Time Series Analysis

Ulrich Stadtmüller

Contents

Introduction

Classical statistics: iid experiments. Usually we observe data $x_1, \ldots, x_n$ as a realisation of iid random variables $X_1, \ldots, X_n$.

Time series analysis. We are looking at only one experiment but observe the outcome over time. We want to unveil the time dependence. The goal is usually to forecast model outcomes.

EXAMPLE. (See distributed graphs)

(1) $x(t) = \cos(.2t + \pi/3)$.
(2) Plot of 30 draws from $U_{[-5,5]}$. Are these really independent?
(3) Population growth between 1780 and 1980 in the USA.
(4) Monthly Accidental Deaths in the USA (1973-1978). The mean seems to be almost constant.
(5) Wölfer Sunspot Numbers (1770-1869). This dataset also exhibits time dependence.
(6) Power generation in Germany. Data are not stable over time. The mean seems to increase faster than lineraly. The variance is also increasing, could be a function of the mean. There is also a clear periodicity in the data.
(7) DAFOX log returns 1993-1996, White Noise, and a GARCH(1,1) model.
(8) Bancruptcies in the US.

Say we observe $X_1, ..., X_n$ and want to forecast X_{n+1}. What is the best forecast? In terms of a L_2-distance its the conditional expectation

$$E[X_{n+1}|X_1, ..., X_n] = \begin{cases} E(X_{n+1}) = \mu & \text{if the } X_i \text{ are iid} \\ \underbrace{g(X_1, \ldots, X_n)}_{\stackrel{e.g.}{=} \alpha + \beta_1 X_1 + \ldots + \beta_n X_n} & \text{otherwise - sample models needed} \end{cases}.$$

.

CHAPTER 1

Basics

1.1. Stochastic processes, stationarity

DEFINITION 1.1.1. A _stochastic process_ is a family of random variables $(X_t, t \in T) = (X_t)_{t \in T}$, T some index set, on a probability space (Ω, Σ, P). The family $(X_t(\omega^*))_{t \in T}$ for some fixed $\omega^* \in \Omega$ is called a realization or path of the stochastic process $(X_t)_{t \in T}$.

REMARK 1.1.2. The most important index sets are $T = \mathbb{R}$ (as interval) or $T = \mathbb{Z}$ (or $\mathbb{N}$) for "time discrete" stochastic process.

If we talk about a _time series_, we mean on the theoretical side a time discrete stochastic process, if we think of data, it is a realization of such a process, or a section of such a realization, w.l.o.g. $(X_t)_{0 \leq t \leq T}$ with some $T \in \mathbb{N}$.

EXAMPLE 1.1.3.

(1) $X_t = A(\omega)\sin(t\nu + B(\omega))$, $t \in \mathbb{Z}$, $\nu \in \mathbb{R}$, and A and B are random variables. A realization of this process is given by :
$X_t(\omega^*) = \alpha \sin(t\nu + \beta^*)$,$\alpha^* = A(\omega^*)$, $\beta^* = B(\omega^*)$.

(2) X_t iid random variables, $t \in \mathbb{N}_0$:

$$X_t = \begin{cases} 1 & p = 1/2 \\ -1 & q = 1/2 \end{cases}, (\varepsilon_t)_{t \in \mathbb{N}_0} \overset{d}{=} U_{[-1,1]} \overset{d}{=} N_{0,\sigma^2} = \text{"white noise"}.$$

(3) $X_t = \Sigma_\nu^t \varepsilon_\nu$, $t \in \mathbb{N}_0$, $(\varepsilon_t)_{t \in \mathbb{N}_0}$ a white noise.
(4) $X_t = \alpha X_{t-1} + \varepsilon_t$, $\quad -1 < \alpha < 1$.

NOTATION 1.1.4. In the following a stochastic process $(X_t)_{t \in T}$ always has $T = \mathbb{Z}$ or $T = \mathbb{N}_0$ as time index set.

DEFINITION 1.1.5. Consider a stochastic process $(X_t)_{t \in T}$ with $E\left(X_t^2\right) < \infty \ \forall t \in T$, then $\mu_X : T \to \mathbb{R}$ defined by $\mu_X(t) = E\left(X_t\right)$ is called the _mean function_ of $(X_t)_{t \in T}$.
$\gamma_X : T \times T \to \mathbb{R}$ defined by $\gamma_X(s,t) = \text{Cov}(X_s, X_t) = \gamma_X(t,s)$, $s, t \in T$ is called the _autocovariance function_ of $(X_t)_{t \in T}$.
$\rho_X : T \times T \to \mathbb{R}$ defined by $\rho_X(s,t) = \text{Cov}\left(X_s, X_t\right)$, $s, t \in T$ is called the _autocorrelation function_ (acf) of $(X_t)_{t \in T}$.

DEFINITION 1.1.6. A stochastic process $(X_t)_{t \in T}$ is called (weakly) _stationary_ if $E\left(X_t^2\right) < \infty$, $\forall t \in T$ and

(1) $\mu_X(t) = \mu$, $\forall t \in T$
(2) $\gamma_X\left(s, t\right) = \gamma_X\left(s + r, t + r\right)$, $\forall s, t, r \in T$.

REMARK 1.1.7.

(1) Stationarity means that there exist functions
$\gamma_X(h) = \gamma_X(h, 0) = \gamma_X(s + h, s), \ h \in \mathbb{Z}, s \in T$
$\rho_X(h) = \rho_X(h, 0) = \rho_X(s + h, s), \ h \in \mathbb{Z}, s \in T$

(2) In particular we have:
$\mathrm{Var}(X_t) = \gamma_X(0)$ is constant over time.
$\gamma_X(-h) = \gamma_X(h)$ and $\rho_X(-h) = \rho_X(h), \ h \in \mathbb{N}.$

(3) Strong stationarity means: $F_{X_{t_1}, \ldots X_{t_n}} = F_{X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_n+s}} \ \forall t_1, \ldots, t_n, s \in T, \forall n$ implies weak stationarity

(4) A time series is called Gaussian if all finite dimensional marginal distribution functions are normal. In this case the distribution functions are completely determined if mean and covariance function are given and strong and weak stationarity coincide.

EXAMPLE 1.1.8.

(1) Let (ε_t) be independent (uncorrelated) random variables with $E(\varepsilon_t) = 0 \ \forall t$ and $E(\varepsilon_t^2) = \sigma_t^2, \ \forall t$. This process is stationary with $\mu_X(t) = 0$,

$$\gamma_X(h) = \begin{cases} \sigma_t^2, & h = 0 \\ 0, & h \in \mathbb{N} \end{cases} = \begin{cases} 1 & h = 0 \\ 0 & h \in \mathbb{N} \end{cases}.$$

Such a process is called a "white noise"-process.

(2) With $(\varepsilon_t)_{t \in \mathbb{Z}}$ as in 1. we consider $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$ with some $\theta \in \mathbb{R}$. Then $\mu_X(t) = 0$,

$$E(X_t \cdot X_{t+h}) = E((\varepsilon_t + \theta \varepsilon_{t-1})(\varepsilon_{t+h} + \theta \varepsilon_{t+h-1}))$$

$$= \begin{cases} \sigma_t^2(1 + \theta^2) & h = 0 \\ \theta \sigma_t^2 & h = \pm 1 \\ 0 & |h| > 1 \end{cases},$$

and $\rho_X(h) = \begin{cases} \frac{\theta}{1+\theta^2} & h = 1 \\ 0 & h \geq 2 \end{cases}$. Therefore, this process is stationary.

(3) The process $X_t = \begin{cases} \varepsilon_t, & |t| \text{ odd} \\ \varepsilon_{t+1} & |t| \text{ even} \end{cases}, t \in \mathbb{Z}$, is nonstationary, since

$$E(X_t, X_{t+1}) = \begin{cases} 0 & |t| \text{ odd} \\ \sigma_t^2 & |t| \text{ even} \end{cases}.$$

(4) $X_t = \sum_{v=0}^t \varepsilon_v, \ t \in \mathbb{N}_0$, is nonstationary since $\mathrm{Var}(X_t) = (t+1)\sigma_t^2$, $\mathrm{Cov}(X_t, X_{t+h}) = (t+1)\sigma_t^2, h \geq 1$. But $X_t - X_{t-1}$ is stationary.

(5) $X_t = \sigma_t \varepsilon_t, \ \sigma_t^2$ a stochastic process obeying $\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 \ \alpha, \beta > 0$ and $\mathcal{F}_t = \sigma(\varepsilon_{t-1}, \varepsilon_t)$.
$$E(X_t) = E(E(X_t | \mathcal{F}_{t-1})) = 0.$$

$$E(X_t X_{t+h}) = E(E(\sigma_t \varepsilon_t \sigma_{t+h} \varepsilon_{t+h} | \mathcal{F}_{t+h-1}))$$

$$= E(\sigma_t \sigma_{t+h} E(\varepsilon_t \varepsilon_{t+h} | \mathcal{F}_{t+h-1})) = \begin{cases} E(\sigma_t^2 \sigma_s^2) = (\alpha + \beta \sigma_t^2)\sigma_s^2 & h = 0 \\ 0 & h \geq 1 \end{cases}.$$

New model: $\sigma_t^2 = \alpha + \beta X_{t-1}^2 + \gamma \sigma_{t-1}^2 \ \alpha, \beta > 0, \ 0 < \gamma < 1$

PROBLEM. The theory will mainly discuss stationary time series, but the data are in most cases not stationary, since:

(1) $\mu_X(t)$ is not constant, but second moments reflect stationary behaviour.
(2) $\mu_X(t)$ is constant, but the second moments are not "OK".
(3) $\mu_X(t)$ is not constant and second moments are not "OK".

Ad situation 1: Here we can use regression in order to get a hand on $\mu_X(t)$: Estimate $\mu_X(t)$, take it out and work with the residuals.

Ad situation 2 and 3: We have no general recipe.

Typically the mean behaviour has two parts.

- continuous and monotone changes in the mean (global trend).
- periodic or cyclic trends (seasonal trends). The length of the periodicity can be seen from the nature of the data.

Sometimes there are also sudden changes in the mean function (changepoint analysis).

Our regression model for the time series is

$$X_t = m_t + s_t + \varepsilon_t,$$

where m_t represents the global trend, s_t seasonal trends and ε_t the residuals, hopefully a stationary process. Tasks: Estimate m_t, (s_t, d), (d the period). How do we do this?

(1) Suggest a model by looking at the data.
(2) Preprocess data to see more.
   (a) To see the global trend, do some local smoothing:

$$W_t = \frac{1}{2q+1} \sum_{\nu=-q}^{q} X_{t+\nu} = \frac{1}{2q+1} \sum_{\nu=-q}^{q} \left(m_{t+\nu} + \underbrace{s_{t+\nu}}_{\approx 0} + \underbrace{\varepsilon_{t+\nu}}_{\approx 0} \right)$$
$$\approx m(t) + 0 + 0$$

   if m is smooth and q is not too large. Typically one would choose q such that the periodicity is smoothed out. (Usually polynomial or exponential regression are suitable for economics data).
   (b) To model $m(t)$, apply regression, take $\hat{m}(t)$ out of the data and attack $s(t)$.
(3) $Z_t = X_t - \hat{m}_t \approx s_t + \varepsilon_t$. Periodicity d can be guessed from situation.
   - model s_t directly.
   - or estimate s_t: $\hat{s}_0 = \frac{1}{N} \sum_{\nu=0}^{N} X_{\nu d}$, $\hat{s}_1 = \frac{1}{N} \sum_{\nu=1}^{N} X_{\nu d+1}$.

1.2. Stationary Processes

ASSUMPTION 1.2.1. *Now we assume that X_t is stationary with existing second moment. Therefore $\mu_X(h) = \mu$ and $\rho_X(h)$ $h \geq 0$, $\gamma_X(h) = \sigma_X^2$. This is all we have using the definition of stationarity.*

Sometimes it is usefull to consider the same quantities, but in the space of Fourrier transforms.

DEFINITION 1.2.2. A $\left\{ \begin{array}{c} \text{real} \\ \text{complex} \end{array} \right\}$ sequence $(z_\nu)_{\nu \in \mathbb{Z}}$ is called _positive semidefinite_ if

$$\sum_{\nu,\mu=1}^{n} w_\nu w_\mu z_{(k_\nu - k_\mu)} \geq 0 \quad \forall (w_\nu)_{\nu-1}^n \in \left\{ \begin{array}{c} \mathbb{R}^n \\ \mathbb{C}^n \end{array} \right\}, \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} \in \mathbb{Z}$$

LEMMA 1.2.3.

(1) Let $\gamma_X(.)$ be the autocorrelation function of a stationary process. Then $\gamma_X(0) \geq 0$, $|\gamma_X(h)| \leq \gamma_X(0)$, and $\gamma_X(h) = \gamma_X(-h)$, $h \in \mathbb{N}$.
(2) A real sequence $(\gamma(h))_{h \in Z}$ is an autocorrelation function of some stationary process iff
    (a) $\gamma(h) = \gamma(-h)$ and
    (b) $\gamma(h)$ is positive semidefinite.

PROOF.

(1) $\gamma_X(0) = \text{Var}(X) \geq 0$
   $|\gamma_X(h)| \leq \gamma_X(0)$ is easy to show with Cauchy-Schwartz-Inequality.
   $\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_{t'-h}, X_{t'})$
(2) ($\Rightarrow$)
    (a) necessary for stationarity, see remark 1.1.7.
    (b) $0 \leq \text{Var}\left(\sum_{\nu=1}^n w_\nu X_{k_\nu}\right) = \sum_{\nu,\mu=1}^n w_\nu w_\mu \text{Cov}(X_{k_\nu}, X_{k_\mu})$
    ($\Leftarrow$)
    Let (X_t) be a Gaussian process with $E(X_t) = 0$ and covariance matrix of

$$\text{Cov}\begin{pmatrix} X_t \\ \vdots \\ X_{t+h} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(h) \\ \gamma(1) & \ddots & & \gamma(h-1) \\ \vdots & & \ddots & \vdots \\ \gamma(h) & \gamma(h-1) & \cdots & \gamma(0) \end{pmatrix}$$

This matrix is positive semidefinte, because of b). By the Kolmogorov existence theorem such a sequence of random variables exist.

$\square$

THEOREM 1.2.4. (Herglotz)
A complex sequence $(z_k)_{k \in \mathbb{Z}}$ is positive semidefinite iff there exists a positive, non-decreasing function $F : [-\pi, \pi] \rightarrow [0, \infty)$ such that

$$z_k = \int_{-\pi}^{\pi} e^{ikx} dF(x)$$

PROOF. No proof.
Idea: ($\Rightarrow$) integrate a Fourier series.
($\Leftarrow$) $\sum w_\nu w_\mu z_{k_\nu} z_{k_\mu} = \int |.|^2 dF \geq 0$ $\qquad \square$

REMARK 1.2.5.

(1) If we require $F(-\pi) = 0$, F is r.h. continuous then F is unique. (We will assume this from now on).

(2) z_k is real iff F is symmetric to 0, i.e. $z_k = 2 \int_{0+}^{\pi} \cos(kx) dF(x) + F(0) - F(0-)$, if F is absolutely continuous, then $z_k = 2 \int_{0+}^{\pi} \cos(kx) f(x) dx$.

