4 Financial time series

4.1 Fundamentals of time series analysis

For more details on time series, consider Brockwell and Davis (1991) and Brockwell and Davis (2002).

Interlude: Conditional expectations

Definition 4.1 (Conditional expectation, conditional probability)
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\boldsymbol{X} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ (i.e., $\boldsymbol{X} : \Omega \to \mathbb{R}^d$, $\boldsymbol{X}$ is $\mathcal{F}$-measurable (i.e., $\boldsymbol{X}^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R}^d)$) and $\mathbb{E}|\boldsymbol{X}| < \infty$) and $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Then any rv $\boldsymbol{Y}$ such that

1) $\boldsymbol{Y} \in \mathcal{G}$ ($\boldsymbol{Y}$ is $\mathcal{G}$-measurable);
2) $\mathbb{E}|\boldsymbol{Y}| < \infty$; and
3) $\mathbb{E}[\boldsymbol{Y} \mathbb{1}_G] = \int_G \boldsymbol{Y} \, d\mathbb{P} = \int_G \boldsymbol{X} \, d\mathbb{P} = \mathbb{E}[\boldsymbol{X} \mathbb{1}_G]$ for all $G \in \mathcal{G}$

is called *conditional expectation of $\boldsymbol{X}$ given $\mathcal{G}$* and denoted by $\mathbb{E}[\boldsymbol{X} \,|\, \mathcal{G}]$.
$\mathbb{P}(A \,|\, \mathcal{G}) = \mathbb{E}[\mathbb{1}_A \,|\, \mathcal{G}]$ is called *conditional probability of A given $\mathcal{G}$*.

Example 4.2 (Do ordinary conditional expectations fit in this setup?)

Let $X \in \{x_1, \dots, x_n\}$, $Z \in \{z_1, \dots, z_m\}$ be rvs (w.l.o.g. all values are attained with non-zero probability).

- Let $\mathbb{P}(X = x_i \,|\, Z = z_j) = \frac{\mathbb{P}(X=x_i, Z=z_j)}{\mathbb{P}(Z=z_j)}$ and $\mathbb{E}[X \,|\, Z = z_j] = \sum_{i=1}^n x_i \mathbb{P}(X = x_i \,|\, Z = z_j)$. Then the *ordinary conditional expectation of X given Z* is defined by
$$\mathbb{E}[X \,|\, Z] = \sum_{j=1}^m \mathbb{E}[X \,|\, Z = z_j] \mathbb{1}_{\{Z=z_j\}}.$$
- To see that this definition coincides with our more general defintion, consider
$$\begin{aligned}\mathcal{G} = \sigma(Z) &= \{Z^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\} = \{\omega \in \Omega : Z(\omega) \in B \in \mathcal{B}(\mathbb{R})\} \\ &= \bigcup_{A \subseteq \{z_1,\dots,z_m\}} \{\omega \in \Omega : Z(\omega) \in A\} = \bigcup_{J \subseteq \{1,\dots,m\}} \{\omega \in \Omega : Z(\omega) \in (z_j)_{j \in J}\} \\ &= \bigcup_{J \subseteq \{1,\dots,m\}} \bigcup_{j \in J} G_j, \quad G_j = \{\omega \in \Omega : Z(\omega) = z_j\},\end{aligned}$$

and let $Y = \sum_{j=1}^{m} \mathbb{E}[X \,|\, Z = z_j] \mathbb{1}_{\{Z = z_j\}}$. Then Y is constant on G_j ($\Rightarrow Y \in \mathcal{G}$) and $\mathbb{E}|Y| < \infty$. Furthermore,

$$\begin{aligned}
\int_{G_j} Y \, d\mathbb{P} &= \int_{G_j} \sum_{k=1}^{m} \mathbb{E}[X \,|\, Z = z_k] \mathbb{1}_{\{Z = z_k\}} \, d\mathbb{P} = \int_{G_j} \mathbb{E}[X \,|\, Z = z_j] \, d\mathbb{P} \\
&= \int_{G_j} \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i \,|\, Z = z_j) \, d\mathbb{P} \\
&= \Bigg(\sum_{i=1}^{n} x_i \mathbb{P}(X = x_i \,|\, Z = z_j) \Bigg) \mathbb{P}(Z = z_j) \\
&= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i, \ Z = z_j) = \int_{G_j} X \, d\mathbb{P}.
\end{aligned}$$

Since every $G \in \mathcal{G}$ is a disjoint union of G_j's, we have that $\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}$ for all $G \in \mathcal{G}$. So $\mathbb{E}[X \,|\, Z]$ corresponds to the ordinary conditional expectation here.

The following property of conditional expectations is used frequently and known as *tower property*.

Lemma 4.3 (Tower property; the smallest σ-algebra remains)
If $\mathcal{G} \subseteq \mathcal{F}$, then $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{F}] = \mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[\mathbb{E}[X \,|\, \mathcal{F}] \,|\, \mathcal{G}]$.

Idea of proof. Let $G \in \mathcal{G} \subseteq \mathcal{F}$. Applying Definition 4.1 Part 3) to $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{F}]$ and then to $\mathbb{E}[X \,|\, \mathcal{G}]$ implies that $\mathbb{E}[\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{F}] \mathbb{1}_G] = \mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \mathbb{1}_G] = \mathbb{E}[X \mathbb{1}_G]$. □

The next property of conditional expectations shows their importance.

Lemma 4.4 (Best $\mathcal{G}$-measurable L^2 approx./predictor of X)
Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Then

$$\min_{Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}[(X - Y)^2] = \mathbb{E}[(X - \mathbb{E}[X \,|\, \mathcal{G}])^2]$$

Proof.

$$
\begin{aligned}
\mathbb{E}[(X-Y)^2] &= \underbrace{\mathbb{E}[(X-\mathbb{E}[X\,|\,\mathcal{G}])^2]}_{\text{independent of } Y} + \underbrace{\mathbb{E}[(\mathbb{E}[X\,|\,\mathcal{G}]-Y)^2]}_{=0\;\Leftrightarrow\; Y=\mathbb{E}[X|\mathcal{G}]} \\
&\quad + 2\underbrace{\mathbb{E}[(X-\mathbb{E}[X\,|\,\mathcal{G}])(\mathbb{E}[X\,|\,\mathcal{G}]-Y)]}_{\underset{\text{property}}{\overset{\text{tower}}{=}} \mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X\,|\,\mathcal{G}])(\mathbb{E}[X\,|\,\mathcal{G}]-Y)\,|\,\mathcal{G}]] \underset{\text{what is known}}{\overset{\text{take out}}{=}} 0}
\end{aligned}
$$

□

For more details on conditional expectations, see Williams (1991).

