4 Financial time series

4.1 Fundamentals of time series analysis

4.1.1 Basic definitions

A *stochastic process* is a family of rvs $(X_t)_{t \in I}$, $I \subseteq \mathbb{R}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A *time series* is a discrete-time ($I \subseteq \mathbb{Z}$) stochastic process.

Definition 4.1 (Mean function, autocovariance function)
Assuming they exist, the *mean function* $\mu(t)$ and the *autocovariance function* $\gamma(t, s)$ of $(X_t)_{t \in \mathbb{Z}}$ are defined by

$$\begin{aligned} \mu(t) &= \mathbb{E}(X_t), \quad t \in \mathbb{Z}, \\ \gamma(t, s) &= \mathrm{cov}(X_t, X_s) = \mathbb{E}((X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)), \quad t, s, \in \mathbb{Z}. \end{aligned}$$

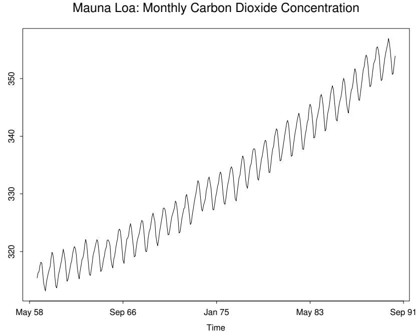
Definition 4.2 ((Weak/strict) stationarity)

1) $(X_t)_{t\in\mathbb{Z}}$ is *(weakly/covariance) stationary* if $\mathbb{E}(X_t^2) < \infty$, $\mu(t) = \mu \in \mathbb{R}$ and $\gamma(t,s) = \gamma(t+h, s+h)$ for all $t, s, h \in \mathbb{Z}$.
2) $(X_t)_{t\in\mathbb{Z}}$ is *strictly stationary* if $(X_{t_1}, \dots, X_{t_n}) \stackrel{\text{d}}{=} (X_{t_1+h}, \dots, X_{t_n+h})$ for all $t_1, \dots, t_n, h \in \mathbb{Z}$, $n \in \mathbb{N}$.

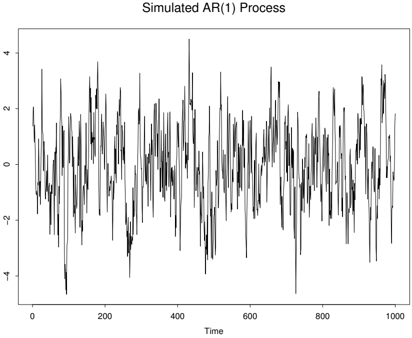
Remark 4.3

1) Both types of stationarity formalize that $(X_t)_{t\in\mathbb{Z}}$ behaves similarly in any epoch.
2)
   - Strict stationarity $\not\Rightarrow$ stationarity: $\mathbb{E}(X_t^2)$ doesn't have to exist (e.g. GARCH processes). If it does, strict stationarity implies stationarity.
   - Stationarity $\not\Rightarrow$ strict stationarity: $\mathbb{E}(|X_t|^p)$, $p > 2$, could change
3) $(X_t)_{t\in\mathbb{Z}} \Rightarrow \gamma(t-s, 0) = \gamma(t,s) = \gamma(s,t) = \gamma(s-t, 0)$, so $\gamma(t,s)$ only depends on the *lag* $h = |t-s|$. We can thus use $\gamma(h) := \gamma(|h|, 0)$, $h \in \mathbb{Z}$.

Stationary?



Mauna Loa: Monthly Carbon Dioxide Concentration
320
330
340
350
May 58
Sep 66
Jan 75
May 83
Sep 91
Time



Simulated AR(1) Process
−4
−2
0
2
4
0
200
400
600
800
1000
Time

(Partial) autocorrelation in stationary time series

Definition 4.4 (ACF)

The *autocorrelation function (ACF)* (or *serial correlation*) of a stationary time series $(X_t)_{t\in\mathbb{Z}}$ is defined by $\rho(h) = \mathrm{corr}(X_h, X_0) = \gamma(h)/\gamma(0), \quad h \in \mathbb{Z}$.

The study of autocorrelation is known as *analysis in the time domain*. Another important quantity is the partial autocorrelation function. The *partial autocorrelation function (PACF)* ϕ is defined by

$$\phi(h) = \mathrm{corr}(X_0 - P_{\mathcal{H}_{h-1}}X_0, X_h - P_{\mathcal{H}_{h-1}}X_h),$$

where $P_{\mathcal{H}_{h-1}}X_t$ denotes the best approximation/prediction of X_t from an element of $\mathcal{H}_{h-1} = \{\sum_{k=1}^{n-1} \alpha_k X_{n-k} : \alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}\}$. Note that $\phi(1) = \phi_{1,1} = \gamma(1)/\gamma(0) = \rho(1)$.

- The PACF is the corr between X_0 and X_h with the linear dependence of $X_1, \dots, X_{h-1}$ removed.

- It can be computed with the Durbin-Levinson algorithm.
- It is mainly used for model identification of $\mathrm{AR}(p)$ processes similarly to how the ACF is used for $\mathrm{MA}(q)$ processes (see later).

White noise processes

Definition 4.5 ((Strict) white noise)

1) $(X_t)_{t\in\mathbb{Z}}$ is a *white noise* process if it is stationary with $\rho(h) = I_{\{h=0\}}$ (no serial correlation). If $\mu(t) = 0$, $\gamma(0) = \sigma^2$, $(X_t)_{t\in\mathbb{Z}}$ is denoted by $\mathrm{WN}(0, \sigma^2)$.
2) $(X_t)_{t\in\mathbb{Z}}$ is a *strict white noise* process if it is a sequence of iid rvs with $\gamma(0) = \sigma^2 < \infty$. If $\mu(t) = 0$, we write $\mathrm{SWN}(0, \sigma^2)$.

One further noise concept is the following (see GARCH processes later).

Let $(X_t)_{t\in\mathbb{Z}}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. A sequence $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ of σ-algebras is called *filtration* if $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$, $t \in \mathbb{Z}$. If $\mathcal{F}_t = \sigma(\{X_s : s \le t\})$, we call $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ the *natural filtration* of $(X_t)_{t\in\mathbb{Z}}$. $(X_t)_{t\in\mathbb{Z}}$ is *adapted* to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ if $X_t \in \mathcal{F}_t$, $t \in \mathbb{Z}$ (X_t is $\mathcal{F}_t$-measurable).

Definition 4.6 (MGDS)

$(X_t)_{t\in\mathbb{Z}}$ is a *martingale-difference sequence (MGDS)* w.r.t. $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ if

i) $\mathbb{E}|X_t| < \infty$ for all t;

ii) $(X_t)_{t\in\mathbb{Z}}$ is adapted to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$; and

iii) $\mathbb{E}(X_{t+1} \,|\, \mathcal{F}_t) = 0$ for all $t \in \mathbb{Z}$.

- If $\mathbb{E}(X_{t+1}|F_t) = X_t$ a.s., then (X_t) is a *martingale* and $\varepsilon_t = X_t - X_{t-1}$ is a MGDS (winnings in rounds of a *fair game*).
- One can show that a MGDS $(\varepsilon_t)_{t\in\mathbb{Z}}$ with $\sigma^2 = \mathbb{E}(\varepsilon_t^2) < \infty$ satisfies $\rho(h) = 0$, $h \neq 0$, so $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$. Furthermore, $\mathbb{E}(\varepsilon_{t+1+k} \,|\, \mathcal{F}_t) = \mathbb{E}(\mathbb{E}(\varepsilon_{t+1+k} \,|\, \mathcal{F}_{t+k}) \,|\, \mathcal{F}_t) = 0$, $k \in \mathbb{N}$.