DEFINITION 1.2.6. If $(\gamma(h))_{h \in \mathbb{Z}}$ is the autocovariance function of a stationary stochastic process (X_t), then the distribution function F defined through theorem 1.2.4 is called the _spectral distribution function_, and in the case of existence, its density f is called the _spectral density_.

REMARK 1.2.7.

(1) A spectral density exists if $\gamma(h) \in \mathcal{L}^2$, i.e. $\sum_{k \in \mathbb{Z}} |\gamma(k)|^2 < \infty$. If $\gamma(h) \in \mathcal{L}^1$, then

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\lambda} = \frac{\gamma_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma_k \cos(k\lambda)$$

and the convergence is uniform on $[-\pi, \pi]$ and f is continuous.
(2) The spectral density explains which frequencies are more or less important in the stochastic process (X_t). We will explain this with the following example:

EXAMPLE 1.2.8.

(1) $X_t = \sum_{j=1}^{n} A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)$ with A_j, B_j, $1 \leq j \leq n$ random variables being pointwise uncorrelated, having mean 0 and variance $\text{Var}(A_j) = \text{Var}(B_j) = \sigma_j^2$, $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_n < \pi$.

$$\gamma(h) = \sum_{j=1}^{n} \sigma_j^2 \left(\cos(\lambda_j t) \cos(\lambda_j(t+h)) + \sin(\lambda_j t) \sin(\lambda_j(t+h)) \right)$$

$$= \sum_{j=1}^{n} \sigma_j^2 \cos(\lambda_j h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda),$$

with $F(\lambda) = \frac{\sigma^2}{2} + \begin{cases} 0, & 0 \leq \lambda < \lambda_1 \\ \frac{\sigma^2}{2} & \lambda_1 \leq \lambda < \lambda_2 \\ \vdots & \\ \frac{\sigma_n^2}{2} & \lambda_n \leq \lambda \end{cases}$, $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$. Hence the spectral distribution function reveals which frequencies show up in the process and how important they are (large variance implies more relevance).

(2) If $\gamma(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \geq 1 \end{cases}$, this is white noise and $f(\lambda) = \frac{\sigma^2}{2\pi}$, $-\pi \leq \lambda \leq \pi$, i.e. in the white noise, all frequencies occur and have the same importance.

The spectral distribution function contains the same information as the autocorrelation function. Sometimes we will need one more interesting quantity: Given random variables $Y, X_1, \ldots, X_n \in L_2(\Omega, \Sigma, P)$ we are interested in $\alpha_o^*, \ldots \alpha_n^*$ minimizing $E(Y - \alpha_0 - \sum_{\nu=1}^{n} \alpha_\nu X_\nu)^2$ (i.e. the best linear prediction of Y by $X_1, \ldots, X_n$ in the L_2 sense. If $E(Y) = E(X_\nu) = 0 \ \forall \nu$, then $\alpha_0 = 0$, otherwise $\alpha_0 = \mu_y - \sum_{\nu=1}^{n} \alpha_\nu \mu_{X_\nu}$. The quantities $\alpha_1^*, \ldots, \alpha_n^*$ depend only on the covariances of the random variables. The $\alpha_1^*, \ldots, \alpha_n^*$ exist, since $\mathcal{L}_2(X_1, \ldots, X_n)$ is a closed subspace of the Hilbert space $L_2(\Omega, \Sigma, P)$.

Now assume $(X_t)_{t\in\mathbb{Z}}$ a stationary process centered at 0.

DEFINITION 1.2.9. The sequence $(\phi(k))_{k\in\mathbb{N}}$ defined through $\phi(k) = \phi_{kk}$, $k \in \mathbb{N}$, where ϕ_{kk} is the coefficient of X_k, in the best linear prediction of X_{k+1} by $X_1, \ldots, X_k$ i.e. $\hat{X}_{k+1} = \sum_{j=1}^{k} \phi_{kj} X_{k+1-j}$, is called the _partial autocorrelation function_ (pacf).[1]

REMARK 1.2.10.

(1) $\phi(1) = \rho(1)$, $E\left(X_2 - \alpha X_1\right)^2 = \sigma^2\left(1 - 2\alpha\rho(1) + \alpha^2\right) \Rightarrow \alpha^* = \rho(1)$.

(2) The ϕ_{kj} $1 \le j \le k$ are the coefficients of linear regression of X_{k+1} on $X_k, X_{k-1}, \ldots, X_1$. These coefficients are shift invariant and depend only on $\rho(1), \ldots, \rho(k)$. (Variances can be divided out.) The ordering of $X_k, \ldots, X_1$ has a reason which we will see later.

(3) Next we look at the regression of X_0 on $X_1, \ldots, X_k$ then we conclude from the remarks above that $\hat{X}_0 = \sum_{j=1}^{n} \phi_{kj} X_j$ (note $E(X_0 X_j) = E(X_k X_{k+1-j}) = \gamma(j)$

(4) Furthermore we can derive the following notation:

$$\phi(k+1) = \frac{E\left(\left(X_{k+1} - \hat{X}_{k+1}\right)\left(X_0 - \hat{X}_0\right)\right)}{\sqrt{E\left(\left(X_{k+1} - \hat{X}_{k+1}\right)^2\right) E\left(\left(X_0 - \hat{X}_0\right)^2\right)}}$$

i.e. the correlation between X_{k+1}- best linear prediction of X_{k+1} from $X_1, \ldots X_k$ and X_0 - best linear predictor of X_0 from $X_1, \ldots X_k$.

LEMMA 1.2.11. _(Yule-Walker equation)_
If $\gamma(h) \to 0$, $k \to \infty$, then for all $k \ge 1$ the linear system of equations

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(k-1) \\ \gamma(1) & \ddots & & \vdots \\ \vdots & & \ddots & \gamma(1) \\ \gamma(k-1) & \ldots & \gamma(1) & \gamma(0) \end{pmatrix} \vec{x} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \vdots \\ \gamma(k-1) \end{pmatrix}$$

in short $T_{k-1}\vec{x} = \gamma_{k-1}$, has the solution $\vec{x}^T = (\phi_{k1}, \ldots, \phi_{kk})$.

REMARK. The autocovariances may be replaced by autocorrelations.

PROOF. Given $(\Omega, \Sigma, \mathbb{P})$, centered random variables X, $E\left(X^2\right) < \infty$, $L_2\left(\Omega, \Sigma, \mathbb{P}\right)$, $< X, Y >= E(XY)$, $\|X\| = \sqrt{E\left(X^2\right)}$ a normed linear space.
Define $H_k := \mathcal{L}\left(X_1, \ldots, X_k\right) = \mathrm{span}(X_1, \ldots, X_k)$ a subspace of $L_2\left(\Omega, \Sigma, \mathbb{P}\right)$

---

[1]It might be of interest to look at the dependence between X_0 and X_{k+1}. The problem is that dependence cannot be described by linear expressions, but as soon as one takes into account the entire dependence structure, things become very complicated. Therefore one considers $\phi(k) = \phi_{kk}$, where the ϕ_{kk} are defined by

$$\hat{X}_{k+1} = \sum_{\nu=1}^{k} \phi_{k\nu} X_{k+1-\nu}$$

$$\hat{X}_0 = \sum_{\nu=1}^{k} \phi_{k\nu} X_\nu.$$

$\min_{X \in H_k} \|Y - X\|$ for $Y \in L_2(\Omega, \Sigma, P)$, is given by $Y^* = P_{H_k} \cdot Y$, where P_{H_k} is the projection onto the space H_k. Then $Y^* - Y \in H_k^\perp$.

$\hat{X}_{k+1} = P_{H_k} X_{k+1} = \sum_{\nu=1}^k \phi_{k\nu} X_{k+1-\nu}$, and the operator $I - P_{H_k}$ projects into $H_k^\perp$, thus $X_{k+1} - \hat{X}_{k+1} \in H_k^\perp$.

$E\left(\left(X_{k+1} - \hat{X}_{k+1}\right) X_{k+1-j}\right) = 0, \quad 1 \le j \le k$

$\gamma(j) - \sum_{\nu=1}^k \phi_{k\nu} E\left(X_{k+1-\nu} X_{k+1-j}\right) = 0, \quad 1 \le j \le k$

$\gamma(j) = \sum_{\nu=1}^k \phi_{k\nu} \gamma(\nu - j) = 0, \quad 1 \le j \le k$

$(\phi_{k1}, \ldots, \phi_{kk})^T$ is a solution.

Next we show that all T_k are regular.

Assume $\exists k_0 : T_{k_0}$ is not regular. T_{k_0} is the covariance matrix of $X_1, \ldots, X_{k_0}$.

With some $r < k_0$ we have $X_1, \ldots X_{k_0}$ are with probability 1 in a subspace of $\mathbb{R}^{k_0}$ of dimension r.

$\overset{e.g.}{\Rightarrow} X_{r+1} = \sum_{\nu=1}^r \alpha_j X_j \overset{\text{Stationarity}}{\Longrightarrow} X_{r+l+1} = \sum_{j=1}^r \alpha_j X_{j+1}, \quad \forall l \ge 0$

$= \ldots = \sum_{j=1}^r \alpha_{l,j} X_j$.

$\forall n : \exists a_n \in \mathbb{R}^r, X_n = a_n^T \vec{X}_r, \quad \forall n \ge r$

$$1 = \frac{\text{Var}(X_n)}{\gamma(0)} = \frac{a_n^T T_r a_n}{\gamma(0)} = \frac{a_r^T U_r^T \Lambda_r U_r a_n}{\gamma(0)} \ge \frac{\|a_n\|^2}{\gamma(0)} \lambda_1,$$

where $\Lambda_r = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix}$ and λ_1 is the smallest eigenvalue of T_r. This implies

$\|a_n\| \le \frac{\gamma(0)}{\lambda_1}, \ \forall n$ and

$\gamma(0) = \text{Var}(X_n) = E\left(X_n a_n^T \vec{X}_r\right) = \sum_{j=1}^r \underbrace{a_{nj}}_{\le c} \underbrace{\gamma(n-j)}_{\to 0} \to 0, \ n \to \infty.$ This is a

contradiction. $\square$

ALGORITHM 1.2.12. *(Durin-Levinson Algorithm)*

For computational purposes, it is more efficient is to use a recursion:

$v_0 = \gamma(0), \ v_1 = v_0\left(1 - \phi_{11}^2\right) = v_0(1 - \rho(1)^2), \ \phi(1) = \rho(1).$

Now for $k \ge 2$

- $\phi(k) = \left(\gamma(k) - \sum_{\nu=1}^{k-1} \phi_{k-1,\nu} \gamma(k - \nu)\right)/v_{k-1}$
- $v_k = v_{k-1}\left(1 - \phi(k)^2\right)\left(= Var\left(X_{k+1} - \hat{X}_{k+1}\right)\right)$
- $\begin{pmatrix} \phi_{k1} \\ \vdots \\ \phi_{kk-1} \end{pmatrix} = \begin{pmatrix} \phi_{k-1,1} \\ \vdots \\ \phi_{k-1,k-1} \end{pmatrix} - \phi_{kk}\begin{pmatrix} \phi_{k-1,k-1} \\ \vdots \\ \phi_{k-1,1} \end{pmatrix}$ *(cf. Brockwell-Davis)*

EXAMPLE 1.2.13.

(1) White noise (ε_t) $\rho(k) = 0, \ k \ne 0, \ \phi(k) = 0, \ \forall k \ge 1, \ (T_k = \sigma_k^2 I_k, \ \vec{\gamma}_k = 0)$

(2) $X_t = \varepsilon_t + \theta\varepsilon_{t-1}, \theta \neq \pm 1$

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \frac{\theta}{1+\theta^2} & k = 1 \\ 0 & k \geq 2 \end{cases}, \phi(k) = -\frac{(-\theta)^k\left(1-\theta^2\right)}{1-(\theta^2)^{k+1}} \left(\sim -\left(1-\theta^2\right)(-\theta)^k \quad |\theta| < 1 \right)$$

$$\begin{pmatrix} 1 & \frac{\theta}{1+\theta^2} & & \\ \frac{\theta}{1+\theta^2} & \ddots & \ddots & \\ & \ddots & \ddots & \\ & & & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} \frac{\theta}{1+\theta^2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \frac{\theta}{1+\theta^2}x_{\nu-1} + x_\nu + \frac{\theta}{1+\theta^2}x_{v+1} =$$

$0, \quad \nu = 2, \ldots, k-1$, constants from limes 1 and k. $X_\nu = c_1 \cdot Z_1^\nu + c_2 \cdot Z_2^\nu$, the Z's are zeros of $\frac{\theta}{1+\theta^2}Z^2 + Z + \frac{\theta}{1+\theta^2}$.

(3) We look for solutions of $X_t - \alpha X_{t-1} = \varepsilon_t, (\varepsilon_t)$ given white noise, $|\alpha| < 1$.
$\mathcal{F}_k = \sigma\left(\varepsilon_k, \varepsilon_{k-1}, \ldots\right) = \sigma\left(X_k, X_{k-1}, \ldots\right)$

$$E\left(X_{k+1}|\mathcal{F}_k\right) = E\left(\underbrace{X_{k+1} - \alpha X_k}_{=\varepsilon_{k+1}} |\mathcal{F}_k \right) + \alpha X_k = \alpha X_k \text{ the best prediction}$$

and hence the best linear prediciton of X_{k+1} from the past.
$X_{k+1} = \sum_{\nu=1}^{k} \phi_{k\nu}X_{k+1-\nu} = \alpha X_k + 0 \cdot X_{k-1} + \ldots + 0 \cdot X_1 = \alpha X_k$.
$\phi_{11} = \alpha = \phi(1) = \rho(1)$
$\phi_{kk} = 0 = \phi(k), \ k \geq 2$.

CHAPTER 2

Linear Processes

2.1. Linear Processes and filters

DEFINITION 2.1.1. A stochastic process of form $X_t = \sum_{\nu \in \mathbb{Z}} \alpha_\nu \varepsilon_{t-\nu}$, $t \in \mathbb{Z}$ with a real sequence $(\alpha_\nu) \in \ell_1$ (i.e. $\sum_{\nu \in \mathbb{Z}} |\alpha_\nu| < \infty$) and a white noise process $(\varepsilon_t)_{t \in \mathbb{Z}}$ with variance σ_ε^2 is called a _two-sided linear process_. If $\alpha_\nu = 0$, $\nu < 0$ the process is called one-sided or _causal_.

REMARK 2.1.2.

(1) The infinite series converges almost surely, since
$$E\left(\left|\sum \alpha_\nu \varepsilon_{t-\nu}\right|\right) \leq \sum_\nu |\alpha_\nu| \underbrace{E\left(|\varepsilon_{t-\nu}|\right)}_{\leq \sqrt{\sigma_\varepsilon^2}} < \infty$$
converges even absolutely almost surely.
(2) The so-called Wald decomposition theorem says, that all mean zero stationary processes can be written as a sum of a linear process and a predictable process.

LEMMA 2.1.3. _A linear process is stationary and has the following parameters:_

(1) $\mu = E(X_t) = 0$
(2) $\gamma_X(k) = \sigma_\varepsilon^2 \sum_\nu \alpha_\nu \alpha_{\nu+k}$, $k \in \mathbb{Z}$
(3) _It has a spectral density_ $f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left|\sum_{\nu \in \mathbb{Z}} \alpha_\nu e^{-i\lambda\nu}\right|^2$, $\lambda \in [-\pi, \pi]$

REMARK 2.1.4.

