+1}^i \eta_{n+1}^i$,
  - for $k = 2, \cdots, h$, recursively set
  $$(\sigma_{n+k}^i)^2 = \hat{\omega} + \hat{\alpha}(\varepsilon_{n+k-1}^i)^2 + \hat{\beta}(\sigma_{n+k-1}^i)^2;\ \varepsilon_{n+k}^i = \sigma_{n+k}^i \eta_{n+k}^i).$$
- Determine the empirical quantile of simulations $\varepsilon_{n+h}^i, i = 1, ..., N$.

11.4 Exercises on second part

15. Proof of Proposition 6.2.

16. (i) Proof of Proposition 6.3.
(ii) Proof of Proposition 6.4.

17. Proof of Proposition 6.6.

18. In case of a linear model, X being an AR(1) process, $X_t = \mu + \rho X_{t-1} + \varepsilon_t$, prove recursively:

$$\forall h > 0,\ E[X_t/\mathcal{F}_{t-h}] = \mu + \rho E[X_{t-1}/\mathcal{F}_{t-h}] = \mu \left(\frac{1-\rho^h}{1-\rho}\right) + \rho^h X_{t-h}.$$

19. Proof of Proposition 6.8.

20. Let a GARCH(p,q) process ε. Prove that it is an $ARMA(\sup(p,q), p)$ process.

21. Let a GARCH(1,1) process with η a Gaussian white noise: $\varepsilon_t = \eta_t \sigma_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. Prove that $\varepsilon_t \in L^4$ only if $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1$.
In this case prove that $kurtosis = 3\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}$.

22. Give the log likelihood of an ARCH(1) process $\varepsilon_t = (\sqrt{\omega + \alpha \varepsilon_{t-1}^2})\eta_t$, where (η_t) are iid, standard Gaussian law.

23. Let an EGARCH process: $\varepsilon_t = \sigma_t \eta_t$, $\log \sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i g(\eta_{t-i}) + \sum_{j=1}^p \beta_j \log \sigma_{t-j}^2$,

where $g(\eta_{t-i}) = \theta\eta_{t-i} + \zeta(|\eta_{t-i}| - E(|\eta_{t-i}|)$, and $\omega, \beta, \theta, \zeta \in \mathbb{R}$. Prove that the volatility σ has a multiplicative dynamics.

24. Proof of Th. 10.2 in case $p = 1$, $|\beta| < 1$, $E[(\log \eta_t^2)^2] < \infty$ and $G = E[g^2(\eta_t)] < \infty$.

25. TGARCH(1,1) model: $\varepsilon_t = \eta_t\sigma_t$, $\sigma_t = \omega + a(\eta_{t-1})\sigma_{t-1}$, with $a(z) = \alpha_+ z^+ + \alpha_- z^- + \beta$, assume $E[a^m(\eta_t)] < 1$.
(i)Prove that the assumption $\|a(\eta)\|_2 < 1$ implies the condition $E[\ln a(\eta)] < 0$ in case of $\beta = 1$ and symmetrical law for the η_t, ;
(ii) Compute the moments of η_t to provide skewness and kurtosis $\kappa_\varepsilon = 3\frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2}$. Cf. [6] p. 252.

26. APAGARCH: In the case $\beta_1 = 0$ and when the law of η_t is symmetric, express the condition $E[\log(\alpha_1(|\eta_t| - \zeta_1\eta_t) + \beta_1)^\delta] < 0$, cf. [6] (10.24) page 257.