4.1.2 ARMA processes

Definition 4.7 ($\mathrm{ARMA}(p, q)$)
Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$. $(X_t)_{t\in\mathbb{Z}}$ is a *zero-mean $\mathrm{ARMA}(p, q)$ process* if it is stationary and satisfies, for all $t \in \mathbb{Z}$,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}. \qquad (6)$$

$(X_t)_{t\in\mathbb{Z}}$ is $\mathrm{ARMA}(p, q)$ with *mean μ* if $(X_t - \mu)_{t\in\mathbb{Z}}$ is a zero-mean $\mathrm{ARMA}(p, q)$.

Remark 4.8

- If the *innovations* $(\varepsilon_t)_{t\in\mathbb{Z}}$ are $\mathrm{SWN}(0, \sigma^2)$, then $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary (follows from the representation as a linear process below).
- The defining equation (6) can be written as

$$\phi(B)X_t = \theta(B)\varepsilon_t, \quad t \in \mathbb{Z},$$

where

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p,$$
$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$
$$B : \ B^k X_t = X_{t-k}, \quad k \in \mathbb{Z} \quad (\textit{backshift operator})$$

Causal processes

For practical purposes, it suffices to consider *causal* ARMA processes, that is, $(X_t)_{t\in\mathbb{Z}}$ satisfying (6) which can be represented as

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \quad \text{(depends on the past/present, not the future)}$$

for $\sum_{k=0}^{\infty} |\psi_k| < \infty$ (*absolute summability condition*; guarantees $\mathbb{E}|X_t| < \infty$).

Proposition 4.9 (ACF for causal processes)

Any process $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ such that $\sum_{k=0}^{\infty} |\psi_k| < \infty$ is stationary with

$$\rho(h) = \frac{\sum_{k=0}^{\infty} \psi_k \psi_{k+|h|}}{\sum_{k=0}^{\infty} \psi_k^2}, \quad h \in \mathbb{Z}.$$

Theorem 4.10 (Stationary and causal ARMA solutions)

Let $(X_t)_{t\in\mathbb{Z}}$ be an $\mathrm{ARMA}(p, q)$ process for which $\phi(z), \theta(z)$ have no roots in common. Then

$$(X_t)_{t\in\mathbb{Z}} \text{ is stationary and causal} \quad \Leftrightarrow \quad \phi(z) \neq 0 \quad \forall\, z \in \mathbb{C} : |z| \leq 1.$$

In this case, $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ for $\sum_{k=0}^{\infty} \psi_k z^k = \theta(z)/\phi(z)$, $|z| \leq 1$.

- See the appendix for an idea of the proof.
- Note that if $\theta(z) \neq 0$, $|z| \leq 1$ (known as *invertibility condition*), we can recover ε_t from $(X_s)_{s\leq t}$ via $\varepsilon_t = \phi(B)X_t/\theta(B)$.

- An $\mathrm{ARMA}(p, q)$ process with mean μ can be written as

$$X_t = \mu_t + \varepsilon_t$$

$$\mu_t = \mu + \sum_{k=1}^{p} \phi_k (X_{t-k} - \mu) + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k}.$$

- If $(X_t)_{t \in \mathbb{Z}}$ is invertible then ε_{t-k} can be expressed in terms of $(X_s)_{s \le t-k}$, hence μ_t can be expressed by $(X_s)_{s \le t-1}$. It follows that μ_t is $\mathcal{F}_{t-1}$-measurable where $\mathcal{F}_{t-1} = \sigma(\{X_s : s \le t-1\})$.
- If $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a MGDS w.r.t. $(\mathcal{F}_t)_{t \in \mathbb{Z}}$, then $\mu_t = \mathbb{E}(X_t \,|\, \mathcal{F}_{t-1})$.

  $\Rightarrow$ An ARMA process puts a particular structure on the conditional mean μ_t given the past. As we will see, a GARCH process puts a certain structure on the conditional variance $\sigma_t^2 = \mathrm{var}(X_t \,|\, \mathcal{F}_{t-1})$.

Example 4.11

1) $\mathrm{MA}(q) = \mathrm{ARMA}(0,q)$: $X_t = \varepsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} \underset{\theta_0=1}{=} \sum_{k=0}^{q} \theta_k \varepsilon_{t-k}$ (causal, absolute summability condition fulfilled).

- ACF: Proposition 4.9 $\Rightarrow \rho(h) = \frac{\sum_{k=0}^{q-|h|} \theta_k \theta_{k+|h|}}{\sum_{k=0}^{q} \theta_k^2}$, $|h| \in \{1, \dots, q\}$, and $\rho(h) = 0$ for all $|h| > q \Rightarrow$ ACF cuts off after lag q.
- PACF: One can show that for an $\mathrm{MA}(q)$, $\phi(h)$ does not cut off but $|\phi(h)|$ is bounded by an exponentially decreasing function in h.

2) $\mathrm{AR}(p) = \mathrm{ARMA}(p,0)$: $X_t - \sum_{k=1}^{p} \phi_k X_{t-k} = \varepsilon_t$. ACF: As for general ARMA processes, the ACF can be computed in several ways; see Brockwell and Davis (1991, Section 3.3), e.g. via $X_t = \theta(B)\varepsilon_t/\phi(B) = \psi(B)\varepsilon_t$ from $\rho(h)$ as in Proposition 4.9.

By Theorem 4.10, an $\mathrm{AR}(1)$ has a stationary and causal solution if and only if $1 - \phi_1 z \neq 0$ for all $z \in \mathbb{C} : |z| \le 1$, so $|\phi_1| < 1$. In this case,

$$X_t = \phi_1 X_{t-1} + \varepsilon_t = \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \dots$$

$$= \phi_1^n X_{t-n} + \sum_{k=0}^{n-1} \phi_1^k \varepsilon_{t-k} \to \sum_{k=0}^{\infty} \phi_1^k \varepsilon_{t-k},$$

so $\psi_k = \phi_1^k$, $k \in \mathbb{N}_0$. By Proposition 4.9,

$$\rho(h) = \frac{\sum_{k=0}^{\infty} \phi_1^{2k+|h|}}{\sum_{k=0}^{\infty} \phi_1^{2k}} = \phi_1^{|h|}, \quad h \in \mathbb{Z}.$$

and the ACF decreases exponentially. For $\mathrm{AR}(p)$, one can show this from a general form of ψ_k (see Brockwell and Davis (1991, p. 92)), possibly with damped $\sin$ waves. Furthermore, one can show that the PACF of an $\mathrm{AR}(p)$ cuts off after lag p; it can be computed with the Durbin–Levinson algorithm.