4.1.1 Basic definitions

Definition 4.5 (Mean function, autocovariance function)
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. 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

$$\mu(t) = \mathbb{E}[X_t], \quad t \in \mathbb{Z},$$
$$\gamma(t,s) = \mathrm{Cov}[X_t, X_s] = \mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)], \quad t, s, \in \mathbb{Z}.$$

Definition 4.6 ((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 $(X_{t_1}, \dots, X_{t_n}) \stackrel{\mathrm{d}}{=} (X_{t_1+h}, \dots, X_{t_n+h})$ for all $t_1, \dots, t_n, h \in \mathbb{Z}$, $n \in \mathbb{N}$.

Remark 4.7

1) Both types of stationarity formalize that $(X_t)_{t\in\mathbb{Z}}$ behaves similarly in any epoch.

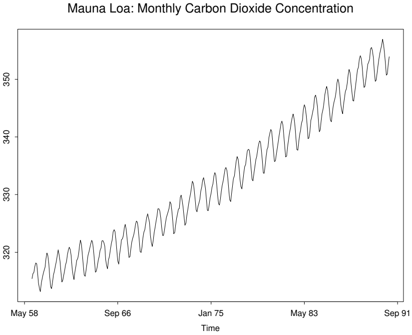
2) Hölder's inequality: $\|XY\|_1 \le \|X\|_p\|Y\|_q$ for $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$, $p,q\in[1,\infty] : \frac{1}{p}+\frac{1}{q}=1$. Under stationarity, $|\gamma(t,s)| \underset{\text{ineq.}}{\overset{\text{triangle}}{\le}} \mathbb{E}[|X_t - \mathbb{E}X_t|\cdot|X_s-\mathbb{E}X_s|] \underset{\text{CSI}}{\le} (\mathbb{E}[(X_t-\mathbb{E}X_t)^2]\mathbb{E}[(X_s-\mathbb{E}X_s)^2])^{1/2} < \infty$, so $\gamma(t,s)$ exists.

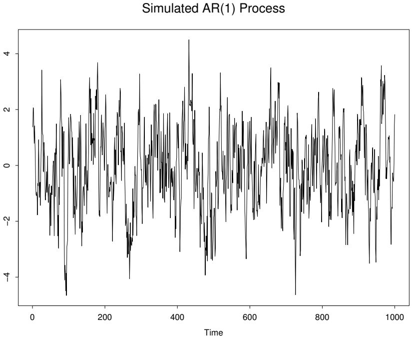
3) Strict stationarity $\underset{\text{ii)}}{\overset{\text{i)}}{\substack{\not\Rightarrow\\ \not\Leftarrow}}}$ stationarity
   - i) $\mathbb{E}[X_t^2]$ doesn't have to exist (e.g., (G)ARCH processes). If it does, strict stationarity implies stationarity.
   - ii) $\mathbb{E}[|X_t|^p]$, $p>2$, could be changing (e.g., Pearson type VII).

4) $(X_t)_{t\in\mathbb{Z}} \Rightarrow \gamma(t-s,0)=\gamma(t,s)=\gamma(s,t)=\gamma(s-t,0)$, so $\gamma(t,s)$ only depends on the emphlag $|t-s|$. We can thus consider $\gamma(h):=\gamma(h,0)$,

$h \in \mathbb{Z}$. For $s = 0$, we obtain $\gamma(t) = \gamma(t-s,0) = \gamma(s-t,0) = \gamma(-t)$, $t \in \mathbb{Z}$, so it suffices to know $\gamma(h)$ for $h \in \mathbb{N}_0$ or $h \in -\mathbb{N}_0$. In particular, if we have shown that $\gamma(h) = f(h)$ for some f and all $h \in \mathbb{N}_0$, we also know $\gamma(-h) = \mathbb{E}[(X_t-\mu)(X_{t-h}-\mu)] = \mathbb{E}[(X_{t-h}-\mu)(X_t-\mu)] = \gamma(h)$, so that $\gamma(h) = f(|h|)$, $h \in \mathbb{Z}$.

Stationary?





(Partial) autocorrelation in stationary time series

Definition 4.8 (ACF)

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

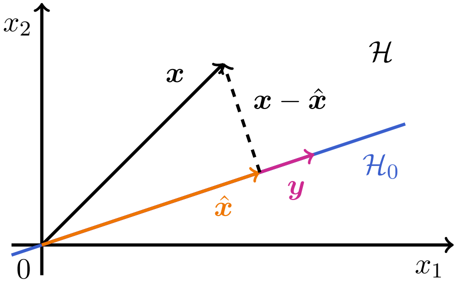
- The study of autocorrelation is known as *analysis in the time domain*. Another important quantity is the partial autocorrelation function. For introducing it, we need some tools.
- *Hilbert's Project Theorem* (see Brockwell and Davis (1991, p. 51)): If $\mathcal{H}_0$ is a closed subspace of the Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$, then:
  - i) There exists a unique $\hat{x} \in \mathcal{H}_0 : \|x - \hat{x}\| = \inf_{y\in\mathcal{H}_0}\|x - y\|$;
  - ii) $\hat{x} \in \mathcal{H}_0$, $\|x - \hat{x}\| = \inf_{y\in\mathcal{H}_0}\|x - y\|$ if and only if $\hat{x} \in \mathcal{H}_0$, $x - \hat{x} \in \mathcal{H}_0^{\perp} = \{x \in \mathcal{H} : \langle x, y\rangle = 0 \text{ for all } y \in \mathcal{H}_0\}$.

Note:

- $\hat{x}$ is the *(orthogonal) projection of* x *onto* $\mathcal{H}_0$, denoted by $P_{\mathcal{H}_0}x$.
- $\hat{x} = P_{\mathcal{H}_0}x$ is the unique element: $\langle x - \hat{x}, y \rangle = 0 \; \forall \, y \in \mathcal{H}_0$ (*prediction equations*; $P_{\mathcal{H}_0}x$ is the best approximation/prediction of x in $\mathcal{H}_0$).

Example 4.9

$\boldsymbol{x} \in \mathcal{H} = \mathbb{R}^2$, $\mathcal{H}_0 = \operatorname{span}\{\boldsymbol{y}\}$



- Yule–Walker equations. Let $X_1, \dots, X_{n-1}, X_n$ be elements of a stationary time series $(X_t)_{t\in\mathbb{Z}}$ with $\mu(t) = 0$, $t \in \mathbb{Z}$. Suppose we would like to find $\hat{X}_n = \sum_{k=1}^{n-1} \phi_{n-1,k} X_{n-k}$ such that

$$\mathbb{E}[(X_n - \hat{X}_n)^2] \to \min_{(\phi_{n-1,k})_{k=1}^{n-1}} .$$

$\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with $\langle X, Y\rangle = \mathbb{E}[XY]$ and $\mathcal{H}_{n-1} = \mathrm{span}\{X_1, \dots, X_{n-1}\} = \{\sum_{k=1}^{n-1} \alpha_k X_{n-k} : \alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}\}$ is a subspace. Therefore, $\hat{X}_n = P_{\mathcal{H}_{n-1}} X_n$ satisfies the prediction equations

