MS ARMA Processes

Collection of results on
scalar ARMA processes

CONTENTS

0 Introduction

(rd. MS Asymptotics for stoch. processes first)

> An ARMA(p,q)-process ("AutoRegressive Moving Average") is a stochastic process that satisfies a law of motion (LoM) of the form

$$X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

> Such processes (although there is an (important!) distinction btwn LoM and process, we will use the label "ARMA-process" to refer both to the LoM and to the actual process that solves it, for convenience — more below) are the generalization of AR & MA-processes and quite general; they appear everywhere in theoretical modelling, but their popularity in empirical work is mainly due to 2 reasons

a) they're fairly easy to estimate

b) any covariance-stationary process can be represented as MA(q) ($q \in \mathbb{N}_0 \cup \{+\infty\}$) and thus (provided stability, see below) estimated as AR!

> ARMA-processes are tractable & versatile!

1 Elementary definitions

> Let's begin with some elementary definitions.

DEF 1.1 (Variants of white noise). Consider a process (X_t). We call (X_t)...

... white noise, $(X_t) \sim WN(0,\sigma^2)$, if $(X_t) \in \mathcal{L}^2(\mathbb{P})$ with $\mathbb{E}X_t = 0$, $\mathbb{E}X_t^2 = \sigma^2$, $\mathbb{E}X_t X_{t-h} = 0 \ \forall h \in \mathbb{Z}\setminus\{0\}$

... strong / independent white noise, if $(X_t) \sim WN(0,\sigma^2) \wedge \{X_t\}_t \perp\!\!\!\perp$

... iid white noise, if $(X_t) \sim WN(0,\sigma^2) \wedge \mathcal{L}((X_t)) = P_X^{\otimes\infty}$

... Gaussian white noise, if $(X_t) \sim WN(0,\sigma^2) \wedge X_t \sim \mathcal{N}(0,\sigma^2) \ \forall t$
(note: since this doesn't imply that (X_t) are jointly Gaussian, this does not yield independence of (X_t) !!)

... Gaussian strong white noise, if $(X_t) \sim \mathcal{N}(0,\sigma^2)^{\otimes\infty}$.

> white noise is the basic building block of ARMA-processes as we will shortly see; before, we introduce an instrument of invaluable practicability in ARMA-analysis (& beyond): the lag operator

The L-operator & its algebra

Fix $(\Omega, \mathcal{A}, \mathbb{P})$. The set of stochastic processes on $((\Omega,\mathcal{A},\mathbb{P}), \mathbb{Z})$ with point-

wise multiplication by real scalars and pointwise addition of processes (remember: these are mappings $\Omega \times \mathbb{Z} \to \mathbb{R}$!) constitutes a real vector space.
[This isn't important in itself but nice to know]

On this vector space, consider the mapping

$$\boxed{\mathbb{L} : (X_t) \mapsto (\mathbb{L}X_t) \equiv (X_{t-1}).}$$

This mapping we call "lag operator". We define several properties of this new operator:

(i) <u>inverse / negative power</u>: $\mathbb{L}^{-1} : X_t \mapsto X_{t+1}$ (depending on context, we sometimes rather define $\mathbb{L}^{-1}X_t \equiv \mathbb{E}(X_{t+1} | \mathcal{F}_t)$ where (X_t) is adapted to $(\mathcal{F}_t)$)

(ii) <u>stacking / powers</u>:

$$\mathbb{L}^n X_t := \underbrace{\mathbb{L}\mathbb{L}\dots\mathbb{L}}_{n \text{ times}} X_t = X_{t-n}$$

$$\mathbb{L}^{-n} X_t := \underbrace{\mathbb{L}^{-1}\mathbb{L}^{-1}\dots\mathbb{L}^{-1}}_{n \text{ times}} X_t$$

$\Longrightarrow \mathbb{L}^i \mathbb{L}^j = \mathbb{L}^{i+j} \quad \forall i,j \in \mathbb{Z}$

$\Longrightarrow \mathbb{L}^0 X_t \equiv X_t$ (an $\mathbb{L}$-identity exists)

(iii) <u>distributivity</u>:

$$(\mathbb{L}^i + \mathbb{L}^j) X_t := X_{t-i} + X_{t-j}$$

$$\mathbb{L}(X_t + Y_t) := X_{t-1} + Y_{t-1} .$$

From these properties, it can be established that the set of <u>lag-operator-polynomials</u> (or "filters"),

$$\{ \textstyle\sum_{i=0}^n c_i \mathbb{L}^i \mid n \in \mathbb{N}_0, c_i \in \mathbb{R} \ \forall i \}$$

is well-defined; and when endowed with the operation of multiplication...

$$\cdot : \left(\textstyle\sum_{i=0}^{n_1} a_i \mathbb{L}^i , \sum_{i=0}^{n_2} b_i \mathbb{L}^i \right) \longmapsto \sum_{i=0}^{n_1 \vee n_2} c_i \mathbb{L}^i$$

$$\text{with} \quad c_i \equiv \textstyle\sum_{j=0}^{i \wedge n_1 \wedge n_2} a_j b_{i-j} \quad \forall i \in \{0, \dots, n_1 \vee n_2\}$$

i.e. the set is an $\mathbb{R}$-vector-space endowed with a bilinear multiplication of vectors, namely the just-defined

i.e. there exist an Isomorphism (bijective linear map) btwn $\mathbb{L}$-polyn.s and $\mathbb{C}$-polyn.s

... it forms an <u>algebra over $\mathbb{R}$</u> that is <u>isomorphic</u> to the algebra of $\mathbb{C} \to \mathbb{C}$-polynomials over $\mathbb{R}$!

this isomorphism maps $\sum_{i=0}^n c_i \mathbb{L}^i \mapsto \sum_{i=0}^n c_i z^i$

Loosely speaking, any operation applied to some lag-polynomial $\Phi(\mathbb{L}) \equiv \sum_{i=0}^n c_i \mathbb{L}^i$, can equivalently (i.e. yielding the same result, in terms of coefficients) be applied to $\sum_{i=0}^n c_i z^i$, the corresponding $\mathbb{C} \to \mathbb{C}$ polynomial;

The reason this is so important is: any $\mathbb{C} \to \mathbb{C}$ polynomial can be written in factorized form (Fundamental theorem of algebra!)

$\Longrightarrow$ For any lag-polynomial,

$$\boxed{\Phi(\mathbb{L}) \equiv \textstyle\sum_{i=0}^n c_i \mathbb{L}^i \equiv \prod_{j=1}^{\overset{\equiv n}{r}} (1 - \lambda_j \mathbb{L})^{m_j}}$$

more below in solution-sect.

<u>where $\{\lambda_j^{-1}\}$ are the roots of</u>

<u>$\mathbb{C} \to \mathbb{C} : z \mapsto \sum_{i=0}^n c_i z^i$ with multipl.s m_j</u>

It is for this reason that we pose the question of invertibility of $\Phi(\mathbb{L})$ only for $n=1$.

Now, given $\Phi(\mathbb{L}) = (1-\rho\mathbb{L})$, how can we find it's inverse, ie. find a $\mathbb{L}$-polyn. $\Phi(\mathbb{L})^{-1}$ s.th.

$$\Phi(\mathbb{L})\Phi(\mathbb{L})^{-1} = \mathbb{L}^0 \quad ?$$

To answer this, we concentrate on the subspace of weakly stationary processes. On this subspace we may show:

in the $\mathcal{L}^2$-sense

(a) $\boxed{\rho \in (-1,1): \quad (1-\rho\mathbb{L})^{-1} = \sum_{i=0}^{+\infty} \rho^i \mathbb{L}^i}$ $\left(\overset{\text{def}}{:} (X_t) \mapsto \left(\lim_{n\to\infty} \sum_{i=0}^{n} \rho^i \mathbb{L}^i X_t \right) \right)$

Proof: We first show that $(1-\rho\mathbb{L})^{-1}$ defined as such actually exists (ie. the series converges in an appropr. sense). To this end, recall that we apply this only to w. stat. processes. Thus, for $(X_t) \in \mathcal{L}^2(\mathbb{P})$ we want to show that

$$\lim_{n\to\infty} \sum_{i=0}^{n} \rho^i X_{t-i} \quad \text{exists as the } \mathcal{L}^2\text{-limit of } \left(\sum_{i=0}^{n} \rho^i X_{t-i}\right)_n .$$

We can show this by the Cauchy-criterion. Fix $t\in\mathbb{Z}$ and define

$$y_n := \sum_{i=0}^{n} \rho^i X_{t-i}, \text{ and consider for } m>n,$$

$$\begin{aligned}
\mathbb{E}\left[|y_m - y_n|^2\right]^{1/2} &= \mathbb{E}\left[\left|\sum_{i=0}^{m} \rho^i X_{t-i} - \sum_{i=0}^{n} \rho^i X_{t-i}\right|^2\right]^{1/2} \\
&= \mathbb{E}\left[\left|\sum_{i=n+1}^{m} \rho^i X_{t-i}\right|^2\right]^{1/2} \\
&\underset{\Delta\text{-inequ. for } \|\cdot\|_{\mathcal{L}^2}}{\leq} \sum_{i=n+1}^{m} |\rho|^i \underbrace{\mathbb{E}\left[|X_{t-i}|^2\right]^{1/2}}_{=\|X_t\|_{\mathcal{L}^2} < +\infty \text{ by } (X_t) \text{ w. stat.y}} \\
&\leq \|X_t\|_{\mathcal{L}^2} \cdot \underbrace{\sum_{i=n+1}^{\infty} |\rho|^i}_{<+\infty \text{ by } |\rho|<1} \\
&\to 0 \text{ as } n\to\infty \text{ by Cauchy-criterion for series.}
\end{aligned}$$

Hence (y_n) is Cauchy in $\mathcal{L}^2$, hence $\exists y \in \mathcal{L}^2(\mathbb{P}): y_n \xrightarrow{\mathcal{L}^2} y$ as $n\to\infty$.

Thus $\left(\sum_{i\geq 0} \rho^i \mathbb{L}^i\right) X_t$ actually exists for (X_t) stationary. Now we may use that for an $\mathcal{L}^2$-convergent series, every rearrangement converges (with the limit unaltered) to show for (X_t) stationary:

$$(1-\rho\mathbb{L}) \cdot \left(\sum_{i\geq 0} \rho^i \mathbb{L}^i\right) X_t = X_t - \rho X_{t-1} + \rho X_{t-1} - \rho^2 X_{t-2} + \rho^2 X_{t-2} - \ldots \overset{\text{by rearrangement!}}{=} X_t$$

that is: $(1-\rho\mathbb{L}) \cdot \left(\sum_{i\geq 0} \rho^i \mathbb{L}^i\right) = \mathbb{L}^0$. $\square$

(b) $\boxed{\rho \in (-1,1)^C: \; (1-\rho\mathbb{L})^{-1} = -(\rho\mathbb{L})^{-1} \sum_{i\geq 0} \rho^{-i} \mathbb{L}^{-i}}$ (again: $\lim_{n\to\infty} \sum_{i=} \ldots$)

Proof: every step up to the rearrangement is identical to the above. Hence just compute,

$$(1-\rho\mathbb{L})\left(1+(\rho\mathbb{L})^{-1}+(\rho\mathbb{L})^{-2}+\dots\right)=\cancel{1}-\rho\mathbb{L}+\cancel{(\rho\mathbb{L})^{-1}}-\cancel{1}+(\rho\mathbb{L})^{-2}-\cancel{(\rho\mathbb{L})^{-1}}+(\rho\mathbb{L})^{-3}+\dots$$
$$=-\rho\mathbb{L}\ .\ \square$$

> Notice that (a) & (b) actually correspond to backward- & forward-iteration respectively!
Also, employing these definitions, we will always obtain a stationary process by inverting $(1-\rho\mathbb{L})$! Sometimes we're interested in explosive processes, though; we will however see in the next section that we can solve expl. LoM for the explosive solution by choosing appropriate pass-through-conditions (& then choosing the coeff.s of the homog. solution appr.ly)

> Now let's revisit $\mathbb{L}$-pol.s of higher order; how to invert $\Phi(\mathbb{L})\equiv\sum_{i=0}^{n}c_i\mathbb{L}^i$?

THM 1.2 (Lag-polynomial inversion). Consider some lag polynomial
$$\Phi(\mathbb{L})\equiv\sum_{i=0}^{n}c_i\mathbb{L}^i\ .$$
If no root of $\mathbb{C}\to\mathbb{C}: z\mapsto\Phi(z)$ is one, $\Phi(\mathbb{L})$ has an inverse and it holds that
$$\Phi(\mathbb{L})^{-1}\equiv\prod_{j=1}^{r}(1-\lambda_j\mathbb{L})^{-m_j}$$
with $(1-\lambda_j\mathbb{L})^{-1}$ defined as in (a) $\vee$ (b) depending on whether $|\lambda_i|\gtrless 1$.
In particular, if $|\lambda_i|<1\ \forall i$, this inverse takes the form of a power series in $\mathbb{L}$:
$$\forall i,\ |\lambda_i|<1\ \Rightarrow\ \Phi(\mathbb{L})^{-1}=\sum_{i\geq 0}\psi_i\mathbb{L}^i,$$
the coeff.s ψ_i can be found by noticing that $\Phi(\mathbb{L})\Phi(\mathbb{L})^{-1}\equiv 1$ induces the recursion
$$\begin{cases}\psi_0c_0=1,\ \psi_0c_1+c_0\psi_1=0,\ \psi_0c_2+\psi_1c_1+\psi_2c_0=0,\dots,\sum_{j=0}^{n}\psi_jc_{n-j},\sum_{j=0}^{n}\psi_{j+1}c_{n-j},\dots,\\ \sum_{j=0}^{n}\psi_{j+k}c_{n-j},\dots\end{cases}$$
which can be solved recursively. Notice in particular that if $|\lambda_i|<1\ \forall i$ we have that the coeff.s of $\Phi(\mathbb{L})^{-1}$ are absolutely summable:
$$\sum_{i\geq 0}|\psi_i|<+\infty.$$

<u>Proof</u> (sketch): The first claim is immediate from the fact that $\Phi(\mathbb{L})$ may be factorized, and that for each factor an inverse exists [Just write $\Phi(\mathbb{L})$ in factorized form and apply $(1-\lambda_i\mathbb{L})^{-1}$ one at a time.]
For the rest of the proof assume $|\lambda_i|<1\ \forall i$. The claim
$$\Phi(\mathbb{L})^{-1}\equiv\sum_{i\geq 0}\psi_i\mathbb{L}^i$$ follows again from the fact that $\mathbb{L}$-polynomials behave like $\mathbb{C}\to\mathbb{C}$-polynomials and the fact that the product of two $\mathbb{C}\to\mathbb{C}$ power series is again a power series [on the intersection of the convergence radii.] The recursion may be established using that $a\cdot\sum_{i\geq 0}b_i=\sum_i a\cdot b_i$, $\sum_{i\geq 0}a_i+\sum_{i\geq 0}b_i=\sum_{i\geq 0}(a_i+b_i)$ for convergent series.
Finally, consider the series $\sum_{i\geq 0}|\psi_i|$. By induction and since $\sum_{i\geq 0}\psi_i \overset{\text{Cauchy product (E700, I.12.4)}}{=} \prod_i\left(\sum_k\lambda_i^k\right)$

it suffices to show that $\sum_{i\geq 0}|\delta_i|<+\infty$ for $\sum_i|\delta_i| = \lim_n \sum_{i=0}^n |\delta_i|$ where $\delta_i := \underbrace{\sum_{k=0}^i \alpha_k \beta_{i-k}}_{\text{Cauchy-product!}}$, if we have $\sum_{i\geq 0}|\alpha_i|, \sum_{i\geq 0}|\beta_i| < +\infty$.