3) $\mathrm{ARMA}(1,1)$: $X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ for $|\phi_1| < 1$ (by Theorem 4.10 this has a stationary and causal solution). For determining the ACF, we first rewrite the process as $X_t = \psi(B)\varepsilon_t$, where

$$\psi(z) = \frac{\theta(z)}{\phi(z)} = \frac{1+\theta_1 z}{1-\phi_1 z} = (1+\theta_1 z) \sum_{k=0}^{\infty} (\phi_1 z)^k$$

$$= \sum_{k=0}^{\infty} \phi_1^k z^k + \sum_{k=1}^{\infty} \theta_1 \phi_1^{k-1} z^k = 1 + \sum_{k=1}^{\infty} \phi_1^{k-1}(\phi_1 + \theta_1) z^k,$$

hence $\psi_0 = 1$ and $\psi_k = \phi_1^{k-1}(\phi_1 + \theta_1)$, $k \geq 1$. It follows that

$$\sum_{k=0}^{\infty} \psi_k \psi_{k+h} \underset{h \geq 1}{=} \underbrace{\psi_0 \psi_h}_{= \phi_1^{h-1}(\phi_1+\theta_1)} + \underbrace{\sum_{k=1}^{\infty} \phi_1^{k-1+k+h-1}(\phi_1 + \theta_1)^2}_{= (\phi_1+\theta_1)^2 \phi_1^h \sum_{k=0}^{\infty} \phi_1^{2k}}$$

$$= \phi_1^{h-1}(\phi_1 + \theta_1)(1 + (\phi_1 + \theta_1)\phi_1/(1 - \phi_1^2))$$

$$= \frac{\phi_1^{h-1}}{1 - \phi_1^2}(\phi_1 + \theta_1)(1 + \phi_1 \theta_1).$$

Proposition 4.9 then implies that

$$\rho(h) = \phi_1^{h-1} \frac{(\phi_1 + \theta_1)(1 + \phi_1 \theta_1)}{1 + 2\phi_1 \theta_1 + \theta_1^2} = \phi_1^{h-1} \rho(1) \underset{(h \to \infty)}{\searrow} 0,$$

so that $\rho(h) = \phi_1^{|h|-1} \rho(1)$ for all $h \in \mathbb{Z} \setminus \{0\}$. The PACF can be computed from the Durbin–Levinson algorithm.

Remark 4.12

$(X_t)_{t\in\mathbb{Z}}$ is an ARIMA(p,d,q) (Integrated) process if

$$\underbrace{\phi(B)}_{\text{order } p}\ \underbrace{\overbrace{(1-B)^d}^{\text{integrated part}}}_{\text{order } d}\ X_t = \underbrace{\theta(B)}_{\text{order } q}\,\varepsilon_t, \quad t\in\mathbb{Z}.$$

We see that this is also an $\mathrm{ARMA}(d+p,q)$ process. Extensions to SARIMA (Seasonal) models are available.

4.1.3 Analysis in the time domain

Correlogram

A *correlogram* is a plot of $(h,\hat\rho(h))_{h\ge 0}$ for the sample ACF

$$\hat\rho(h) = \frac{\sum_{t=1}^{n}(X_{t+h}-\bar X_n)(X_t-\bar X_n)}{\sum_{t=1}^{n}(X_t-\bar X_n)^2}, \quad h\in\{0,\dots,n\}.$$

The sample PACF can be computed from $\hat\rho(h)$ via the DL algorithm.

Theorem 4.13

Let $X_t - \mu = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$ and $(Z_t) \sim \mathrm{SWN}(0, \sigma^2)$. Under suitable conditions,

$$\sqrt{n}\left(\begin{pmatrix}\hat{\rho}(1)\\ \vdots \\ \hat{\rho}(h)\end{pmatrix} - \begin{pmatrix}\rho(1)\\ \vdots \\ \rho(h)\end{pmatrix}\right) \underset{(n\to\infty)}{\overset{\mathrm{d}}{\to}} \mathrm{N}_h(\mathbf{0}, W), \quad h \in \mathbb{N},$$

for some covariance matrix W depending on ρ; see McNeil et al. (2015, Theorem 4.13).

If the ARMA process is SWN itself, then $\sqrt{n}\begin{pmatrix}\hat{\rho}(1)\\ \vdots \\ \hat{\rho}(h)\end{pmatrix} \underset{(n\to\infty)}{\overset{\mathrm{d}}{\to}} \mathrm{N}_h(\mathbf{0}, I_h)$, $h \in \mathbb{N}$, so that with probability $1-\alpha$,

$$\hat{\rho}(k) \underset{(n \text{ large})}{\in} \left[-\frac{q_{1-\alpha/2}}{\sqrt{n}}, \frac{q_{1-\alpha/2}}{\sqrt{n}}\right] = I_{\alpha,n}, \quad k \in \{1, \dots, h\},$$

where $q_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$. $I_{0.05,n}$ is typically displayed in the correlogram. If more than 5% of $\hat{\rho}(k)$, $k \in \{1, \dots, h\}$, lie outside $I_{0.05,n}$, this is evidence against the (iid) hypothesis of $\mathrm{SWN} \Rightarrow$ serial correlation.

Portmanteau tests

- As a formal test of this hypothesis (SWN), one can use the Ljung–Box test with test statistic

$$T = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}(k)^2}{n-k} \underset{n \text{ large}}{\sim} \chi_h^2; \quad \text{reject if } T > {\chi_h^2}^{-1}(1-\alpha).$$

- If $(X_t)_{t\in\mathbb{Z}}$ is SWN, $(|X_t|)_{t\in\mathbb{Z}}$ is also iid It is a good idea to also apply the correlogram and Ljung–Box tests to $(|X_t|)_{t\in\mathbb{Z}}$ as a further test.

4.1.4 Statistical analysis of time series

The Box–Jenkins approach

Approach for the statistical analysis of $(X_t)_{t\in\mathbb{Z}}$:

1) Preliminary analysis

   i) Plot the time series $\Rightarrow$ Does it look stationary?

   ii) If necessary, clean the (e.g. high-frequency) data and plot it again.

iii) Make it stationary by removing trend and seasonality (regime switches etc.). A typical decomposition is

$$X_t = \underbrace{\mu_t}_{\text{trend}} + \underbrace{s_t}_{\text{seasonal component}} + \underbrace{\varepsilon_t}_{\text{residual process}}.$$

- A trend μ_t can be estimated via smoothing with local averages:

$$\begin{aligned}\tilde{X}_t &= \frac{1}{2h+1}\sum_{k=-h}^{h} X_{t+k}\\ &= \underbrace{\sum_{k=-h}^{h}\frac{\mu_{t+k}}{2h+1}}_{\approx\mu_t} + \underbrace{\sum_{k=-h}^{h}\frac{s_{t+k}}{2h+1}}_{\approx 0} + \underbrace{\sum_{k=-h}^{h}\frac{\varepsilon_{t+k}}{2h+1}}_{=\tilde{\varepsilon}_t}\end{aligned}$$

  or exponentially weighted moving averages (see `HoltWinters(, beta=FALSE, gamma=FALSE)`).

- A seasonal component s_t can be estimated similarly, simply consider $(\tilde{X}_s)_{s=1}^S$ (for monthly data, $S = 12$) with

$$\tilde{X}_s = \frac{1}{N} \sum_{k=0}^{N-1} X_{s+kS}, \quad s \in \{1, \dots, S\},\ N = \left\lfloor \frac{n}{S} \right\rfloor.$$

Removing μ_t, s_t can be done non-parametrically (see R's `stl()`) or via regression or by taking differences.

2) Analysis in the time domain

   i) Plot ACF, PACF and use the Ljung–Box test for $(X_t)_{t\in\mathbb{Z}}$ (hints at an ARMA) and $(|X_t|)_{t\in\mathbb{Z}}$ (hints at an GARCH). If the SWN hypothesis cannot be rejected, fit a (static) distribution.

   ii) Do ACF (MA) or PACF (AR) cut off? (determines the order(s))

3) Model fitting

   i) Identify the order (if possible; see above);

   ii) Fit various (low-order) ARMA models (various ways; often (conditional) MLE);

iii) Model-selection criterion (e.g. AIC, BIC) $\Rightarrow$ select "best" model; see also the automatic procedure by Tsay and Tiao (1984).

4) Residual analysis

i) Consider the residuals

$$\hat{\varepsilon}_t = X_t - \hat{\mu}_t, \quad \hat{\mu}_t = \hat{\mu} + \sum_{k=1}^{p} \hat{\phi}_k (X_{t-k} - \hat{\mu}) + \sum_{k=1}^{q} \hat{\theta}_k \hat{\varepsilon}_{t-k},$$

typically recursively computed (e.g. by letting the first q $\hat{\varepsilon}$'s be 0 and the first p X's be $\bar{X}_n$)

ii) Check (plots, ACF, Ljung–Box, ...) the model assumptions.