$$\begin{aligned} & \langle X_n - \hat{X}_n, Y\rangle = 0, \quad \forall\, Y \in \mathcal{H}_{n-1} \\ \Leftrightarrow\ & \underbrace{\langle X_n - \hat{X}_n, \sum_{k=1}^{n-1} \alpha_k X_{n-k}\rangle}_{=\sum_{k=1}^{n-1} \alpha_k \langle X_n - \hat{X}_n, X_{n-k}\rangle} = 0, \quad \forall\, \alpha_1, \dots, \alpha_{n-1} \in \mathbb{R} \\ \Leftrightarrow\ & \underbrace{\langle X_n - \hat{X}_n, X_l\rangle}_{\substack{=\mathbb{E}[(X_n - \sum_{k=1}^{n-1} \phi_{n-1,k} X_{n-k}) X_l] \\ = \mathbb{E}[X_n X_l] - \sum_{k=1}^{n-1} \phi_{n-1,k} \mathbb{E}[X_{n-k} X_l]}} = 0, \quad \forall\, l \in \{1, \dots, n-1\} \end{aligned}$$

$$\Leftrightarrow \ \gamma(n-l) = \sum_{k=1}^{n-1} \gamma(n-k-l)\phi_{n-1,k}$$

$$\underset{\text{station.}}{\Leftrightarrow} \gamma(h) = \sum_{k=1}^{n-1} \gamma(h-k)\phi_{n-1,k}, \quad \forall\, h \in \{1,\dots,n-1\}$$

$$\Leftrightarrow \ \Gamma_{n-1}\boldsymbol{\phi}_{n-1} = \boldsymbol{\gamma}_{n-1}, \quad (\textit{Yule–Walker equations})$$

where

$$\boldsymbol{\phi}_{n-1} = (\phi_{n-1,1},\dots,\phi_{n-1,n-1}),$$
$$\boldsymbol{\gamma}_{n-1} = (\gamma(1),\dots,\gamma(n-1)),$$
$$\Gamma_{n-1} = (\gamma(|i-j|))_{i,j=1}^{n-1}.$$

Hilbert's Projection Theorem ii) $\Rightarrow$ there exists at least one solution $\boldsymbol{\phi}_{n-1}$ and all of them lead to the same $\hat{X}_n$ (unique by i)). If Γ_{n-1} is regular (invertible), $\boldsymbol{\phi}_{n-1}$ is unique. This holds, e.g., if $\gamma(0) > 0$, $\gamma(h) \to 0$ $(h \to \infty)$; see Brockwell and Davis (1991, p. 167).

- ϕ_n can be computed with the *Durbin–Levinson algorithm*: Let $(X_t)_{t\in\mathbb{Z}}$ be stationary with $\mu(t) = 0$, $t \in \mathbb{Z}$, $\gamma(0) > 0$, $\gamma(h) \to 0$ ($h \to \infty$). Then, for all $n \in \mathbb{N}$,

$$\phi_{n,n} \underset{(*)}{=} \frac{\gamma(n) - \sum_{k=1}^{n-1} \gamma(n-k)\phi_{n-1,k}}{\gamma(0) - \sum_{k=1}^{n-1} \gamma(n-k)\phi_{n-1,n-k}} = \frac{\rho(n) - \sum_{k=1}^{n-1} \rho(n-k)\phi_{n-1,k}}{1 - \sum_{k=1}^{n-1} \rho(n-k)\phi_{n-1,n-k}},$$

$$\begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} \underset{(**)}{=} \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{n,n} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}.$$

Proof. The Yule–Walker equations hold if and only if

$$
\begin{pmatrix} \gamma(0) & \cdots & \gamma(n-2) & \gamma(n-1) \\ \vdots & \ddots & \vdots & \vdots \\ \cdots & \cdots & \gamma(0) & \vdots \\ \cdots & \cdots & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \\ \phi_{n,n} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n-1) \\ \gamma(n) \end{pmatrix}
$$

$$
\Leftrightarrow \quad \Gamma_{n-1} \begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} + \phi_{n,n} \underbrace{\begin{pmatrix} \gamma(n-1) \\ \vdots \\ \gamma(1) \end{pmatrix}}_{\underset{\text{YW}}{=} \Gamma_{n-1} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}} = \underbrace{\begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n-1) \end{pmatrix}}_{\underset{\text{YW}}{=} \Gamma_{n-1} \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix}},
$$

and $\sum_{k=1}^{n-1} \gamma(n-k)\phi_{n,k} + \phi_{n,n}\gamma(0) \underset{(***)}{=} \gamma(n)$. Multiplying with Γ_{n-1}^{-1} leads $(**)$. For $(*)$, use the kth row $\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n}\phi_{n-1,n-k}$ in $(***)$ and solve w.r.t. $\phi_{n,n}$. □

Definition 4.10 (PACF)

The *partial autocorrelation function (PACF)* of a stationary time series $(X_t)_{t\in\mathbb{Z}}$ with $\mu(t)=0$, $t\in\mathbb{Z}$, $\gamma(0)>0$, $\gamma(h)\to 0$ $(h\to\infty)$ is

$$\begin{aligned}\phi(h) &= \mathrm{Cor}[X_0 - P_{\mathcal{H}_{h-1}}X_0, X_h - P_{\mathcal{H}_{h-1}}X_h]\\ &= \frac{\mathbb{E}[X_0(X_h - P_{\mathcal{H}_{h-1}}X_h)]}{\mathbb{E}[(X_h - P_{\mathcal{H}_{h-1}}X_h)(X_h - P_{\mathcal{H}_{h-1}}X_h)]}\\ &= \frac{\mathbb{E}[X_0X_h] - \sum_{k=1}^{h-1}\phi_{h-1,k}\mathbb{E}[X_0X_{h-k}]}{\mathbb{E}[X_h(X_h - P_{\mathcal{H}_{h-1}}X_h)]}\\ &\underset{\text{station.}}{=} \frac{\gamma(h) - \sum_{k=1}^{h-1}\gamma(h-k)\phi_{h-1,k}}{\gamma(0) - \sum_{k=1}^{h-1}\gamma(k)\phi_{h-1,k}} \underset{\text{DL algo.}}{=} \phi_{h,h}, \quad h\in\mathbb{Z}.\end{aligned}$$

The PACF is...

- the Cor between X_0 and X_h with the linear dependence of $X_1,\dots,X_{h-1}$ removed;

- $\phi_{h,h}$ (obtained from the Durbin-Levinson algorithm); and
- the coefficient of X_1 in the best linear L^2-approximation of X_h by $X_1, \dots, X_{h-1}$.

Note that $\phi(1) = \phi_{1,1} = \gamma(1)/\gamma(0) = \rho(1)$.

White noise processes

Definition 4.11 ((Strict) white noise)

1) $(X_t)_{t\in\mathbb{Z}}$ is a *white noise* process if it is stationary with $\rho(h) = \mathbb{1}_{\{h=0\}}$ (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 i.i.d. 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).