(1) All quantities exist since $(\alpha_\nu) \in \ell_1$.
(2) $f_X(\lambda) = f_\varepsilon(\lambda) \cdot |\text{Trig series with coeff } \alpha_\nu|^2$. The function $\sum_\nu \alpha_\nu e^{-i\lambda\nu}$ is called the _transfer function_ of the so-called linear filter $\sum_{\nu \in \mathbb{Z}} \alpha_\nu B^\nu$, where B is the so-called shift operator: $B^\nu Y_t = Y_{t-\nu}$. $\left(\sum \alpha_\nu B^\nu\right) \varepsilon_t = \sum \alpha_\nu \varepsilon_{t-v} = X_t$. The absolute value squared of the transfer function is called the power function of the linear filter.

PROOF. All series are pointwise absolutely convergent, which allows limits and sums to be interchanged(Fubini).

(1)
$$\mu_X(t) = E\left(\sum_\nu \alpha_\nu \varepsilon_{t-\nu}\right) = \sum_\nu \alpha_\nu E\left(\varepsilon_{t-\nu}\right) \equiv 0$$

13

(2)

$$\gamma_X(h) = E\left(X_t X_{t+h}\right)$$

$$= \sum_\nu \sum_\mu \alpha_\nu \alpha_\mu \underbrace{E\left(\varepsilon_{t-\nu}\varepsilon_{t+h-\mu}\right)}_{=\sigma_\varepsilon^2 \delta_{t-\nu, t+h-\mu}}$$

$$= \sigma_\varepsilon^2 \sum_\nu \alpha_\nu \alpha_{\nu+h}$$

(3)

$$\int_{-\pi}^{\pi} e^{i\lambda k} f_X(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{i\lambda k} \frac{\sigma_\varepsilon^2}{2\pi} \sum_\nu \alpha_\nu e^{-i\lambda\nu} \sum_\mu \alpha_\mu e^{i\lambda\mu} d\lambda$$

$$= \frac{\sigma_\varepsilon^2}{2\pi} \sum_{\nu,\mu} \alpha_\nu \alpha_\mu \underbrace{\int_{-\pi}^{\pi} e^{i\lambda(k+\mu-\nu)} d\lambda}_{=2\pi\delta_{0, k+\mu-\nu}}$$

$$= \sigma_\varepsilon^2 \sum_\mu \alpha_\mu \alpha_{\mu+k} = \gamma_X(k)$$

$\square$

REMARK 2.1.5. The same calculations show more generally:

Given a stationary process (X_t) with spectral distribution function F_X and given a linear filter $\sum_{\nu \in \mathbb{Z}} \alpha_\nu B^\nu$, $(\alpha_\nu) \in \ell_1$, then the process

$$Y_t = \left(\sum_\nu \alpha_\nu B^\nu\right) X_t = \sum_{\nu \in \mathbb{Z}} \alpha_\nu X_{t-\nu}$$

is again stationary with spectral distribution function

$$F_Y = \int_{-\pi}^{\pi} A(\nu) dF_X(\nu), \quad \lambda \in [-\pi, \pi]$$

where $A(\nu) = \left|\sum_\nu \alpha_\nu e^{-i\lambda\nu}\right|^2$ is the power transfer function, the operator B is given by $BX_t = X_{t-1}$, $BX_t = B^{j-1}\left(BX_t\right) = X_{t-j}$, $j \in \mathbb{N}$, $B^0 = \text{id}$, $B^{-1}X_t = X_{t+1}$, $B : L_2 \to L_2$, $\|B\| = 1$ on stationary random variables. If f_X exists, then $f_Y(\lambda) = A(\lambda) f_X(\lambda)$.

2.2. ARMA-Models

Many situations in economics can approximatively be described by some difference equation disturbed by some unexpected noise. Thus let a white noise $(\varepsilon_t)_t$, $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$ be given. We seek stochastic processes which are stationary and satisfy the following difference equation

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \ldots - \alpha_p X_{t-p} = \varepsilon_t + \sum_{\mu=1}^{q} \beta_\mu \varepsilon_{t-\mu}, \ t \in \mathbb{Z}, \quad (*)$$

$q \in \mathbb{N}$, $\alpha_\mu, \beta_\mu \in \mathbb{R}$, $\alpha_p, \beta_q \neq 0$.

DEFINITION 2.2.1. A stationary stochastic process (X_t) satisfying $(*)$ for $t \in \mathbb{Z}$ with a given white noise process with i.i.d. random variables (ε_j), $E(\varepsilon_j) = 0$, $E(\varepsilon_j^2) = \sigma_\varepsilon^2 \in [0, \infty)$ is called an <u>$ARMA(p,q)$-process</u> (**A**uto **R**egressive **M**oving **A**verage).

REMARK 2.2.2.

(1) We are here more restrictive than necessary for many things concering the white noise.
(2) A stationary stochastic process (X_t) is called an ARMA(p,q)-process with mean μ if $(X_t - \mu)$ is an ARMA(p,q) process.

NOTATION 2.2.3. We denote the following polynomials

$$\Phi(z) = \sum_{\nu=0}^{p} a_\nu z^\nu, \ a_\nu = \begin{cases} 1 & \nu = 0 \\ -\alpha_\nu & 1 \le \nu \le p \end{cases},$$

$$\Psi(z) = \sum_{\mu=0}^{q} b_\mu z^\mu, \ b_\mu = \begin{cases} 1 & \mu = 0 \\ \beta_\mu & 1 \le \mu \le q \end{cases}$$

and can rewrite equation $(*)$ as $\Phi(B)X_t = \Psi(B)\varepsilon_t$.

Special cases.

Moving average process of order q (MA(q)). In this case $\Phi(z) \equiv 1$ and equation $(*)$ can be written as

$$X_t = \varepsilon_t + \sum_{\mu=1}^{q} \varepsilon_{t-\mu} \beta_\mu.$$

LEMMA 2.2.4. *Such a moving average process has the following characteristics:*

(1)

$$\mu_x \equiv 0, \quad \gamma_X = \begin{cases} \sigma_\varepsilon^2 \sum_{\mu=0}^{q-|k|} b_\mu b_{\mu+|k|}, & 0 \le |k| \le q \\ 0 & |k| > q \end{cases}$$

(2)

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{\mu=0}^{q} b_\mu e^{-i\lambda\mu} \right|^2 = \frac{1}{2\pi} \sum_{-q}^{q} \gamma_X(k) e^{-i\lambda k}$$

(3) *The partial autocorrelation function $(|\phi(k)|)$ decays exponentially fast to zero, provided $\Psi(z) \ne 0$ for $|z| = 1$.*

EXAMPLE. Consider a MA(1) process $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$, $\Psi(z) = 1 + \theta z$, $z = -\frac{1}{\theta}$, $\gamma_X(k) = \begin{cases} (1+\theta^2)\sigma_\varepsilon^2 & k = 0 \\ \theta\sigma_\varepsilon^2 & k = 1 \\ 0 & k \ge 2 \end{cases}$, $\phi(k) = -\frac{(-\theta)^k(1-\theta^2)}{1-(\theta^2)^{k+1}}$, $|\theta| \ne 1$, $f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| 1 + \theta e^{-i\lambda} \right|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \left(1 + 2\theta \cos\lambda + \theta^2 \right)$

PROOF. Follows directly from Lemma 2.1.3 $\square$

Autoregressive process of order p (AR(p)). In this case $\Psi(z) \equiv 1$ and equation $(*)$ can be written as

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = \varepsilon_t.$$

EXAMPLE 2.2.5. Consider an AR(1)-process (X_t): $X_t - \alpha X_{t-1} = \varepsilon_t$, $\Phi(z) = 1 - \alpha z$, $z = -\frac{1}{\alpha}$

$$X_t = \alpha X_{t-1} + \varepsilon_t = \alpha^2 X_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t$$

$$= \alpha^{n+1} X_{t+n+1} + \sum_{\nu=0}^{n} \alpha^\nu \varepsilon_{t-\nu} \overset{\mathbb{P} \text{ a.s.}}{\underset{(n \to \infty)}{\to}} 0 + \sum_{\nu=0}^{\infty} \alpha^\nu \varepsilon_{t-\nu}.$$

$(1 - \alpha B) X_t = \varepsilon_t$ and therefore $X_t = (1 - \alpha B)^{-1} \varepsilon_t = (\sum_{\nu=0}^{\infty} \alpha^\nu B^\nu) \varepsilon_t = \sum_{\nu=0}^{\infty} \alpha^\nu \varepsilon_{t-\nu}$ (v. Neumann series)

To see the uniqueness, consider two solutions $z_t^{(1)}$ and $z_t^{(2)}$. Then

$$z_t^{(1)} - z_t^{(2)} - \alpha \left(z_{t-1}^{(1)} - z_{t-1}^{(2)} \right) = 0.$$

For $y_t = z_t^{(1)} - z_t^{(2)}$ we obtain $y_t - \alpha y_{t-1} = 0 \Rightarrow y_t = c \cdot \alpha^t \overset{\mu=0}{\Rightarrow} c = 0$ and thus $z_t^{(1)} = z_t^{(2)}$.

$\mu = 0$, $\gamma_X(k) = \frac{\sigma_\varepsilon^2}{1-\alpha^2} \alpha^k$, $\rho_X(k) = \alpha^k$, $\phi(k) = \begin{cases} \alpha, & k = 1 \\ 0, & k \geq 2 \end{cases}$ [1]

$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{\nu=0}^{\infty} \alpha^\nu e^{-i\lambda\nu} \right|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|1-\alpha e^{-i\lambda}|^2} = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|\Phi(e^{-i\lambda})|^2} = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{(1-2\alpha\cos\lambda+\alpha^2)}$.

In the case that $\alpha = \pm 1$ we have no stationary solution. If $|\alpha| > 1$ then $X_t = \sum_{\nu=0}^{\infty} \left(\frac{1}{\alpha} \right)^\nu \varepsilon_{t+\nu}$ which means that the stationary solution is non-causal (since it depends on future ε_t), this case is not interesting for us.

General Case.

THEOREM 2.2.6. *Let be given the difference equation $(*)$ with polynomials $\Phi(z)$ and $\Psi(z)$, having no zero in common.*

(1) *There exists a stationary and causal solution of $(*)$ if and only if $\Phi(z) \neq 0$ for $|z| \leq 1$.*

(2) *If $\Phi(z) \neq 0$ on $|z| \leq 1$ then the unique solution of $(*)$ is given by*

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \quad t \in \mathbb{Z},$$

where the coefficients (c_j) are the Taylor coefficients of $\Psi(z)/\Phi(z)$ at $z_0 = 0$. The radius of convergence is $R = \min\{|z|, \Phi(z) = 0\} > 1$. Hence $|c_j| \leq \kappa_q q^{-j}$ for any $q \in (1, R)$ with some $\kappa_q > 0$. Therefore $(c_j) \in \ell_1$.

---

[1]This is part of the reason that the partial auto-correlation function was introduced. For AR processes it breaks down at the order of the process.

(3) *In this case we have*

$$\gamma_X(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} c_j c_{j+k}, \quad k \in \mathbb{N}_0$$

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{n=0}^{\infty} c_n e^{-i\lambda n} \right|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{\Psi\left(e^{-i\lambda}\right)}{\Phi\left(e^{-i\lambda}\right)} \right|^2$$

(4) *If $\Psi(z) = 1$, then $\phi(k) = 0$ for $k > p$, otherwise $\phi(k) \to 0$, $k \to \infty$ geometrically fast (ϕ is the partial autocorrelation function).*

REMARK.

(1) From (2) it follows that the $|\rho_X(k)| \le \tilde{\kappa}_q q^{-k}$ for all $q \in (1, k)$ with suitable $\tilde{\kappa}_q > 0$.

(2) If $\Phi(z) \ne 0$ on $|z| = 1$, then $X_t = \sum_{j=-\infty}^{\infty} d_j \varepsilon_{t-j}$, where d_j are the coefficients of the Laurent expansion of $\frac{\Psi(z)}{\Phi(z)} = \sum_{j=-\infty}^{\infty} d_j z^j$.

PROOF. Reminder:
$A(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu$, $B(z) = \sum_{\nu=0}^{\infty} b_\nu z^\nu$, $|z| < R$. Then $C(z) = A(z)B(z) = \sum_{\rho=0}^{\infty} z^\rho \sum_{\nu+\mu=\rho} a_\nu b_\mu$, $|z| < R$, $\frac{1}{A(z)} = \sum_{\nu=0}^{\infty} a_\nu^- z^\nu$ in $U_\varepsilon(0)$ if $A(0) \ne 0$. $C(z)/A(z) = \sum_{\rho=0}^{\infty} z^\rho \sum c_\nu a_\mu^-$.

Necessity:
$r = \min\{|z_1|, \ldots, |z_p|\} > 1$, z_i zero of $\Phi(z)$.

$$\frac{1}{\Phi(z)} = \sum_{n=0}^{\infty} a_n^- z^n,$$

$|z| < r, |a_j^-| < \kappa_q q^{-j}$, $q \in (1, r)$
$Y_t := \Psi(B)\varepsilon_t$ is an MA(q) process $\Phi(z) = \prod_{j=1}^{p}(1 - \alpha_j z)$ with $\alpha_j = \frac{1}{z_j}$

$$\begin{aligned}
Z_t &= \prod_{j=1}^{p} (1 - \alpha_j B)^{-1} Y_t \\
&= \prod_{j=1}^{p-1} (1 - \alpha_j B)^{-1} \underbrace{\left((1 - \alpha_p B)^{-1} Y_t \right)}_{Y_t^{(1)} = \sum_{\nu=0}^{\infty} \alpha_p^\nu Y_{t-\nu}} \\
&= \sum_{n=0}^{\infty} a_n^- B^n \underbrace{Y_t}_{\Psi(B)\varepsilon_t}.
\end{aligned}$$

$Y_t^{(1)} = \sum_{\nu=0}^{\infty} \alpha_p^\nu Y_{t-\nu}$ is a stationary, causal process. (Z_t) is a solution of $(*)$, since $\Phi(B)Z_t = \Phi(B) \sum_{n=0}^{\infty} a_n^- B^n Y_t = Y_t = \sum B^\rho \sum_{\nu+\mu=\rho} a_\nu a_\mu^-$.

(2) and (3) follow from (1) and the above.

Sufficiency:

$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \ (\psi_j) \in \ell_1.$

$$\Phi(B)X_t = \left(\sum_{n=0}^{p} a_n B^n \sum_{j=0}^{\infty} \psi_j B^j \right) \varepsilon_t$$

$$= \left(\sum_{\rho=0}^{\infty} B^\rho \sum_{\nu+\mu=\rho} a_\nu \psi_\mu \right) \varepsilon_t$$

$$\overset{!}{=} \sum_{\rho=0}^{q} b_\rho B^\rho \varepsilon_t,$$

where we have set $b_\rho = \begin{cases} \sum_{v+\mu=\rho} a_\nu \psi_\mu, & 0 \leq \rho \leq q \\ 0, & \text{otherwise} \end{cases}$.