4.1.5 Prediction

Let $X_{t-n+1}, \dots, X_t$ denote the data available at time t and suppose we want to compute $P_t X_{t+1}$. Assume we have the history $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$ of the underlying ARMA model available (including today t).

Conditional expectation ($\mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t)$ is best L^2 approx. to X_{t+h})

Let the ARMA $(X_t)_{t\in\mathbb{Z}}$ be invertible and $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a MGDS w.r.t. $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. Since $\mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t)$ minimizes $\mathbb{E}((X_{t+h} - \cdot)^2)$, $P_t X_{t+h} = \mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t)$ $\Rightarrow$ Compute $\mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t)$ recursively in terms of $\mathbb{E}(X_{t+h-1} \,|\, \mathcal{F}_t)$. Use that $\mathbb{E}(\varepsilon_{t+h} \,|\, \mathcal{F}_t) = 0$ and that $(X_s)_{s\le t}$, $(\varepsilon_s)_{s\le t}$ are "known" at time t (invertibility insures that ε_t can be written as a function of $(X_s)_{s\le t}$).

Example 4.14 (Prediction in the $\mathrm{ARMA}(1,1)$ model)

$\mathrm{ARMA}(1,1)$: $X_t - \mu = \phi_1(X_{t-1} - \mu) + \varepsilon_t + \theta_1\varepsilon_{t-1}$. Then

$$\begin{aligned}
\mathbb{E}(X_{t+1} \,|\, \mathcal{F}_t) &= \mu + \phi_1(X_t - \mu) + \theta_1\varepsilon_t + \underbrace{\mathbb{E}(\varepsilon_{t+1} \,|\, \mathcal{F}_t)}_{\underset{\text{MGDS}}{=}0};\\
\mathbb{E}(X_{t+2} \,|\, \mathcal{F}_t) &= \mu + \phi_1\mathbb{E}(X_{t+1} \,|\, \mathcal{F}_t) - \phi_1\mu\\
&\qquad + \theta_1 \underbrace{\mathbb{E}(\varepsilon_{t+1} \,|\, \mathcal{F}_t)}_{=0} + \underbrace{\mathbb{E}(\varepsilon_{t+2} \,|\, \mathcal{F}_t)}_{=0}\\
&= \mu + \phi_1(\mathbb{E}(X_{t+1} \,|\, \mathcal{F}_t) - \mu) = \mu + \phi_1^2(X_t - \mu) + \phi_1\theta_1\varepsilon_t;\\
\mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t) &= \cdots = \mu + \phi_1^h(X_t - \mu) + \phi_1^{h-1}\theta_1\varepsilon_t \underset{(h\to\infty)}{\to} \mu.
\end{aligned}$$

Exponentially weighted moving averages

- Used for prediction and trend estimation;
- Assume there is no deterministic seasonal component;
- Typically directly applied to price series;
- Prediction

$$P_t X_{t+1} = \sum_{k=0}^{n-1} \alpha(1-\alpha)^k X_{t-k} = \alpha X_t + (1-\alpha) P_{t-1} X_t.$$

Increasing $\alpha \in (0,1)$ puts more weight on the last observation.

4.2 GARCH models for changing volatility

- (G)ARCH = (generalized) autoregressive conditionally heteroscedastic
- They are the most important models for daily risk-factor returns.

4.2.1 ARCH processes

Definition 4.15 ($\mathrm{ARCH}(p)$)
Let $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0,1)$. $(X_t)_{t\in\mathbb{Z}}$ is an $\mathrm{ARCH}(p)$ *process* if it is strictly stationary and satisfies

$$X_t = \sigma_t Z_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k}^2,$$

where $\alpha_0 > 0$, $\alpha_k \geq 0$, $k \in \{1, \dots, p\}$.

Typical examples: $Z_t \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$ or $Z_t \overset{\text{ind.}}{\sim} t_\nu(0,\sigma^2)$ for $\sigma^2 = (\nu-2)/\nu$.

Remark 4.16

1) σ_{t+1} is $\mathcal{F}_t$-measurable $\Rightarrow \mathbb{E}(X_{t+1}\,|\,\mathcal{F}_t) = \sigma_{t+1}\mathbb{E}(Z_{t+1}\,|\,\mathcal{F}_t) = \sigma_{t+1}\mathbb{E}(Z_{t+1})$ $= 0$. Thus, $\mathrm{ARCH}(p)$ processes are MGDSs w.r.t. the natural filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. If they are stationary, they are white noise since

$$\gamma(h) = \mathbb{E}(X_t X_{t+h}) \underset{\text{property}}{\overset{\text{tower}}{=}} \mathbb{E}(\mathbb{E}(X_t X_{t+h}\,|\,\mathcal{F}_{t+h-1}))$$
$$= \mathbb{E}(X_t\mathbb{E}(X_{t+h}\,|\,\mathcal{F}_{t+h-1})) = 0, \quad h \in \mathbb{N}.$$

This also applies to GARCH processes; see below.

2) If $(X_t)_{t\in\mathbb{Z}}$ is stationary, then $\mathrm{var}(X_{t+1}\,|\,\mathcal{F}_t) = \mathbb{E}((\sigma_{t+1}Z_{t+1})^2\,|\,\mathcal{F}_t) =$ $\sigma_{t+1}^2\mathbb{E}(Z_{t+1}^2\,|\,\mathcal{F}_t) = \sigma_{t+1}^2\mathbb{E}(Z_{t+1}^2) = \sigma_{t+1}^2$.

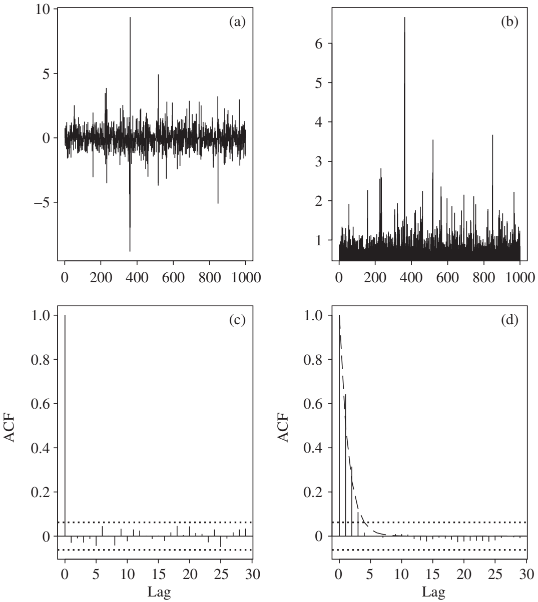
$\Rightarrow$ *Volatility* ($= \sigma_t$, the conditional standard deviation) is changing in time, depending on past values of the process. This is where "autoregressive conditionally heteroscedastic" comes from. ARCH models can thus capture *volatility clustering* (if one of $|X_{t-1}|, \dots, |X_{t-p}|$ is large, X_t is drawn from a distribution with large variance).

Example 4.17 ($\mathrm{ARCH}(1)$)

- One can show that an $\mathrm{ARCH}(1)$ process $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary $\Leftrightarrow \mathbb{E}(\log(\alpha_1 Z_t^2)) < 0$. In this case, $X_t^2 = \alpha_0 \sum_{k=0}^{\infty} \alpha_1^k \prod_{j=0}^{k} Z_{t-j}^2$.
- $(X_t)_{t\in\mathbb{Z}}$ is stationary $\Leftrightarrow \alpha_1 < 1$. In this case, $\mathrm{var}(X_t) = \alpha_0/(1-\alpha_1)$.

