Empirical Methods
Time Series Econometrics I & II
Stationary Processes

Martin Wagner

Department of Economics, University of Klagenfurt

Winter Term 2025

This version: July 30, 2025

Content of this Chapter

- Stochastic Processes
- Autoregressive Processes
- The Backward Shift Operator
- Autoregressive Moving Average Processes
- The Wold Representation
- Multivariate ARMA Processes
- Deterministic Components

- Let $x_t : \Omega \to \mathbb{R}^n$ denote a random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. [Sometimes we write sloppily $x_t \in \mathbb{R}^n$.]

Definition

A stochastic process is an (indexed) family of random variables $\{x_t\}_{t \in \mathcal{T}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- In time series econometrics $\mathcal{T}$ often is considered to be $\mathbb{N}, \mathbb{Z}$ (discrete-time) or $[0, 1]$, $\mathbb{R}$ (continuous time).
- For some given $\omega \in \Omega$, the sequence $\{x_t(\omega)\}_{t \in \mathcal{T}}$ is referred to as realization, trajectory or sample path of the stochastic process $\{x_t\}_{t \in \mathcal{T}}$.
- An observed time series is thus a finite subset of a trajectory.

Definition

Let $\mathbf{T}$ be the set of all vectors of the form:

$$\mathbf{T} = \{t = (t_1, \ldots, t_n)' \in \mathcal{T}^n : t_1 < t_2 < \ldots < t_n, n \in \mathbb{N}\}$$

Then the finite-dimensional distribution functions of $\{x_t\}_{t \in \mathcal{T}}$ are the functions $\{F_t(\cdot), t \in \mathbf{T}\}$ defined for $t = (t_1, \ldots, t_n)'$ by:

$$F_t(z) = \mathbb{P}\{x_{t_1} \leq z_1, \ldots, x_{t_n} \leq z_n\},$$

for $z = (z_1, \ldots, z_n)' \in \mathbb{R}^n$.

Theorem

The probability distribution functions $\{F_t(\cdot), t \in \mathbf{T}\}$ are the finite-dimensional distribution functions of some stochastic process if and only if for any $n \in \mathbb{N}$, $t = (t_1, \ldots, t_n)'$ and $1 \leq i \leq n$ it holds that:

$$\lim_{z_i \to \infty} F_t(z) = F_{t(i)}(z(i)),$$

where $t(i)$ and $z(i)$ are the $(n-1)$-dimensional component vectors obtained by deleting the i-th components of t and z respectively.

- The condition in the theorem merely states that the marginal distributions of $F_t(\cdot)$ correspond to the lower dimensional distribution functions.
- A proof of the result is given, e. g., in Davidson (1994, Theorem 12.4).

Definition

Let $\{x_t\}_{t\in\mathcal{T}}$ be a stochastic process such that $\mathbb{E}(x_t'x_t) < \infty$ for all $t \in \mathcal{T}$. The autocovariance function (ACF) $\Gamma_x(r,s)$ of $\{x_t\}_{t\in\mathcal{T}}$ is defined as:

$$\Gamma_x(r,s) := \mathbb{E}((x_r - \mathbb{E}(x_r))(x_s - \mathbb{E}(x_s))') \quad \text{for} \quad r,s \in \mathcal{T}$$

Definition

The stochastic process $\{x_t\}_{t\in\mathcal{T}}$ is called (weakly) stationary, if:

(i) $\mathbb{E}(x_t'x_t) < \infty$ for all $t \in \mathcal{T}$

(ii) $\mathbb{E}(x_t) = \mu$ for all $t \in \mathcal{T}$

(iii) $\mathbb{E}(x_r x_s')$ only depends upon $r - s$, i. e.,:

$$\Gamma_x(r,s) = \Gamma_x(r+v, s+v)\ \forall\ r,s,v \text{ such that } r,s,r+v,s+v \in \mathcal{T}$$

- Weak stationarity allows to redefine:

$$\Gamma_x(r, s) = \Gamma_x(r - s, 0) = \Gamma_x(h)$$

  for $h := r - s$.

- This means that the ACF can be redefined as a function of $h = r - s$, i. e., the time difference or lag only.

- When considering a weakly stationary process, typically $\mathcal{T} = \mathbb{Z}$ (or $\mathbb{R}$ or $[0, 1]$).

Definition

The stochastic process $\{x_t\}_{t\in\mathbb{Z}}$ is called strictly stationary if the joint distribution functions of $(x_{t_1},\ldots,x_{t_k})'$ and $(x_{t_1+h},\ldots,x_{t_k+h})'$ are identical for all k, for all $t_1,\ldots,t_k$ and for all $h\in\mathbb{Z}$.