With $H(z) = \sum_{n=0}^{\infty} \psi_n z^n$ we have $\Phi(z)H(z) = \Psi(z)$ and thus $H(z) = \Psi(z)/\Phi(z)$ in $U_\varepsilon(0)$, since $H(z)$ is a power series converging in $|z| \leq 1$. Because Φ and Ψ have no common zeros, $\Psi(z) \neq 0$ in $|z| \leq 1$.

$\square$

REMARK. In case $\Psi(z) \neq 0$ in $|z| \leq 1$ we may write $\varepsilon_t = \sum_{n=0}^{\infty} \tilde{c}_j X_{t-j}$.

2.3. ARIMA and SARIMA-processes

Goal: Extend ARMA-models to cover trends and seasonality effects.

$$\Phi(B)X_t = \Psi(B)\varepsilon_t \qquad (*)$$

2.3.1. ARIMA-processes. $\Phi(z) = (1-z)\tilde{\Phi}(z)$, $\tilde{\Phi}(z) \neq 0$ in $|z| \leq 1$ of order p. So

$$\Phi(B)X_t = \tilde{\Phi}(B) \underbrace{(1-B)X_t}_{Y_t} = \tilde{\Phi}(B)Y_t = \Psi(B)\varepsilon_t,$$

where (Y_t) satisfies an ARMA(p,q) difference equation.

$Y_t = X_t - X_{t-1} \hookrightarrow X_t = X_0 + \sum_{\nu=1}^{t} Y_\nu$ for given X_0. (X_t) is a nonstationary process and if Y_t has mean μ then X_t has a linear trend. In this case we say that (X_t) follows an ARIMA(p,q,1) process (**A**uto **R**egressive **I**ntegrated **M**oving **A**verage). One can iterate this, look for example at

$$\Phi(z) = (1-z)^2 \tilde{\Phi}(z)$$

then $Y_t = X_t - 2X_{t-1} + X_{t-2}$ is an ARMA(p,q) process with mean μ. $X_t = X_0 + t(X_1 - X_0) + \sum_{\mu=2}^{t} Y_\mu (1 - \mu + 1), \quad t \geq 0$, given X_0 and X_1, (quadratic trend, $and X_t$ is an ARIMA(p,q,2) process).

2.3.2. SARIMA-processes. Consider
$$\Phi(z) = (1 - z^d)\tilde{\Phi}(z) \hookrightarrow \Phi(B)X_t = \Psi(B)\varepsilon_t,$$
$\Phi(B)X_t = \tilde{\Phi}(B)(X_t - X_{t-1}) = \Psi(B)\varepsilon_t$ and $(X_t - X_{t-1}) =: Y_t$. Knowing $X_0, \dots, X_{d-1}$ we get $X_t = \sum_{k=0}^m Y_{kd+l}$, if $t = md+l$ with $0 \le l \le d$. For the (Y_t) process we apply ARMA(p,q). This often leads to large coefficients p and q. It is more efficient to apply

$$\Phi_1\left(B^d\right)\left(1 - B^d\right)^l X_t = \Psi_1\left(B^d\right)U_t \quad \oplus$$

where $\Phi_1(z) \ne 0$ in $|z| \le 1$ and to then apply an ARIMA(p,q,m) model to (U_t) :
$$\Phi_2(B)(1 - B)^m U_t = \Psi_2(B)\varepsilon_t,$$

with $\Phi_2(z) \ne 0$ on $|z| \le 1$.

$$\underbrace{\Phi_2(B)\Phi_1\left(B^d\right)}_{\tilde{\Phi}(B)}(1 - B)^m\left(1 - B^d\right)^l X_t = \Psi_1\left(B^d\right)\Psi_2(B)\varepsilon_t$$

this is a SARIMA$(p, q, m) \times (\tilde{p}, \tilde{q}, \tilde{m})$ model.

2.4. Forecasting

2.4.1. Some theory. In case $(X_s)_{s \le t}$ is known, we are looking for the best linear forecast[2] for $X_{t+\nu}$, $\nu \in \mathbb{N}$, denoted by $\left[\hat{X}_{t+\nu}\right]$ and put $\mathfrak{M}_t^X = \mathcal{L}(X_x, \ s \le t)$. In case only $(X_s)_{t-m \le s \le t}$ is known, we are looking for the best linear forecast for $X_{t+\nu}$, $\nu \in \mathbb{N}$ denoted by $\left[\hat{X}_{t+\nu,m}\right]$ and put $\mathfrak{M}_t^X = \mathcal{L}(X_{t-m}, \dots X_t)$ which is a closed subset of $L_2(\Omega, \Sigma, P)$.

In the first case we have $\hat{X}_{t+\nu} = \text{argmin}_{\tilde{X} \in \bar{\mathfrak{M}}_t^X}\left\|X_{t+\nu} - \tilde{X}_{t+\nu}\right\|^2$, $\left\|X_{t+\nu} - \hat{X}_{t+\nu}\right\| = \sigma_\nu^2$, while in the second case $\hat{X}_{t+\nu,m} = \text{argmin}_{\tilde{X} \in \bar{\mathfrak{M}}_{t,m}^X}\left\|X_{t+\nu} - \tilde{X}\right\|^2$, $\left\|X_{t+\nu} - \hat{X}_{t+\nu}\right\| = \sigma_{\nu,m}^2 \ge \sigma_\nu^2$.

If the processes are stationary, none of the quantities depend on t. Since t is not important we put $t - n$, $m - n - 1$. We know $X_1, \dots, X_n$. Denote by $H_n = \mathcal{L}(X_1, \dots, X_n)$. Analogously to the calculations for the partial autocorrelation function we obtain

$$\hat{X}_{n+1} - \hat{X}_{n+1,1} = P_{H_n}X_{n+1} \overset{\text{see PACF}}{=} \sum_{\nu=1}^n \phi_{n,\nu}X_{n+1-\nu} \qquad (\diamond)$$

and $\sigma_{1,n-1}^2 = v_n = E\left(X_{n+1} - \hat{X}_{n+1}\right)^2$ with coefficients $\vec{\phi}_n = (\phi_{n,1}, \dots, \phi_{n,n})^T$. The coefficients $\phi_{n,\nu}$ can be obtained via $T_n \cdot \vec{\phi}_n = \vec{\gamma}_n$.

Goal: Write $H_n = H_{n-1} \oplus W_n$, i.e. the orthogonal sum of the span of $X_1, \dots, X_{n-1}$ and the orthogonal compliment.[3]

---

[2]One could also look for the best L_2 approximation, but the term $E[X_{t+\nu} \mid \sigma(X_s, \ s \le t)]$ is usually complicated.

[3]$\oplus$ represents the orthogonal sum between vector spaces $H_{n-1} \perp W_n$

Let $\hat{X}_1 = 0$ (=mean), and define $\tilde{X}_\nu := X_\nu - \hat{X}_\nu = \underbrace{\left(\text{id} - P_{H_{\nu-1}}\right)}_{\text{projects into } H_{\nu-1}^\perp} \cdot X_\nu$

$\underbrace{\tilde{X}_\nu}_{\in H_{n-1}^\perp} \underbrace{\tilde{X}_\mu}_{\in H_{n-1}} \quad 1 \le \mu \le \nu - 1$

$H_n = \underbrace{\mathcal{L}\left(X_1 - \hat{X}_1\right) \oplus \mathcal{L}\left(X_2 - \hat{X}_2\right) \oplus \ldots}_{=H_{n-1}} \oplus \underbrace{\mathcal{L}\left(X_n - \hat{X}_n\right)}_{=W_n}$

Hence $\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{nj} \tilde{X}_{n+1-j}$.

LEMMA 2.4.1. *Assume (X_t) denotes a centered, square integrable process. Furthermore, we assume that its covariance matrix*

$$\kappa(i,j) = E\left(X_i \cdot X_j\right), \; 1 \le i, j \le n$$

is not singular for all $n \in \mathbb{N}$. Then the quantities from above can be calculated recursively:

$$v_0 = \kappa(1,1) \; (= \gamma_X(0) \; \textit{in the stationary case})$$

then for $n = 1, 2, \ldots$ we do

$$\theta_{n,n-k} = \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j\right) / v_k, \quad k = 0, 1, \ldots, n-1$$

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j \qquad \textit{Innovation algorithm.}$$

PROOF.

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{n,j}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right)$$

multiplying both sides by $(X_{k+1} - \hat{X}_{k+1})$, $0 \le k \le n-1$ and taking expectations, we obtain

$$E\left(\underbrace{\left(\hat{X}_{n+1} - X_{n+1} + X_{n+1}\right)}_{\tilde{X}_{n+1}} \underbrace{\left(X_{k+1} - \hat{X}_{k+1}\right)}_{\tilde{X}_{k+1}}\right) = \sum_{j=1}^{n} \theta_{n,j} E\left(\left(X_{n+1-j} - \hat{X}_{n+1-j}\right)\left(X_{k+1} - \hat{X}_{k+1}\right)\right)$$

for all $0 \le k \le n-1$. Observe that

$$E\left(\left(X_{n+1-j} - \hat{X}_{n+1-j}\right)\left(X_{k+1} - \hat{X}_{k+1}\right)\right) = \begin{cases} v_k & n+1-j = k+1 \\ 0 & \text{indices different} \end{cases}.$$

Now using $\tilde{X}_{n+1} \perp \tilde{X}_{k+1}$ and $\hat{X}_{k+1} = \sum_{j=1}^{k} \theta_{k,j}\left(X_{k+1-j} - \hat{X}_{k+1-j}\right)$ and $\nearrow$, we get

$\kappa(n+1, k+1) - \sum_{j=1}^{k} \theta_{k,j} \underbrace{E\left(X_{n+1}\left(X_{k+1-j} - \hat{X}_{k+1-j}\right)\right)}_{\theta_{n,n-j} v_j \text{ by } (*)} = \theta_{n,n-k} v_k.$

Now put $k - j = \nu$

$$\kappa\left(n+1,k+1\right)-\sum_{\nu=0}^{k-1}\theta_{k,k-\nu}\underbrace{E\left(X_{n+1}\left(X_{\nu+1}-\hat{X}_{\nu+1}\right)\right)}_{\theta_{n,n-v}v_{\nu}}$$

$v_n\left\|X_{n+1}-\hat{X}_{n+1}\right\|^2=\left\|X_{n+1}\right\|^2-\left\|\hat{X}_{n+1}\right\|^2=\kappa(n+1,n+1)-\sum_{j=1}^{n}\theta_{n,j}^2 v_{n-j}^2$ since $X_{n+1}-\hat{X}_{n+1}\perp\hat{X}_{n+1}$ by Pythagoras. $\qquad\square$

2.4.2. Multistep forecast. To forcast $\hat{X}_{n+k}^{(k)}=P_{H_n}\cdot X_{n+k}$, with $k\geq 1$ again we calculate $\hat{X}_{n+k}^{(k)}=\sum_{j=1}^{n}\phi_n^{(k)}X_{n+1-j}$ and the same calculations as for the Yule-Walker equation lead to

$$\Gamma_n\vec{\phi}_n^{(k)}=\begin{pmatrix}\gamma_X(k)\\\vdots\\\gamma_X(n+k-1)\end{pmatrix}.$$

Another possibility is to use the innovation algorithm.

$\hat{X}_{n+k}^{(k)}=P_{H_n}\cdot X_{n+k}=P_{H_n}\cdot P_{H_{n+k-1}}\cdot X_{n+k}=P_{H_n}\cdot\sum_{j=1}^{n+k-1}\theta_{n+k-1,j}\tilde{X}_{n+k-j}=$

$\sum_{j=k}^{n+k-1}\theta_{n+k-1,j}\left(X_{n+k-j}-\hat{X}_{n+k-j}\right)$, since $\tilde{X}_{n+k-j}=\begin{cases}\perp H_n & \text{if } 1\leq j\leq k-1\\\in H_n & j\geq k\end{cases}$.

Then we just calculate $\theta_{m,j}$ for $1\leq m\leq n+k-1$ by the innovation algorithm.

$v_{n+k}^{(k)}=\left\|X_{n+k}-\hat{X}_{n+k}^{(k)}\right\|^2=\kappa(n+k,n+k)-\sum_{j=k}^{n+k-1}\theta_{n+k-1,j}^2 v_{n+k-1-j}$

2.4.3. ARMA process.

$$\Phi(B)\cdot X_t=\Psi(B)\cdot\varepsilon_t\qquad(*)$$

(1) MA(q)
    (a) $q=1$ $X_t=\varepsilon_t+\beta\varepsilon_{t-1}$, $t\in\mathbb{Z}$

$$\kappa(i,j)=\begin{cases}\sigma_\varepsilon^2(1+\beta^2), & |i-j|=0\\\beta\sigma_\varepsilon^2, & |i-j|=1\\0 & \text{otherwise}\end{cases}$$

        $v_0=\sigma_\varepsilon^2(1+\beta^2)$, $r_0=\frac{v_0}{\sigma_\varepsilon^2}$

        $\theta_{11}=\kappa(2,1)/v_0=\frac{\beta\sigma_\varepsilon^2}{(1+\beta^2)\sigma_\varepsilon^2}=\frac{\beta}{r_0}$,

        $v_1=\kappa(2,2)-\theta_{1,1}^2 v_0=\dots$

        $r_1=r_0-\frac{\beta^2}{r_0}$

        $\theta_{2,2}=0,\theta_{2,1}=\left(\kappa(3,2)-\theta_{11}\theta_{22}v_1\right)/v_1=\frac{\beta}{r_1}$, $r_2=r_0-\frac{\beta^2}{r_1}$
        by induction:

        $\theta_{n,1}=\frac{\beta}{r_{n-1}}$, $\theta_{n,k}=0$, $k\leq 2\leq n$, $r_n=r_0-\frac{\beta^2}{r_{n-1}}$, $n\geq 3$. $r_n\to 1$
        (exponentially fast), $v_n\to\sigma_\varepsilon^2$

        $\hat{X}_1=0$, $\hat{X}_2=\frac{\beta}{r_0}X_1$, $\hat{X}_3=\frac{\beta}{r_1}\left(X_2-\hat{X}_2\right)=\frac{\beta}{r_1}\left(X_2-\frac{\beta}{r_0}X_1\right)$ ...

        $\hat{X}_{n+1}=\frac{\beta}{r_{n-1}}\left(X_n-\hat{X}_n\right)$.