  Proof of necessity. $X_t^2 = \sigma_t^2 Z_t^2 = (\alpha_0 + \alpha_1 X_{t-1}^2) Z_t^2 \Rightarrow \sigma_X^2 = \mathbb{E}(X_t^2) = \alpha_0 + \alpha_1 \mathbb{E}(X_{t-1}^2 Z_t^2) = \alpha_0 + \alpha_1 \sigma_X^2 \Rightarrow \sigma_X^2 = \frac{\alpha_0}{1-\alpha_1}$, $\alpha_1 < 1$. $\square$

  For sufficiency, see McNeil et al. (2015, Proposition 4.18).
- Provided that $\mathbb{E}(Z_t^4) < \infty$ and $\alpha_1 < (\mathbb{E}(Z_t^4))^{-1/2}$, one can show that $\kappa(X_t) = \frac{\mathbb{E}(X_t^4)}{\mathbb{E}(X_t^2)^2} = \frac{\kappa(Z_t)(1-\alpha_1^2)}{(1-\alpha_1^2\kappa(Z_t))}$. If $\kappa(Z_t) > 1$, $\kappa(X_t) > \kappa(Z_t)$. For Gaussian or t innovations, $\kappa(X_t) > 3$ (leptokurtic).
- Parallels with the AR(1) process: If $\mathbb{E}(X_t^4) < \infty$, $\alpha_1 < 1$ and $\varepsilon_t = \sigma_t^2(Z_t^2 - 1)$, one can show that $(X_t^2)_{t\in\mathbb{Z}}$ is an $\mathrm{AR}(1)$ of the form $X_t^2 - \frac{\alpha_0}{1-\alpha_1} = \alpha_1(X_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1}) + \varepsilon_t$.



a) Realization ($n = 1000$) of an $\mathrm{ARCH}(1)$ process with $\alpha_0 = 0.5$, $\alpha_1 = 0.5$ and Gaussian innovations;

b) Realization of the volatility $(\sigma_t)_{t\in\mathbb{Z}}$;

c) Correlogram of $(X_t)_{t\in\mathbb{Z}}$, compare with Remark 4.16 1);

d) Correlogram of $(X_t^2)_{t\in\mathbb{Z}}$ (AR(1)); dashed line = true ACF

4.2.2 GARCH processes

Definition 4.18 ($\mathrm{GARCH}(p,q)$)
Let $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0,1)$. $(X_t)_{t\in\mathbb{Z}}$ is a $\mathrm{GARCH}(p,q)$ *process* if it is strictly stationary and satisfies

$$X_t = \sigma_t Z_t,$$

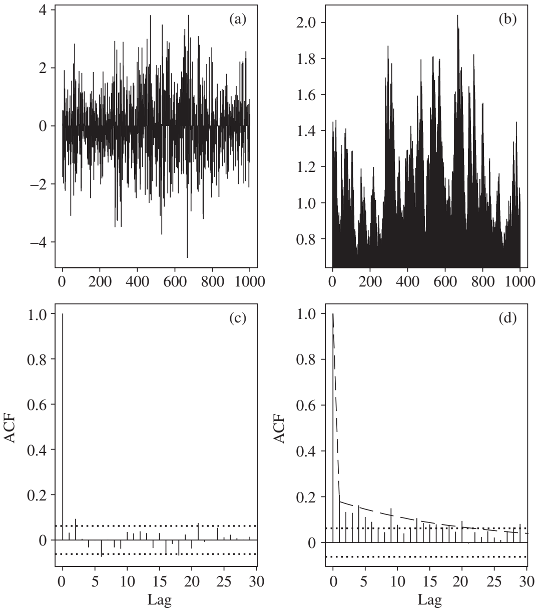
$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2,$$

where $\alpha_0 > 0$, $\alpha_k \ge 0$, $k \in \{1,\dots,p\}$, $\beta_k \ge 0$, $k \in \{1,\dots,q\}$.

If one of $|X_{t-1}|,\dots,|X_{t-p}|$ or $\sigma_{t-1},\dots,\sigma_{t-q}$ is large, X_t is drawn from a distribution with (persistently) large variance. Periods of high volatility tend to be more persistent.

Example 4.19 ($\mathrm{GARCH}(1,1)$)

- One can show (via stoch. recurrence relations) that a $\mathrm{GARCH}(1,1)$ process $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary if $\mathbb{E}(\log(\alpha_1 Z_t^2+\beta_1))<\infty$. In this case, $X_t = Z_t\sqrt{\alpha_0(1+\sum_{k=1}^{\infty}\prod_{j=1}^{k}(\alpha_1 Z_{t-j}^2+\beta_1))}$.
- $(X_t)_{t\in\mathbb{Z}}$ is stationary $\Leftrightarrow \alpha_1+\beta_1<1$. In this case, $\mathrm{var}(X_t)=\frac{\alpha_0}{1-\alpha_1-\beta_1}$.
- Provided that $\mathbb{E}((\alpha_1 Z_t^2+\beta_1)^2)<1$ (or $(\alpha_1+\beta_1)^2<1-(\kappa(Z_t)-1)\alpha_1^2$), one can show that $\kappa(X_t)=\frac{\kappa(Z_t)(1-(\alpha_1+\beta_1)^2)}{1-(\alpha_1+\beta_1)^2-(\kappa(Z_t)-1)\alpha_1^2}$. If $\kappa(Z_t)>1$ (Gaussian, scaled t innovations), $\kappa(X_t)>\kappa(Z_t)$.
- Parallels with the ARMA(1,1) process: If $\mathbb{E}(X_t^4)<\infty$, $\alpha_1+\beta_1<1$ and $\varepsilon_t=\sigma_t^2(Z_t^2-1)$, one can show that $(X_t^2)_{t\in\mathbb{Z}}$ is an $\mathrm{ARMA}(1,1)$ of the form $X_t^2-\frac{\alpha_0}{1-\alpha_1-\beta_1}=(\alpha_1+\beta_1)\big(X_{t-1}^2-\frac{\alpha_0}{1-\alpha_1-\beta_1}\big)+\varepsilon_t-\beta_1\varepsilon_{t-1}$.



a) Realization ($n = 1000$) of a $\mathrm{GARCH}(1,1)$ process with $\alpha_0 = 0.5$, $\alpha_1 = 0.1$, $\beta_1 = 0.85$ and Gaussian innovations;

b) Realization of the volatility $(\sigma_t)_{t\in\mathbb{Z}}$;

c) Correlogram of $(X_t)_{t\in\mathbb{Z}}$, compare with Remark 4.16 1);

d) Correlogram of $(X_t^2)_{t\in\mathbb{Z}}$ (ARMA(1,1)); dashed line = true ACF

Prediction of GARCH(1,1)

Assume $(X_t)_{t\in\mathbb{Z}}$ is a stationary $\mathrm{GARCH}(1,1)$ with $\mathbb{E}(X_t^4)<\infty$.