This may be shown as follows: for any $n\in\mathbb{N}$,

$$\sum_{i=0}^n|\delta_i| = \sum_{i=0}^n\left|\sum_{k=0}^i \alpha_k\beta_{i-k}\right| \underset{\Delta\text{-ineq.}}{\leq} \sum_{i=0}^n\sum_{k=0}^i |\alpha_k||\beta_{i-k}|.$$

Hence,

$$\lim_{n\to\infty}\sum_{i=0}^n|\delta_i| \leq \lim_{n\to\infty}\sum_{i=0}^n\sum_{k=0}^i|\alpha_k|\cdot|\beta_{i-k}| \overset{(*)}{=} \left(\sum_{i\geq 0}|\alpha_i|\right)\cdot\left(\sum_{i\geq 0}|\beta_i|\right) \overset{(**)}{<} +\infty$$

where equ. $(*)$ follows from the definition of the Cauchy-product applied to the series $(\sum|\alpha_i|)$ and $(\sum|\beta_i|)$, which exists (hence $(**)$). That the Cauchy-product exists may be verified in Rudin, p. 72-75. □

2 AR, MA, ARMA -equations & -processes

DEF 2.1 (AR, MA, ARMA). Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim WN(0,\sigma^2)$. We call the following equations ...

- AR(p)-LoM: $X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t$

  $\Leftrightarrow \phi(\mathbb{L})X_t = c + \varepsilon_t$, with $\phi(\mathbb{L}) = 1 - \phi_1\mathbb{L} - \ldots - \phi_p\mathbb{L}^p$

- MA(q)-LoM: $X_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}$

  $\Leftrightarrow X_t = c + \theta(\mathbb{L})\varepsilon_t$, with $\theta(\mathbb{L}) = 1 + \theta_1\mathbb{L} + \ldots + \theta_q\mathbb{L}^q$

- ARMA(p,q)-LoM: $X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}$

  $\Leftrightarrow \phi(\mathbb{L})X_t = c + \theta(\mathbb{L})\varepsilon_t$.

We call a process (X_t) an AR(p)-process (resp. MA(q), ARMA(p,q)) or a solution to AR(p) (resp....) if it satisfies the LoM.
Conversely, given any process (X_t), we call any LoM that is satisfied by (X_t) [one process can satisfy multiple different LoM! More later] a 'representation of (X_t)'.

> It is worth noticing that any process satisfying a MA(q)-LoM is already in solution-form, that is, it is represented in a non-recursive manner:

'solution to ARMA $(\underbrace{p}_{>0},\underbrace{q}_{\geq 0})$-LoM' $\overset{\text{def}}{\Longleftrightarrow}$ non-recursive representation of a process that satisfies the given LoM.

> It is important to keep in mind the conceptual difference btwn process/solution and LoM — especially since we/the literature frequently refer to 'the process' by writing the LoM [usually, the reference is then to the stationary solution of the ARMA-LoM

> The next few pages clarify how we can obtain a solution / de-recursify

II.3 Solving linear stochastic DE.s (LSDE.s) [a cookbook-entry, supplemented with light comments on background]

> This section briefly recaps linear SDE.s

> Formulation: a linear SDE (of degree $p \in \mathbb{N}$) is an equation of the sort

$$(*) \begin{cases} \underbrace{\underline{x}_t}_{\in \mathbb{R}^d} = \underline{\nu} + \underset{\sim}{A}_1 \underline{x}_{t-1} + \ldots + \underset{\sim}{A}_p \underline{x}_{t-p} + \underline{\varepsilon}_t \\ \text{where } (\underline{\varepsilon}_t) \text{ is some exogeneous stochastic process (e.g. white noise, or MA, or...)} \end{cases}$$

(with the implicit understanding that there is an underlying space $(\Omega, \mathcal{A}, P, (\mathcal{F}_t)_{t\in J}), \underbrace{J \subseteq \mathbb{Z} : \ldots}_{|J| \geq p})$

which may be concisely written as

$$\overbrace{(\underset{\sim}{I} - \underset{\sim}{A}_1 \mathbb{L} - \underset{\sim}{A}_2 \mathbb{L}^2 - \ldots - \underset{\sim}{A}_p \mathbb{L}^p)}^{\equiv \underset{\sim}{p}(\mathbb{L})} \underline{x}_t = \underline{\nu} + \underline{\varepsilon}_t .$$

Importantly, since $\underline{\varepsilon}_t$ is exog. given, and $(\mathcal{F}_t)$ is adapted to $(\underline{x}_t, \underline{\varepsilon}_t)$, $\underline{x}_t$ is backward-looking (II.1)

> Solution: How solve (*)? Befor tackling this, need to have clear understanding of what we're after;

We call the process $(\underline{x}_t^*)_{t\in J}$ a solution to (*) if $(\underline{x}_t^*)$ satisfies (*),

$$\underset{\sim}{p}(\mathbb{L})\, \underline{x}_t^* = \underline{\nu} + \underline{\varepsilon}_t \quad \forall t \in \underbrace{J}_{\subseteq \mathbb{Z} : |J| \geq p} \text{ a.s.}$$

(and if it is represented non-recursively)

> The question arises: does (*) have a solution? If yes how many?

Ans: In general, (*) has a continuum of solutions & the form a vector subspace of $(\mathbb{R}^d)^J$ (space of processes $(\underline{x}_t)_{t\in J}$) of dimension p. If we enrich (*) with requirements on the trajectories of solutions (e.g. $\underline{x}_0^* \equiv \underline{c}$ a.s.) we get a 'boundary value problem' and i.G. we get a unique solution if we add p such conditions.

> The theory of solving (*): first, notice that we may express any degree-p LSDE as an appr. written degree-1-SDE

↳ i.e. for $\underline{X}_t \equiv (\underline{x}_t' \ldots \underline{x}_{t-p+1}')'$ can write (*) as

"companion form"

$$(**) \quad \underline{X}_t = \underline{V} + \underset{\sim}{A} \underline{X}_{t-1} + \underline{E}_t \quad \text{(cf. MTSA, I.8)}$$

hence, for giving general ideas we first consider (*) for $d=1$ and for $d>1$ refer to the above degree-1-system and its associated solution routine

↳ i.e. if you have $p>1, d>1$, do essentially the same things but on (**) and with matrices (since in (**) $p=1$, solving homog. eqn. is straightforward)

Consider (*) for $p \geq 1$, $d=1$. First important concept:

"Net response of system (i.e. here x_t) caused by multiple stimuli is sum of responses to individual stimuli"

Superposition principle: If $(x_t^{(1)})$ and $(x_t^{(2)})$ are solutions to (*), then $(x_t^{(1)} - x_t^{(2)})$ solve the homogeneous equation

$$p(\mathbb{L})(x_t^{(1)} - x_t^{(2)}) = 0 \quad (H)_{(*)}$$

(this directly follows from distr.ty of $\mathbb{L}$: $\phi(\mathbb{L})(X_t^{(1)} - X_t^{(2)}) = \phi(\mathbb{L})X_t^{(1)} - \phi(\mathbb{L})X_t^{(2)} = \nu + \varepsilon_t - \nu - \varepsilon_t = 0$)

Or equivalently, <u>any</u> solution X_t^* to (*) can be written as

$$X_t^{(*)} = \tilde{X}_t^{(H)} + \tilde{X}_t^{(P)}$$

where $\tilde{X}_t^{(P)}$ is some particular solution to (*) and $\tilde{X}_t^{(H)}$ is some solution to $(H)_{(*)}$

Hence, the <u>family of solutions</u> to (*) can be written as
("the General solution")

$$\boxed{X_t^{(G)} = X_t^{(H)} + X_t^{(P)}}$$

this is yet to be established

where

$X_t^{(H)}$ is a parametric family of processes, each member of which solves $(H)_{(*)}$ (in practice, this will be some term involving undetermined constants which are the parameters of the family)

$X_t^{(P)} \equiv \underbrace{\phi(\mathbb{L})^{-1}}_{\text{backward or forward iteration}} [\nu + \varepsilon_t]$

backward or forward iteration, or both, depending on roots of $\phi(\cdot)$, see later; for now take existence for granted here.

if $J \neq \mathbb{Z}$, then eventually iteration leads out of J. In this case, we just define $\varepsilon_t = 0$ $\forall t \notin J$. (Usually, $J = \mathbb{N}_0$)

is the canonical particular solution to (*)

This <u>General solution</u> can give rise to a unique <u>particular solution</u>
family of stoch. proc.s
one stoch. proc.
if we have enough side-conditions on $(X_t^{(G)})$ [e.g. initial conditions, pass-along-conditions, terminal conditions, ...] to pin down all parameters in $(X_t^{(H)})$ [$X_t^{(P)}$ is unique by construction].

> Finally: how do we obtain $X_t^{(P)}$, $X_t^{(H)}$?

[$X_t^{(P)}$] Clearly, it is n&s to know how to compute $\phi(\mathbb{L})^{-1}$. This can be reasoned like so:

1) $\phi(\mathbb{L}) \equiv 1 - a_1\mathbb{L} - \dots - a_p\mathbb{L}^p$ is a Polynomial in $\mathbb{L}$.

Lag-polynomial is algebraically eqvt. to $\mathbb{C}$-poly!

Due to the algebraic properties of $\mathbb{L}$ sketched in II.2, this polynomial behaves just as if it was a polynomial over $z \in \mathbb{C}$!

In particular, completely analogous to the polynomial $\phi(z) \equiv 1 - a_1 z - \dots - a_p z^p$, it can be <u>factorized</u>!

Fundamental theorem of Algebra!

$$\boxed{\phi(\mathbb{L}) = \left(1 - \frac{1}{z_1}\mathbb{L}\right)^{m_1} \cdot \ldots \cdot \left(1 - \frac{1}{z_n}\mathbb{L}\right)^{m_n}}$$

where $z_1, \dots, z_n$ ($n \le p$) are the roots of $\phi(z) = 0$ with multiplicities $m_1, \dots, m_n$ ($\sum_i m_i = p$)

Mini-Excurse: Characteristic Polynomial

The equation $\phi(z) \equiv 1 - a_1 z - \ldots - a_p z^p = 0$ is of such central importance to solving (*) for $X_t^{(G)}$ that it was dubbed "characteristic equation / polynomial" to (*). W/o going into theory, fundamental theorem of algebra ensures that

$$\phi : \mathbb{C} \to \mathbb{C} : z \mapsto 1 - a_1 z - \ldots - a_p z^p$$

has exactly p roots (some of them repeated) and may be rewritten as

$$\phi(z) = \left(1 - \frac{1}{z_1} z\right)^{m_1} \cdots \left(1 - \frac{1}{z_n} z\right)^{m_n} \qquad \left(\textstyle\sum_i m_i = p\right)$$

where $z_1, \ldots, z_n$ are the roots to $\phi(z) = 0$ with (algebraic) multiplicities $m_1, \ldots, m_n$.

Usually, when we refer to the characteristic roots, though, of (*), we refer to

$$\lambda_i = z_i^{-1} \quad , \quad i \in \{1, \ldots, n\}$$

which are the roots to

$$1 - a_1 \lambda^{-1} - \ldots - a_p \lambda^{-p} = 0$$

(since polynomial in neg. expon.s, is hard to solve but)

$$\overset{\cdot \lambda^p}{\Leftrightarrow} \quad \boxed{\lambda^p - a_1 \lambda^{p-1} - \ldots - a_p = 0}$$

($\Rightarrow a_p \neq 0$ ensures 0 is not a root!)

sometimes the above is referred to as "characteristic polynomial".

2) Now, while tedious in practice, in theory inversion of $\phi(\mathbb{L})$ is straightforward:

$$\phi(\mathbb{L})^{-1} = \underbrace{(1 - \lambda_1 \mathbb{L})^{-m_1}}_{} \cdots (1 - \lambda_n \mathbb{L})^{-m_n}$$

$$= \underbrace{(1 - \lambda_1 \mathbb{L})^{-1} \cdots (1 - \lambda_1 \mathbb{L})^{-1}}_{m_1 \text{ times}}$$

how compute? Compute each inverse sep.ly (geometric power series), then compute each product sep.ly using Cauchy-product for series

computed as in II.2, (b) & (c) depending on $\underline{|\lambda_1| \gtrless 1}$!

this gives $X_t^{(P)}$.

NOTE: by construction, this inversion automatically yields a stationary process! (We can also backwards-iterate for an explosive λ – just then we need to stop after finitely many steps)

$[X_t^{(H)}]$ Now consider

$$(H)_{(*)} \qquad \phi(\mathbb{L}) X_t = 0$$

$$\Leftrightarrow \quad (1 - \lambda_1 \mathbb{L})^{m_1} \cdots (1 - \lambda_n \mathbb{L})^{m_n} X_t = 0$$

First consider the case of distinct roots: $n = p$, $m_i = 1 \ \forall i$. Then, it's fairly clear to see from the factorization that

$$X_t = c \cdot \lambda_i^t \quad \text{for any } c \in \mathbb{R} \quad \forall i \in \{1, \ldots, p\}$$

solves $(H)_{(*)}$

Pf: $\phi(\mathbb{L}) c \lambda_i^t = \dfrac{\phi(\mathbb{L})}{(1 - \lambda_i \mathbb{L})} \underbrace{(1 - \lambda_i \mathbb{L}) c \lambda_i^t}_{= c (1 - \lambda_i \mathbb{L}) \lambda_i^t = c \cdot (\lambda_i^t - \lambda_i \cdot \lambda_i^{t-1})} = 0 \quad \square$

Pass-through at $t = \tau$

Note: $c \in \mathbb{R}$ arbitrary here! In particular it is instructive to consider $c = \tilde{c} \cdot \lambda_i^{-\tau}$ so that the solution to $(H)_{(*)}$ takes the form $\tilde{c} \cdot \lambda_i^{t-\tau}$. This already suggests in its notation that $\tilde{c}$ is pinned down by a pass-through condition in $t = \tau$; we'll need this momentarily.