    (b) $q>1$. Since $\gamma_X(k)=0$, $k>q$ we have $\theta_{n\nu}=0$, $q+1\leq\nu\leq n$, $\forall n$.
        Consequently:

        $\hat{X}_{n+1}=\sum_{j=1}^{q}\theta_{n,j}\left(X_{n+1-j}-\hat{X}_{n+1-j}\right)$

$$\hat{X}_{n+1}^{(k)} = \sum_{j=k}^{q} \theta_{n+k-1,j} \left(X_{n+k-j} - \hat{X}_{n+k-j} \right) = \begin{cases} 0 & \text{if } k > q \end{cases}$$

(2) General ARMA(p,q)

Reduces essentially to a MA-process by the following procedure: With $m = p \vee q$ define

$$W_t := \begin{cases} \sigma_\varepsilon^{-1} X_t & , t = 1, \dots, m \\ \sigma_\varepsilon^{-1} \Phi(B) X_t & , t \geq m+1 \end{cases} \left(= \Psi(B) \frac{\varepsilon_t}{\sigma_\varepsilon} \right).$$

We have $H_n = \mathcal{L}(X_1, \dots, X_n) = \mathcal{L}(W_1, \dots, W_n)$, as usual we put $\hat{X}_1 = \hat{W}_1 = 0$. (W_t) is a nonstationary with mean zero and covariances

$$\kappa(i,j) = \begin{cases} \sigma_\varepsilon^{-2} \gamma_X(i-j) & , 1 \leq i,j \leq m \\ \sigma_\varepsilon^{-2} \left(\gamma_X(i-j) - \sum_{\nu=1}^{p} \alpha_\nu \gamma_X(\nu - (i-j)) \right) & , \min\{i,j\} \leq m \leq \max\{i,j\} \leq 2m \\ \sum_{\nu=0}^{q} b_\nu b_{v+|i-j|} & , m+1 \leq \min\{i,j\} \\ 0 & , \text{otherwise} \end{cases}$$

Apply innovation algorithm to $(W_t) \curvearrowright \theta_{n,j}^W, v_n^W, v_0^W = \frac{\gamma_X(0)}{\sigma_\varepsilon^{-2}}$

$$\hat{W}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{n,j}^W \left(W_{n+1-j} - \hat{W}_{n+1-j} \right) & , 1 \leq n \leq m-1 \\ \sum_{j=1}^{q} \theta_{n,j}^W \left(W_{n+1-j} - \hat{W}_{n+1-j} \right) & , n \geq m \end{cases}$$

$$\hat{W}_{n+k}^{(k)} = \begin{cases} \sum_{j=k}^{n+k-1} \theta_{n+k-1,j}^W \left(W_{n+k-j} - \hat{W}_{n+k-j} \right) & , 1 \leq n+k-1 \leq m-1 \\ \sum_{j=k}^{q} \theta_{n+k-1,j}^W \left(W_{n+k-j} - \hat{W}_{n+k-j} \right) & , n+k-1 \geq m \end{cases}$$

$$\hat{W}_t = P_{H_{t-1}} \cdot W_t = \begin{cases} P_{H_{t-1}} \left(\sigma_t^{-1} X_t \right) = \sigma_\varepsilon^{-1} \hat{X}_t & , 1 \leq t \leq m \\ \underbrace{P_{H_{t-1}} \left(\sigma_\varepsilon^{-1} \Phi(B) X_t \right)}_{\sigma_\varepsilon^{-1} \left(\hat{X}_t - \sum_{j=1}^{p} \alpha_j X_{t-j} \right)} & , t \geq m+1 \end{cases}$$

we thus have $(X_t = \sum_{\nu=1}^{p} \alpha_\nu X_{t-\nu} + \sigma_\varepsilon W_t, \ t \geq 1)$, $X_t - \hat{X}_t = \sigma_\varepsilon \left(W_t - \hat{W}_t \right)$, for all $t \geq 1$.

Finally:

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{n,j}^W \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) & , 1 \leq n \leq m-1 \\ \sum_{j=1}^{p} \alpha_j X_{n+1-j} + \sum_{j=1}^{q} \theta_{n,j}^W \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) & , n \geq m \end{cases}$$

$v_n^X = v_n^W \sigma_\varepsilon^2$ and

$$\hat{X}_{n+k}^{(k)} = \begin{cases} \sum_{j=k}^{n+k-1} \theta_{n+k-1}^W \left(X_{n+k-j} - \hat{X}_{n+k-j} \right) & n \leq m-k \\ \sum_{j=1}^{(k-1)\wedge p} \alpha_j \hat{X}_{n+k-j} + \sum_{j=k}^{p} a_j X_{n+k-j} + \sum_{j=k}^{q} \theta_{n+k-1,j}^W \left(X_{n+k-j} - \hat{X}_{n+k-j} \right) & n \geq m-k+1 \end{cases}$$

Special case: $q = 0$, i.e. we have an AR(p). Then

$$\hat{X}_{n+1} = \sum_{j=1}^{p} \alpha_j X_{n+1-j}, \ n \geq p$$

and

$$\hat{X}_{n+1}^{(k)} = \sum_{j=1}^{(k-1)\wedge p} \alpha_j \hat{X}_{n+k-j} + \sum_{j=k}^{p} \alpha_j X_{n+k-j}, \ n \geq p + k - 1$$

(Handouts distributed)

2.5. Kalman Filter

In software tools, forecasting is very often accomplished by Kalman filters. Kalman filters apply to a more general situation of dynamic processes described by linear equations for $t \geq 1$. These system depend on two equations, namely the *"system equation"*

$$Z_{t+1} = A_t Z_t + B_t \xi_t,$$

with given matrices $A_t \in \mathbb{R}^{p \times p}$, $B_t \in \mathbb{R}^{p \times m}$ and random vectors (ξ_t) with $E(\xi_t) \equiv 0$, $\mathrm{Cov}(\xi_t) = Q_t \in \mathbb{R}^{m \times m}$. However it is possible that one cannot observe Z_t directly, but only a relate variable Y_t (Z_t can be interpreted as a state space variable and Y_t as an observation varaible), which is given by the so-called *"observation equation"*:

$$Y_t = C_t Z_t + \eta_t,$$

with matrices $C_t \in \mathbb{R}^{q \times p}$ and random vectors η_t with $E(\eta_t) \equiv 0$, $\mathrm{Cov}(\eta_t) = R_t \in \mathbb{R}^{q \times q}$, η_t and η_{t+h} are uncorrelated for $h \neq 0$, η_t, ξ_t are independent and also independent of Z_1.

EXAMPLE.

(1) AR(p) process

$$Z_t = \begin{pmatrix} X_t \\ \vdots \\ X_{t-p+1} \end{pmatrix}, A_t = A = \begin{pmatrix} \alpha_1 & \cdots & \cdots & \alpha_p \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}, \xi_t = \begin{pmatrix} \varepsilon_{t+1} \\ 0 \end{pmatrix},$$

$$Q_t = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix} \ Y_t = (1, 0, \ldots, 0) Z_t = X_t, \text{ thus } C_t = C =$$

$(1, 0, \ldots, 0)$, $\eta_t = 0$, $R_t = 0$. System equation $X_{t+1} = \sum_{\nu=1}^{p} \alpha_\nu X_{t+1-\nu} + \varepsilon_{t+1}$, observation equation $Y_t = X_t$. Searching for $\hat{X}_{t+1}$ from $Y_1, \ldots, Y_t$. If all matrices are constant we are in the stationary case. Interesting for us is the stable case, where we have a limit behaviour (eigenvalues of A are all $|.| < 1$, ($\hat{=} \Phi(z) \neq 0$ in $|z| \leq 1$ in the example), that means the influence of the earlier values diminishes over time and we get

$$Z_{t+1} = \sum_{\nu=0}^{\infty} A^\nu \xi_{t-\nu}.$$

(2) MA(q) process

$$Z_t = \begin{pmatrix} \varepsilon_t \\ \vdots \\ \varepsilon_{t-q} \end{pmatrix} \in \mathbb{R}^{q+1} \; \xi_t = \begin{pmatrix} \varepsilon_{t+1} \\ 0 \\ \\ 0 \end{pmatrix} \in \mathbb{R}^{q+1} \; A = \begin{pmatrix} 0 & & & \\ 1 & & \ddots & \\ & \ddots & & \ddots \\ & & & 1 & 0 \end{pmatrix} \in$$

$\mathbb{R}^{(q+1)\times(q+1)} \Rightarrow Z_{t+1} = AZ_t + \xi_t, \; Y_t = (1, \beta_1, \ldots, \beta_q)\, Z_t = X_t.$

(3) ARMA(p,q) process

$Z_t = (X_t, X_{t-1}, \ldots, X_{t-p}, \varepsilon_t, \ldots, \varepsilon_{t-q+1})^T, \; \xi_t = (\varepsilon_{t+1}, 0, \ldots, 0)^T, \; A =$

$$\begin{pmatrix} \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ 1 & 0 & & & & \\ & \ddots & \ddots & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 1 & 0 \end{pmatrix}, \; Y_t = (1, 0, \ldots, 0)\, Z_t = X_t.$$ This repre-

sentation is not unique.

(4) Missing values

If one observes only $X_{t_1}, \ldots, X_{t_r}$ with $1 \leq t_1 < t_2 < \ldots < t_r \leq n$ from all n values. The system equation describes the process and $Y_t = C_t Z_t$, where C_t pickes the non-missing values.

Goal: Best linear approximation in the L_2-sense for Z_{t+1}

(1) from $Y_1, \ldots, Y_{t+1}$ "filtering problem"
(2) from $Y_1, \ldots, Y_t$ "prediction problem"
(3) from $Y_0, \ldots, Y_n$ "smoothing problem" (missing values)

We concentrate on the prediction problem (the others are slightly different but follow the same arguments). We know that $\hat{Z}_{t+1}$ is the best linear prediction based on $Y_1, \ldots, Y_t$ for Z_{t+1} iff

$$\text{Cov}\left(Z_{t+1} - \hat{Z}_{t+1}, Y_s\right) = 0 \quad 1 \leq s \leq t,$$

since $Z_{t+1} - \hat{Z}_{t+1} \perp \mathcal{L}(Y_1, \ldots, Y_t)$. If we assume that $\hat{Z}_t$ is the best linear prediction for Z_t from $Y_1, \ldots, Y_{t-1}$ how do we get $\hat{Z}_{t+1}$?

(1) Predictor Step:

$\tilde{Z}_{t+1} = A_t$ best linear predition of Z_{t+1} from $Y_1, \ldots, Y_t$ if $\hat{Z}_t$ is the best linear prediction of Z_t from $Y_1, \ldots, Y_{t-1}$. $\tilde{Y}_t = C_t \hat{Z}_t$ best linear predition of Y_t from $Y_1, \ldots, Y_{t-1}$. $\Rightarrow Y_t - \tilde{Y}_t$ is the new information, knowing Y_t ($\perp \mathcal{L}(Y_1, \ldots, Y_{t-1})$).

(2) Corrector Step:

$\hat{Z}_{t+1} = \tilde{Z}_{t+1} + K_t\left(Y_t - \tilde{Y}_t\right) = A_t \hat{Z}_t + K_t\left(Y_t - C_t \hat{Z}_t\right) = A_t \hat{Z}_t + K_t C_t\left(Z_t - \hat{Z}_t\right)$ with a suitable matirx K_t to be chosen subject to $\text{Cov}(Z_{t+1} - \tilde{Z}_{t+1}, Y_s) \stackrel{!}{=} 0, 1 \leq s \leq t$ (the optimality condition $Z_{t+1} - \tilde{Z}_{t+1} \perp \mathcal{L}(Y_1, \ldots, Y_t)$) which in turn is $\text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1} - K\left(Y_t - \tilde{Y}_t\right), Y_s - \tilde{Y}_s\right) \stackrel{!}{=} 0, 1 \leq s \leq t$, where we have subtracted $\tilde{Y}_s$ from thr right term of the covariance, which does not alter the validity of the equation. Now $Z_{t+1} - \tilde{Z}_{t+1}$ as well

as $Y_t - \tilde{Y}_t$ is in $\mathcal{L}\left(Y_1, \ldots, Y_{t-1}\right)^{\perp}$, so that the only equation which adds any information is where $s = t$. This evaluates to

$$\text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1}, Y_t - \tilde{Y}_t\right) - K_t \text{Cov}\left(Y - \tilde{Y}_t, Y_t - \tilde{Y}_t\right) \stackrel{!}{=} 0.$$

With $\hat{P}_t = \text{Cov}\left(Z_{t+1} - \hat{Z}_{t+1}, Z_{t+1} - \hat{Z}_{t+1}\right)$ (the "predition error") and $\tilde{P}_{t+1} = \text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1}, Z_{t+1} - \tilde{Z}_{t+1}\right) = A_t \hat{P}_t A_t^T + Q_t$ we find

$$\text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1}, C_t\left(Z_{t+1} - \hat{Z}_{t+1}\right) + \eta_t\right) - K_t \text{Cov}\left(C_t\left(Z_t - \tilde{Z}_t\right) + \eta_t, \ldots\right) = 0$$

$$\Leftrightarrow \text{Cov}\left(A_t\left(Z_t - \hat{Z}_t\right) + \xi_t, C_t\left(Z_t - \hat{Z}_t\right) + \eta_t\right) - K_t\left(C_t \tilde{P}_t C_t^T + R_t\right) = 0$$

$$A_t \hat{P}_t C_t^T - K_t\left(C_t \tilde{P}_t C_t^T + R_t\right) = 0$$

$$\Rightarrow K_t = A_t \hat{P}_t C_t^T \left(C_t \tilde{P}_t C_t^T + R_t\right)^{-1}$$

if the inverse does not exist, the Moore-Penrose inverse (fulfilling $A^- A A^- = A^-$) will do the job.

To find $\hat{P}_{t+1}$ not that

$$\hat{P}_{t+1} = \text{Cov}\left(Z_{t+1} - \hat{Z}_{t+1}, Z_{t+1} - \hat{Z}_{t+1}\right)$$

$$= \text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1} - K_t\left(Y_t - \tilde{Y}_t\right), Z_{t+1} - \tilde{Z}_{t+1} - K_t\left(Y_t - \tilde{Y}_t\right)\right)$$

$$= \text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1}, Z_{t+1} - \tilde{Z}_{t+1}\right) - 2\text{Cov}\left(Z_{t+1} - \tilde{Z}_{t+1}, K_t\left(Y_t - \tilde{Y}_t\right)\right)$$

$$+ K_t \text{Cov}\left(Y_t - \tilde{Y}_t, Y_t - \tilde{Y}_t\right) K_t^T$$

$$= \tilde{P}_{t+1} - 2 A_t \hat{P}_t C_t^T K_t^T + K_t\left(C_t \hat{P}_t C_t^T + R_t\right) K_t$$

$$= \tilde{P}_{t+1} - A_t \hat{P}_t C_t^T K_t^T$$

In total we get an recursive algorithm in two steps. Given $\hat{Z}_t, \hat{P}_t, \tilde{P}_t$:

(1) Predition step: from $\hat{Z}_t, \hat{P}_t$ $\tilde{Z}_{t+1} = A_t \hat{Z}_t, \tilde{Y}_t = C_t \hat{Z}_t, \tilde{P}_{t+1} = A_t \hat{P}_t A_t^T + Q_t$.
(2) Correction step: $K_t = A_t \hat{P}_t C_t \left(C_t \tilde{P}_t C_t + R_t\right)^-, \hat{Z}_{t+1} = \tilde{Z}_{t+1} + K_t\left(Y_t - \tilde{Y}_t\right)$
$\hat{P}_{t+1} = \tilde{P}_{t+1} - A_t \hat{P}_t C_t^T K_t^T$,

with starting values $\hat{Z}_1 = 0, \hat{P}_1 = \sigma I = \tilde{P}_1$ where σ^2 refelcts the uncertainty in the model.