Definition 4.12 (MGDS)

Let $(X_t)_{t\in\mathbb{Z}}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. A sequence $(\mathcal{F}_{t\in\mathbb{Z}})$ (*accrual of information* over time) of σ-algebras is called *filtration* if $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$ for all $t \in \mathbb{Z}$. If $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$, 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 $X_t \in \mathcal{F}_t$ for all $t \in \mathbb{Z}$ (X_t is $\mathcal{F}_t$-measurable). $(X_t)_{t\in\mathbb{Z}}$ 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 (discrete-time) *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 $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$.

4.1.2 ARMA processes

Definition 4.13 ($\mathrm{ARMA}(p,q)$)

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 $t \in \mathbb{Z}$,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}. \quad (8)$$

$(X_t)_{t\in\mathbb{Z}}$ 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 4.14

- R, Brockwell and Davis (1991): ✓
  S-Plus: θ_k's have opposite signs.
- 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 (8) can be written as

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

  where

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

Causal processes

For practical purposes, it suffices to consider *causal* ARMA processes, that is, $(X_t)_{t\in\mathbb{Z}}$ satisfying (8) 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*).

This condition implies that

$$\mathbb{E}\left[\left|\sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k}\right|\right] = \mathbb{E}\left[\lim_{n\to\infty}\left|\sum_{k=0}^{n}\psi_k\varepsilon_{t-k}\right|\right] \underset{\text{ineq.}}{\overset{\text{triangle}}{\leq}} \mathbb{E}\left[\lim_{n\to\infty}\sum_{k=0}^{n}|\psi_k\varepsilon_{t-k}|\right]$$

$$\underset{\text{convergence}}{\overset{\text{monotone}}{=}} \lim_{n\to\infty}\mathbb{E}\left[\sum_{k=0}^{n}|\psi_k\varepsilon_{t-k}|\right] = \lim_{n\to\infty}\sum_{k=0}^{n}|\psi_k|\mathbb{E}|\varepsilon_{t-k}|$$

$$\underset{\text{noise}}{\overset{\text{white}}{=}} \mathbb{E}|\varepsilon_t|\cdot\lim_{n\to\infty}\sum_{k=0}^{n}|\psi_k| < \infty,$$

so that $\sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k}$ converges (absolutely) a.s. It also converges in L^2 to the same limit (see Brockwell and Davis (2002, p. 83)).

Proposition 4.15 (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

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

Proof. Since $\varepsilon_t \perp \varepsilon_{t+h}$ for all $h \neq 0$ and by the absolute summability condition, $\mathbb{E}[X_t^2] = \sigma^2 \sum_{k=0}^{\infty} \psi_k^2 < \infty$, $t \in \mathbb{Z}$. Furthermore, $\mu(t) = 0$, $t \in \mathbb{Z}$. Finally, $\gamma(h) = \mathbb{E}[X_t X_{t+h}] = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \mathbb{E}[\varepsilon_{t-k}\varepsilon_{t+h-l}] = \sigma^2 \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \mathbb{1}_{\{l=k+h\}} = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+h}$, $h \in \mathbb{N}_0$. Remark 4.7 4) implies that $\gamma(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+|h|}$, $h \in \mathbb{Z}$, from which the claim follows. □

Theorem 4.16 (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 $\sum_{k=0}^{\infty} \psi_k z^k = \theta(z)/\phi(z)$, $|z| \leq 1$.

Proof. Idea; see Brockwell and Davis (1991, p. 85) for more details.

"$\Leftarrow$" $\phi(z) \neq 0$, $|z| \leq 1 \Rightarrow 1/\phi(z)$ holomorphic on $|z| < 1 + \varepsilon$ for some $\varepsilon > 0 \Rightarrow 1/\phi(z) = \sum_{k=0}^{\infty} a_k z^k$, $a_k(1+\varepsilon/2)^k \to 0$ $(k \to \infty)$ $\Rightarrow \exists\, c > 0 : |a_k| < c(1+\varepsilon/2)^{-k}$, $k \in \mathbb{N}_0 \Rightarrow \sum_{k=0}^{\infty} |a_k| < \infty$. Proposition 4.15 $\Rightarrow \varepsilon_t/\phi(B)$ is stationary. $\phi(B)X_t = \theta(B)\varepsilon_t \Rightarrow X_t = \frac{1}{\phi(B)}\phi(B)X_t = \theta(B)\varepsilon_t/\phi(B)$ is stationary (and causal).

"$\Rightarrow$" $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} = \psi(B)\varepsilon_t$, $\sum_{k=0}^{\infty} |\psi_k| < \infty \Rightarrow \theta(B)\varepsilon_t = \phi(B)X_t = \eta(B)\varepsilon_t$ for $\eta(B) = \phi(B)\psi(B)$. Let $\eta(z) = \phi(z)\psi(z) = \sum_{k=0}^{\infty} \eta_k z^k$, $|z| \le 1$. With $\theta_0 = 1$, it follows that $\sum_{k=0}^{q} \theta_k \varepsilon_{t-k} = \sum_{k=0}^{\infty} \eta_k \varepsilon_{t-k}$. Applying $\mathbb{E}[\cdot\, \varepsilon_{t-j}]$ $(\langle \cdot, \varepsilon_{t-j}\rangle)$ and using that $(\varepsilon_t) \sim \mathrm{WN}(0,\sigma^2)$, we obtain $\eta_k = \theta_k$, $k \in \{0,\dots,q\}$, and $\eta_k = 0$, $k > q$. This implies that $\theta(z) = \eta(z) = \phi(z)\psi(z)$ for all $|z| \le 1$. Assume $\phi(z_0) = 0$ for some $|z_0| \le 1$. Then $0 \ne \theta(z_0) = 0 \cdot \psi(z_0)$. Since $|\psi(z)| \le \sum_{k=0}^{\infty} |\psi_k| < \infty$ for all $|z| \le 1$, we obtain a contradiction. Thus $\phi(z) \ne 0$ for all $|z| \le 1$. □

Note that if $\theta(z) \ne 0$, $|z| \le 1$ (known as *invertibility condition*), we can recover ε_t from $(X_s)_{s \le t}$ 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 (G)ARCH process puts a certain structure on the conditional variance $\sigma_t^2 = \mathrm{Var}[X_t \,|\, \mathcal{F}_{t-1}]$.

Example 4.17

1) MA(q) = 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 ✓).

- ACF: Proposition 4.15 $\Rightarrow \rho(h) = \frac{\sum_{k=0}^{q-|h|} \theta_k \theta_{k+|h|}}{\sum_{k=0}^{q} \theta_k^2}$, $|h| \in \{1, \dots, q\}$, and $\rho(h) = 0$ for all $|h| > q \Rightarrow$ ACF cuts off after lag q.
- PACF (for MA(1)): Let $\theta = \theta_1$. Yule–Walker equations $\Leftrightarrow P_h \boldsymbol{\phi}_h = \boldsymbol{\rho}_h$. One can show by induction (or the Durbin–Levinson algorithm) that

$$\phi_{h,h} = -\frac{(-\theta)^h (1-\theta^2)}{1-\theta^{2(h+1)}}, \quad h \in \mathbb{N},$$

$$\phi_{h,h-k} = (-\theta)^{-k} \Big(\frac{1-\theta^{2k}}{1-\theta^2}\Big) \phi_{h,h}, \quad k \in \{1, \dots, h-1\}.$$

  In particular, $\phi(h) = \phi_{h,h} \searrow 0$ exponentially. One can show that for an MA(q), $\phi(h)$ does not cut off but $|\phi(h)|$ is bounded by an exponentially decreasing function in h.