(ii) $\forall i$, give some pass-through condition at time $\tau \in \mathbb{Z}$; then, for $|\lambda_i| < 1$, let $\tau \to -\infty$, for $|\lambda_i| > 1$, let $\tau \to +\infty$ $\Rightarrow c \equiv 0$ (cf. Note "pass-through at $t = \tau$")

Note $x_t^{(H)} \equiv 0$. From time to time, we will want to achieve $x_t^{(H)} \equiv 0 \ \forall t$ (e.g. when we want the solution to be stationary). This can be achieved by setting $c_{i,j} = 0 \ \forall i,j$ but the question remains what the intuition is; Ans: we can think of $c \equiv 0$ as requiring either

(i) $(\varepsilon_t)_{t \leq 0} \equiv 0 \ \wedge \ x_0 = \dots = x_p = \Phi^{-1}(1)\nu$ $(\Rightarrow c \equiv 0)$, or

Moreover, can see that for any two solutions to $(H)_{(*)}$, $x_t^{(1)}, x_t^{(2)}$ any of their linear comb.s is again a solution:

$$\Phi(\mathbb{L})\left(h_1 x_t^{(1)} + h_2 x_t^{(2)}\right) = 0$$

(this follows directly from dist.y of $\mathbb{L}$)

Finally, since $\lambda_1 \neq \dots \neq \lambda_p$ (by assumption) the sequences $(\lambda_1^t)_{t \in J}, \dots, (\lambda_p^t)_{t \in J}$ are linearly independent

$\Rightarrow$ The set of solutions to $(H)_{(*)}$ is a vector subspace of $\mathbb{R}^J$ with dimension p

since $\{(\lambda_i^t)_t \mid i \in \{1, \dots, p\}\}$ is the basis. Why not larger basis / more solutions? This is bcs of the factorization $\Phi(\mathbb{L}) = 0$ iff at least one factor is zero. Moreover, $(1 - \lambda_i \mathbb{L}) x_t = 0$ iff $x_t = c\lambda_i^t$!

$\Rightarrow$ We can achieve any solution to $(H)_{(*)}$ by choosing appropriately the constants in the sum $\sum_{i=1}^p c_i \lambda_i^t$.

$$\Rightarrow \boxed{x_t^{(H)} = \left\{ \sum_{i=1}^p c_i \lambda_i^t \;\middle|\; c_i \in \mathbb{R} \ \forall i \right\}}$$

distinct roots

For the case of repeated roots, $n < p$, it is possible to show the same result with the modification that now

note: $\sum_i m_i = p$

See Neusser's script "Difference Equations for Economists", section 2.4.1 (pp. 55 ff. in PDF)

$$\boxed{x_t^{(H)} = \left\{ \sum_{i=1}^n \left[\sum_{j=0}^{m_i - 1} c_{ij} \, t^j \right] \lambda_i^t \;\middle|\; c_{ij} \in \mathbb{R} \ \forall i,j \right\}}$$

repeated roots

In any case, the set of solutions to $(H)_{(*)}$ is a vector subspace of dimension p of $\mathbb{R}^J$, hence the set of General solutions to $(*)$ is a vector subspace of dim p of the space of $\mathbb{R}$-valued processes with index set J.

> Lastly, one remark on cases with $d > 1$. In such cases it is usually beneficial to write the system in companion form and apply completely analogous principles to the above to determine the general solution as

$$\boxed{\begin{aligned} \underset{\sim}{X}_t^{(G)} &= \underset{\sim}{X}_t^{(H)} + \underset{\sim}{X}_t^{(P)} \\ \text{where} \quad \underset{\sim}{X}_t^{(P)} &\equiv (\underset{\sim}{I} - \underset{\sim}{A}\mathbb{L})^{-1}(\underset{\sim}{E}_t + \underset{\sim}{V}) \\ \text{and} \quad \underset{\sim}{X}_t^{(H)} &\equiv \left\{ \underset{\sim}{A}^t \cdot \underset{\sim}{c} \;\middle|\; \underset{\sim}{c} \in \mathbb{R}^{d \cdot p} \right\} \end{aligned}}$$

note, since can write $\underset{\sim}{A} \equiv \underset{\sim}{Q} \underset{\sim}{J} \underset{\sim}{Q}^{-1}$ ($\underset{\sim}{J}$ Jordan matrix, cf. DynMo, I.1) we get a similar form like above $x_t^{(H)}$ for the elements in $[\underset{\sim}{X}_t]_{1:d}$

> In the following sections, using these results we will only work with degree-1 vector-valued LSDE.s

> We see: the solution to an ARMA(p,q)-LoM takes the form

$$X_t = \underbrace{\phi(1)^{-1}c + \phi(\mathbb{L})^{-1}\theta(\mathbb{L})\varepsilon_t}_{\substack{\text{canonical particular solution} \\ \text{(always exists in } \mathcal{L}^2(\mathbb{P})\text{, by THM 1.2)}}} + \underbrace{\sum_{i=1}^{n}\left(\sum_{j=0}^{m_i-1} c_{i,j}\, t^j\right)\lambda_i^t}_{\substack{\text{homogeneous solution,} \\ \text{parameterized by } \{c_{i,j}\}}}$$

and, since $\phi(\mathbb{L})^{-1}\theta(\mathbb{L})$ is measurable mapping, properties like weak/strict stat.y & ergodicity are inherited from (ε_t) for this term.

if

"$\phi(\mathbb{L})$ is stable"

(a) all roots of $\mathbb{C}\to\mathbb{C}: z\mapsto\phi(z)$ are <u>outside</u> the unit circle $\Leftrightarrow |\lambda_i|<1$ $\forall i$, we have that $\phi(\mathbb{L})^{-1}$ is a power series in $\mathbb{L}$

(b) $c_{i,j}=0$ $\forall i,j$ $\leftarrow$ see above for interpretation (Sidenote "$X_t^{(H)}\equiv 0$")

we have that

$$X_t = \phi(1)^{-1}c + \sum_{i\geq 0}\psi_i\varepsilon_{t-i}\ ;\ \text{for } q=0,\ (\psi_i) \text{ given as in THM 1.2:}$$

$$q=0 \Rightarrow \begin{cases} \psi_0 c_0 = 1,\ \psi_0 c_1 + c_0\psi_1 = 0,\ \psi_0 c_2 + \psi_1 c_1 + \psi_2 c_0 = 0, \dots, \sum_{j=0}^{n}\psi_j c_{n-j}, \sum_{j=0}^{n}\psi_{j+1}c_{n-j}, \dots, \\ \sum_{j=0}^{n}\psi_{j+k}c_{n-j}, \dots \end{cases}$$

in which case we refer to (X_t) as an MA(∞)-process.
We will <u>show</u> momentarily that such an MA(∞) is weakly stationary.
follows already from $\phi(\mathbb{L})^{-1}\theta(\mathbb{L})$ being measurable mapping, but we'll show it explicitly

Hence: for $\phi(\mathbb{L})$ stable, the unique stationary solution to ARMA(p,q) is given as an MA(∞).

> From time to time the question also arises whether we can think of (ε_t) as a linear combination of past (X_t) [e.g. we have a time series of (X_t) and want to know if we can back out (ε_t) from it]; this is indeed not always the case!
To see this, write the ARMA(p,q):

$$\phi(\mathbb{L})X_t = c + \theta(\mathbb{L})\varepsilon_t \quad\Longrightarrow\quad \varepsilon_t \overset{\text{an AR}(\infty)\text{ in } X_t,\varepsilon_t\text{ ! :)}}{=} \theta(\mathbb{L})^{-1}\phi(\mathbb{L})X_t + \theta(1)^{-1}c$$

and we see that $\theta(\mathbb{L})^{-1}\phi(\mathbb{L})$, again by THM 1.2, is a power series in $\mathbb{L}$ <u>iff</u> all roots of $z\mapsto\theta(z)$ are outside the unit circle!

In such a case, ε_t can be written as a series over <u>past</u> (X_t); we often sloppily say "ε_t belongs to X_t's past", or more formally (ε_t) is fundamental to (X_t).
If $\theta(\mathbb{L})^{-1}$ involves <u>negative</u> powers of $\mathbb{L}$ ($\Leftrightarrow$ some roots of $z\mapsto\theta(z)$ inside unit circle) then this 'fundamentality' doesn't obtain; if $\theta(\mathbb{L})^{-1}$ involves <u>only</u> negative powers of $\mathbb{L}$, we say (ε_t) belongs to the 'future of (X_t)'.

DEF 2.2 (Invertibility, fundamentality). Consider some ARMA(p,q) ($p\geq 0$, $q\geq 1$). We call the MA-polynomial $\theta(\mathbb{L})$ invertible if $z\mapsto\theta(z)$ has its roots outside the unit circle.
On a related point, for a stoch. process $(y_t)_{t\in\mathbb{Z}}$ and some $t\in\mathbb{Z}$, we define $\mathcal{H}(y^t) := \{z_t \mid z_t = \sum_{i\geq 0} a_i y_{t-i} \wedge \sum_{i\geq 0} a_i^2 < +\infty\}$ to be the Hilbert space of square summable linear combinations of the one-sided infinite history of the random variable y_t.
We call a process (X_t) fundamental to (y_t) if $\forall t\in\mathbb{Z}$, $\mathcal{H}(x^t)\subseteq\mathcal{H}(y^t)$.

Intuitively, $\mathcal{H}(x^t) \subseteq \mathcal{H}(y^t)$ means that any random variable obtained from the past of (x_t) can be obtained from the past of (y_t); in particular $\mathbb{E}[z_t|\sigma(y^t)]$ is known a.s. $\forall z_t \in \mathcal{H}(x^t)$!

3 Moments of ARMA(p,q)

> Given a particular ARMA(p,q)-LoM, we may characterize the behavior of its solution (usually its stationary solution) by considering its moments

D&P 3.1 (Moments of a stochastic process). Consider some stochastic process (X_t). We denote, as usual, $\mathbb{E}X_t$, if it exists, as the expectation of the random variable X_t.
Suppose $X_t \in \mathcal{L}^2(\mathbb{P})$. We define for $h \in \mathbb{Z}$

$$\gamma_h^t \equiv \mathbb{E}[(X_t - \mathbb{E}X_t)(X_{t-h} - \mathbb{E}X_{t-h})] \quad (\text{from which } \gamma_0^t \equiv Var(X_t)).$$

and call it the autocovariance of X_t at lag $h \in \mathbb{Z}$ if $h<0$ we say "autocov at lead $-h$"

$\rho_h^t \equiv \frac{\gamma_h^t}{\gamma_0^t}$ is the autocorrelation at lag h.

Notice that $\forall h \in \mathbb{Z}$, $\gamma_h^t = \gamma_{-h}^t$. Now let (X_t) be weakly stationary. Provided that $\sum_{h\in\mathbb{Z}} |\gamma_h| < +\infty$, we define

$$g_X : \mathbb{C} \to \mathbb{C} : z \mapsto \sum_{h\in\mathbb{Z}} \gamma_h \cdot z^h$$

and call it the autocovariance-generating function. g_X's restriction to the complex unit circle, and divided by 2π ...!

$$S_X : \mathbb{R} \to \mathbb{C} : \omega \mapsto \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma_h \exp(-i \cdot \omega \cdot h)$$

is called the population spectrum of (X_t). It may be used for frequency domain analysis — but this is not done here.

> Let's analyze these moments for a couple of processes

THM 3.2 (Moments of MA(q), q<+∞). Consider a (X_t) that satisfies the MA(q)-LoM

note: $\theta_0 = 1$ by convention.

$$X_t = c + \theta(\mathbb{L})\varepsilon_t, \quad \varepsilon_t \sim WN(0,\sigma^2).$$

Then we have $(X_t) \in \mathcal{L}^2(\mathbb{P})$, (X_t) is weakly stationary with

$$\mathbb{E}X_t = c, \quad \text{and } \forall h \geq 0$$

$$\gamma_h = \begin{cases} 0 & \text{for } h > q \\ \sigma^2 \cdot \sum_{i=h}^{q} \theta_i \theta_{i-h} & \text{for } h \in \{0,1,\dots,q\} \end{cases}$$

Proof: Since $(\varepsilon_t) \in \mathcal{L}^2(\mathbb{P})$ and by linearity $\mathbb{E}(X_t) = c \ \forall t$. Now also for $h > 0$

$$\mathbb{E}([X_t - c][X_{t-h} - c]) = \mathbb{E}\Big[\underbrace{\textstyle\sum_{i=0}^q \theta_i \varepsilon_{t-i} \cdot \sum_{i=0}^q \theta_i \varepsilon_{t-h-i}}_{= \sum_{i,j=0}^q \theta_i \theta_j \varepsilon_{t-i} \varepsilon_{t-h-j}}\Big]$$

$$= \sum_{i,j=0}^q \theta_i \theta_j \underbrace{\mathbb{E}[\varepsilon_{t-i}\varepsilon_{t-h-j}]}_{= \sigma^2 \cdot \mathbb{1}\{i = h+j\}} \qquad = \sigma^2 \sum_{i=0}^q \sum_{j=0}^q \mathbb{1}\{i = h+j\} \theta_i \theta_j$$

$$= \sigma^2 \cdot \begin{cases} 0 & \text{if } h > q \quad (\Rightarrow i < h+j \ \forall i,j \in \{1, \dots, q\}) \\ \sum_{i=h}^q \theta_i \theta_{i-h} & \text{if } h \in \{1, \dots, q\}. \end{cases}$$

□

THM 3.3 (Moments of stable AR(p) / of MA(∞)). Consider a process (X_t) that satisfies the AR(p)-LoM

$$\phi(\mathbb{L}) X_t = c + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

where $\phi(\mathbb{L})$ is stable (all characteristic roots smaller than 1 in modulus, cf. above). Then, taking the canonical particular solution, $X_t \equiv \phi(1)^{-1} c + \phi(\mathbb{L})^{-1} \varepsilon_t$,

(i) (X_t) is stationary

(ii) $\mathbb{E}X_t = \phi(1)^{-1} c = (1 - \phi_1 - \dots - \phi_p)^{-1} \cdot c$

(iii) $\forall h \geq 0$,

keep in mind: $\gamma_{-k} = \gamma_k \ \forall k > 0$

$$\gamma_h = \begin{cases} \sum_{i=1}^p \phi_i \gamma_{h-i} & \text{for } h \geq 1 \\ \sigma^2 + \sum_{i=1}^p \phi_i \gamma_i & \text{for } h = 0 \end{cases} \qquad \text{with } \sum_{h \in \mathbb{Z}} |\gamma_h| < +\infty$$

(iv) note these are defined recursively — we can compute $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ as:

$$(\gamma_0, \dots, \gamma_{p-1})' = \left[(\underset{\sim}{I}_{p^2} - \underset{\sim}{A} \otimes \underset{\sim}{A})^{-1} \operatorname{vec}(\underset{\sim}{\Sigma})\right]_{1:p} = \left[\sigma^2 (\underset{\sim}{I}_{p^2} - \underset{\sim}{A} \otimes \underset{\sim}{A})^{-1}\right]_{1:p}$$

$$\underset{\sim}{A} \equiv \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix}_{p \times p}, \quad \underset{\sim}{\Sigma} \equiv \begin{bmatrix} \sigma^2 & \underset{\sim}{0}_{p-1}' \\ \underset{\sim}{0}_{p-1} & \underset{\sim}{0}_{p-1} \end{bmatrix}$$

Proof: Firstly, notice that $X_t \equiv \phi(1)^{-1} c + \phi(\mathbb{L})^{-1} \varepsilon_t$ exists as an $\mathcal{L}^2$-limit:

$$\sum_{i=0}^n \psi_i \varepsilon_{t-i} \xrightarrow[n \to \infty]{\mathcal{L}^2} \phi(\mathbb{L})^{-1} \varepsilon_t \quad \text{with } (\psi_i)_{i \in \mathbb{N}_0} \text{ defined in THM 1.2.}$$

Recall that the space $\mathcal{L}^2(\mathbb{P})$ is a complete normed vector space — hence $\phi(\mathbb{L})^{-1}\varepsilon_t \in \mathcal{L}^2(\mathbb{P})$ and the expectation, variance and all covariances exist! For weak stationarity, we need just establish their time-independence, so let's do that.