If we start with $\hat{X}_{t+1}$ in the algorithm for the AR(p) process, we see that we are in a fixed point, as it should be $\left(\hat{P}_t = \left(\begin{array}{c}\sigma_{\varepsilon}^2 \\ \end{array}\right), \tilde{Y}_t = Y_t\right)$. Simulation and theoretical results show that (under certain assumptions) the influence of the starting value diminishes quickly and the algorithm converges to the desired solution.

Multistep prediciton is also possible: $\tilde{Z}_{t+h} = A_{t+h-1}\tilde{Z}_{t+h-1} = A_{t+h-1} \cdot \ldots \cdot A_t \tilde{Z}_t$ and $\tilde{Z}_t = \hat{Z}_t$. The prediciton error can also be calculated.

REMARK.

(1) Another approach shows the above via the Gauss Markov theorem.
(2) It is also possible to include control variables U.

Statistical Analysis

3.1. Estimation in the Time Domain

Given is a stationary time series (X_t) with mean μ, autocovariance function $\gamma_X(.)$ and partial autocorrelation function $\phi_X(.)$. Assume we have observed $X_1, \ldots, X_n$ (data $x_1 = X_1(\omega), \ldots, x_n = X_n(\omega)$).

Statistics:

(1) $\hat{\mu} = \bar{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$

(2) $\hat{\gamma}_X(h) = \frac{1}{n} \sum_{j=1}^{n-|h|} (X_j - \bar{X})(X_{j+|h|} - \bar{X})$, $0 \le h \le n-1$

(3) $\hat{\rho}(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$

These are the classical estimates from statistics. However, we are no longer dealing with independent observations. Will this work in a dependend case?

for the partial autocorrelation function we have the Yule-Walker equation

$$\Gamma_k \vec{\phi}_k = \vec{\gamma}_k,$$

for estimation we solve $\hat{\Gamma}_k \vec{x} = \hat{\vec{\gamma}}_k$, lemma 1.3 says this is uniquely solvable, so $\hat{\vec{\phi}}_k = \hat{\Gamma}_k^{-1} \hat{\vec{\gamma}}_k$. We find $\hat{\vec{\phi}}_k \xrightarrow{P} \vec{\phi}_k$, $n \to \infty$, since $\hat{\vec{\gamma}}_k \to \vec{\gamma}_k$, $n \to \infty$, $\hat{\Gamma}_k \xrightarrow{P} \Gamma_k$ for $n \to \infty$ and the same holds for Γ_k^{-1}.

Furthermore

$$\sqrt{n}\left(\hat{\vec{\phi}}_k - \vec{\phi}_k\right) = \sqrt{n}\hat{\Gamma}_k^{-1}\left(\hat{\vec{\gamma}}_k - \hat{\Gamma}_k \vec{\phi}_k\right)$$

$$= \sqrt{n}\hat{\Gamma}_k^{-1}\begin{pmatrix} \vdots \\ \hat{\gamma}(j) - \gamma(j) - \sum_{\nu=1}^{k}(\hat{\gamma}(j-\nu) - \gamma(j-\nu))\phi_{k\nu} \\ \vdots \end{pmatrix}$$

$$= \sqrt{n}\hat{\Gamma}_k^{-1} A\left(\hat{\vec{\gamma}}_k - \vec{\gamma}_k\right)$$

with $A = \begin{pmatrix} & \vdots & \\ 0 & & I_k \\ & \vdots & \end{pmatrix} - \left\{ \begin{pmatrix} \phi_{k1} & 0 & \cdots & \cdots & 0 \\ \phi_{k2} & \phi_{k1} & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \phi_{kk} & \cdots & \cdots & \phi_{k1} & 0 \end{pmatrix} + \begin{pmatrix} 0 & \phi_{k2} & \cdots & \cdots & \phi_{kk} & 0 \\ \vdots & \phi_{k3} & \cdots & \phi_{kk} & 0 & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & \phi_{kk} & 0 & & \vdots & \vdots \\ 0 & \cdots & & \cdots & 0 & 0 \end{pmatrix} \right\}$

$$\sqrt{n}\left(\hat{\vec{\phi}}_k - \vec{\phi}_k\right) \xrightarrow{d} \mathcal{N}_{0, \Gamma_k^{-1} A V A^T \Gamma_k^{-1}}$$

3.2. Estimation in ARMA-Models

Our goal here is to estimate the unknown parameers $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma_\varepsilon^2$.

3.2.1. AR(p)-process.

$$X_t - \alpha_1 X_{t-1} - \ldots - \alpha_p X_{t-p} = \varepsilon_t \quad (*)$$

with iid white noise (ε_t), $E(\varepsilon_1) = 0$, $\text{Var}(\varepsilon_1) = \sigma_\varepsilon^2$. $\Phi(z) \neq 0$ in $|z| \leq 1$. Multiplying by X_t and taking expectations in $(*)$ we obtain

$$\gamma_X(0) - \alpha_1 \gamma_X(1) - \ldots - \alpha_p \gamma_X(p) = E(\varepsilon_t X_t) = \sigma_\varepsilon^2 \qquad (\alpha)$$

(since $X_t = \sum_{\nu=0}^{\infty} c_\nu \varepsilon_{t-\nu}$, $c_0 = 1$).

Yule-Walker equation for $k = p$:

$$\Gamma_p \vec{\alpha}_p = \vec{\gamma}_p \qquad (\beta)$$

Hence we obtain: $\hat{\vec{\alpha}}_p = \hat{\Gamma}_p^{-1} \hat{\vec{\gamma}}_p = \hat{R}_p^{-1} \hat{\vec{\rho}}_p$, $\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_X(0) - \sum_{\nu=1}^{p} \hat{\alpha}_\nu \hat{\gamma}_X(\nu)$

REMARK. One can show that $\hat{\Phi}(z) \neq 0, |z| \leq 1$ and hence the $\hat{\alpha}_i$ generate again a stationary process.

THEOREM 3.2.1. *Under the additional assumption $E(\varepsilon_1^4) < \infty$ we find for $p' \geq p$ and $\vec{\alpha}_p' = (\alpha_1, \ldots, \alpha_p, 0, \ldots, 0)$*

$$\sqrt{n}\left(\hat{\vec{\alpha}}_p - \vec{\alpha}_{p'}\right) \xrightarrow{d} \mathcal{N}_{0, \sigma_\varepsilon^2 \Gamma_{p'}^{-1}}$$

PROOF. Follows from above for $k = p'$ and the fact that

$$\Gamma_{p'}^{-1} A V A \Gamma_{p'}^{-1} = \sigma_\varepsilon^2 \Gamma_{p'}^{-1},$$

in particular $\sigma_\varepsilon^2 \left(\Gamma^{-1}\right)_{p',p'} = 1$ for $p' > p$. $\qquad\qquad \square$

We could use least square principle ot estimate α's.

$$\sum_{t=2}^{n}\left(X_t - \sum_{\nu=1}^{p} \alpha_\nu X_{t-\nu}\right)^2 = \left\|\vec{X} - \tilde{A}\vec{\alpha}\right\|^2 \to \min_{\vec{\alpha}}$$

$$\tilde{A} = \begin{pmatrix} X_1 & & & \\ X_2 & X_1 & & \\ \vdots & & \ddots & \\ X_{n-1} & \cdots & \cdots & X_{n-p} \end{pmatrix}_{n-1 \times p}, \vec{X} = \begin{pmatrix} X_2 \\ \\ X_n \end{pmatrix}, \tilde{\alpha} = \left(\tilde{A}^T \tilde{A}\right)^{-1} \tilde{A}^T \vec{X}$$

$$\sqrt{n}\left(\tilde{\alpha} - \vec{\alpha}\right) = \left(\frac{\tilde{A}^T \tilde{A}}{n}\right)^{-1}\left(\frac{\tilde{A}^T \vec{X}}{\sqrt{n}} - \frac{\left(\tilde{A}^T \tilde{A}\right)}{\sqrt{n}}\vec{\alpha}\right) = \left(\frac{\tilde{A}^T \tilde{A}}{n}\right)^{-1} \tilde{A}^T\left(\frac{\vec{X} - \tilde{A}\vec{x}}{\sqrt{n}}\right) = \hat{\Gamma}_p^{-1} \underbrace{\left(\frac{\sum_{j=k+1}^{n} X_{j-k}\varepsilon_j}{\sqrt{n}}\right)}_{\xrightarrow{d} \mathcal{N}_{0, \Gamma_p \sigma_\varepsilon^2}}$$

What about MLE? Assume $\varepsilon_t \sim \mathcal{N}_{0,\sigma_\varepsilon^2}$

$$f_{X_1,\ldots,X_n}(\vec{x}) = f_{X_n | \underbrace{X_{n-1},\ldots,X_1}_{X_{n-1},\ldots,X_{n-p}}}(x_n \mid x_{n-1},\ldots,x_1) f_{X_{n-1},\ldots,X_1}$$

$$= \prod_{j=p+1}^{n} f_{X_j | X_{j-1},\ldots,X_{j-p}} f_{X_p,\ldots,X_1}$$

$$= \prod_{j=p+1}^{n} \mathfrak{n}_{\theta,\sigma_\varepsilon^2}\left(x_j - \sum_{\nu=1}^{p} \alpha_\nu x_{j-\nu}\right) f_{X_p,\ldots,X_1}$$

$$\Rightarrow \quad L(X_1,\ldots,X_n) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}^{n-p}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=p+1}^{n}\left(X_j - \sum_{\nu=1}^{p} \alpha_\nu X_{j-\nu}\right)^2\right) f_{X_p,\ldots,X_1}.$$

$$\to \max$$

(For details see Procter and Davis)

THEOREM 3.2.2. *If $m > p$ then we have provided that $E\left(\varepsilon_1^4\right) < \infty$*

$$\sqrt{n}\left(\hat{\vec{\phi}}_m - \vec{\phi}_m\right) \xrightarrow{d} \mathcal{N}_{0,\Gamma_m^{-1}\sigma_\varepsilon^2}$$

and in particular for $m > p$ $\sqrt{n}\left(\hat{\phi}(m) - \phi(m)\right) \xrightarrow{d} \mathcal{N}_{0,1}$.

REMARK.

(1) This yields an asymptotic test on $H_{0,m} : \phi(m) = 0$. If $\sqrt{n}\left|\hat{\phi}(m)\right| > \mathcal{N}_{0,1}^{-1}\left(1 - \frac{\alpha}{2}\right)$ then reject the hypothesis. (This explains the bounds seen when plotting the partial autocovariance function for example in R).

(2) The quatities $\hat{\phi}_{m,\nu}$ can be calculated using the Durbin-Levinson algorithm using $\hat{\gamma}_X(.)$ instead of $\gamma_X(.)$.

(3) $n(\hat{\vec{\alpha}}-\vec{\alpha})^T\Gamma_p(\hat{\vec{\alpha}}_n-\vec{\alpha})/\sigma_\varepsilon^2 \xrightarrow{d} \chi_p^2$, by Slutzky's theorem, we may replace Γ_p and σ_ε^2 by their estimates.

3.2.2. MA(q)-process.

$$X_t = \varepsilon_t + \sum_{\nu=1}^{q} \beta_\nu \varepsilon_{t-\nu}, \quad t \in \mathbb{Z} \quad (*)$$

as before $b_\nu = \begin{cases} 1 & \nu = 0 \\ \beta_\nu & 1 \leq \nu \leq q \end{cases}$. To estimate the b_ν's σ_ε^2 we have

$$\gamma_X(0) = \sum_{\nu=0}^{q} b_\nu^2 \sigma_\varepsilon^2$$

$$\gamma_X(1) = \sum_{v=0}^{q-1} b_v b_{\nu+1} \sigma_\varepsilon^2$$

$$\gamma_X(q) = b_0 b_q \sigma_\varepsilon^2,$$

$q + 1$ nonlinear equations for $q + 1$ parameters. Is this solvable and is the solution unique? Is there an algorithm?

Instead the following asymptotic procedure is used:

$$\hat{X}_{m+1} = \sum_{\nu=1}^{m} \theta_{m\nu} \underbrace{\left(X_{m+1-\nu} - \hat{X}_{m+1-\nu} \right)}_{\approx \varepsilon_{m+1-\nu} \text{ if } m \text{ is large.}}$$

$$\approx \sum_{\nu=1}^{m} \theta_{m\nu} \varepsilon_{m+1-\nu} \overset{!}{=} \sum_{\nu=1}^{q} b_\nu \varepsilon_{m+1-\nu},$$

$v_m \approx \sigma_\varepsilon^2$ if m is large. In the innovation algorithm $\theta_{m\nu}$, v_m were calculated from γ_X's. $\hat{\theta}_{m,\nu}$ $\hat{v}_m$ from $\hat{\gamma}_X(.) \hookrightarrow \hat{b}_\nu = \hat{\theta}_{m,\nu}$.

Let us consider a causal linear process $X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$, $c_0 = 1$, $(c_j) \in \ell_1$, (ε_t) iid white noise, $E(\varepsilon_1) = 0$, $\text{Var}(\varepsilon_1) = \sigma_\varepsilon^2$.

THEOREM 3.2.3. *Assume that we have in addition $E(\varepsilon_1^4) < \infty$, then we have for any sequence of integers $m(n)$ with*

$$m(n) \uparrow \infty, \quad m(n)/n^{1/3} \to 0, \quad n \to \infty$$

and any integer $k \in \mathbb{N}$ that

$$\sqrt{n} \left(\hat{\theta}_{m(n),1} - c_1, \ldots, \hat{\theta}_{m(n),k} - c_k \right)^T \overset{d}{\to} \mathcal{N}_{0,A}$$

with $a_{ij} = \sum_{\nu=1}^{i \wedge j} c_{i-\nu} c_{j-\nu}, 1 \leq i, j \leq k$. and $\hat{v}_m \overset{p}{\to} \sigma_\varepsilon^2$, $n \to \infty$.

REMARK.

(1) In case of MA(q), choose $k = q$, $c_\nu = b_\nu$, $1 \leq \nu \leq k$,

$$\hat{\theta}_{m(n),\nu} \overset{P}{\to} b_\nu,$$

$1 \leq \nu \leq q$ but $\hat{\theta}_{l,\nu} \ncong b_\nu$ for some fixed l. If m is small we have a strong bias, if m is very large we get a large variance.

(2) Recipe: Start with the calculations of $\hat{\theta}_{m,\nu}$ $m = 1, 2, \ldots$ and stop after $\hat{\theta}_{m,\nu}$ does not change much when m is increased.