2) AR(p) = 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 4.15.

  For $\mathrm{AR}(1)$, it is an exercise to show that $\rho(h) \searrow 0$ exponentially. For $\mathrm{AR}(p)$, one can show from a general form of ψ_k (see Brockwell and Davis (1991, p. 92)) that $\rho(h) \searrow 0$ exponentially, possibly with damped $\sin$ waves.

- PACF: For $h > p$, let $Y \in \mathcal{H}_{h-1} = \mathrm{span}\{X_1, \dots, X_{h-1}\}$. Since $(X_t)_{t\in\mathbb{Z}}$ is causal, $Y \in \mathrm{span}\{\varepsilon_s : s \le h-1\}$. Thus,

$$\left\langle X_h - \sum_{k=1}^{p} \phi_k X_{h-k}, Y \right\rangle = \langle \varepsilon_t, Y \rangle = 0.$$

  Prediction equations $\Rightarrow \sum_{k=1}^{p} \phi_k X_{h-k}$ is the best linear approximation in the L^2-sense to X_h from $X_1, \dots, X_{h-1}$, so $\sum_{k=1}^{p} \phi_k X_{h-k} =$

$P_{\mathcal{H}_{h-1}} X_h$. Hence,

$$\phi(h) = \mathrm{Cor}[\underbrace{X_0 - P_{\mathcal{H}_{h-1}} X_0}_{\in \operatorname{span}\{X_0,\dots,X_{h-1}\}}, \underbrace{X_h - P_{\mathcal{H}_{h-1}} X_h}_{=\varepsilon_h}] \underset{\text{causality}}{=} 0,$$

i.e., the PACF of an $\mathrm{AR}(p)$ cuts off after lag p. For $1 \le h \le p$, one can use the Durbin–Levinson algorithm to compute $\phi(h)$.

3) $\mathrm{ARMA}(1,1)$: $X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, $|\phi_1| < 1$ (Theorem 4.16 $\Rightarrow$ 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 4.15 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 $h \in \mathbb{Z}\backslash\{0\}$. The PACF can be computed from the Durbin–Levinson algorithm.

SARIMA models

$(X_t)_{t\in\mathbb{Z}}$ is a SARIMA$(p,d,q)\times(\tilde{p},\tilde{d},\tilde{q})_s$ (Seasonal; Integrated (i.e., may be made stationary by differencing)) process if

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

We see that this is also an $\mathrm{ARMA}(d+p+s(\tilde{d}+\tilde{p}),q+s\tilde{q})$ process. (Seasonal) "differences" are taken to get data from a stationary model.

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 4.18

Let $X_t - \mu = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$, $\sum_{k=0}^{\infty} |\psi_k| < \infty$ and $(Z_t) \sim \mathrm{SWN}(0, \sigma^2)$. If either $\mathbb{E}[X_t^4] < \infty$ or $\sum_{k=0}^{\infty} k\psi_k^2 < \infty$, then

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

where $W_{ij} = \sum_{k=1}^{\infty}(\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))(\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k))$, $i, j \in \{1, \dots, h\}$.

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 (i.i.d.) 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 i.i.d. 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 detected via smoothing with local averages:

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

- A seasonal component s_t (from X_1 to X_S; e.g., Jan–Dec) can be detected similarly, simply consider $(\tilde{X}_s)_{s=1}^{S}$ with

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

Removing μ_t, s_t can be done non-parametrically (see R's `stl()`) or via regression (e.g., $X_t = a_0 + a_1 t + a_2 t^2 + a_3 t \sin(\frac{2\pi}{12} t + 4) + \varepsilon_t$ for the "airline" dataset) or by taking differences (SARIMA).

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 (G)ARCH). 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.

iii) In a multivariate setting, determine a multivariate model for the residuals ($\Rightarrow$ dependence between time series); see, e.g., `demo(copula_garch)` in the R package `copula`.

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).

Exponential smoothing

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

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

$\alpha \in (0,1) \uparrow \Rightarrow$ more weight is put on the last observation.

Conditional expectation

Let the ARMA $(X_t)_{t\in\mathbb{Z}}$ be invertible and $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a MGDS w.r.t. $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. By Lemma 4.4, $P_t X_{t+h} = \mathbb{E}[X_{t+h} \,|\, \mathcal{F}_t]$ ($\mathbb{E}[X_{t+h} \,|\, \mathcal{F}_t]$ minimizes $\mathbb{E}[(X_{t+h} - \cdot)^2]$) $\Rightarrow$ For $h \in \mathbb{N}$, 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 4.19 (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

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

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 4.20 ($\mathrm{ARCH}(p)$)
Let $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0,1)$ (often: $Z_t \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$ or $t_\nu(0,(\nu-1)/\nu)$).
$(X_t)_{t\in\mathbb{Z}}$ 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 $\alpha_0 > 0$, $\alpha_k \geq 0$, $k \in \{1,\dots,p\}$.

Remark 4.21

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 4.22 ($\mathrm{ARCH}(1)$)

- One can show (via stoch. recurrence relations) 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, $\mathrm{Var}[X_t] = \frac{\alpha_0}{1-\alpha_1}$.

Proof.

"$\Rightarrow$" $X_t^2 = \sigma_t^2 Z_t^2 = (\alpha_0+\alpha_1 X_{t-1}^2)Z_t^2 \Rightarrow \mathbb{E}[X_t^2] = \alpha_0+\alpha_1 \underbrace{\mathbb{E}[X_{t-1}^2 Z_t^2]}_{=\mathbb{E}[X_{t-1}^2]=\sigma_X^2}$

$\Rightarrow \sigma_X^2 = \alpha_0 + \alpha_1\sigma_X^2 \Rightarrow \sigma_X^2 = \frac{\alpha_0}{1-\alpha_1}$, $\alpha_1 < 1$.