27. In the following example, $\omega_h = h\omega$, $a_h(z) = 1 - h\delta + \sqrt{h}(\rho z + \sqrt{\zeta - \rho^2}\eta')$, η' being independent of η, η and $\eta' \in L^{2(1+\delta)}$, check the assumptions:
$h^{-1}\omega_h \to \omega$; $h^{-1}(1 - E[a_h(\eta_n^h)]) \to \delta$; $h^{-1}Var[a_h(\eta_n^h)] \to \zeta$; $h^{-1/2}Cov(a_h(\eta_n^h), \eta_n^h) \to \rho$, $\limsup_{h\to 0} h^{-1-\delta}E\left[(a_h(\eta_n^h) - 1)^{2(1+\delta)}\right] < +\infty$,

28. Feymann-Kac formula for the option price: $C(S,t) = e^{-r(t-T)}E_Q[g(S_T)/\mathcal{F}_t]$: prove that $\sigma \to C(S,\sigma,t)$ is increasing (cf. Jeanblanc-Yor) so at least numerically invertible.

29. Look at Black-Scholes model, $Z_t = \log S_t - \log S_{t-1} = \mu - \frac{1}{2}\sigma^2 + \sigma\varepsilon_t$, one step SDF is defined as $B(t,t+1) = e^{-r}$, $M_{t,t+1} = \exp(a + bZ_{t+1})$.
With the constraint $B(t,t+1) = E[M_{t,t+1}/\mathcal{F}_t]$ (16) we get $e^{-r} = E[\exp(a + bZ_{t+1})/\mathcal{F}_t]$, and $S_t = E[S_{t+1}M_{t,t+1}/\mathcal{F}_t]$ means $1 = E[e^{a+(b+1)Z_{t+1}}/\mathcal{F}_t]$.
Using that the law of Z_{t+1} given $\mathcal{F}_t$ is the Gaussian law $(\mu - \frac{1}{2}\sigma^2, \sigma^2)$, prove the existence of a and b. Then define the risk neutral probability with its characteristic function $E_\pi[e^{uZ_{t+1}}]$
$E(e^{aX}) = e^{ma+\frac{1}{2}a^2\sigma^2}$ if X Gaussian (m, σ^2)

30. GARCH-type model: $Z_t = \log S_t - \log S_{t-1} = \mu_t + \varepsilon_t$, $\varepsilon_t = \sigma_t\eta_t$ where η is a white noise. Suppose that the filtrations generated by ε, Z, η are the same. Once again $B(t,t+1) = e^{-r}$, and the SDF $M_{t,t+1} = \exp(a_t + b_t\eta_{t+1})$, where the $\mathcal{F}$-adapted processes a and b are to be provided.

31. Prove that the expected shortfall satisfies $ES_{t,h}(\alpha) = \frac{1}{\alpha}\int_0^\alpha VaR_{t,h}(u)du$.
(Exercise 12.16 in [6], look also at Example 12.11 page 332).

References

[1] R. AZENCOTT, D. DACUNHA-CASTELLE : “Séries d'observations irrégulières”, Masson, Paris, 1984.

[2] Luc BAUWENS, Christian M. HAFNER, Sébastien LAURENT, Handbook of Volatility Models and Their Applications, Wiley 2012.

[3] T.BOLLERSEV, Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* 31, 307-327, 1986.

[4] BOX AND JENKINS : "Times series analysis", Holden-Day, San Francisco, 1976.

[5] D. COX: "Prediction by exponentially weighted moving average and related methods", Journal of the Royal Statistical Society B,23, p 414-422, 1961.

[6] Christian FRANCQ and Jean-Michel ZAKOIAN, GARCH Models: Structure, Statistical Inference and Financial Applications, Wiley, 2010.

[7] C. GOURIEROUX et A. MONFORT : "Séries temporelles et modèles dynamiques",Economica, Paris,1990.

[8] E.J. HANNAN : "Multiple time series", Wiley, New York, 1970.

[9] A.N. SHIRYAEV: "Essential of Stochastic finance", Chapter II, Advanced Series on Statistical Science and Appl. Vol 3, World Scientific, 1999.

[10] Timo Teräsvirta, An Introduction to Univariate GARCH Models, pp. 17-42, Chapter in "Handbook on Financial Time Series", Springer, 2009.

[11] T.H. WONNACOTT and R.J. WONNACOTT : Introductory Statistics, Wiley, 1969. *perhaps on line...http://www.pdfsdocuments.com/wonnacott-introductory-statistics.pdf*