3.2.3. General ARMA(p,q) process.

$$X_t - \sum_{\nu=1}^{p} \alpha_\nu X_{t-\nu} = 1 + \sum$$

$X_t = \sum_{n=0}^{\infty} c_n \varepsilon_{t-n}$, $c_0 = 1$, c's are the Taylor coefficients of $\frac{\Psi(z)}{\Phi(z)}$ at $z_0 = 0$.

$$\sum_{\nu=0}^{\infty} c_\nu \varepsilon_{t-\nu} = X_t = \varepsilon_t + \sum_{\nu=1}^{q} \beta_\nu \varepsilon_{t-\nu} + \sum_{\mu=1}^{p} \alpha_p \sum_{j=0}^{\infty} c_j \varepsilon_{t-\mu-j}$$

$$= \varepsilon_t + \sum_{\nu=1}^{q} \beta_\nu \varepsilon_{t-\nu} + \sum_{\rho=1}^{\infty} \varepsilon_{t-\rho} \sum_{l=1}^{p \wedge \rho} \alpha_\nu \varepsilon_{\rho-\nu}.$$

Comparing coefficients (all coefficents are zero if indices are out of their domain)

$$\begin{pmatrix} c_{q+1} \\ c_{q+2} \\ \vdots \\ c_{q+p} \end{pmatrix} = \begin{pmatrix} c_q & c_{q-1} & \cdots & c_{q-p+1} \\ c_{q+1} & c_q & \cdots & c_{q-p+2} \\ \vdots & & \ddots & \vdots \\ c_{q+p-1} & & & c_q \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \vdots \\ \alpha_p \end{pmatrix}, \quad c_0 = 1$$

$$\begin{pmatrix} c_1 \\ \vdots \\ \vdots \\ c_q \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_q \end{pmatrix} + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ c_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ c_{q-1} & \cdots & c_{q-p+1} & c_{q-p} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \vdots \\ \alpha_p \end{pmatrix}$$

For the estimation of $\bar{\alpha}_p$ use the system

$$\begin{pmatrix} \hat{\theta}_{m,q+1} \\ \vdots \\ \hat{\theta}_{m,q+p} \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{m,q} & \hat{\theta}_{m,q-1} & \cdots & \hat{\theta}_{m,q-1+p} \\ \vdots & & & \\ \hat{\theta}_{m,q+p-1} & \cdots & \cdots & \hat{\theta}_{m,q} \end{pmatrix} \hat{\bar{\alpha}}_p$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \vdots \\ \hat{\beta}_q \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{m,1} \\ \vdots \\ \vdots \\ \hat{\theta}_{m,q} \end{pmatrix} - \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \hat{\theta}_{m,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ \hat{\theta}_{m,q-1} & \cdots & \hat{\theta}_{m,q-p+1} & \hat{\theta}_{m,q-p} \end{pmatrix}$$

with $\hat{\theta}_{m,r}$ is as above. Asymptotically $\begin{pmatrix} \vec{\alpha} \\ \vec{\beta} \end{pmatrix}$ is normal under some assumptions.

Are there MLE if $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ iid?

The Maximum Likelihood function

$$L_n\left(\vec{X}_n; \vec{\alpha}, \vec{\beta}, \sigma_\varepsilon^2\right) = \frac{1}{\left((2\pi)^n \det \Gamma_n\right)^{1/2}} \exp\left(-\frac{1}{2}\vec{X}_n^T \Gamma_n^{-1} \vec{X}_n\right)$$

$$\hat{X}_j = E\left(X_j \mid X_1, \ldots, X_{j-1}\right) = \sum_{\nu=1}^{j-1} \theta_{j-1,r}(X_{j-r}; \underbrace{\hat{X}_{j-r}}_{M}) \quad j \geq 1, \hat{X}_1 = 0$$

$$\hat{\vec{X}}_n = \left(\hat{X}_1, \ldots, \hat{X}_n\right)^T.$$

The calculation of Γ_n^{-1} can be avoided by

$$\tilde{\theta} = \begin{cases} \theta_{ij} & j = 1, \ldots, i, \ i = 1, 2, 3, \ldots \\ 1 & j = 0, \ i = 0, 1, 2, \ldots \\ 0 & j > i, \ i = 0, 1, 2, \ldots \end{cases}$$

$C = (\theta_{i,j})_{i,j=0}^{n-1}$ $D = \text{diag}(v_0, \ldots, v_{n-1})$

Then $\hat{X}_j = \sum_{\mu=1}^{j} \left(\tilde{\theta}_{j-1,j-\mu} - \delta_{i,\mu}\right)\left(X_\mu - \hat{X}_\mu\right)$

$\hookrightarrow \hat{\vec{X}} = (C - I)(\vec{X} - \hat{\vec{X}})$

$\hat{\vec{X}} = \vec{X} - \hat{\vec{X}} + \hat{\vec{X}} = C(\vec{X} - \hat{\vec{X}})$

By the definition of $v_{j-1} = \text{Var}(X_j - \hat{X}_j)$ and independence of the innovations, we find

$\text{Cov}(\vec{X} - \hat{\vec{X}}) = D$, $\Gamma_n = \text{Cov}(\vec{X}) = E(\vec{X}_n \vec{X}_n^T) = C\text{Cov}(\vec{X} - \hat{\vec{X}})C^T = CDC^T$.

$\det \Gamma_n = \det C \cdot \det D \cdot \det C^T = 1 \cdot \prod_{j=0}^{n-1} v_j \cdot 1 = \prod_{j=0}^{n-1} v_j$

$-\log L_n(\vec{X}_n; \vec{\alpha}, \vec{\beta}, \sigma_\varepsilon^2) = \frac{n}{2}\log 2\pi + \frac{1}{2}\log(v_0 \cdot \ldots \cdot v_{n-1}) + \frac{1}{2}\sum_{j=1}^n (X_j - \hat{X}_j)^2/v_{j-1}$

$r_j = \frac{v_j}{\sigma_\varepsilon^2} \Rightarrow -\log L_n(\vec{X}_n; \ldots) = \frac{n}{2}\log 2\pi + \frac{n}{2}\log \sigma_\varepsilon^2 + \frac{1}{2}\log(r_0 \cdot \ldots \cdot r_{n-1}) + \frac{1}{2\sigma_\varepsilon^2}\sum_{j=1}^n (X_j - \hat{X}_j)/r_{j-1}$ Minimize with respect to $\hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon^2$

$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \hat{X}_j)^2/r_{j-1}$. Equivalently we have to minimize wrt $\hat{\alpha}, \hat{\beta}$

$\ell_n(\vec{X}_n, \vec{\alpha}, \vec{\beta}) = \frac{n}{2}\log \hat{\sigma}_\varepsilon^2 + \frac{1}{2}\log(r_0 \cdot \ldots \cdot r_{n-1})$

Start with $\vec{\alpha}, \hat{\beta} \hookrightarrow$ calculate $\hat{X}_j$, $r_j \hookrightarrow \ell_n(\vec{X}_n, \vec{\alpha}, \vec{\beta})$. In case of an AR(p) process, it is simple:

$f_{X_1, \ldots, X_n} = f_{X_n|X_{n-1}, \ldots, X_{n-p}} \cdot \ldots \cdot f_{X_{p+1}|X_p, \ldots, X_1} \cdot f_{X_1, \ldots, X_p} = \prod_{j=1}^{n-p} f_{X_{n+1-j}|X_{n-j}, \ldots, X_{n-j-p+1}}$.

$f_{X_1, \ldots, X_p}$ and $f_{x_{n+1-j}|X_{n-j}, \ldots, X_{n-p-j+1}} = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left(X_j - \sum_{\nu=1}^p \alpha_\nu X_{j-\nu}\right)^2\right\}$

3.3. Model Choice and Model Checks

A choice of a suitable ARMA(p,q) model

Check ACF, PACF (Assume that we have a stationary model

(1) If ACF has a clear cutoff point at q, meaning $\rho(q) \neq 0$, $\rho(h) = 0$ for $h > q$ then choose a MA(q) model (Actual test on $\rho(q+1), \rho(q+2), \ldots = 0$).
(2) If PACF has a clear cutoff point at p, meaning $\phi(p) \neq 0$, $\phi(h) = 0$ for $h > p$ then choose an AR(p) model (Actual test on $\phi(p+1), \phi(p+2), \ldots = 0$).
(3) If neither the ACF nor the PACF have a clear cutoff point, think of an ARMA(p,q) model. But which p, q have to be chosen? If we fit a model, maximize the log-likelihood function (depending on p and q) is likely to have a very large order.

REMARK.

(1) If p and q are "large", then the model fits the data, but we have estimates of $\vec{\alpha}, \vec{\beta}$ with large "variance".
(2) If p and q are "small", then the model fit may not be great, but our estimates have less variance.

So look for a criterion to minimize wrt $\vec{\alpha}, \vec{\beta}$. For AR(p) we would like to choose p. Our first idea is to minimize $\hat{\sigma}_\varepsilon^2$ $(p = 1, 2, \ldots, P)$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{j=p+1}^n \left(X_j - \sum_{\nu=1}^p \hat{\alpha}_\nu X_{j-\nu}\right)^2,$$

where the $\hat{\alpha}_j$ are the ML-estimates. This quantity will be decreasing (in general) in p. So this quantity is not a suitable approximation of my total MSE.

(Akaike '69) Let (Y_t) be an independent copy of (X_t)

$$E\left((Y_t - \sum_{j=1}^p \hat{\alpha}_j Y_{t-j})^2 \mid X_1, \dots, X_n\right) = E\left[(Y_t - \sum_{j=1}^p \alpha_j Y_{t-j})^2\right]$$

$$\neq (\hat{\vec{\alpha}}_p - \vec{\alpha}_p)\Gamma_p(\hat{\vec{\alpha}}_p - \vec{\alpha}_p) + \underbrace{0}_{\text{mixed terms (YW-equations)}}$$

We can estimate this quantity by

$$(\circledast) = \frac{1}{n}\sum_{j=p+1}^n \left(X_j - \sum_{\nu=1}^p \alpha_\nu X_{j-\nu}\right)^2 + Q_1$$

$$= \frac{1}{n}\left(\sum_{j=p+1}^n X_j - \sum_{\nu=1}^p \hat{\alpha}_\nu X_{j-\nu}\right)^2 + Q_1 + (\hat{\vec{\alpha}} - \vec{\alpha})^T \hat{\Gamma}_p(\hat{\vec{\alpha}} - \vec{\alpha})$$

$\sqrt{n}(\hat{\vec{\alpha}} - \hat{\alpha}) \to \mathcal{N}(0, \sigma_\varepsilon^2 \Gamma_p^{-1})$, $\frac{nQ_1}{\sigma_\varepsilon^2} \to \chi_p^2$, $E(\frac{Q_1}{\sigma_\varepsilon^2} \cdot n) \to p$

Finite Predictor Eror $\text{FPE}(p) = \hat{\sigma}_\varepsilon^2(p)(1 + \frac{2p}{n}) \to$ minimize wrt p $\to \hat{p}$ (does not tend in general to p) (AIC - Akaike Information Criterion)

REMARK.

(1) $\log \text{FPE}(p) = -\frac{2}{n}\log L_n + \frac{p}{n} + C_n$

(2) For ARMA(p,q) model $\text{AIC}(p,q) = -\frac{2}{n}\log L_n + 2 \underbrace{\frac{p+q+1}{n}}_{\text{"punishment term"}} + C_n$

$\text{AIC}(p,q) = -\frac{2}{n}\log L_n(\vec{X}_n, \hat{\vec{\alpha}}, \hat{\vec{\beta}}, \hat{\sigma}_\varepsilon^2) + 2\frac{p+q+1}{n} \to \min_{p,q}$.

Further criteria:

$\text{AICC}(p,q) = -\frac{2}{n}\log L_n(\vec{X}_n, \hat{\vec{\alpha}}, \hat{\vec{\beta}}, \hat{\sigma}_\varepsilon^2) + 2\frac{p+q+1}{n-p-q-2} \to \min_{p,q}$.

$\text{BIC}(p,q) = -\frac{2}{n}\log L_n(\vec{X}_n, \hat{\vec{\alpha}}, \hat{\vec{\beta}}, \hat{\sigma}_\varepsilon^2) + 2\log n \cdot \frac{p+q+1}{n} \to \min_{p,q}$

After choosing a model, one should check whether it is appropriate. We may look at empirical residuals $\hat{\varepsilon}_t$ and check whether they have the white noise property.

How to get $\hat{\varepsilon}_t$?

$X_t - \hat{X}_t(\hat{\vec{\alpha}}, \hat{\vec{\beta}}) = \hat{\varepsilon}_t$ $t = 1, \dots, n$ (works for large t), where $\hat{X}_t$ is the best linear predictor for X_t.

$\hat{\varepsilon}_t = \hat{\Psi}^{-1}(B)\Phi(B)X_t$, where $\hat{\Psi}^{-1}$ has to be approxmiated. In an autoregressive model AR(p) $\hat{\varepsilon}_t = (X_t - \sum_{\nu=1}^p \hat{\alpha}_\nu X_{t-\nu})$, $t = p+1, \dots, n$

Portmanteau-Test:

$T_n = n \cdot \sum_{\nu=1}^m \hat{\rho}_{\hat{\varepsilon}}^2(\nu) \xrightarrow{d} \chi_{m-p-q}^2$, reject if the test statistic exceeds the $1-\alpha$ quantile.

(modified test statistic: $T_n^* = n(n+2)\sum_{\nu=1}^m \frac{\hat{\rho}_{\hat{\varepsilon}}^2(\nu)}{n-\nu} \xrightarrow{d} \chi_{m-p-q}^2$), the Portmanteau-Test is very conservative.

3.4. Some Remarks on the Estimation of the Spectral Density

ASSUMPTION. $(\gamma_X(h)) \in \ell_1 \hookrightarrow then\ f \in C[-\pi, \pi],\ symmetric$
$f(\lambda) = \sum_{k \in \mathbb{Z}} \gamma_X(k) e^{-ik\lambda},\ \lambda \in [0, \pi].$

Idea: Get an estimate by $\gamma_X(.) || \hat{\gamma}_X(.)$ (based on the sample $X_1, \ldots, X_n$)
This leads to the so-called centered Periodogram:

$$I_{n,\bar{X}}(\lambda) = \sum_{|k|<n} \hat{\gamma}_X(k) e^{-ik\lambda}$$

Note that the frequencies $\lambda_j = \frac{2\pi j}{n}$, $j = -\left[\frac{n-1}{2}\right], \ldots, -1, 0, 1, \ldots, \left[\frac{n}{2}\right]$ are special,
the vectors $\vec{b}_j = \frac{1}{\sqrt{n}} \left(e^{-i\lambda_j 1}, e^{-i\lambda_j 2}, \ldots, e^{-i\lambda_j n} \right) \in \mathbb{C}^n$, $j = -\left[\frac{n-1}{2}\right], \ldots, -1, 0, 1, \ldots, \left[\frac{n}{2}\right]$,
form an orthonormal basis on $\mathbb{C}^n$.
$X_t = \frac{1}{\sqrt{n}} \sum_{j=-\left[\frac{n-1}{2}\right]}^{\left[\frac{n}{2}\right]} c_j e^{it\lambda_j}$, $t = 1, 2, \ldots, n$, $c_j = \frac{1}{\sqrt{n}} \sum_{\nu=1}^{n} X_\nu e^{-i\nu\lambda_j}$ (FFT)
A simple calculation shows:

$$I_{n,\bar{X}}(\lambda_j) = \left| \frac{1}{\sqrt{n}} \sum_{\nu=1}^{n} \left(X_\nu - \bar{X} \right) e^{-i\lambda_j \nu} \right|^2$$

$$\overset{j \neq 0}{=} \left| \frac{1}{\sqrt{n}} \sum_{\nu=1}^{n} X_\nu e^{-i\lambda_j \nu} \right|^2 = I_n(\lambda_j)$$

Now it can be shown:

$$E(I_n(\lambda_j)) = \begin{cases} n\mu^2 + 2\pi f(0) + o(1) & j = 0 \\ 2\pi f(\lambda_j) + o(1), & j \neq 0 \end{cases}, \quad n \to \infty$$

So $\frac{1}{2\pi} I_n(\lambda)$ is asymptotically unbiased, however, if $X_t = \varepsilon_t, \varepsilon_t$ an iid white noise
with $E\left(\varepsilon_1^4\right) = \mu_4$, $E(\varepsilon_1^3) = 0$, then

$$\mathrm{Var}(I_n(\lambda_j)) = \begin{cases} 2\sigma_\varepsilon^4 + \frac{1}{n}(\mu_4 - 3\sigma^4) & \lambda_j = 0 \text{ or } \pm\pi \\ \sigma_\varepsilon^4 + \frac{1}{n}(\mu_4 - 3\sigma^4) & \text{otherwise} \end{cases}$$

Even in this simple case $(X_t = \varepsilon_t, f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi})$ we do not have a consistent estimator.
Way out: Tapering: use a weight sequence $w_{k,n}$ weighting the $\hat{\gamma}_X(k)$ down for large
k.

$$w_{k,n} = \left(1 - \frac{|k|}{m_n} \right) 1_{[-m_n, m_n]}(k)$$

$$\hat{f}_n(\lambda_j) = \sum_{|k|<n} w_{k,n} \hat{\gamma}_X(k) e^{-ik\lambda_j}, \quad j = \ldots$$

asymptotically consistent if e.g. $m_n \to \infty$, $m_n/n \to 0$, "smoothing parameter".