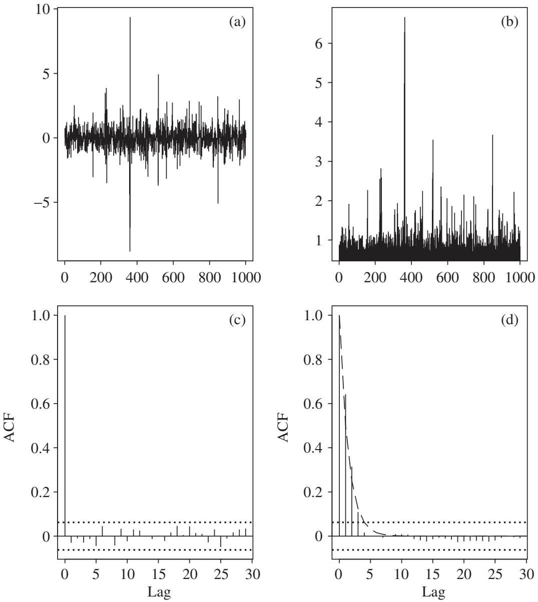
"$\Leftarrow$" $\mathbb{E}[\log(\alpha_1 Z_t^2)] \underset{\text{Jensen}}{\leq} \log \mathbb{E}[\alpha_1 Z_t^2] = \log(\alpha_1 \mathbb{E}[Z_t^2]) = \log \alpha_1 < 0.$

$\Rightarrow \mathbb{E}[X_t^2] \overset{\text{see above}}{\underset{\text{i.i.d.}}{=}} \alpha_0 \sum_{k=0}^{\infty} \alpha_1^k \prod_{j=1}^{k} \mathbb{E}[Z_{t-j}^2] = \frac{\alpha_0}{1-\alpha_1}$

$\Rightarrow (X_t)_{t\in\mathbb{Z}}$ is a MGDS with finite, constant variance, hence a white noise process (see above). $\square$

- One can show that for $\beta \geq 1$, the strictly stationary $\mathrm{ARCH}(1)$ process $(X_t)_{t\in\mathbb{Z}}$ satisfies $\mathbb{E}[X_t^{2\beta}] < \infty$ if and only if $\mathbb{E}[Z_t^{2\beta}] < \infty$ and $\alpha_1 < (\mathbb{E}[Z_t^{2\beta}])^{-1/\beta}$. From this result one can show that $\mathrm{kurt}(X_t) = \frac{\mathbb{E}[X_t^4]}{\mathbb{E}[X_t^2]^2} = \frac{\mathrm{kurt}(Z_t)(1-\alpha_1^2)}{(1-\alpha_1^2\,\mathrm{kurt}(Z_t))}$. If $\mathrm{kurt}(Z_t) > 1$, $\mathrm{kurt}(X_t) > \mathrm{kurt}(Z_t)$. For Gaussian or t innovations, $\mathrm{kurt}(X_t) > 3$ (leptokurtic).
- Parallels with the AR(1) process: Let $\alpha_1 < 1$ and $\varepsilon_t = \sigma_t^2(Z_t^2 - 1)$. Then $X_t^2 = \sigma_t^2 Z_t^2 = \sigma_t^2 + \varepsilon_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \varepsilon_t$, where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a MGDS since
  - $\mathbb{E}|\sigma_t^2(Z_t^2 - 1)| \leq \mathbb{E}[\sigma_t^2]\mathbb{E}[Z_t^2] = \mathbb{E}[\sigma_t^2] < \infty$;
  - $\sigma_t^2(Z_t^2 - 1) \in \mathcal{F}_t$ for all t;
  - $\mathbb{E}[\sigma_{t+1}^2(Z_{t+1}^2 - 1)\,|\,\mathcal{F}_t] = \sigma_{t+1}^2(\mathbb{E}[Z_{t+1}^2] - 1) = 0$.

  If $\mathbb{E}[X_t^4] < \infty$, then $\mathbb{E}[\varepsilon_t^2] = \mathbb{E}[(X_t^2 - \alpha_0 - \alpha_1 X_{t-1}^2)^2] < \infty$. Hence, $(X_t^2)_{t\in\mathbb{Z}}$ is an $\mathrm{AR}(1)$ of the form $X_t^2 - \frac{\alpha_0}{1-\alpha_1} = \alpha_1\Big(X_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1}\Big) + \varepsilon_t$.



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 4.21 1);

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

4.2.2 GARCH processes

Definition 4.23 ($\mathrm{GARCH}(p,q)$)
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 $\alpha_0 > 0$, $\alpha_k \geq 0$, $k \in \{1,\dots,p\}$, $\beta_k \geq 0$, $k \in \{1,\dots,q\}$.

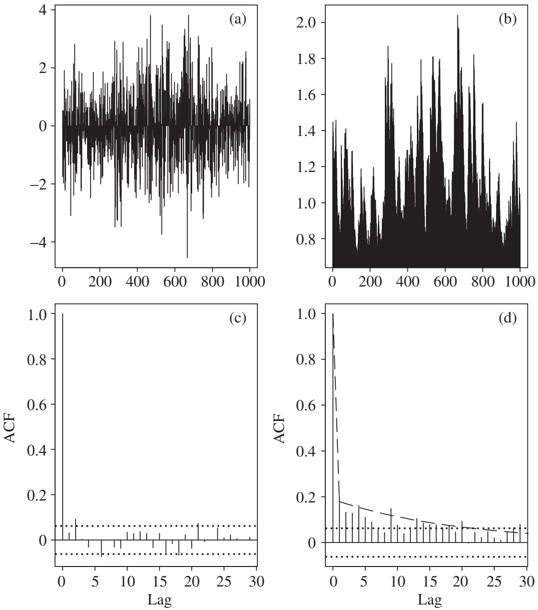
If one of $|X_{t-1}|,\dots,|X_{t-p}|$ or $\sigma_{t-1},\dots,\sigma_{t-q}$ is large, X_t is drawn from a distribution with (persistently) large variance. Periods of high volatility tend to be more persistent.

Example 4.24 ($\mathrm{GARCH}(1,1)$)

- 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, $X_t = Z_t\sqrt{\alpha_0\bigl(1+\sum_{k=1}^{\infty}\prod_{j=1}^{k}(\alpha_1 Z_{t-j}^2+\beta_1)\bigr)}$.

- $(X_t)_{t\in\mathbb{Z}}$ is stationary $\Leftrightarrow \alpha_1+\beta_1<1$. In this case, $\mathrm{Var}[X_t]=\frac{\alpha_0}{1-\alpha_1-\beta_1}$.
- One can show that $\mathbb{E}[X_t^4]<\infty$ if and only if $\mathbb{E}[(\alpha_1 Z_t^2+\beta_1)^2]<1$ (or $(\alpha_1+\beta_1)^2<1-(\mathrm{kurt}(Z_t)-1)\alpha_1^2$). From this result one can show that $\mathrm{kurt}(X_t)=\frac{\mathrm{kurt}(Z_t)(1-(\alpha_1+\beta_1)^2)}{1-(\alpha_1+\beta_1)^2-(\mathrm{kurt}(Z_t)-1)\alpha_1^2}$. If $\mathrm{kurt}(Z_t)>1$ (Gaussian, scaled t innovations), $\mathrm{kurt}(X_t)>\mathrm{kurt}(Z_t)$.
- Parallels with the ARMA(1,1) process: Let $\alpha_1+\beta_1<1$ and $\varepsilon_t=\sigma_t^2(Z_t^2-1)$ (MGDS), and assume $\mathbb{E}[X_t^4]<\infty$. Then $\sigma_{t-1}^2=X_{t-1}^2-\varepsilon_{t-1}$ implies that $X_t^2=\sigma_t^2 Z_t^2=\sigma_t^2+\varepsilon_t^2=\alpha_0+\alpha_1 X_{t-1}^2+\beta_1\sigma_{t-1}^2+\varepsilon_t=\alpha_0+(\alpha_1+\beta_1)X_{t-1}^2+\varepsilon_t-\beta_1\varepsilon_{t-1}$ which can be rewritten as $X_t^2-\frac{\alpha_0}{1-\alpha_1-\beta_1}=(\alpha_1+\beta_1)\bigl(X_{t-1}^2-\frac{\alpha_0}{1-\alpha_1-\beta_1}\bigr)+\varepsilon_t-\beta_1\varepsilon_{t-1}$, i.e., a $\mathrm{GARCH}(1,1)$ is an $\mathrm{ARMA}(1,1)$ for (X_t^2).