Since $\forall t \in \mathbb{Z}$, the sequence $(\sum_{i=0}^n \psi_i \varepsilon_{t-i})_{n\in\mathbb{N}_0}$ converges in $\mathcal{L}^2(\mathbb{P})$ to $\phi(\mathbb{L})^{-1}\varepsilon_t$, we may compute for $t\in\mathbb{Z}$ arbitrary

$$\mathbb{E}\left(\phi(\mathbb{L})^{-1}\varepsilon_t\right) = \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=0}^n \psi_i \varepsilon_{t-i}\right]$$

using that for a random sequence (y_n): $y_n \xrightarrow{\mathcal{L}^2} y$, $\mathbb{E}y_n \to \mathbb{E}y$ (cf. E703, II.1)

now using linearity of $\mathbb{E}(\cdot)$ & $\varepsilon_t \sim WN(0,\sigma^2)$

$$\Rightarrow \mathbb{E}(\phi(\mathbb{L})^{-1}\varepsilon_t) = \lim_{n\to\infty} \sum_{i=0}^n \psi_i \underbrace{\mathbb{E}(\varepsilon_{t-i})}_{=0} = 0.$$

$$\Rightarrow \mathbb{E}(X_t) = \phi(1)^{-1}c \quad \forall t.$$

Now consider the variance of X_t.
Following the same strategy as above, consider for $n\in\mathbb{N}$

$$\mathbb{E}\left(\left|\sum_{i=0}^n \psi_i \varepsilon_{t-i}\right)^2\right) = \mathbb{E}\left[\sum_{i=0}^n \sum_{j=0}^n \psi_i\psi_j \varepsilon_{t-i}\varepsilon_{t-j}\right] = \sum_i\sum_j \psi_i\psi_j \sigma^2 \mathbb{1}\{i=j\}$$
$$= \sigma^2 \sum_{i=0}^n \psi_i^2$$

Hence, $\operatorname{Var}(\phi(\mathbb{L})^{-1}\varepsilon_t) = \lim_{n\to\infty} \sigma^2 \sum_{i=0}^n \psi_i^2 = \sigma^2 \sum_{i\geq 0}^n \psi_i^2 < +\infty$

by $\sum_{i\geq 0}|\psi_i| < +\infty$ (cf. THM 12)

again: $\mathbb{E}$ & lim interchangeable by $\mathcal{L}^2$-convergence

Finally, consider for $h>0$ and for $n\in\mathbb{N}$

$$\mathbb{E}\left[\sum_{i=0}^n \psi_i \varepsilon_{t-i} \cdot \sum_{i=0}^n \psi_i \varepsilon_{t-h-i}\right] = \sum_{i=0}^n \sum_{j=0}^n \psi_i\psi_j \sigma^2 \mathbb{1}\{i=h+j\}$$
$$= \sigma^2 \sum_{i=h}^n \psi_i \psi_{i-h}$$

$$\Rightarrow \mathbb{E}\left(\phi(\mathbb{L})^{-1}\varepsilon_t \cdot \phi(\mathbb{L})^{-1}\varepsilon_{t-h}\right) = \lim_{n\to\infty} \sigma^2 \sum_{i=h}^n \psi_i\psi_{i-h} = \sigma^2 \sum_{i\geq h} \psi_i \psi_{i-h} < +\infty$$

again lim & $\mathbb{E}$ interchangeable by $\xrightarrow{\mathcal{L}^2}$

$\sum_{i\geq h}\psi_i\psi_{i-h} = \sum_{i\geq 0}\psi_{i+h}\psi_i$ converges absolutely (o/w get $\perp$ with $\exists$ of moment which is est.ed already; cf. below also)

since $\sum_{i\geq 0}|\psi_i| < +\infty$, we also have $\boxed{\sum_{j\in\mathbb{Z}}|\gamma_j| < +\infty}$; proof: first, we show $\sum_{k\geq 0}\sum_{h\geq 0}|\psi_{k+h}|\cdot|\psi_k| < +\infty$: pick $n\in\mathbb{N}$, and then $\sum_{k=0}^n\sum_{h\geq 0}|\psi_{k+h}||\psi_k| = \sum_{k=0}^n |\psi_k| \underbrace{\sum_{h\geq 0}|\psi_{k+h}|}_{<+\infty} = \sum_{k=0}^n|\psi_k|\sum_{h\geq k}|\psi_h| \leq \sum_{h\geq 0}|\psi_h|\cdot\sum_{k=0}^n|\psi_k| \leq \sum_{h\geq 0}|\psi_h|\cdot\sum_{k=0}^\infty|\psi_k| < +\infty$. Hence $(\sum_{k=0}^n\sum_{h\geq 0}|\psi_{k+h}||\psi_k|)_n$ is a bounded, increasing sequ., thus convergent; furthermore, since the double series converges, we can apply Fubini-Tonelli to deduce $\sum_{h\geq 0}\sum_{k\geq 0}|\psi_{k+h}||\psi_k| < +\infty$. The main claim now follows by: $\sum_{j\in\mathbb{Z}}|\gamma_j| \leq 2\sum_{j\geq 0}|\gamma_j|$
$\leq 2\sigma^2\sum_{j\geq 0}|\sum_{i\geq 0}\psi_{i+j}\psi_i| \underset{\Delta\text{-ineq}}{\leq} 2\sigma^2\sum_{j\geq 0}\sum_{i\geq 0}|\psi_{i+j}\psi_i| < +\infty.$ est.ed above

We see, all first 2^{nd} moments of (X_t) exist and are time-invariant — i.e. it is stationary! This proves (i), and (ii) was shown along the way.

Now as a way to compute γ_h in dependence of the AR-parameters, consider (iii). Take the LoM (putting $\mu \equiv \phi(1)^{-1}c = \mathbb{E}X_t$)

$$X_t - \mu = \phi_1(X_{t-1}-\mu) + \ldots + \phi_p(X_{t-p}-\mu) + \varepsilon_t$$

and multiply with $(X_{t-h}-\mu)$, then take $\mathbb{E}$; using stationarity of (X_t) we obtain

$$\gamma_h = \phi_1\gamma_{h-1} + \ldots + \phi_p\gamma_{h-p} + \sigma^2\mathbb{1}\{h=0\}$$

which is the claimed equation. The question remains, since this equation is recursive, how to actually compute these γ_h!
As with any (stable) recursion, the answer lies in finding initial conditions — i.e. here in computing $\gamma_0,\ldots,\gamma_{p-1}$ and then simply following the recursion.
(iv): the clue to prove this point is to notice that the LoM is eqvt to

$$\begin{cases} X_t = c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t \\ X_{t-1} = X_{t-1} \\ \vdots \\ X_{t-p} = X_{t-p} \end{cases} \iff \underline{y}_t = \underline{d} + \underset{\sim}{A}\, \underline{y}_{t-1} + \begin{pmatrix} 1 \\ \underline{0}_{p-1} \end{pmatrix} \varepsilon_t ,$$

$$\underline{y}_t \equiv (X_t, \ldots, X_{t-p})', \ \underline{d} = (c, \underline{0}_{p-1}')', \ \underset{\sim}{A} \text{ as in prop.}$$

This is a simple AR(1) in vector-form and we can readily compute

$$(\underline{y}_t - \underbrace{\mathbb{E}\underline{y}_t}_{= (\underset{\sim}{I} - \underset{\sim}{A})^{-1} \underline{d}}) = \underset{\sim}{A} (\underline{y}_{t-1} - \mathbb{E}\underline{y}_t) + \begin{pmatrix} 1 \\ \underline{0}_{p-1} \end{pmatrix} \varepsilon_t$$

$$\Rightarrow \underset{\sim}{\Gamma}_0 = \underset{\sim}{A} \underbrace{\underset{\sim}{\Gamma}_{-1}}_{\equiv \mathbb{E}[(\underline{y}_{t-1} - \mathbb{E}\underline{y}_{t-1})(\underline{y}_t - \mathbb{E}\underline{y}_t)'] = \underset{\sim}{\Gamma}_1'} + \underbrace{\underset{\sim}{\Sigma}}_{\text{as in prop.}} , \quad \underset{\sim}{\Gamma}_1 = \underset{\sim}{A} \underbrace{\underset{\sim}{\Gamma}_0}_{= \underset{\sim}{\Gamma}_0'}$$

$$\Rightarrow \underset{\sim}{\Gamma}_0 = \underset{\sim}{A} \underset{\sim}{\Gamma}_0 \underset{\sim}{A}' + \underset{\sim}{\Sigma} \iff \operatorname{vec}(\underset{\sim}{\Gamma}_0) = (\underset{\sim}{A} \otimes \underset{\sim}{A}) \operatorname{vec}(\underset{\sim}{\Gamma}_0) + \operatorname{vec}(\underset{\sim}{\Sigma})$$

$$\iff \operatorname{vec}(\underset{\sim}{\Gamma}_0) = (\underset{\sim}{I}_{p^2} - \underset{\sim}{A} \otimes \underset{\sim}{A})^{-1} \operatorname{vec}(\underset{\sim}{\Sigma})$$

Finally, the claim follows from noticing

$$\underset{\sim}{\Gamma}_0 \equiv \mathbb{E}\left[(\underline{y}_t - \mathbb{E}\underline{y}_t)(\underline{y}_t - \mathbb{E}\underline{y}_t)'\right] = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_{-1} & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \\ \gamma_{1-p} & \gamma_{2-p} & \cdots & \gamma_0 \end{bmatrix} \quad \text{and } \gamma_h = \gamma_{-h} .$$

$\square$

> This concludes the initial analysis of ARMA.s [if you ask yourself how to compute the AC.s of an ARMA(p,q), it is straightforward to follow the above steps, mutatis mutandis, provided the AR-polynomial is stable.]

> One last point: we could ask what to do about MA/AR/ARMA-LoM.s with unstable/non-invertible polynomials

- The clue here is that by THM 1.2, a stationary solution always exists (we just might have to iterate forward)
- From this stationary solution, we can define a new white noise sequence that, together with the stationary solution satisfies an ancillary LoM that is stable, resp. invertible!
- Cf. Hamilton, Sect. 3.7 (p. 40f. in PDF) for MA(1)
- Cf. E806, Assignment 2, Q2 for AR(1):

Consider the covariance-stationary solution to

$$y_t = \phi y_{t-1} + u_t, \quad u_t \sim WN(0, \sigma^2), \ |\phi| > 1, \tag{3}$$

denoted also as $(y_t)_{t \in \mathbb{Z}}$. The claim is that this can be represented as

$$y_t = \phi^{-1} y_{t-1} + \tilde{u}_t, \quad \tilde{u}_t \sim WN(0, \tilde{\sigma}^2),$$

which is a stable law of motion involving a sequence of white noise that is to be determined.

Start by finding the stationary solution to (3). Since $|\phi| > 1$, this can be done by forward iteration, which yields

$$y_t = -\sum_{i\geq 1} \phi^{-1} u_{t+i}.$$

Now simply put $\tilde{u}_t := y_t - \phi^{-1} y_{t-1}$. This reduces the question to showing that this $(\tilde{u}_t)_{t\in\mathbb{Z}}$ is indeed white noise, and finding its variance, $\tilde{\sigma}^2$.

To achieve this, first compute the autocovariance of y_t at lag $h \geq 0$:[6]

$$\gamma_h := \mathbb{E}(y_t y_{t-h}) = \mathbb{E}\left(\sum_{i\geq 1}\phi^{-i}u_{t+i} \cdot \sum_{i\geq 1}\phi^{-i}u_{t+i-h}\right)$$

$$= \mathbb{E}\left(\sum_{i\geq 1}\sum_{j\geq 1}\phi^{-i-j}u_{t+i}u_{t+j-h}\right) = \sum_{i\geq 1}\sum_{j\geq 1}\phi^{-i-j}\mathbb{E}\left(u_{t+i}u_{t+j-h}\right)$$

$$= \sum_{i\geq 1}\sum_{j\geq 1}\phi^{-i-j}\sigma^2 \cdot \underbrace{\mathbb{1}\{j = i+h\}}_{(*)} = \sigma^2\sum_{i\geq 1}\phi^{-h-2i}$$

$$= \sigma^2\phi^{-h-2}\sum_{i\geq 0}\phi^{-2i} = \frac{\sigma^2\phi^{-h-2}}{1-\phi^{-2}}$$

$$\Rightarrow \gamma_h = \frac{\sigma^2\phi^{-h}}{\phi^2 - 1}.$$

(Note: it is at $(*)$ that we need $h \geq 0$.)