CHAPTER 4

ARCH- and GARCH-Processes

4.1. ARCH- and GARCH-Processes

Process W_t price of a stock, stock-index of a currency rate (here US \$/UK pound)

log returns: $X_t = \log \frac{W_t}{W_{t-1}} = \log \left(1 + \frac{W_t - W_{t-1}}{W_{t-1}}\right) \approx \frac{W_t - W_{t-1}}{W_{t-1}}$

Looking at data sets, we find that the ACF looks like one for iid random variables. But the ACF's for $\left(X_t^2\right)$ of $(|X_t|)$ do not look like they should for iid random variables. $\Rightarrow$ ARMA models are not suitable. We have to look for a new class of models, which should allow the following patterns we observed in the data.

(1) The data themselves show practically no correlation, but the modulus and the squares show significant correlation.
(2) The variance (the volatility) shows periods of larger and smaller values.
(3) The data look heavy tailed (look at QQ-plots comparing with exponential distribution)
(4) The time points where data are above a high threshold come in clusters.

ACF of X_t^2 looks like an AR(p) process that led Engle '82 (Nobel price laureate) to introduce the following model (most simple case)

$X_t = \sigma_t \varepsilon_t$ with a white noise iid random variables (ε_t), $E(\varepsilon_1) = 0$, $E(\varepsilon_1^2) = 1$ (often even $\mathcal{N}_{0,1}$).

$\sigma_t^2 = \beta + \lambda X_{t-1}^2$, $\beta, \lambda > 0$.

Generalizing this idea we come to

DEFINITION. (GARCH(p,q), ARCH(p) time series)

Let be given $p, q \geq 0$. A real time series (X_t) being defined by $X_t = \sigma_t \varepsilon_t$ with iid random variables (ε_t) having moments $E(\varepsilon_1) = 0$ and $E(\varepsilon_1^2) = 1$ and a stochastic process $\left(\sigma_t^2\right)$ satisfying with coefficients $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots \beta_q \geq 0$, s.t. $\alpha_p \cdot \beta_q > 0$ the difference equation

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{\nu=1}^{q} \beta_\nu \sigma_{t-\nu}^2$$

is called an GARCH(p,q) process (Generalized Auto-Regressive Conditionally Heteroskedastic). $\left(\sigma_t^2\right)$ is called the volatility process. In case $q = 0$ (X_t) is called an ARCH(p) process.

ASSUMPTION. *We denote by $\mathcal{F}_t = \sigma\left(\varepsilon_s, s \leq t\right)$, $t \in \mathbb{Z}$ and we will assume from now on that σ_t^2 is $\mathcal{F}_{t-1}$ is measurable for all t.*

REMARK. We will be interested in processes (σ_t^2) being strongly stationary (i.e. the entire distribution function is time-shift independent) and the -as we will see- our assumption can be verified. The also the proces (X_t) will be strongly stationary

$$F_{\sigma_t, \varepsilon_t}(x) = \int F_{\sigma_t}(x \cdot \nu) dF_{\varepsilon_t}(\nu) = F_{\sigma_{t+h}, \varepsilon_{t+h}}(x) = \int F_{\sigma_{t+h}}(x \cdot \nu) dF_{\varepsilon_{t+h}}(\nu)$$

4.2. Some Properties of the GARCH-Processes

Let us consider the process (X_t^2)

$$
\begin{aligned}
X_t^2 \quad &= \quad \sigma_t^2 \varepsilon_t^2 = \sigma_t^2 + \underbrace{\sigma_t^2(\varepsilon_t^2 - 1)}_{=:\eta_t = X_t^2 - \sigma_t^2} \\
&= \quad \alpha_0 + \sum_{\nu=1}^{p} \alpha_\nu X_{t-\nu}^2 + \sum_{\mu=1}^{q} \beta_\mu \sigma_{t-\mu}^2 + \eta_t \\
\overset{\sigma_t^2 = X_t^2 - \eta_t}{=} \quad &\alpha_0 + \sum_{\nu=1}^{p \vee q} (\alpha_\nu + \beta_\nu) X_{t-\nu}^2 - \sum_{\mu=1}^{q} \beta_\mu \eta_{t-\mu} + \eta_t
\end{aligned}
$$

Putting $\tilde{p} = p \vee q$ then X_t^2 follows an $\text{ARMA}(\tilde{p}, q)$ process with mean, provided the (η_t) are an uncorrelated white noise.

$$E(\eta_t) = E(E(\sigma_t^2(\varepsilon_t^2 - 1) \mid \mathcal{F}_{t-1}) = E(\sigma_t^2 E(\varepsilon_t^2 - 1 \mid \mathcal{F}_{t-1})) = 0 \quad \forall t$$

For $s \leq t$

$$
\begin{aligned}
E(\eta_s \eta_t) &= E(E(\sigma_s^2(\varepsilon_s^2 - 1)\sigma_t^2(\varepsilon_t^2 - 1) \mid \mathcal{F}_{t-1})) \\
&= \begin{cases} 0 & s < t \\ E(\sigma_t^4 E((\varepsilon_t^2 - 1)^2) = E(\sigma_t^4) E((\varepsilon_t^2 - 1)^2) & s = t \end{cases}
\end{aligned}
$$

If $E(\sigma_t^4) < \infty$, $E(\varepsilon_t^4) < \infty$, then (η_t) form a white noise, note that if (σ_t^2) is strongly stationary, $E(\eta_t^2)$ is constant.

Next: $E(X_t) = E(\sigma_t \varepsilon_t) = 0$, $\forall t$.

For $s \leq t$ $E(X_s \cdot X_t) = E(\sigma_s \varepsilon_s \sigma_t \varepsilon_t) = \begin{cases} 0 & s < t \\ E(\sigma_t^2) \cdot 1 & s = t \end{cases}$. If $E(\sigma_t^2) < \infty$, then (X_t) is uncorrelated and weakly stationary if $E(\sigma_t^2) \equiv \text{const.}$

Important: To have a parameter constellation such that (σ_t^2) is strictly stationary and has some moments.

For simplicity we restrict ourselves to GARCH(1,1) processes for a moment.

(1) $$X_t = \sigma_t \cdot \varepsilon_t$$

(2) $$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

THEOREM 4.2.1. *Then both* (1) *and* (2) *have a strongly stationary solution iff*

(*) $$E(\log(\alpha_1 \varepsilon_1^2 + \beta_1)) < 0.$$

Then

$$(+) \qquad \sigma_t^2 = \alpha_0 \left(1 + \sum_{j=1}^{\infty} \prod_{\nu=1}^{j} (\alpha_1 \varepsilon_{t-\nu}^2 + \beta_1) \right),$$

hence σ_t^2 is $\mathcal{F}_{t-1}$ measurable and the only strongly stationary solution of (2). *Furthermore, if $\alpha_1 + \beta_1 < 1$ then* (*) *holds and*

$$E(X_t^2) = E(\sigma_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} < \infty.$$

REMARK.

(1) (X_t) will have much less existing moments than (ε_t).
(2) Iteration of (2) leads to

$$\sigma_t^2 = \alpha_0 \left(1 + \sum_{\nu=1}^{k} \prod_{\mu=1}^{\nu} (\alpha_1 \varepsilon_{t-\mu} + \beta_1) \right) + \prod_{\nu=1}^{k+1} (\alpha_1 \varepsilon_{t-\nu}^2 + \beta_1) \sigma_{t-k-1}^2.$$

(3) (Repetition) $D \ x_n = \alpha_0 + \alpha_1 x_{n-1} \ D_h : x_n = \alpha_1 x_{n-1} \ n \in \mathbb{N}$
D_h has set of solutions $x = c\alpha_1^n, \ c \in \mathbb{R}$.
D has set of solutions $x_n = x_n^{(0)} + c\alpha_1^n \ c \in \mathbb{R} \ x_n^{(0)}$ some solution of D_h.
If $\alpha_1 \neq 1$ then $x_n^{(0)} = \frac{\alpha_0}{1-\alpha_1}$ is the only stationary solution.
Stability: If $|\alpha_1| < 1$ then then $x_n \overset{n \to \infty}{\to} \frac{\alpha_0}{1-\alpha_1}$.
For us:
$(D_{G,h}) \ Y_t = (\alpha_1 \varepsilon_{t-1}^2 + \beta_1) Y_{t-1} \ (D_G) \ \sigma_t^2 = \alpha_0 + (\alpha_1 \varepsilon_{t-1}^2 + \beta_1) \sigma_{t-1}^2$
$Y_t = \prod_{j=1}^{t} (\alpha_1 \varepsilon_{t-j}^2 + \beta_1) Y_0 \overset{a.s.}{\underset{(*)}{\to}} 0$ "Perpetuties".

PROOF. Assuming that (+) holds and $\alpha_1 + \beta_1 < 1$ then

$$E(\log(\alpha_1 \varepsilon_1^2 + \beta_1)) \overset{\text{Jensen}}{\leq} \log E(\alpha_1 \varepsilon_1^2 + \beta_1) = \log(\alpha_1 + \beta_1) < 0$$

which implies (*).

$$E(X_t^2) = E(\sigma_t^2) \overset{(+)}{\underset{\text{Tonelli}}{=}} \alpha_0 \left(1 + \sum_{\nu=1}^{\infty} \prod_{\mu=1}^{\nu} \underbrace{E(\alpha_1 \varepsilon_1^2 + \beta_1)}_{=\alpha_1+\beta_1} \right)$$

$$= \alpha_0 \left(\sum_{\nu=0}^{\infty} (\alpha_1 + \beta_1)^{\nu} \right) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}.$$

There exists some $\rho > 1$ such that $\log \rho + E(\log(\alpha_1 \varepsilon_1^2 + \beta_1)) < 0$

$$\frac{1}{n} \sum_{\nu=0}^{n-1} \underbrace{(\log \rho + \log(\alpha_1 \varepsilon_\nu^2 + \beta_1))}_{iid} \overset{a.s.}{\underset{\text{SLLN}}{\to}} \log \rho + E(\log(\alpha_1 \varepsilon_1^2 + \beta_1)) < 0.$$

i.e. j=n-v

$$\frac{1}{n}\sum_{j-1}^{n}(\log\rho + \log(\alpha_1\varepsilon_{n-j}^2 + \beta)) \underset{\text{SLLN}}{\overset{a.s.}{\to}} \log\rho + E(\log(\alpha_1\varepsilon_1^2 + \beta_1)) < 0.$$

$$\log\left(\rho^n\prod_{j=1}^{n}(\alpha_1\varepsilon_{n-j}^2 + \beta_1)\right) \overset{a.s.}{\to} -\infty$$

i.e. $\rho^t\prod_{j=1}^{t}(\alpha_1\varepsilon_{t-j}^2 + \beta_1) \overset{a.s.}{\to} 0, t \to \infty$ and this implies that $\prod_{j=1}^{t}(\alpha_1\varepsilon_{t-j}^2 + \beta_1) \leq c_\omega\rho^{-t}$ so the product converges to 0 exponentially fast. Hence $Y_t \overset{a.s.}{\to} 0$ $t \to \infty$ for any Y_0 and the formula (+) makes sense. also solves (2) which can be checked by plugging it in. It is a strongly stationary solution. The reason is that the ε_t are iid and the distribution is not changed by changing their index.

Assume that σ_t^2 and $\tilde{\sigma}_t^2$ are two strongly stationary solutions. Then $\sigma_t^2 - \tilde{\sigma}_t^2$ statisfy $D_{G,h}$, hence $\overset{a.s.}{\to} 0$ $t \to \infty$ $\sigma_t^2 - \tilde{\sigma}_t^2$ are strictly stationary i.e. $\sigma_t^2 - \tilde{\sigma}_t^2 = 0$ a.s.

Now assume that (*) is violated. Remark (2) shows that

$$\sigma_t^2 \geq \alpha_0\left(1 + \sum_{j=1}^{k}\prod_{\nu=1}^{j}(\alpha_1\varepsilon_{t-\nu}^2 + \beta_1)\right)$$

If $E(\log(\alpha_1\varepsilon_1^2 + \beta_1)) > 0$ then the arguments from above show that the r.h.s. $\overset{a.s.}{\to} \infty$. A more detailed analysis with random walk theory the l.h.s. has $\limsup_{t\to\infty}(.) = \infty$. $\qquad\square$

4.2.1. Two special processes.

4.2.1.1. ARCH(1)

. $X_t = \sigma_t\varepsilon_t$ $\varepsilon_t \sim \mathcal{N}_{0,1}$ (1)

$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$ (2)

and equation (*) reads now

$$(\log(\alpha_1\varepsilon_1^2)) = \log\alpha_1 + \frac{4}{\sqrt{2\pi}}\int_0^\infty \log t\, e^{\frac{-t^2}{2}}\,dt$$

$$= \log\alpha_1 - (c + \log 2),$$

where $c = 0.5772...$ the Euler Mascharorie Constant. $(*) < 0 \iff 0 < \alpha_1 < 2e^C$ Then we get strongly stationary solutiosn of (1) and (2) with $E(X_t^2) < \infty$, $E(\sigma_t^2) < \infty$.

What else can we say about the distribution function of X_0?

$h(u, \alpha_1) = E((\alpha_1\varepsilon_1^2)^u) = \alpha_1^u E(e^{uW}), W = \log\varepsilon_1^2$.

$h(0, \alpha_1) = 1$, h is convex, $h'(0, \alpha_1) < 0$ if $\alpha_1 \in (0, 2e^C)$

$h(u, \alpha_1) \to \infty$ if $\alpha_1 > 0$.

$$\kappa(\alpha_1) = \begin{cases} > 1 & \alpha_1 \in (0, 1) \\ = 1 & \alpha_1 = 1 \\ < 1 & \alpha_1 \in (1, 2e^C) \end{cases}$$

$P(X_0 > x) \sim dx^{-2\kappa(\alpha_1)}$, $x \to \infty$ with some $d > 0$. $E(X_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2$.