- Considering a set $\mathcal{T}\neq\mathbb{Z}$ requires the above property for all values in $\mathcal{T}$.
- Clearly, if $\mathbb{E}(x_t'x_t)<\infty$ and $\{x_t\}_{t\in\mathbb{Z}}$ is strictly stationary, then it is also weakly stationary.
- Weak stationarity, in general, does not imply strong stationarity, with the exception of Gaussian processes, i. e., stochastic processes with Gaussian finite dimensional distributions.

- White Noise $\{\varepsilon_t\} \sim \mathrm{WN}(0, \Sigma)$:
  - $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma$
  - $\mathbb{E}(\varepsilon_t) = 0$
  - $\mathbb{E}(\varepsilon_t \varepsilon_s') = 0 \; \forall \; t \neq s$
- Moving Average Process of order q $[\mathrm{MA}(q)$ Process$]$:
  - $x_t = \varepsilon_t + B_1 \varepsilon_{t-1} + \ldots + B_q \varepsilon_{t-q}$, with $\{\varepsilon_t\} \sim \mathrm{WN}(0, \Sigma)$, $B_i \in \mathbb{R}^{n \times n}$ and $B_q \neq 0$
  - It can easily be shown that $\{x_t\}_{t \in \mathbb{Z}}$ is weakly stationary.
  - It holds that $\Gamma_x(h) = 0$ for all $|h| > q$.

Autoregressive Processes

$\mathrm{AR}(p)$ Process

A stochastic process $\{y_t\}_{t\in\mathcal{T}}$ is called autoregressive process of order p, $\mathrm{AR}(p)$, if for every $t \in \mathcal{T}$ it holds that:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t,$$

where $\{\varepsilon_t\}_{t\in\mathcal{T}} \sim \mathrm{WN}(0, \Sigma)$ with $\Sigma > 0$ and $A_i \in \mathbb{R}^{n\times n}$ with $A_p \neq 0$.

- The above equation is a linear stochastic difference equation with constant coefficients and the process $\{y_t\}_{t\in\mathcal{T}}$ is defined as the solution to it.
- Whether $\{y_t\}_{t\in\mathcal{T}}$ is stationary or not depends upon the parameters and which solution to the above equation is considered.
- We start by considering the special case $n = p = 1$.

- The input sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is considered to be "given" (this does not mean observed).
- We also consider for the moment y_0 as given.
- Note that this is, after all, a difference equation, so we may need to consider a starting value.

The forward part of the solution for $t > 0$ is obtained by recursive substitution:

$$
\begin{aligned}
y_1 &= \rho y_0 + \varepsilon_1 \\
y_2 &= \rho y_1 + \varepsilon_2 \\
&= \rho(\rho y_0 + \varepsilon_1) + \varepsilon_2 \\
&= \rho^2 y_0 + \varepsilon_2 + \rho\varepsilon_1 \\
&\;\;\vdots \\
y_t &= \rho^t y_0 + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}, \quad t > 0
\end{aligned}
$$

The backward part of the solution for $t < 0$ is also obtained by recursive substitution:

$$\begin{aligned} y_0 &= \rho y_{-1} + \varepsilon_0 \\ y_{-1} &= \rho^{-1} y_0 - \rho^{-1} \varepsilon_0 \\ y_{-2} &= \rho^{-1}(y_{-1} - \varepsilon_{-1}) \\ y_{-2} &= \rho^{-2} y_0 - \rho^{-2} \varepsilon_0 - \rho^{-1} \varepsilon_1 \\ &\vdots \\ y_t &= \rho^t y_0 - \sum_{j=t}^{-1} \rho^j \varepsilon_{t-j}, \quad t < 0 \end{aligned}$$

- Now, with the solution obtained for given y_0 we can address the question whether the solution is stationary (on $\mathbb{N}$, on $\mathbb{Z}$):
  - The first necessary condition is: $\mathbb{E}(y_0) = 0$, which implies $\mathbb{E}(y_t) = 0$ for all t.
  - In order to have finite variance we need $\mathbb{E}(y_0^2) < \infty$.
  - Thus, we have $\mathbb{E}(y_0) = 0$, $\mathbb{E}(y_0^2) = \sigma_0^2$ as necessary conditions for stationarity of the solution.
- Checking stationarity, e. g., on $\mathbb{Z}$ requires to calculate mean, variance and ACF of $\{y_t\}_{t \in \mathbb{Z}}$.