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 4.21 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, $\mathbb{E}[X_{t+h} \,|\, \mathcal{F}_t] = 0,\ h \in \mathbb{N}$.
- $\mathbb{E}[X_{t+1}^2 \,|\, \mathcal{F}_t] = \sigma_{t+1}^2 \mathbb{E}[Z_{t+1}] = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2$.
  For $h \geq 2$, X_{t+h}^2 and σ_{t+h}^2 are rvs, and

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

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

The GARCH(p,q) model

- Higher-order (G)ARCH models have the same general behavior 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). A squared IGARCH(1,1) resembles an ARIMA(0,1,1) model.

4.2.3 Simple extensions of the GARCH model

Consider stationary GARCH processes as white noise for ARMA processes.

Definition 4.25 ($\mathrm{ARMA}(p_1, q_1)$ 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

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

$$\mu_t = \mu + \sum_{k=1}^{p_1} \phi_k (X_{t-k} - \mu) + \sum_{k=1}^{q_1} \theta_k (X_{t-k} - \mu_{t-k}),$$

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p_2} \alpha_k (X_{t-k} - \mu_{t-k})^2 + \sum_{k=1}^{q_2} \beta_k \sigma_{t-k}^2,$$

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

- For the ARMA process to be a causal and invertible linear process, as before, the polynomials $\tilde{\phi}(z) = 1 - \phi_1 z - \cdots - \phi_{p_1} z^{p_1}$ and $\tilde{\theta}(z) = 1 + \theta_1 z + \cdots + \theta_{q_1} z^{q_1}$ should have no common roots and no roots in $\{z \in \mathbb{C} : |z| \le 1\}$.
- 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:
  - *GARCH with leverage*. 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 GARCH with leverage 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.

4.2.4 Fitting (G)ARCH 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

$$f_{X_0,\dots,X_n}(X_0,\dots,X_n) = f_{X_0}(X_0)\prod_{t=1}^{n} f_{X_t\,|\,X_{t-1},\dots,X_0}(X_t\,|\,X_{t-1},\dots,X_0)$$

$$= f_{X_0}(X_0) \prod_{t=1}^{n} f_{X_t | X_{t-1}}(X_t \,|\, X_{t-1})$$
$$= f_{X_0}(X_0) \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\Big(\frac{X_t}{\sigma_t}\Big),$$

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

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

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

$$L(\alpha_0, \alpha_1, \beta_1; X_0, \dots, X_n) = f_{X_1,\dots,X_n | X_0,\sigma_0}(X_1, \dots, X_n \,|\, X_0, \sigma_0)$$

$$= \prod_{t=1}^{n} f_{X_t | X_{t-1},\dots,X_0,\sigma_0}(X_t \,|\, X_{t-1}, \dots, X_0, \sigma_0) = \prod_{t=1}^{n} f_{X_t|\sigma_t}(X_t \,|\sigma_t)$$

$$= \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\Big(\frac{X_t}{\sigma_t}\Big), \quad \text{where } \sigma_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2}.$$

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.

- $\mathrm{GARCH}(p, q)$. Suppose we have data $X_{-p+1}, \dots, X_0, X_1, \dots, X_n$. Evaluate the likelihood conditional on the (observed) $X_{-p+1}, \dots, X_0$ as well as the (unobserved) $\sigma_{-q+1}, \dots, \sigma_0$ (choose starting values as above).

- 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 $\hat{\boldsymbol{\theta}}_n$.

Properties of (Q)MLEs

- We consider two situations: The model which has been fitted...

  1) ... has been correctly specified;

2) ... has the correct dyanmics but the the innovation distribution is erroneously assumed to be Gaussian (in this case the MLE is known as *quasi-maximum likelihood estimator (QMLE)*).

- The asymptotic results for GARCH models are similar to the results in the i.i.d. case; they have been derived in a series of papers. We only treat pure GARCH models, the form of the results will apply more generally (e.g., to ARMA models with GARCH errors).
- Under 1), one can show that for a $\mathrm{GARCH}(p, q)$ model with Gaussian innovations,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{(n\to\infty)}{\overset{\mathrm{d}}{\to}} \mathrm{N}_{p+q+1}(\mathbf{0}, I(\boldsymbol{\theta})^{-1}),$$

where

$$I(\boldsymbol{\theta}) := \mathbb{E}\Big[\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)^{\top}\Big] = -\mathbb{E}\Big(\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\Big) =: J(\boldsymbol{\theta})$$

is the *Fisher (or: expected) information* matrix. Thus we have a consistent and asymptotically normal estimator.

- In practice, the $I(\boldsymbol{\theta})$ is often approximated by an *observed information matrix*. Two candidates are
$$\bar{I}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\Big(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)^{\top}\Big) \quad \text{and} \quad \bar{J}(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2},$$
where the former has *outer-product* and the latter has *Hessian* form. Evaluating them at the MLEs leads to $\bar{I}(\hat{\boldsymbol{\theta}}_n)$ or $\bar{J}(\hat{\boldsymbol{\theta}}_n)$; in practice, the derivatives are often approximated using first and second-order differences. Under 1), $\bar{I}(\hat{\boldsymbol{\theta}}_n) \approx \bar{J}(\hat{\boldsymbol{\theta}}_n)$. One could also take the *sandwich estimator* $\bar{J}(\hat{\boldsymbol{\theta}}_n)\bar{I}(\hat{\boldsymbol{\theta}}_n)^{-1}\bar{J}(\hat{\boldsymbol{\theta}}_n)$.
- Under 2), one still obtains a consistent estimator. If the true innovation distribution has finite fourth moment, then
$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{(n\to\infty)}{\overset{\text{d}}{\to}} \mathrm{N}_{p+q+1}(\mathbf{0}, J(\boldsymbol{\theta})^{-1}I(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}),$$
Note that $I(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ typically differ in this case. $J(\boldsymbol{\theta})^{-1}I(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}$ can be estimated by the sandwich estimator.

- If model checking suggests that the dynamics have been adequately described by the GARCH model, but the Gaussian assumption seems doubtful, then standard errors for parameter estimates should be computed based on this covariance matrix estimate.