- $X_t=\sigma_t Z_t \Rightarrow \mathbb{E}(X_t\,|\,\mathcal{F}_{t-1})=\sigma_t\mathbb{E}(Z_t)=0$, so $(X_t)_{t\in\mathbb{Z}}$ is MGDS and thus, by the tower property, $\mathbb{E}(X_{t+h}\,|\,\mathcal{F}_t)=0,\ h\in\mathbb{N}$.
- $\mathbb{E}(X_{t+1}^2\,|\,\mathcal{F}_t)=\sigma_{t+1}^2\mathbb{E}(Z_{t+1})=\alpha_0+\alpha_1X_t^2+\beta_1\sigma_t^2$.
For $h\ge 2$, X_{t+h}^2 and σ_{t+h}^2 are rvs, and

$$\begin{aligned}\mathbb{E}(X_{t+h}^2\,|\,\mathcal{F}_t) &\underset{(*)}{=} \mathbb{E}(\sigma_{t+h}^2\,|\,\mathcal{F}_t)\mathbb{E}(Z_t^2)=\alpha_0+\alpha_1\mathbb{E}(X_{t+h-1}^2\,|\,\mathcal{F}_t)\\ &\quad+\beta_1\underbrace{\mathbb{E}(\sigma_{t+h-1}^2\,|\,\mathcal{F}_t)}_{\underset{(*)}{=}\mathbb{E}(X_{t+h-1}^2|\mathcal{F}_t)}=\alpha_0+(\alpha_1+\beta_1)\mathbb{E}(X_{t+h-1}^2\,|\,\mathcal{F}_t)\\ &=\ldots=\alpha_0\sum_{k=0}^{h-1}(\alpha_1+\beta_1)^k+(\alpha_1+\beta_1)^{h-1}(\alpha_1X_t^2+\beta_1\sigma_t^2).\end{aligned}$$

$$\Rightarrow \mathbb{E}(\sigma_{t+h}^2\,|\,\mathcal{F}_t)=\mathbb{E}(X_{t+h}^2\,|\,\mathcal{F}_t)\underset{(h\to\infty)}{\overset{\text{a.s.}}{\to}}\frac{\alpha_0}{1-\alpha_1-\beta_1}=\mathrm{var}(X_t).$$

The GARCH(p,q) model

- Higher-order GARCH models have the same general behaviour as $\mathrm{ARCH}(1)$ and $\mathrm{GARCH}(1,1)$ models, but their mathematical analysis becomes more tedious.
- One can show that $(X_t)_{t\in\mathbb{Z}}$ is stationary $\Leftrightarrow \sum_{k=1}^{p} \alpha_k + \sum_{k=1}^{q} \beta_k < 1$.
- A squared $\mathrm{GARCH}(p,q)$ process has the structure

$$X_t^2 = \alpha_0 + \sum_{k=1}^{\max(p,q)} (\alpha_k + \beta_k) X_{t-k}^2 + \varepsilon_t - \sum_{k=1}^{q} \beta_k \varepsilon_{t-k},$$

  where $\varepsilon_t = \sigma_t^2 (Z_t^2 - 1)$, $\alpha_k = 0$, $k \in \{p+1, \dots, q\}$ if $q > p$, or $\beta_k = 0$ for $k \in \{q+1, \dots, p\}$ if $p > q$. This resembles the $\mathrm{ARMA}(\max(p,q), q)$ process and is formally such a process provided $\mathbb{E}(X_t^4) < \infty$.
- There are also *IGARCH models* (i.e. non-stationary $\mathrm{GARCH}(p,q)$ models with $\sum_{k=1}^{p} \alpha_k + \sum_{k=1}^{q} \beta_k = 1$; infinite variance).

4.2.3 Simple extensions of the GARCH model

Consider stationary GARCH processes as white noise for ARMA processes.

Definition 4.20 ($\mathrm{ARMA}(p_1, q_1)$ with $\mathrm{GARCH}(p_2, q_2)$ errors)
Let $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0,1)$. $(X_t)_{t\in\mathbb{Z}}$ is an $\mathrm{ARMA}(p_1, q_1)$ *process with* $\mathrm{GARCH}(p_2, q_2)$ *errors* if it is stationary and satisfies

$$\begin{aligned}
X_t &= \mu_t + \sigma_t Z_t,\\
\mu_t &= \mu + \sum_{k=1}^{p_1} \phi_k (X_{t-k} - \mu) + \sum_{k=1}^{q_1} \theta_k (X_{t-k} - \mu_{t-k}),\\
\sigma_t^2 &= \alpha_0 + \sum_{k=1}^{p_2} \alpha_k (X_{t-k} - \mu_{t-k})^2 + \sum_{k=1}^{q_2} \beta_k \sigma_{t-k}^2,
\end{aligned}$$

where $\alpha_0 > 0$, $\alpha_k \geq 0$, $k \in \{1, \ldots, p_2\}$, $\beta_k \geq 0$, $k \in \{1, \ldots, q_2\}$, $\sum_{k=1}^{p_2} \alpha_k + \sum_{k=1}^{q_2} \beta_k < 1$.

- ARMA models with GARCH errors are quite flexible models. It is easy to see that the conditional mean of $(X_t)_{t\in\mathbb{Z}}$ is $\mu_t = \mathbb{E}(X_t \,|\, \mathcal{F}_{t-1})$ and that the conditional variance of $(X_t)_{t\in\mathbb{Z}}$ is $\sigma_t^2 = \mathrm{var}(X_t \,|\, \mathcal{F}_{t-1})$.
- Other extensions not futher discussed here:
  - *GJR-GARCH*. These models introduce a parameter in the volatility equation in order for the volatility to react asymmetrically to recent returns (bad news leading to a fall in the equity value of a company tends to increase volatility, the so-called *leverage effect*).
  - *Threshold GARCH (TGARCH)*. More general models (than GJR-GARCH) in which the dynamics at time t depend on whether X_{t-1} (or Z_{t-1}; sometimes even a coefficient) was below/above a threshold.
  - Note that one could also use an asymmetric innovation distribution with mean 0 and variance 1, e.g. from the generalized hyperbolic family or skewed t distribution.

4.2.4 Fitting GARCH models to data

Building the likelihood

- The most widely used approach is maximum likelihood. We first consider $\mathrm{ARCH}(1)$ and $\mathrm{GARCH}(1,1)$ models, the general case easily follows.
- $\mathrm{ARCH}(1)$. Supose we have data $X_0, X_1, \dots, X_n$. The joint density can be written as

$$\begin{aligned} f_{X_0,\dots,X_n}(X_0,\dots,X_n) &= f_{X_0}(X_0)\prod_{t=1}^n f_{X_t\,|\,X_{t-1},\dots,X_0}(X_t\,|\,X_{t-1},\dots,X_0) \\ &= f_{X_0}(X_0)\prod_{t=1}^n f_{X_t\,|\,X_{t-1}}(X_t\,|\,X_{t-1}) \\ &= f_{X_0}(X_0)\prod_{t=1}^n \frac{1}{\sigma_t} f_Z\Big(\frac{X_t}{\sigma_t}\Big), \end{aligned}$$

where $\sigma_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2}$ and f_Z denotes the density of the innovations $(Z_t)_{t\in\mathbb{Z}}$ (mean 0, variance 1; typically $\mathrm{N}(0,1)$ or $t_\nu(0,\frac{\nu-2}{\nu})$). The problem is that f_{X_0} is not known in tractable form. One thus typically considers the conditional likelihood given X_0

$$\begin{aligned} L(\alpha_0, \alpha_1; X_0, \dots, X_n) &= f_{X_1,\dots,X_n \mid X_0}(X_1, \dots, X_n \mid X_0) \\ &= \frac{f_{X_0,\dots,X_n}(X_0, \dots, X_n)}{f_{X_0}(X_0)} = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\Big(\frac{X_t}{\sigma_t}\Big). \end{aligned}$$

Similarly for $\mathrm{ARCH}(p)$ models, one considers the likelihood conditional the first p values.

- $\mathrm{GARCH}(1,1)$. Here we construct the joint density of $X_1, \dots, X_n$ conditional on both X_0 and σ_0, so

$$\begin{aligned} L(\alpha_0, \alpha_1, \beta_1; X_0, \dots, X_n) &= f_{X_1,\dots,X_n \mid X_0,\sigma_0}(X_1, \dots, X_n \mid X_0, \sigma_0) \\ = \prod_{t=1}^{n} f_{X_t \mid X_{t-1},\dots,X_0,\sigma_0}(X_t \mid X_{t-1}, \dots, X_0, \sigma_0) &= \prod_{t=1}^{n} f_{X_t \mid \sigma_t}(X_t \mid \sigma_t) \\ = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\Big(\frac{X_t}{\sigma_t}\Big), \quad \text{where } \sigma_t &= \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2}. \end{aligned}$$

Note that σ_0^2 is not observed. One typically chooses the sample variance of $X_1, \dots, X_n$ (or 0) as starting values.