Now we can show the white noise property. For $h > 0$ (for $h < 0$ simply use $\gamma_h = \gamma_{-h}$)

$$\mathbb{E}(\tilde{u}_t \cdot \tilde{u}_{t-h}) = \mathbb{E}\left((y_t - \phi^{-1}y_{t-1})\cdot(y_{t-h} - \phi^{-1}y_{t-h-1})\right)$$

$$= \gamma_h - \phi^{-1}\gamma_{h-1} - \phi^{-1}\gamma_{h+1} + \phi^{-2}\gamma_h$$

now using $\phi^{-j}\gamma_h = \gamma_{h+j}$, $\forall h, j > 0$, as can be checked above

$$= 0.$$

What about $h = 0$? We can show, using $\gamma_h = \gamma_{-h}$ and $\phi^{-j}\gamma_h = \gamma_{h+j}$, $\forall h, j > 0$:

$$\tilde{\sigma}^2 = \mathbb{E}(\tilde{u}_t^2) = \gamma_0 - \phi^{-1}\gamma_1 - \phi^{-1}\gamma_1 + \phi^{-2}\gamma_0$$

$$= (1+\phi^{-2})\gamma_0 - 2\gamma_2 = \sigma^2\left[\frac{1+\phi^{-2}}{\phi^{-2}-1} - 2\frac{\phi^{-2}}{\phi^{-2}-1}\right]$$

$$= \sigma^2\frac{1-\phi^{-2}}{\phi^{-2}-1}$$

$$\Rightarrow \tilde{\sigma}^2 = \phi^{-2}\sigma^2.$$

6 Throughout the computations, I use that $y_t = -\sum_{i\geq 1}\phi^{-1}u_{t+i}$ is the $\mathcal{L}^2(\mathbb{P})$-limit of a sequence of random variables (this is analogously established as part (a) in question 1); this means that it is covariance-stationary (after establishing that the first two moments are also time-invariant), and that we can readily interchange $\mathbb{E}$ and $\sum_{i\geq 0}$, as for $\mathcal{L}^2(\mathbb{P})$-convergent sequences (here the partial sums), the moments (of partial sums) are also convergent, and for finite sums we can interchange sum and expectation.

Generally, for AR(p) with an unstable root λ_i:

$$\prod_{j\neq i}(1-\lambda_j\mathbb{L})\cdot(1-\lambda_i\mathbb{L})\,X_t = \overbrace{\varepsilon_t}^{\sim WN(0,\sigma^2)}$$

$$\Leftrightarrow \underbrace{\prod_{j\neq i}(1-\lambda_j\mathbb{L})\cdot(1-\lambda_i^{-1}\mathbb{L})}_{\text{stable } \mathbb{L}\text{-polyn.}}X_t = \underbrace{(1-\lambda_i^{-1}\mathbb{L})(1-\lambda_i\mathbb{L})^{-1}\varepsilon_t}_{\substack{\text{can show this is white noise!}\\ \text{(in fact we already did left!)}\\ \sim WN(0,\lambda_i^{-2}\sigma^2)}}$$

For MA(q) with unstable root ξ_i:

$$X_t = \mu + \prod_{j\neq i}(1-\xi_j\mathbb{L})\cdot(1-\xi_i\mathbb{L})\varepsilon_t$$

$$= \mu + \underbrace{\prod_{j\neq i}(1-\xi_j\mathbb{L})\cdot(1-\xi_i^{-1}\mathbb{L})}_{\text{stable!}}\cdot\underbrace{(1-\xi_i^{-1}\mathbb{L})^{-1}(1-\xi_i\mathbb{L})\varepsilon_t}_{\sim WN(0,\xi_i^2\sigma^2)}$$

4 Wold's decomposition theorem

> We have seen at the end of the last section that even unstable LoM.s (which may give the impression of producing only non-stationary processes) can be solved for stationary processes and these can be shown to solve stable LoM.s, i.e. have a stable representation

> Herman Wold was among the first to notice just how powerful this idea was — his famous theorem asserts that any weakly stationary process can be written as the sum of a deterministic sequence (e.g. a sine-wave or a mean, etc.) and an MA(q)-process ($q \in \mathbb{N}_0 \cup \{+\infty\}$):

THM 4.1 (Wold's decomposition). Let (X_t) be a weakly stationary process. Then, $\exists (b_i)_{i\in\mathbb{N}_0} \in \mathbb{R}^{\mathbb{N}_0}$ (conventionally $b_0 = 1$), $(\eta_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ and $(\varepsilon_t) \sim WN(0,\sigma^2)$ s.th. $\forall t\in\mathbb{Z}$

$$X_t = \sum_{j\geq 0} b_j\varepsilon_{t-j} + \eta_t, \quad \text{a.s.}$$

with $\sum_{j\geq 0} b_j^2 < +\infty$.

> Wold's theorem underpins the attractiveness of (stable & invertible) ARMA-models: they can fit about any w.stat.y process!!

└ Empirically, suppose (X_t)'s Wold repr. has abs.y summable coeff.s $\sum_{j\geq 0}|b_j| < +\infty$ — then, (X_t) can be represented as an AR(∞) and estimated as an AR(p) with p suff.y large

$\Rightarrow$ ARMA.s are empirically extremely useful.

5 Empirical Analysis I: Estimation & Model Diagnostics

> There are a number of ways to estimate an ARMA(p,q)-model — in this section we will work our way towards one popular such way; after that, we'll look at a number of empirical procedures, including formal tests, to decide which model (in terms of (p,q)) to fit to a given time series & whether at all a given model is a good fit in terms of residual autocorrelation. (If the true DGP was indeed ARMA(p,q), the regressions residuals should be approx.y normal

> First, let's lay the groundwork & present the fundamental assumptions: $(p,q \geq 0)$

For a given sample of data $\{y_t\}_{t \in \{\underbrace{1-p, \dots, 0}_{\text{"pre-sample"}}, \underbrace{1, \dots, T}_{\text{"sample"}}\}}$, we may make the following assumptions

(needed for AR-part. By convention, if $p=0$, this is just empty and $t \in \{1,\dots,T\}$.)

(A0) (i) (y_t) is an ARMA-process, ie.

$$\phi(\mathbb{L})y_t = c + \theta(\mathbb{L})u_t, \quad u_t \sim WN(0,\sigma^2), \quad \text{and}$$

(ii) $\phi(\cdot)$ is stable, and

$\Rightarrow (y_t)$ exists as $MA(\infty)$ and is w. stationary

(iii) $\theta(\cdot)$ is invertible, and

(iv) $\phi(\cdot)$ and $\theta(\cdot)$ have no common roots, ie.

$$\nexists\, b(\cdot) \text{ s.th. } \phi = b \cdot \varphi \;\wedge\; \theta = b \cdot \vartheta.$$

in practice, for univariate analysis this coinc. roots problem is easy to detect, things look a lot like classical multicollinearity!

(This assumption can be thought of like a no-multicollinearity-assumption in cross-sectional regression: were there common roots, then the ARMA(p,q) (y_t) would be non-identified since there would be an ARMA-LoM of strictly lower order also representing (y_t) that we would obtain by canceling $b(\cdot)$ on both sides.)

These assumptions we shall maintain throughout the entire analysis. We may also assume

(A1) On the innovations, we may make one or none of the following assumptions (in addition to (A0).(i)):

Cf. also DEF 1.1!

a) (u_t) is str. stationary & ergodic, in $\mathcal{L}^2(\mathbb{P})$, and a MDS (ie. $\mathbb{E}(u_t | u^{t-1}) = 0$) with $\underbrace{\mathbb{E}(u_t^2 | u^{t-1}) = \sigma^2}_{\text{conditional homoskedasticity}}$ $\forall t$

b) $(u_t) \overset{iid}{\sim} (0, \sigma^2)$, or

c) $(u_t) \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$ (ie. $\mathcal{L}((u_t)) = \mathcal{N}(0,\sigma^2)^{\otimes\infty}$).

Notice that c) $\Rightarrow$ b) $\Rightarrow$ a), but not in reverse.

> Let's now talk about estimation, starting with the simplest case.

AR(p) by OLS

> Consider the above sample & process under (A0) but with $q=0$:

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t \\ &= \underline{x}_t' \underline{\beta} + u_t \quad , \quad \underline{x}_t' \equiv (1, y_{t-1}, \ldots, y_{t-p}) , \ \underline{\beta} \equiv (c, \phi_1, \ldots, \phi_p)'. \end{aligned}$$

The naive OLS-procedure suggests to estimate

$$\hat{\underline{\beta}} = \left(\tfrac{1}{T} \textstyle\sum_{t=1}^T \underline{x}_t \underline{x}_t' \right) \left(\tfrac{1}{T} \textstyle\sum_{t=1}^T \underline{x}_t y_t \right).$$

Indeed, granting (A0) & (A1).a), this estimator is consistent & asy. normal:

$$\hat{\underline{\beta}} \xrightarrow{p} \underline{\beta} \quad \text{as } T \to \infty$$

$$\sqrt{T}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \mathbb{E}(\underline{x}_t \underline{x}_t')^{-1}\right) \quad \text{as } T \to \infty$$

with $\mathbb{E}(\underline{x}_t \underline{x}_t') = \begin{bmatrix} 1 & \mu & \mu & \cdots & \mu \\ \mu & \gamma_0+\mu^2 & \gamma_1+\mu^2 & \cdots & \gamma_{p-1}+\mu^2 \\ \mu & \gamma_1+\mu^2 & \gamma_0+\mu^2 & \cdots & \gamma_{p-2}+\mu^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu & \gamma_{p-1}+\mu^2 & \gamma_{p-2}+\mu^2 & \cdots & \gamma_0+\mu^2 \end{bmatrix}$, with $\mu \equiv \mathbb{E} y_t = \phi(1)^{-1} c$.

Note: the Avar($\hat{\underline{\Phi}}$) can be shown to not depend on σ^2 [since it's present in γ_0, it will cancel out]

$\mathbb{E}\underline{x}_t\underline{x}_t' = \begin{bmatrix} 1 & \mu \underline{1}_p' \\ \mu \underline{1}_p & \underbrace{\underline{V}}_{= Var([y_t \ldots y_{t-p+1}]')} + \mu^2 \underline{1}\underline{1}' \end{bmatrix}$ (recall: $\begin{bmatrix} \underline{A} & \underline{B} \\ \underline{C} & \underline{D} \end{bmatrix}^{-1}$ exists if $\underline{A}^{-1}, (\underline{D} - \underline{C}\underline{A}^{-1}\underline{B})^{-1}$ exist (or with roles of $\underline{A}, \underline{D}$ reversed))

here: $\underline{A} = 1$, $\underline{D} - \underline{C}\underline{A}^{-1}\underline{B} = \underline{V}$

$\Rightarrow (\mathbb{E}\underline{x}_t\underline{x}_t')^{-1} = \begin{bmatrix} 1 + \mu^2 \underline{1}'\underline{V}^{-1}\underline{1} & -\mu \underline{1}'\underline{V}^{-1} \\ -\mu \underline{V}^{-1}\underline{1} & \underline{V}^{-1} \end{bmatrix}$

$\Rightarrow$ Avar($\hat{\underline{\Phi}}$) $= \sigma^2 \underline{V}^{-1}$ ← $\propto \sigma^2$ (not shown here)

Proof: Since asy. norm.y with zero mean implies $\hat{\underline{\beta}} - \underline{\beta} = O_p(T^{-1/2}) = o_p(1)$, we only need to show the 2nd result. First,

$$\hat{\underline{\beta}} - \underline{\beta} = \left(\tfrac{1}{T} \textstyle\sum_t \underline{x}_t \underline{x}_t'\right)^{-1} \left(\tfrac{1}{T} \textstyle\sum_t \underline{x}_t u_t\right) \quad \text{(by usual OLS-algebra.)}$$

Now, $\frac{1}{T}\sum_t \underline{x}_t \underline{x}_t'$ has typical elements

$1 \to 1$, $\frac{1}{T}\sum_t y_{t-k} \xrightarrow{p} \mathbb{E}(y_{t-k}) = \mu$ by (A0) & LLN for w. stat.y processes, and

$\frac{1}{T}\sum_t y_t y_{t-k} \xrightarrow{p} \mathbb{E}(y_t y_{t-k}) = \gamma_k + \mu^2$ by either (A0) & LLN for w. stat.y processes (have to show $(y_t y_{t-k})$ is w. stat.y in $\mathcal{L}^2(\mathbb{P})$ which might not be possible — cf. Hamilton, Ch. 7 for LLN on 2nd moments), or more simply by (A0) & (A1).a) & ergodic thm.

(Or can use LLN for $\mathcal{L}^1$-Mixingale) is nice since you don't need stationarity but just uniform int.y for appr.e processes

$\Rightarrow \frac{1}{T}\sum_t \underline{x}_t \underline{x}_t' \xrightarrow{p} \mathbb{E}(\underline{x}_t \underline{x}_t')$ by element-wise convergence.

Now consider $\sqrt{T}\left(\frac{1}{T}\sum_t \underline{x}_t u_t\right)$. We have:

- $\mathbb{E}(\|\underline{x}_t u_t\|^r) < +\infty$ for $r > 2$

- $\frac{1}{T}\sum_t \mathbb{E}(\underline{x}_t\underline{x}_t' u_t^2) = \mathbb{E}(\underline{x}_t\underline{x}_t')\sigma^2$ by LIE & (A0)
- $\frac{1}{T}\sum_t \underline{x}_t\underline{x}_t' u_t^2 \xrightarrow{p} \mathbb{E}(\underline{x}_t\underline{x}_t' u_t^2) = \mathbb{E}(\underline{x}_t\underline{x}_t')\sigma^2$ by Ergodic thm & (A0) + (A1).a)

and finally, a filtration to which (y_t) & (u_t) are adapted

$$\mathbb{E}(\underline{x}_t u_t \mid \mathcal{F}_t) = \mathbb{E}\left(\begin{matrix} u_t \\ y_{t-1}u_t \\ \vdots \\ y_{t-p}u_t \end{matrix} \middle| \mathcal{F}_t\right) = \underline{0}$$

by $\mathbb{E}(y_{t-k}u_t \mid \mathcal{F}_t) = \mathbb{E}\Big(y_{t-k}\underbrace{\mathbb{E}(u_t \mid \mathcal{F}_{t-k})}_{=\mathbb{E}[\mathbb{E}(u_t \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-k}] = 0} \mid \mathcal{F}_t\Big) = 0.$

tower property

establishes the MDS-property.
The standard MDS-CLT (e.g. MS "Asymptotics for stoch processes", THM 5.3) implies

$$\sqrt{T}\,\frac{1}{T}\sum_t \underline{x}_t u_t \xrightarrow{d} \mathcal{N}\Big(0, \underbrace{\mathbb{E}(\underline{x}_t\underline{x}_t' u_t^2)}_{=\sigma^2\mathbb{E}(\underline{x}_t\underline{x}_t')}\Big)$$

by $\mathbb{E}(u_t^2 \mid y^{t-1})$ & LIE.

$=\sigma(u^{t-1})$ by y_t MA(∞) in u_t

$\square$

> Now as we'll see shortly, we can also estimate an ARMA(p,q) by a 2-step OLS procedure;
before, we look at the classical

ARMA (p,q) by Maximal Likelihood & Quasi-Maximum Likelihood

> ML is a classical estimation approach & quickly explained; suppose we knew the distribution of (y_t) up to the parameters $\phi(\cdot)$, $\theta(\cdot)$, σ^2.
Then, the likelihood could be written as

$$f(\{y_t\}_{t\in\{1-p,\dots,0,1,\dots,T\}} \mid \underline{\phi}, \underline{\theta}, \sigma^2).$$

Again, if $p=0$, we put $t\in\{1,\dots,T\}$.