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 4.25.
- 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

$$\hat{Z}_t = \hat{\varepsilon}_t / \hat{\sigma}_t, \quad \hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{k=1}^{p_2} \hat{\alpha}_k \hat{\varepsilon}_{t-k}^2 + \sum_{k=1}^{q_2} \hat{\beta}_k \hat{\sigma}_{t-k}^2; \tag{9}$$

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 modeling and modeling of shocks that drive the process;
- Practical in higher dimensions.

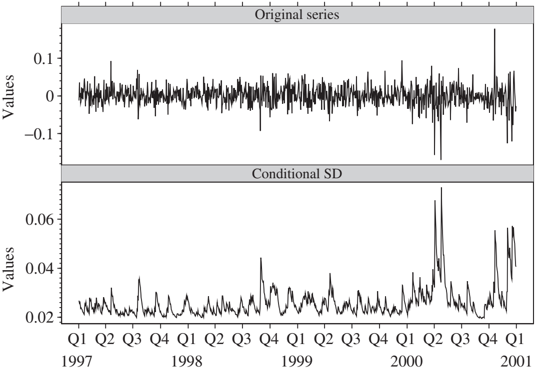
Drawbacks: ARMA fitting errors propagate through to the fitting of innovations (overall error hard to quantify).

Example 4.26 (GARCH model for Microsoft log-returns)

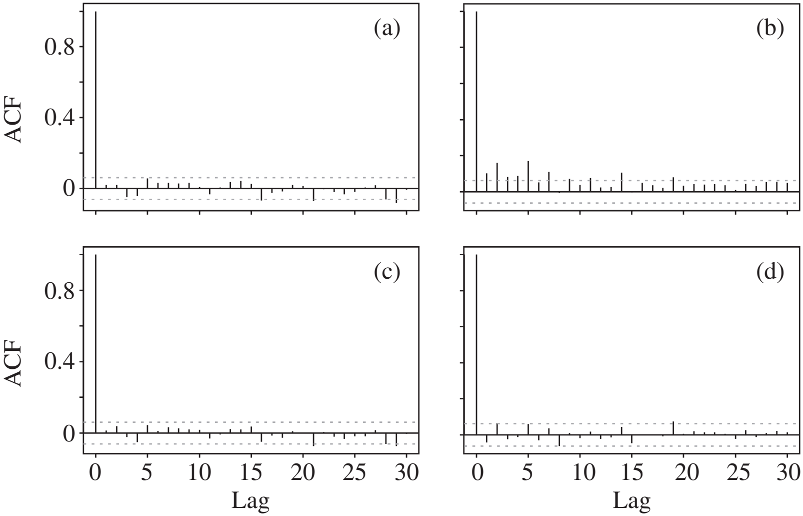
- Consider Microsoft daily log-returns from 1997–2000 (1009 values). The raw returns show no evidence of serial correlation, the absolute values do (Ljung–Box test based on the first 10 estimated correlations fails at the 5% level).
- Various models with t innovations are fitted via MLE: $\mathrm{GARCH}(1,1)$,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

- A model with leverage effect further improves the fit (both raw and absolute standardized residuals show no serieal correlation; Ljung–Box does not reject).

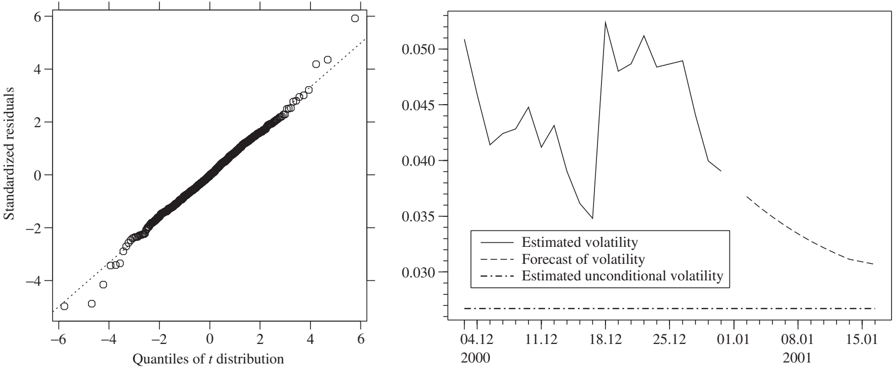
Microsoft log-returns 1997–2000: Data (top) and estimated volatitlity (bottom) from a $\mathrm{GARCH}(1,1)$ with leverage term.



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 (here: without leverage effect).



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 (G)ARCH model or ARMA model with GARCH errors).
- Assume we know $X_{t-n+1}, \dots, X_t$ and want to forecast σ_{t+h}, $h \geq 1$.
- 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: Exponential smoothing and conditional expectations.

Exponential smoothing

- A one-period ahead forecast $P_t X_{t+1}$ of X_{t+1} based on $\mathcal{F}_t$ is given by

$$P_t X_{t+1} = \alpha X_t + (1-\alpha) P_{t-1} X_t. \qquad (10)$$

  Applied to $(X_{t+1} - \mu_{t+1})^2$ leads to

$$P_t (X_{t+1} - \mu_{t+1})^2 = \alpha (X_t - \mu_t)^2 + (1-\alpha) P_{t-1} (X_t - \mu_t)^2. \qquad (11)$$

- Since $\sigma_{t+1}^2 = \mathbb{E}[(X_{t+1} - \mu_{t+1})^2 \,|\, \mathcal{F}_t]$, we can use (11) 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 Section 3). Alternatively, apply exponential smoothing to μ_t via $P_{t-1} X_t$ in (10).

Conditional expectation

The general procedure becomes clear from the following two examples

Example 4.27 (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 4.28 (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 4.19 and 4.27 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 (9).

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
$$F_{X_{t+1}|\mathcal{F}_t}(x)=\mathbb{P}(\mu_{t+1}+\sigma_{t+1}Z_{t+1}\le x\,|\,\mathcal{F}_t)=F_Z\Big(\frac{x-\mu_{t+1}}{\sigma_{t+1}}\Big).$$
- Let $\mathrm{VaR}_\alpha^t=F^{-}_{X_{t+1}|\mathcal{F}_t}(\alpha)$ and let ES_α^t denote the corresponding time-dynamic expected shortfall. We then have
$$\mathrm{VaR}_\alpha^t=\mu_{t+1}+\sigma_{t+1}F_Z^{-}(\alpha),\quad \mathrm{ES}_\alpha^t=\mu_{t+1}+\sigma_{t+1}\,\mathrm{ES}_\alpha(Z).$$
- If we can estimate μ_{t+1}, σ_{t+1} (parametrically/non-parametrically/semi-parametrically), we only have left to estimate $F_Z^{-}(\alpha)$ and $\mathrm{ES}_\alpha(Z)$.
- For GARCH-type models it is easy to calculate $F_Z^{-}(\alpha)$ 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 $F_Z^{-}(\alpha)$ and $\mathrm{ES}_\alpha(Z)$.