- Similarly for ARMA models with GARCH errors. In this case,

$$L(\boldsymbol{\theta}; X_0, \dots, X_n) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\Big(\frac{X_t - \mu_t}{\sigma_t}\Big)$$

for the ARMA specification for μ_t and the GARCH specification for σ_t; all parameters are collected in $\boldsymbol{\theta}$, including unknown parameters of the innovation distribution. The *log-likelihood* is thus given by

$$\ell(\boldsymbol{\theta}; X_0, \dots, X_n) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log\Big(\frac{1}{\sigma_t} f_Z\Big(\frac{X_t - \mu_t}{\sigma_t}\Big)\Big).$$

- Extensions to models with leverage or threshold effects are also possible.
- The log-likelihood ℓ is typically maximized numerically to obtain $\hat{\boldsymbol{\theta}}_n$.

Model checking

- After model fitting, check its residuals. We consider an ARMA model with GARCH errors of the form $X_t - \mu_t = \varepsilon_t = \sigma_t Z_t$; see Definition 4.20.
- We distinguish two kinds of residuals:
  1) *Unstandardized residuals*. These are the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ from the ARMA part of the model, calculated as in Section 4.1.4. Under the hypothesized model they should behave like a realization of a GARCH process.
  2) *Standardized residuals*. These are reconstructed realizations of the SWN which drives the GARCH process. They are calculated from the unstandardized residuals via
  $$\hat{Z}_t = \hat{\varepsilon}_t / \hat{\sigma}_t, \quad \hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{k=1}^{p_2} \hat{\alpha}_k \hat{\varepsilon}_{t-k}^2 + \sum_{k=1}^{q_2} \hat{\beta}_k \hat{\sigma}_{t-k}^2; \qquad (7)$$
  starting values for $\hat{\varepsilon}_t$ are taken as 0 and starting values for $\hat{\sigma}_t$ are taken as the sample variance (or 0); ignore the first few values then.

- The standardized residuals should behave like SWN. Check this via correlograms of $(\hat{Z}_t)$ and $(|\hat{Z}_t|)$ and by applying the Ljung–Box test of strict white noise. In case of no rejection (the dynamics have been satisfactorily captured), the validity of the innovation distribution can also be assessed (e.g. via Q-Q plots or goodness-of-fit tests).

  $\Rightarrow$ *Two-stage analysis* possible: First estimate the dynamics via QMLE (known as *pre-whitening* of the data), then model the innovation distribution using the standardized residuals.

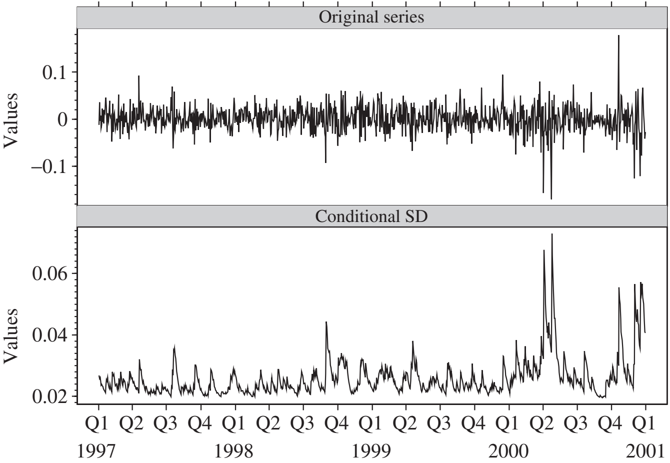
  Advantages:
  - More transparency in model building;
  - Separating of volatility modelling and modelling of shocks that drive the process;
  - Practical in higher dimensions.

  Drawbacks: ARMA fitting errors propagate through to the fitting of innovations (overall error hard to quantify).

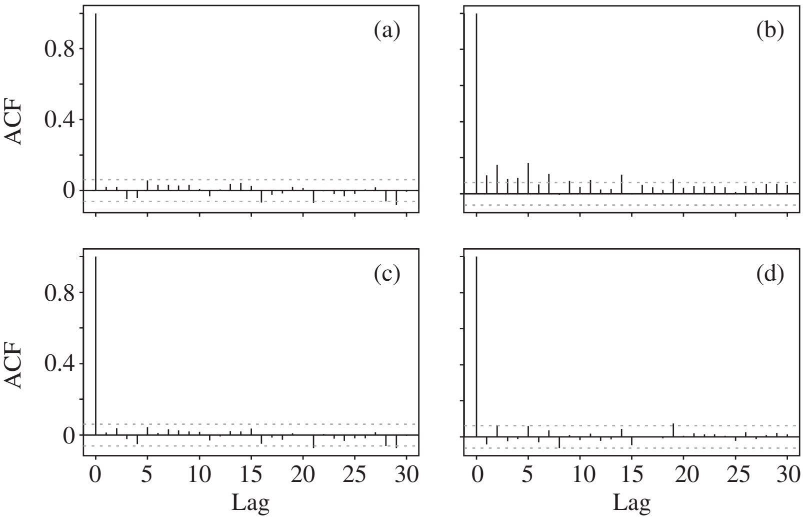
Example 4.21 (GARCH model for Microsoft log-returns)

- Consider Microsoft daily log-returns from 1997–2000 (1009 values). The raw returns show no evidence of serial correlation, the absolute values do (Ljung–Box test based on the first 10 estimated correlations fails at the 5% level).
- Various models with t innovations are fitted via MLE: $\mathrm{GARCH}(1,1)$, $\mathrm{AR}(1)$–$\mathrm{GARCH}(1,1)$, $\mathrm{MA}(1)$–$\mathrm{GARCH}(1,1)$, $\mathrm{ARMA}(1,1)$–$\mathrm{GARCH}(1,1)$. The basic $\mathrm{GARCH}(1,1)$ is favored according to Akaike's information criterion.
- A model GRJ model further improves the fit (both raw and absolute standardized residuals show no serieal correlation; Ljung–Box does not reject).

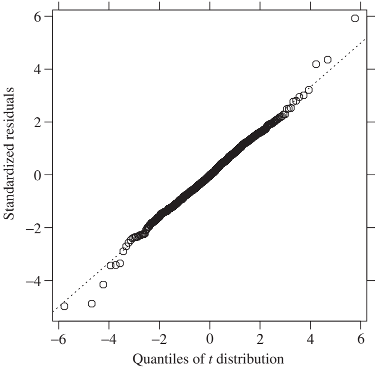
Microsoft log-returns 1997–2000: Data (top) and estimated volatitlity (bottom) from a GJR-$\mathrm{GARCH}(1,1)$.

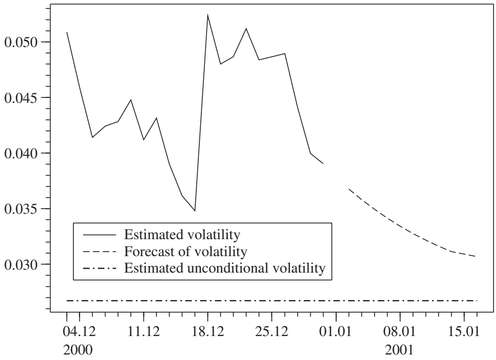


Correlograms of a) (X_t); b) $(|X_t|)$; c) $(\hat{Z}_t)$; and d) $(|\hat{Z}_t|)$



Q-Q plot of the standardized residuals (left); Estimated and predicted volatility (right) for the first 10 days of 2001 based on a $\mathrm{GARCH}(1,1)$ model.