For doing MLE, a central step is to factorize the likelihood using Bayes' law:

$$f(\{y_t\}_{t\in\{1-p,\dots,0,1,\dots,T\}} \mid \underline{\phi}, \underline{\theta}, \sigma^2)$$
$$= f(y_0,\dots,y_{1-p} \mid \underline{\phi}, \underline{\theta}, \sigma^2)$$
$$\cdot \prod_{t=1}^{T} f(y_t \mid y_{t-1},\dots,y_0,\dots,y_{1-p}; \underline{\phi}, \underline{\theta}, \sigma^2)$$

or in logs:

$$\mathcal{L}(\{y_t\} \mid \underline{\phi}, \underline{\theta}, \sigma^2) = \underbrace{\log f(y_0, \dots, y_{1-p} \mid \underline{\phi}, \underline{\theta}, \sigma^2)}_{\text{initialization term}} + \underbrace{\sum_{t=1}^{T} \log f(y_t \mid y^{t-1}; \underline{\phi}, \underline{\theta}, \sigma^2)}_{\text{conditional (on init. val.s for y) likelihood}}$$

In the conditional term, conditioning on y^{t-1} already fixes the values $y_{t-1}, \dots, y_{t-p}$, that appear in

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q};$$

Now provided we know $\mathcal{L}(u_t)$, pinning down $f(y_t \mid y^{t-1}; \underline{\phi}, \underline{\theta}, \sigma^2)$ from this eqn. is feasible — a much more convenient approach in practice, that however slightly deviates from the above 'true' conditional MLE, is to condition also on the initial values of (u_t), $u_0, \dots, u_{1-q}$:

Using conditioning on $u_0, \dots, u_{1-q}$, we can recursively compute the implied u_t exactly: given $y_0, \dots, y_{1-p}, u_0, \dots, u_{1-q}$,

innovation in $t=1$ implied by initial values $y_0, \dots, y_{1-p}, u_0, \dots, u_{1-q}$

we get: $u_1 = u_1^0 := y_1 - (c + \phi_1 y_0 + \dots + \phi_p y_{1-p} + \theta_1 u_0 + \dots + \theta_q u_{1-q})$, exactly!

Then, can inductively compute a 'pseudo-observed' sample $\{u_t\}_{t \in \{1,\dots,T\}}$!

And using this:

density of $y_t \mid y^{t-1} \dots$, e.g. $\mathcal{N}(c + \dots, \sigma^2)$ — density of u_t, e.g. $\mathcal{N}(0, \sigma^2)$

$$f(y_t \mid y^{t-1}, u^0; \underline{\phi}, \underline{\theta}, \sigma^2) \equiv g(u_t^0(y^{t-1}, u^0; \underline{\phi}, \underline{\theta}, \sigma^2))$$

Of course, $u_0, \dots, u_{1-q}$ are not known — typically, they are assumed to be all zero. Notice that, by following the recursion for u_t, the effect of the initial $u_0, \dots, u_{1-q}$ on current u_t will fade to zero, provided $\theta(\cdot)$ is invertible; thus, for ML-estim. of $\underline{\phi}, \underline{\theta}, \sigma^2$ using the below conditional approach requires invertibility of $\theta(\cdot)$ in order to work well!

> Typically, two types of estimators are interesting:

Conditional ML:

$$(\hat{c}, \hat{\underline{\phi}}, \hat{\underline{\theta}}, \sigma^2)_{CMLE} := \underset{(\underline{\varphi}, \underline{\vartheta}, s^2)}{\operatorname{argmax}} \sum_{t=1}^{T} \log f(y_t \mid y^{t-1}, \overset{\substack{=(u_0,\dots,u_{1-q}) \\ =\emptyset \text{ if } q=0}}{u^0}; \underline{\varphi}, \underline{\vartheta}, s^2)$$

this part deviates from the 'actual' conditional conditional LH

Unconditional ML:

$$(\hat{c}, \hat{\underline{\phi}}, \hat{\underline{\theta}}, \sigma^2)_{UMLE} := \underset{(\underline{\varphi}, \underline{\vartheta}, s^2)}{\operatorname{argmax}} \mathcal{L}(\{w_t\} \mid \underline{\phi}, \underline{\theta}, \sigma^2).$$

How compute full likelihood? This can be done by Kalman-filter, since ARMA-LoM has a state-space representation

- Cf. DynMo, I.2 for intro to KF
- Cf. sheet E802—PS4 & KF on ARMA(1,1) for the KF-formulas on a ARMA(1,1).

For $q=0$, can just use the decomposition at top of page, exploiting that $\mathcal{L}(y^0) = \mathcal{N}(\mu, \gamma_0)$ where $\mu = \phi(1)^{-1}c$, γ_0 can be computed using methods of sect. 3.

> In most applications, CMLE is favored — this is essentially for two reasons:

(Num) CMLE might be a lot easier to compute, potentially even allowing closed forms;
E.g. when (A0) + (A1).c) holds, so that $f(\cdot)$ is Gaussian, we have

$$\mathcal{L}(y_t \mid y^{t-1}, u^0) = \mathcal{N}\left(c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \theta_1 u^0_{t-1} + \ldots + \theta_q u^0_{t-q}, \sigma^2\right)$$

where u^0_{t-j} is the fixed value obtained by computing recursively the innovations implied by the initial values $y_0, \ldots, y_{1-p}, u_0, \ldots, u_{1-q}$.
└ u^0_{t-j} is a nonlinear function of $c, \underline{\phi}, \underline{\theta}$, so that still we need to apply numerical optimization
BUT: this additional-conditioning approach is still less comp.y heavy than computing the full likelihood $\mathcal{L}$

(Cons) When the likelihood is misspecified [Usually $\mathcal{L}(u_t)$ assumed Gaussian, but is not really Gaussian], both approaches are actually QMLE (quasi-MLE).
QMLE might still be consistent & asy normal (cf. E703, I.12: QMLE is a special M-estimator; cf. also E803 I.4.1: QMLE minimizes asy.y the Kullback-Leibler-divergence of the $(\underline{\phi}, \underline{\theta}, \sigma^2)$-parametrized missspec. distribution to the true distribution).
A very simple special case is $q=0$, i.e. for the AR(p). Then, CMLE (even the exact one, since there are no MA-terms) yields just OLS:

$$\sum_{t=1}^{T} \log f(y_t \mid y^{t-1}; \underline{\phi}, \underline{\theta}, \sigma^2) = \sum_{t=1}^{T} \log\left(\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y_t - c - \sum_{j=1}^{p} \phi_j y_{t-j})^2}{2\sigma^2}\right]\right)$$

$$= \text{stuff} - \sum_{t=1}^{T} \frac{u_t^2}{2\sigma^2} \leftarrow \text{SSR !}$$

and is thus consistent & asy. normal under (A0) & (A1).a), i.e. even if $\mathcal{L}(u_t) \neq \mathcal{N}$!

Other procedures on ARMA(p,q): 2-step OLS, 3-step-MLE, Yule-Walker

> In any (linear) context, OLS & MLE are always the first two go-to-options for estimation;
For ARMA-models, there are a number of other approaches as well for estimation; here we just briefly look at each one

2-step OLS-estimation for ARMA(p,q).

Consider a sample from an ARMA(p,q) satisfying (A0) & (A1).a)
Dufour & Jouini (2005) suggest the following algorithm:

[1] run AR(h)-regression on sample, with h large (so that residuals are approx.y WN; recall: if $\Theta(\cdot)$ is invertible, ARMA(p,q) can be represented as AR(∞)!) via OLS, & obtain residuals $\hat{u}_t$

[2] estimate ARMA(p,q) on $\{y_t, \hat{u}_t\}$ via OLS,

and they show that the $(\hat{c}, \hat{\underline{\phi}}, \hat{\underline{\theta}})$ so-obtained are consistent and asy. normal at rate $T^{-1/4}$ (slower convergence), provided $q = O(T^{1/2})$.

3-step-MLE.

Proposed by Dufour & Pelletier (2015):

[1]-[2] as above,

[3] use $(\hat{c}, \hat{\underline{\phi}}, \hat{\underline{\theta}}, \hat{\sigma}^2)$ as input to iterative optimization in CMLE (cf. above), e.g. Newton-Raphson, and perform 1 iteration.

$\propto$ SSR

Yule-Walker estimators.

This estimator estimates the autocorrelations & then backs out the parameters from the Yule-Walker-equations;
this is essentially a simple MM-estimator — for AR(p):
the Yule-Walker eqn.s read

$$E\left[(y_{t-k}-\mu)\left(y_t-\mu-\sum_{j=1}^{p}\phi_j(y_{t-j}-\mu)\right)\right]=0 \quad \forall k\in\{0,\dots,p\}$$

$$\Rightarrow \hat{\underline{\phi}}^{YW} = \hat{\underline{\Gamma}}^{-1}\hat{\underline{\gamma}}_p, \quad \hat{\Gamma} = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \cdots & \hat{\gamma}_0 \end{bmatrix}, \quad \hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_p)'$$

and the approach can be extended to ARMA(p,q) — however there the YWE.s are nonlinear, bringing its own issues.

One can show, for AR(p) ☑ (A0) & (A1).a), YW, OLS (and CMLE) are asymptotically equivalent.

Model selection & checking / diagnostics

> When attempting to analyze a given time series $\{y_t\}$, we typically don't know (p,q) [we don't even know whether (y_t) admits an ARMA-representation, but in light of the Wold-decomposition theorem, we turn a blind eye on this issue and assume there exist orders (p,q) s.th. (y_t) has an ARMA(p,q)-representation], so we have to select them somehow

> A traditional approach to this, called "Box-Jenkins-selection-approach" consists of a visual analysis of the Autocorrelation (AC) and partial autocorrelation (PAC) of the given sample:

5. UTSA V: Box-Jenkins-Approach for model selection

> If we have a stationary process (which we need to determine from the data!), then the question arises: which model should we pick to represent it?

> It turns out that nonstationarity of a process and also the specific nature of stationary ones can be inferred from a careful look at sample autocorrelations and partial autocorrelations!

OK

Def. 5.1: (Partial) autocorrelation (& -functions)

- Let $\{y_t\}$ be an arbitrary stationary process.
- We define:

(a) $\gamma_s := E[(y_t - \mu)(y_{t-s} - \mu)] = Cov(y_t, y_{t-s})$ as $\{y_t\}$'s s-th degree <u>autocovariance</u> (this implies $\gamma_0 = Var(y_t)$)

(b) $\rho_s := \dfrac{\gamma_s}{\gamma_0}$ — $\gamma_0 = \sqrt{\gamma_0} \cdot \sqrt{\gamma_0}$ as $\{y_t\}$'s s-th degree <u>autocorrelation</u>

(c) $\varphi_s := (\underline{\varphi})_s$ where $\underline{\varphi} \equiv E[\underline{y}_{L(s)} \underline{y}_{L(s)}']^{-1} E[\underline{y}_{L(s)} y_t]$ where $\underline{y}_{L(s)} := (1, y_{t-1}, \dots, y_{t-s})'$

as the s-th degree <u>partial autocorrelation</u> of $\{y_t\}$.

↳ It's just the coefficient of y_{t-s} in a projection of y_t onto $(y_{t-1}, \dots, y_{t-s})$

· Equivalently, we define

$$ACF(s) := \rho_s \quad , \quad PACF(s) := \varphi_s$$

as the autocorrelation-function and the partial-autocorrelation-function, respectively.

> We can compute γ_s and ρ_s only from the explicit solution of $\{y_t\}$, ie. from its $MA(\infty)$-representation!

↳ therefore (but not only for this!) we need stationarity!

> Also, we can only compute φ_s from $\{y_t\}$'s $AR(\infty)$-representation (obtained from multiplying the equation by $\Theta(L)^{-1}$)

↳ For this we need that $\Theta(L)$ is fully invertible, ie. that it has no unit roots (this condition is called 'inve...... of $\{y_t\}$')

↳ Note: the PACs of $\{y_t\}$ are for $q \geq 1$ not identical to the parameters of y_t's lags!!

> Computing ACF & PACF for ARMA(p,q) with sample moments

↳ if $\{y_t\}$ is stationary, we may compute the $\hat{\rho}_s$ and $\hat{\varphi}_s$ out of sample moments, ie:

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^{T} (y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2} \quad \text{and} \quad \hat{\varphi}_s = \left[\left(\sum_{t=s}^{T} \underline{y}^t_{L(s)} \underline{y}^{t\,\prime}_{L(s)} \right)^{-1} \left(\sum_{t=s}^{T} \underline{y}^t_{L(s)} y_t \right) \right]_s$$

↳ The different values for $\hat{\rho}_s$ and $\hat{\varphi}_s$ depending on s are then presented in a correlogram

↳ For computing the theoretical ρ_s and φ_s (done for identification), we can use the following approach:

↳ ACF: Yule-Walker-equations.

↳ Let $y_t \underset{\text{stat.}}{\sim} ARMA(p,q)$. Then, to get $\gamma_0, \dots, \gamma_s$, do the following:

Take y_t $h+1$ times (with $h = p+1$) and replace ν by $\nu = \mu(1-\alpha_1 - \dots - \alpha_p)$ (follows by $\mu := E[y_t] = \frac{\nu}{1-\alpha_1-\dots-\alpha_p}$), then multiply the $h+1$ equations by y_{t-s} for $s = 0, 1, \dots, h$ to get

$$\begin{cases} y_t \cdot y_t - y_t\mu = \alpha_1(y_t y_{t-1} - y_t\mu) + \dots + \alpha_p(y_{t-p}y_t - y_t\mu) \\ \qquad + \varepsilon_t y_t + \theta_1 \varepsilon_{t-1} y_t + \dots + \theta_q \varepsilon_{t-q} y_t \\ \vdots \\ y_t y_{t-h} - y_{t-h}\mu = \alpha_1(y_{t-1}y_{t-h} - y_{t-h}\mu) + \dots + \alpha_p(y_{t-p}y_{t-h} - y_{t-h}\mu) \\ \qquad + \varepsilon_t y_{t-h} + \theta_1 \varepsilon_{t-1} y_{t-h} + \dots + \theta_q \varepsilon_{t-q} y_{t-h} \end{cases}$$

Then take Expectations to get (recall $E[\varepsilon_t y_s] = 0$ for $t > s$!)

$$(*) \begin{cases} \gamma_0 = \alpha_1 \gamma_1 + \dots + \alpha_p \gamma_p + \sigma^2 + \theta_1(\alpha_1 + \theta_1)\sigma^2 + \dots + \langle \text{s.th. complicated} \rangle \\ \vdots \\ \gamma_h = \alpha_1 \gamma_{h-1} + \dots + \alpha_p \gamma_{h-p} + \dots \end{cases}$$

↳ then divide (*) by γ_0 to get correlations and solve for $\rho_1, \dots, \rho_h$ by iteration (OR: solve the first p+1 equations $(0, \dots, p$ which are lin. indep.!) for $\gamma_0, \dots, \gamma_p$ and then det. γ_{p+s} recursively)

↳ PACF: with ARMA, it's always the case that $\varphi_1 = \rho_1$.