4.2.5 Volatility forecasting and risk measure estimation

- Consider a weakly and strictly stationary time series $(X_t)_{t\in\mathbb{Z}}$ of the form

$$X_t = \mu_t + \sigma_t Z_t$$

  adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, where $\mu_t, \sigma_t \in \mathcal{F}_{t-1}$ and $\mathbb{E}Z_t = 0$, $\operatorname{var} Z_t = 1$, independent of $\mathcal{F}_{t-1}$ (e.g. $(X_t)_{t\in\mathbb{Z}}$ could be a GARCH model or ARMA model with GARCH errors).
- Assume we know $X_{t-n+1}, \dots, X_t$ and want to forecast σ_{t+h}, $h \ge 1$.
- Since $\mathbb{E}(\sigma_{t+h}^2 \,|\, \mathcal{F}_t) = \mathbb{E}((X_{t+h} - \mu_{t+h})^2 \,|\, \mathcal{F}_t)$ our forecasting problem is related to the problem of predicting $(X_{t+h} - \mu_{t+h})^2$.
- Two possible approaches: Conditional expectations and exponentially weighted moving averages.

Conditional expectation

The general procedure becomes clear from the following two examples

Example 4.22 (Prediction in the $\mathrm{GARCH}(1,1)$ model)

- A $\mathrm{GARCH}(1,1)$ model is of type $X_t = \mu_t + \sigma_t Z_t$ for $\mu_t = 0$. Since $\mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t) = 0$, $\hat{\mu}_{t+h} = P_t X_{t+h} = 0$ for all $h \in \mathbb{N}$.
- A natural prediction of X_{t+1}^2 based on $\mathcal{F}_t$ is its conditional mean
$$\mathbb{E}(X_{t+1}^2 \,|\, \mathcal{F}_t) = \sigma_{t+1}^2 = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2.$$
If $\mathbb{E}(X_t^4) < \infty$, this is the optimal squared error prediction.
- We thus obtain the one-step-ahead forecast
$$\hat{\sigma}_{t+1}^2 = \widehat{\mathbb{E}(X_{t+1}^2 \,|\, \mathcal{F}_t)} = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \hat{\sigma}_t^2.$$
- If $h > 1$, σ_{t+h}^2 and X_{t+h}^2 are rvs. Their predictions (coincide and) are
$$\mathbb{E}(\sigma_{t+h}^2 \,|\, \mathcal{F}_t) = \alpha_0 + \alpha_1 \mathbb{E}(X_{t+h-1}^2 \,|\, \mathcal{F}_t) + \beta_1 \mathbb{E}(\sigma_{t+h-1}^2 \,|\, \mathcal{F}_t)$$

$$= \alpha_0 + (\alpha_1 + \beta_1)\mathbb{E}(X_{t+h-1}^2 \,|\, \mathcal{F}_t)$$
$$= \alpha_0 + (\alpha_1 + \beta_1)\mathbb{E}(\sigma_{t+h-1}^2 \,|\, \mathcal{F}_t)$$

so that a general formula is

$$\mathbb{E}(\sigma_{t+h}^2 \,|\, \mathcal{F}_t) = \alpha_0 \sum_{k=0}^{h-1} (\alpha_1 + \beta_1)^k + (\alpha_1 + \beta_1)^{h-1}(\alpha_1 X_t^2 + \beta_1 \sigma_t^2).$$

Note that for $h \to \infty$, $\mathbb{E}(\sigma_{t+h}^2 \,|\, \mathcal{F}_t) \overset{\text{a.s.}}{\to} \frac{\alpha_0}{1-\alpha_1-\beta_1}$, so the prediction of squared volatility converges to the unconditional variance of the process.

Example 4.23 (Prediction in the $\mathrm{ARMA}(1,1)$–$\mathrm{GARCH}(1,1)$ model)

Let $X_t - \mu_t = \sigma_t Z_t =: \varepsilon_t$ as before. It follows from Examples 4.14 and 4.22 that

$$\mathbb{E}(X_{t+h} \,|\, \mathcal{F}_t) = \mu + \phi_1^h (X_t - \mu) + \phi_1^{h-1}\theta_1 \varepsilon_t,$$

$$\mathrm{var}(X_{t+h} \,|\, \mathcal{F}_t) = \alpha_0 \sum_{k=0}^{h-1} (\alpha_1 + \beta_1)^k + (\alpha_1 + \beta_1)^{h-1}(\alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2).$$

For ε_t, σ_t, substitute values obtained from (7).

Exponentially weighted moving averages

- A one-period ahead forecast $P_t X_{t+1}$ of X_{t+1} based on $\mathcal{F}_t$ is given by
$$P_t X_{t+1} = \alpha X_t + (1-\alpha) P_{t-1} X_t. \tag{8}$$
Applied to $(X_{t+1} - \mu_{t+1})^2$ leads to
$$P_t (X_{t+1} - \mu_{t+1})^2 = \alpha (X_t - \mu_t)^2 + (1-\alpha) P_{t-1} (X_t - \mu_t)^2. \tag{9}$$
- Since $\sigma_{t+1}^2 = \mathbb{E}((X_{t+1} - \mu_{t+1})^2 \,|\, \mathcal{F}_t)$, we can use (9) as exponential smoothing scheme for the unobserved squared volatility σ_{t+1}^2. This yields a recursive scheme for the one-step-ahead volatility forecast given by
$$\hat{\sigma}_{t+1}^2 = \alpha (X_t - \hat{\mu}_t)^2 + (1-\alpha) \hat{\sigma}_t^2,$$
which is then iterated.
- α is typically chosen small (e.g. RiskMetrics: $\alpha = 0.06$); $\hat{\mu}_t$ is often chosen as 0 (see Chapter 3). Alternatively, apply exponential smoothing to μ_t via $P_{t-1} X_t$ in (8).

Estimators of VaR_α and ES_α

- Suppose we have losses $X_{t-n+1}, \dots, X_t$ and we would like to estimate VaR_α, ES_α for $F_{X_{t+1}|\mathcal{F}_t}$. Writing F_Z for the df of the innovations (Z_t), the $\mathcal{F}_t$-measurability of μ_{t+1} and σ_{t+1} implies that
$$F_{X_{t+1}|\mathcal{F}_t}(x) = \mathbb{P}(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq x \,|\, \mathcal{F}_t) = F_Z\Big(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\Big).$$
- Let $\mathrm{VaR}_\alpha^t = F^{\leftarrow}_{X_{t+1}|\mathcal{F}_t}(\alpha)$ and let ES_α^t denote the corresponding time-dynamic expected shortfall. We then have
$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1} F_Z^{\leftarrow}(\alpha), \quad \mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\, \mathrm{ES}_\alpha(Z).$$
- If we can estimate μ_{t+1}, σ_{t+1} (parametrically/non-parametrically/semi-parametrically), we only have left to estimate $F_Z^{\leftarrow}(\alpha)$ and $\mathrm{ES}_\alpha(Z)$.
- For GARCH-type models it is easy to calculate $F_Z^{\leftarrow}(\alpha)$ and $\mathrm{ES}_\alpha(Z)$. And if we use exponential smoothing or QMLE to estimate μ_{t+1}, σ_{t+1}, we can use the residuals $\hat{Z}_s = (X_s - \hat{\mu}_s)/\hat{\sigma}_s$, $s \in \{t-n+1, \dots, n\}$ to estimate $F_Z^{\leftarrow}(\alpha)$ and $\mathrm{ES}_\alpha(Z)$.