Then, one can find φ_s by regressing

generally not white noise!

$$\tilde{y}_t = \varphi_1^* \tilde{y}_{t-1} + \varphi_2^* \tilde{y}_{t-2} + \dots + \varphi_s \tilde{y}_{t-s} + e_t \text{ where } \tilde{y}_t \equiv y_t - \mu$$

and φ_s^* is not the PAC of y_s, y_t but an irrelevant auxiliary parameter (→ Keep in mind assumption of invertibility

> Model selection using the Box-Jenkins-Approach:

Ok

- By comparing the sample ACF & PACF (which are computed under the assumption of stationarity & invertibility!) to what we theoretically expect from several types of models, we can find the most appropriate model:

what we'd expect to see in a correllogram, i.e. $h \mapsto AC(h), PAC(h)$:

Process	ACF	PACF
any non-stationary	very slow decay	
WN	$\rho_s = 0 \ \forall s$	$\varphi_s = 0 \ \forall s$
AR(1) w. $a_1 > 0$	direct exp. decay: $\rho_s = a_1^s$	$\varphi_1 = a_1$, $\varphi_s = 0 \ \forall s > 1$
AR(1) w. $\alpha_1 < 0$	oscillating decay: $\rho_s = a_1^s$	s.a.
AR(p)	decay towards zero, coeff. may oscillate	$\varphi_s > 0 \ \forall s \le p$, $\varphi_h = 0 \ \forall h > p$ } drop to zero
MA(1), $\theta_1 > 0$	$\rho_1 = \theta_1 \sigma^2$, $\rho_s = 0 \ \forall s > 1$	oscillating decay, $\varphi_1 > 0$
MA(1), $\theta_1 < 0$	$\rho_1 = \theta_1 \sigma^2$, $\rho_s = 0 \ \forall s > 1$	Geometric decay $\varphi_1 < 0$
ARMA(1,1) $a_1 > 0$	direct decay beginning after ρ_1	Oscillating decay beginning after φ_1
ARMA(1,1) $a_1 < 0$	osc. decay after ρ_1	exp. decay after φ_1
ARMA(p,q)	Decay (direct or osc.) after ρ_q	Decay (direct or osc.) after φ_p

> A more rigorous approach, based on testing: select

$$(p^*, q^*) \in \underset{(p,q) \in \{0,\dots,p_m\} \times \{0,\dots,q_m\}}{\arg\min} \mathcal{C}(p,q)$$

where $\mathcal{C}(p,q) = \log\left(\frac{1}{T} SSR(p,q) \right) + (p+q) \cdot \frac{C_T}{T}$

and $SSR(p,q) = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 \big|_{p,q}$

residuals of $\widehat{ARMA}(p,q)$, using presample $\{p_m - 1, \dots, 0\}$

always use maximal lag for presample so as to not give unfair advantage !

and we define three criteria:

AIC : $\mathcal{C}$ with $C_T = 2$

HQ : $\mathcal{C}$ with $C_T = 2 \log(\log(T))$

SC (or "BIC"): $\mathcal{C}$ with $C_T = \log T$.

A consistent order selection is guaranteed if

(i) $C_T \longrightarrow \infty$ for $T \to \infty$

(ii) $C_T / T \longrightarrow 0$ for $T \to \infty$

which is true for HQ & SC. (AIC asy. overestimates the true order with positive probability.)

An algebraic fact is

$$p^*_{AIC} \underset{\text{for } T \geq 16}{\geq} p^*_{HQ} \underset{\forall T}{\geq} p^*_{BIC}$$

$p^*_{AIC} \geq p^*_{BIC}$ for $T \geq 8$

> Once we have estimated an ARMA(p,q), we do several model diagnostic checks (to make sure we didn't fuck up somewhere...):

1) <u>Checking the "whiteness" of residuals $\{\hat{u}_t\}$.</u>

- Informal analysis
  - graph residuals — should roughly look like white noise with const. variance
  - Compute autocorrelations of $\{\hat{u}_t\}$ & check for <u>in</u>significance
- conduct tests: Portmanteau...

$H_0 : \rho_{u,1} = \dots = \rho_{u,h} = 0$ vs. $H_1 : \rho_{u,j} \neq 0$ for at least $j = 1, \dots h$

$$Q_h = T \sum_{j=1}^{h} \hat{\rho}_{u,j}^2 \overset{T,h\to\infty}{\longrightarrow} \chi^2(h - \underline{p - q})$$

- Reject H_0 if Q_h is large
- Problems
  - if h is too small, test may be strongly size distorted
  - if h is too large, power may be low
  - χ^2 approximation only works well for sufficiently large h

or (modified PM, ie) Ljung-Box :

- Modified Portmanteau Test to improve χ^2-approximation in small samples

$$\text{LB}_h = T(T+2) \sum_{j=1}^{h} \frac{\hat{\rho}_{u,j}^2}{T-j} \overset{d}{\longrightarrow} \chi^2(h-p-q)$$

- Choice for h
  - quarterly data: between 12 and 24 depending on sample size
  - monthly data: between 18 and 36 depending on sample size

or Lagrange-multiplier:

- Consider AR(h) model for error terms

  $u_t = \phi_1 u_{t-1} + \ldots + \phi_h u_{t-h} + \varepsilon_t$

  $H_0 : \phi_1 = \cdots = \phi_h = 0 \quad vs. \quad H_1 : \phi_1 \neq 0 \ \text{ or } \cdots \text{ or } \phi_h \neq 0$

- Test based on auxiliary model including regressors of AR model

  $\hat{u}_t = \nu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \phi_1 \hat{u}_{t-1} + \cdots + \phi_h \hat{u}_{t-h} + e_t$

  $\mathrm{LM}_h = TR^2 \xrightarrow{d} \chi^2(h)$

  - R^2 of auxiliary regression model
  - choose smaller h

Descriptions in MTSA (i.e. for VAR-case) are more thorough:

2) Checking the white noise property of the errors

- For carrying down the theory of methods to achieve this, it is helpful to define the following notation:

	Autocovariance	Autocorrelation
true values	$C_u(i) \equiv E[u_t u_{t-i}'] \; (= \Gamma_u(i))$	$R_u(i) \equiv \begin{cases} R_u = D_u^{-1} C_u(0) D_u^{-1} & \text{if } i=0 \\ 0 & \text{if } i \neq 0 \end{cases}$, $D_u \equiv \text{diag}\{\sqrt{E[u_{kt}^2]}\}_{k=1,\ldots,K}$, $C_u(0) = \Sigma_u$
est. using true error	$C_i \equiv \frac{1}{T}\sum_{t=1+i}^{T} u_t u_{t-i}'$	$R_i \equiv D^{-1} C_i D^{-1}$, $D \equiv \text{diag}\{\sqrt{\frac{1}{T}\sum_{t=1}^{T} u_{kt}^2}\}_{k=1,\ldots,K}$, $\equiv \hat{\gamma}_{kk}$
est. using residuals	$\hat{C}_i \equiv \frac{1}{T}\sum_{t=1+i}^{T} \hat{u}_t \hat{u}_{t-i}'$	$\hat{R}_i \equiv \hat{D}^{-1} \hat{C}_i \hat{D}^{-1}$

K

↳ Note that it makes no difference whether we can determine

$C_u(i) = 0 \ \forall i \neq 0 \quad$ or $\quad R_u(i) = 0 \ \forall i \neq 0$!

- Now for asymptotics, we define an object that comprises all expressions that we want to test

$C_h^* \equiv (C_1, \ldots, C_h)_{K \times Kh}$ → NOTE that $C_0 \equiv \hat{\Sigma}_u$ is not included!

$C_h = \frac{1}{T}\sum_{t=1+h}^{T} u_t u_{t-h}'$

... and bring it into vector form to obtain

$\text{vec}(C_h^*) \equiv c_h^*$ which is $K^2 h \times 1$

- If $u_t \sim SWN(0, \Sigma_u)$, we can actually show (though the proof is not of prime importance to us)

$(*_1) \quad \sqrt{T}\, c_h^* \xrightarrow{d} \mathcal{N}(0, \Sigma_{c_h^*})$

where we can show that $\Sigma_{c_h^*} \equiv E[c_h^* c_h^{*\prime}]$
$= I_h \otimes \Sigma_u \otimes \Sigma_u$

(using vec(·)-algebra and independence of 2^{nd}-moments of u_t)

- Using $(*_1)$, we can show with $\hat{D} \equiv \text{diag}\{\sqrt{\hat{\gamma}_{ii}}\}_{i=1,\ldots,K}$

$\sqrt{T}\, r_h^* \xrightarrow{d} \mathcal{N}(0, I_h \otimes R_u \otimes R_u)$

$r_h^* \equiv \text{vec}([R_1 \ldots R_h])$

and if we only take $\text{vec}(R_i)$:

$\sqrt{T}\, \text{vec}(R_i) \xrightarrow{d} \mathcal{N}(0, R_u \otimes R_u)$

→ Thus, if we want an individual test on $\rho_{mn}(i)$, i.e. the m-n-th element of R_i, we can simply

test $H_0: \rho_{mn}(i) = 0$ by

$$\frac{|r_{mn,i}|}{\sqrt{[R_u \otimes R_u]_{mn} \cdot \frac{1}{T}}} \geq 1.96 \quad (\text{for } \alpha = 0.05)$$

and for $m=n$ we obtain

$|r_{mm,i}| \geq 1.96/\sqrt{T}$ ← confidence bounds

since $\text{diag}\{R_u \otimes R_u\} = \{1\}_{i=1,\ldots,K^2}$

↳ Note that if we perform sequence of tests, since $\alpha = 0.05$ and since tests on R_i and $R_{j \neq i}$ are independent, we will reject 1 out of 20 times by chance! even if H_0 is correct

→ Generally, however, we don't have the true errors u_t, but just the residuals $\hat{u}_t$

↳ for equivalent definition of $\hat{c}_h^*$ and $\hat{r}_h^*$, we obtain

$\sqrt{T}\, \hat{c}_h^* \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{c}_h^*})$

where $\Sigma_{\hat{c}_h^*} \equiv (I_h \otimes \Sigma_u - \Psi_h) \otimes \Sigma_u$

where $\Psi_{Kh \times Kh}$ is a pos. s.d. correction matrix (exact form not central here, but do notice that by Ψ being psd, $\hat{c}_h^*$ actually has a smaller asympt. variance than c_h^*!)

and

$\sqrt{T}\, \hat{r}_h^* \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{r}_h^*})$

where $\Sigma_{\hat{r}_h^*} \equiv (I_h \otimes R_u - \Theta_h) \otimes R_u$

where again $\Theta_{Kh \times Kh}$ is psd.

→ it is worth noting that $\Psi_h, \Theta_h \to 0$ as $h \to \infty$, so for large h, the difference in the standard errors fades (however notice that in practice, the maximum h is limited by the size of the sample)

→ As before, we can use

$\sqrt{T}\, \text{vec}(\hat{R}_i) \xrightarrow{d} \mathcal{N}(0, \Sigma_R(i))$, $\Sigma_R(i) \equiv (R_u - R^*(i)) \otimes R_u$ ← some block of Θ_h

to do individual testing, but

↳ as before, elements of $\hat{R}_i$ may be correlated

↳ $\hat{R}_i, \hat{R}_{j \neq i}$ may now be correlated too, and individual testing may be misleading!

→ this result begs the question: how can we jointly test $R_u(1), \ldots, R_u(h)$?

→ there are two popular tests:

K

$(*_1)$

(1) Portmanteau-test (PMT):

test $H_0: R_h = (R_u(1), \ldots, R_u(h)) = 0$ ← resid. AC! by

$$
\begin{aligned}
Q_h &:= T \cdot \sum_{i=1}^{h} tr(\hat{R}_i' \hat{R}_u^{-1} \hat{R}_i \hat{R}_u^{-1}) \qquad \hat{R}_u = \hat{D}^{-1} \hat{\Sigma}_u \hat{D}^{-1} \\
&= T \textstyle\sum_i tr(\hat{R}_i' \hat{R}_u^{-1} \hat{R}_i \hat{R}_u^{-1} \hat{D}^{-1} \hat{D}) \\
&= T \textstyle\sum_i tr(\hat{D} \hat{R}_i' \hat{D} \hat{D}^{-1} \hat{R}_u^{-1} \hat{D}^{-1} \hat{D} \hat{R}_i \hat{D} \hat{D}^{-1} \hat{R}_u^{-1} \hat{D}^{-1}) \\
&= T \textstyle\sum_i tr(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}) \\
&= T \textstyle\sum_i \text{vec}(\hat{C}_0^{-1} \hat{C}_i)' \text{vec}(\hat{C}_i \hat{C}_0^{-1}) \qquad \text{using (II).2, 10.} \\
&= T \textstyle\sum_i \text{vec}(\hat{C}_i)' (I \otimes \hat{C}_0^{-1})(I \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_i) \qquad \text{using (II).2, 8.} \\
&= T\, \text{vec}(\hat{C}_h^*)' (I \otimes \hat{C}_0^{-1} \otimes \hat{C}_0^{-1}) \text{vec}(\hat{C}_h^*) \\
&= T \cdot \hat{c}_h^{*\prime} (I \otimes \hat{\Sigma}_u \otimes \hat{\Sigma}_u)^{-1} \hat{c}_h^* \quad \underset{\substack{h \to \infty \\ \frac{h}{T} \to 0}}{\overset{a}{\sim}} \chi^2(K^2(h-p))
\end{aligned}
$$

for every lag up to h, we have K^2 restrictions (R_i is $K \times K$), but used $K^2 \cdot p$ parameters for getting R's!

since correction term not included (so that we can actually still compute Q_h)

→ Note that we need to adjust for the $K^2 \cdot p$ parameters that we already estimated in order to get $\hat{u}_t$

↳ Caveats:

↳ since we don't include the psd. correction matrix Ψ_h, for 'small' h we 'divide' by a 'too large' variance

↳ this makes the test (relative to the benchmark $\chi^2(K^2(h-p))$-distr.) too conservative: it rejects less often than indicated by its significance level!

i.e. size is lower than chosen sign. level

↳ We can correct for some (though not all) of this 'overdivision' by instead relying on

$$\bar{Q}_h := T^2 \sum_{i=1}^{h} (T-i)^{-1}\, tr(\hat{R}_i' \hat{R}_u^{-1} \hat{R}_i \hat{R}_u^{-1})$$

bcs here we inflate the terms at higher lags to correct the downward bias of the covariance estimates...

$\hat{C}_i \equiv \frac{1}{T}\sum_{t=1+i}^{T} \hat{u}_t \hat{u}_{t-i}'$

downward bias bcs we lose more observations, but still divide every i by T

But note that this addresses another issue and only partly fixes the problem!

Actually 2 issues here:

(1) since Ψ_h not included, χ^2 is poor approximation for h small, esp. in small samples, need to increase h, o.w. PMT is undersized, χ^2-approx. works poorly

(2) If, given T, h gets large we add less and less meaningful covariances (look at Q_h: always dividing by T the terms that get smaller in it mechanically (since, given T, add less terms) means we add less than we should)

⇒ power loss if h↑, given fixed T

⇒ poss. solution: use $\bar{Q}_h$ with h large

→ thus, the PMT should only be consulted if h is much larger (by a factor > 4.5) than p, otherwise $\chi^2(K^2(h-p))$ provides poor approx. (to address (1)) and if T is large (such that (2) is not so problematic)

(2) Lagrange-Multiplier-Test:

→ LMT relies on assumption that

$u_t = D_1 u_{t-1} + \ldots + D_h u_{t-h} + v_t$ ← WN

and tests $H_0: D_i = 0 \ \forall i \iff u_t = v_t$

→ Test statistic is based on

$\hat{u}_t = \nu^0 + A_1^0 y_{t-1} + \ldots + A_p^0 y_{t-p} + D_1 \hat{u}_{t-1} + \ldots + D_h \hat{u}_{t-h} + \hat{\varepsilon}_t$

$\neq A_p$ since presence of $\hat{u}_t$'s; $\neq$ since additional error from resids we

and we test with

$\lambda_{LM}(h) = \text{vec}(\hat{D}_1, \ldots, \hat{D}_h)' \hat{\Sigma}_{\hat{D}}^{-1} \text{vec}(\hat{D}_1, \ldots, \hat{D}_h)$

$\xrightarrow{d} \chi^2(K^2 h)$

$= T \cdot \text{vec}(\hat{C}_h^*)' \hat{\Sigma}_{\hat{c}_h^*}^{-1} \text{vec}(\hat{C}_h^*)$ thus λ_{LM} is Q_h with presence of Ψ_h

also included here

should, in practice, not choose h too large since else we estimate too many parameters

→ Caveats:

↳ if T is small, $\chi^2(K^2 h)$ is only poor approx to distr. of λ_{LM}

↳ Test rejects far too often! even for moderately large T, approx can be poor, can use Rao-test instead

⇒ In practice, it is always best to evaluate both tests, do some individual testing and bring intuition into the game!

2) Testing residuals for Normality (Needed for forecast-intervals, cf. below

- **Jarque-Bera-test:**

- Idea: Check whether 3rd and 4th moments of standardized error terms u_t^s are in line with normal distribution

$$H_0 : \mathrm{E}[u_t^s]^3 = 0 \text{ and } \mathrm{E}[u_t^s]^4 = 3 \text{ vs. } \mathrm{E}[u_t^s] \neq 0 \text{ or } \mathrm{E}[u_t^s]^4 \neq 3$$

$$\mathrm{JB} = \frac{T}{6}\left[T^{-1}\sum_{t=1}^{T}(\hat{u}_t^s)^3\right]^2 + \frac{T}{24}\left[T^{-1}\sum_{t=1}^{T}(\hat{u}_t^s)^4 - 3\right]^2 \xrightarrow{d} \chi^2(2)$$

- rejection of non-normality may point to general misspecification

3) Testing (the residuals) for non-normality

- Testing the error terms (estimated by residuals) for normality is important for two reasons:
  - (i) quantifying forecast uncertainty chiefly relies on a normal (or 'Gaussian') error process
  - (ii) Non-normality can be indicative of more general violations of assumptions needed for consistent estimation (eg. t-distr. with low degree of freedom tends to have large to infinite 2nd moment – violating SWN-assumption!)
- The idea to approach the test relies on the fact that if

$x \sim \mathcal{N}(0,1) \Rightarrow E[X^3] = 0 \quad \wedge \quad E[X^4] = 3$
($\mathcal{N}$ is symmetric)

and consequentially,

$X_K \sim \mathcal{N}(0, I_K) \Rightarrow E[X.\hat{}3] = 0_K \wedge E[X.\hat{}4] = 3_K \equiv \begin{bmatrix}3\\ \vdots \\ 3\end{bmatrix}_K$
(each element to the 3rd power)

> Thus, we can test whether these restrictions hold – and if yes this is at least not indicative that we don't have Gaussian errors

> In the process of doing so, we substitute the true errors with residuals from our estimation which does not alter the asymptotic results, however:

(1) we first normalize the residuals:

$\hat{w}_t \equiv \hat{C}^{-1}\hat{u}_t$ where $\hat{u}_t$ is the residual vector and $\hat{C}$ is such that $\hat{\Sigma}_u = \hat{C}\hat{C}'$, ie. its a square root of $\hat{\Sigma}_u$.

and where $\hat{\Sigma}_u \equiv \frac{1}{T-Kp-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$

(Note that there are generally infinitely many orthogonal decompositions of $\hat{\Sigma}_u$ – eg. Choleski, Spectral, etc. ...) – Caution is necessary: the $\hat{w}_t$-vectors and thus test results will generally differ depending on which decomposition we use!
↳ per convention, we use Choleski ($\hat{C}$ is lower triangular; it can also be obtained by spectral decomposition: $\hat{C} \equiv P\Lambda^{1/2}G$ where G is some orthogonal rotation matrix), but it is wise to perform the test using several decompositions and overall reject H_0 if it is rejected with one particular decomposition!

(2) Upon this normalization, and if $u_t \sim \mathcal{N}(0, \Sigma_u)$, we obtain

$\hat{w}_t \xrightarrow{d} \mathcal{N}(0, I_K)$ if $\operatorname{plim}\hat{C} = C$ where $C: \Sigma_u = CC'$

and as we saw we asymptotically have

$E[w_t.\hat{}3] = 0 \wedge E[w_t.\hat{}4] = 3_K$
($w_t \equiv C^{-1}u_t$)

(3) This justifies the estimates

$\hat{b}_1 \equiv (\hat{b}_{11}, \dots, \hat{b}_{K1})', \quad \hat{b}_2 \equiv (\hat{b}_{12}, \dots, \hat{b}_{K2})'$

where $\hat{b}_{k1} := \frac{1}{T}\sum_{t=1}^{T}\hat{w}_{kt}^3$ and $\hat{b}_{k2} := \frac{1}{T}\sum_{t=1}^{T}\hat{w}_{kt}^4$

for which we can obtain (proof involved)

$$\sqrt{T}\begin{pmatrix}\hat{b}_1 \\ \hat{b}_2 - 3_K\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix}6\cdot I_K & 0 \\ 0 & 24\cdot I_K\end{bmatrix}\right)$$

(4) Finally, we test the hypothesis...

$H_0: u_t$ ($\Leftrightarrow y_t$) is normally distr.

... by one of the following tests

$\hat{\lambda}_s = \hat{b}_1'\Sigma_{\hat{b}_1}^{-1}\hat{b}_1 = T\cdot\hat{b}_1'\hat{b}_1/6 \xrightarrow[H_0]{d} \chi^2(K)$ (skewness)

$\hat{\lambda}_k = (\hat{b}_2 - 3_K)'\Sigma_{\hat{b}_2}^{-1}(\hat{b}_2 - 3_K) = T(\hat{b}_2 - 3_K)'(\hat{b}_2 - 3_K)/24 \xrightarrow[H_0]{d} \chi^2(K)$ (kurtosis)

$\hat{\lambda}_{ks} = \hat{\lambda}_s + \hat{\lambda}_k \xrightarrow[H_0]{d} \chi^2(2K)$

3) Testing for structural breaks

- **Split-sample-test / Break-point-test, or Chow-test:**

1

4) Testing for Structural Breaks

> Suppose that for the given model

$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad t \in \{1, \dots, T\}$

the parameters $v, A_1, \dots, A_p$ change over time, or more specifically, they change in period $T_1 < T$
↳ i.e. we actually have two models: one for each sample subperiod $t \in \{1, \dots, T_1\}$ and $t \in \{T_1+1, \dots, T\}$

> If we suspect this situation for a given model, we can test whether this is true by following along these steps

1] Define the partitioned model analogously to the system definition in (I).9: (clumsy notation for "partitioned")

$Y^\pi = [Y_{(1)} \; Y_{(2)}] = B^\pi Z^\pi + U^\pi$

where $Y_{(1)} \equiv [y_1, \dots, y_{T_1}]_{K\times T_1}$, $Y_{(2)} \equiv [y_{T_1+1}, \dots, y_T]_{K\times(T-T_1)}$

$B^\pi \equiv [B_1 \; B_2] = [v_1 \; A_{11} \dots A_{p1} \; v_2 \; A_{12} \dots A_{p2}]_{K\times 2(Kp+1)}$

and

$Z^\pi \equiv \begin{bmatrix} Z_{(1)} & 0 \\ 0 & Z_{(2)} \end{bmatrix}_{2(Kp+1)\times T}$ where $Z_{(1)} \equiv [Z_0 \dots Z_{T_1}]_{(Kp+1)\times T_1}$

2] Test $H_0: B_1 = B_2$ vs $H_1: B_1 \neq B_2$ through a Likelihood-ratio test (→ for info on likelihood-functions, see (II).3):

(1) estimate restricted model (ie. restricted by $H_0: B_1 = B_2$) through OLS (for linear est. is equivalent to (quasi-) ML) to obtain $\hat{B}_1^R = \hat{B}_2^R = YZ'(ZZ')^{-1}$ ← here we use the non-part. objects since we estimate across all $t \in \{1, \dots, T\}$

and $\tilde{\Sigma}_u^R = \frac{1}{T}\hat{U}^R\hat{U}^{R\prime}$ (ML-estimator)

where $\hat{U}^R$ is just the normal residual matrix from estimation over whole sample period.

1/ practical note: often it's a good idea to delete a few observations after T_1 (if affordable!) so as to not make estimates of $B_{(2)}$ influenced by pre-break-values!!

(2) estimate unrestricted model (ie. $H_1: B_1 \neq B_2$ is allowed) through separate OLS regressions 1/

2

→ $\hat{B}_1^{UR} = Y_{(1)}Z_{(1)}'(Z_{(1)}Z_{(1)}')^{-1}$ (analogously for $\hat{B}_2^{UR}$)

→ Caution: there are now different approaches towards $\tilde{\Sigma}_u^{UR}$, since principally covariance of $u_{t\leq T_1}$ can be different from $u_{t>T_1}$! Theoretically, we have

$\tilde{\Sigma}_{u^*}^{UR} = \begin{bmatrix}\tilde{\Sigma}_{u(1)}^{UR} & 0 \\ 0 & \tilde{\Sigma}_{u(2)}^{UR}\end{bmatrix}_{KT\times KT}$ instead of $I_T \otimes \Sigma_u$
($u^* \equiv \operatorname{vec}(U^\pi)$)

↳ algebraically, this can complicate derivation of log-LH-function, so we won't further investigate, but it's important to keep in mind.

(3) construct Likelihood-Ratio-statistic, ie.

$\lambda := 2\cdot(\ln L_0 - \ln L_1) < 0$ (ie. $\frac{L_0}{L_1} \in [0,1]$) (log-LH with parameters under H_0)

$\Leftrightarrow \lambda_{LR} = -2(\ln L_0 - \ln L_1) > 0$

$= 2(\ln L_1 - \ln L_0) = 2(\ln L_{UR} - \ln L_R)$

$= T\cdot(-\log(\det(\tilde{\Sigma}_u^{UR})) + \log(\det(\tilde{\Sigma}_u^R)))$

$\Leftrightarrow \lambda_{LR} = T(\log(|\tilde{\Sigma}_u^R|) - \log(|\tilde{\Sigma}_u^{UR}|))$

this is case for uniform variance in UR

since with a restriction, we can never do better than without restriction – if restr. is true we (asymptotically!) at best achieve same fit as unrestricted

→ consistent: if H_0 false, then $\ln L_0 < \ln L_1$ and test will get arb. large, rejecting with Prob. 1; if H_0 true, $\ln L_0 - \ln L_1 \xrightarrow{P} 0$ and test will never reject.

↳ one can show

$\lambda_{LR} \overset{a}{\sim} \chi^2(\#\text{ of restrict.})$
(# of parameters in B^π and Σ_{u^*} that are allowed to differ between two subsamples)

(4) What are relevant test statistics?
↳ In practice (ie. JMulTi), we see

① Split-sample-test: λ_{SS} in JMulTi
↳ VAR mean-parameters (ie $B_{(1)}, B_{(2)}$) allowed to differ, $\tilde{\Sigma}_u^{UR}$ is computed uniformly (ie. $\tilde{\Sigma}_{u(1)}^{UR} = \tilde{\Sigma}_{u(2)}^{UR}$)
↳ # restrict. $= K(Kp+1)$ (no. of elements in $B_{(1)}, B_{(2)}$)
$H_0: v, A_i$ stable

② Break-Point-test: λ_{BP} in JMulTi
↳ B's and $\tilde{\Sigma}_{u(i)}^{UR}$ computed seperately
↳ # restr. $= K(Kp+1) + \frac{K(K+1)}{2}$ (= # elements in $B_{(i)}$ and # of distinct elements in $\tilde{\Sigma}_{u(i)}^{UR}$)
$H_0: v, A_i, \Sigma_u$ stable

JMulTi implements bootstrapped versions

3

③ Chow forecast test: λ_{CF} in JMulTi
↳ idea: estimate B with $t \in \{1, \dots, T_1\}$ and then use the estimates to forecast into $t \in \{T_1+1, \dots, T\}$ (see (I).11)
↳ two possible tests

(1) estimate the Σ_u based on all $t \in \{1, \dots, T\}$ and reject if difference of $|\tilde{\Sigma}_u|$ and $|\tilde{\Sigma}_{u(1)}|$ is too large ($\chi^2(K^2p + K + (K(K+1))/2)$)
(residuals from T_1+1 onwards are forecast errors!)
(variance matrix only of first T_1 observations)

Version in JMulTi

JMulTi implements bootstrapped version

①-③ implemented simultaneously

(2) use estimated forecast error matrix for the $t \in \{T_1+1, \dots, T\}$ and use it to construct Wald-test with forecast errors ($\chi^2(hK)$, where h is no. of forecasted periods)
(in practice its better to divide by hK and use F-critVals from $F(Kh, T-Kp-1)$)

↳ The forecast tests have great advantage that you can still do them even when size of $\{T_1+1, \dots, T\}$ is not even sufficient for estimation to work (algebraically)!

→ Caution: technically, these tests don't indicate (by rejecting H_0) that there is a break really in the conjectured period, but just that the two sub-samples are different!

⇒ In practice, though, the small-sample distributions of the λ's might vary quite dramatically from their limiting distr.