- First, we consider $t \in \mathbb{N}$ and the forward solution:

$$\begin{aligned}
\mathbb{E}(y_t) &= 0 \\
\text{Var}(y_t) &= \mathbb{E}\left(\rho^t y_0 + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}\right)^2 \\
&= \rho^{2t}\sigma_0^2 + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\varepsilon^2 + 2\rho^t \sum_{j=0}^{t-1} \rho^j \mathbb{E}(y_0 \varepsilon_{t-j})
\end{aligned}$$

- If $\mathbb{E}(y_0\varepsilon_t) = 0$ for all $t > 0$, the above simplifies to:

$$\text{Var}(y_t) = \rho^{2t}\sigma_0^2 + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\varepsilon^2$$

- Thus, the variance is constant over time, if and only if:

$$\sigma_0^2 = \frac{1}{1-\rho^2}\sigma_\varepsilon^2 = \mathrm{Var}(y_t)$$

- Next, we consider the ACF, whose computation is simplified by using:

$$y_t = \rho^h y_{t-h} + \sum_{j=0}^{h-1} \rho^j \varepsilon_{t-j}$$

- Under the already used assumptions on y_0, we obtain for $0 \leq h \leq t$:

$$\begin{aligned}\mathrm{Cov}(y_t, y_{t-h}) &= \mathbb{E}(y_t y_{t-h}) = \mathbb{E}\left[\left(\rho^h y_{t-h} + \sum_{j=0}^{h-1} \rho^j \varepsilon_{t-j}\right) y_{t-h}\right] \\ &= \rho^h \mathbb{E}(y_{t-h}^2) + 0 = \rho^h \frac{1}{1-\rho^2} \sigma_\varepsilon^2\end{aligned}$$

- Thus, it seems we have established stationarity on $\mathbb{N}$, but the question is whether y_t is well-defined for $t \to \infty$.
- In other words, what is the meaning of $\lim_{t\to\infty} \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}$?

Definition

Denote with $L^2(\Omega, \mathcal{F}, \mathbb{P})$ the set of all square integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, i. e., of all random variables $y : \Omega \to \mathbb{R}$ such that $\mathbb{E}(y^2) < \infty$.

- We can define an "inner product" on L^2 as: $\langle y_1, y_2 \rangle = \mathbb{E}(y_1 y_2)$ and an associated "norm" $\|y\| = \sqrt{\mathbb{E}(y^2)}$.
- But now, what is a Hilbert space? A complete metric space whose norm is generated by an inner product.
- The above is not (yet) an inner product: $\mathbb{E}(y^2) = 0$ for all random variables y that are equal to 0 up to $\mathbb{P}$-null sets (rather than only for 0).
- We, therefore, need to consider equivalence classes of random variables, i. e., y_1 is equivalent to y_2 if and only if $\mathbb{P}(y_1 = y_2) = 1$.

- This now allows us to consider a Hilbert space defined on the equivalence classes, which we denote as $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.
- It remains to show completeness, i. e., it remains to show that every Cauchy sequence converges **in** $\mathcal{L}^2$.
- A proof of completeness is contained in Brockwell and Davis (1991, Section 2.10).
- The Cauchy criterion is going to be very useful to show well-"behavedness" of the limit mentioned before.

A Convergence Proposition

Proposition

If $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of random variables such that $\sup_t \mathbb{E}|\varepsilon_t| < \infty$ and if $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the series $\sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$ converges absolutely with probability 1. If, in addition, $\sup_t \mathbb{E}(\varepsilon_t)^2 < \infty$, then the series converges in mean square sense to the same limit.

- Note that in the above result no stationarity assumption has been made.
- The above result can also be used to study the limit behavior of the series considered, if we choose the following starting value:

$$y_0 = \sum_{j=0}^{\infty} \rho^j \varepsilon_{0-j}, \quad |\rho| < 1,$$

  which implies that $y_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$ and that the system was "started in the infinite past".
- Clearly, this yields a stationary solution on $\mathbb{Z}$.

- $\{y_t\}_{t\in\mathbb{Z}}$ as defined above fulfills the assumptions of the previous proposition and defines a stationary autoregressive process of order 1, i. e., an AR(1) process.
- In particular, it holds that $\{y_t\}_{t\in\mathbb{Z}}$ only depends upon ε_j's dated not later than t, i. e., $j \leq t$.
- In this sense, the given sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is causal for $\{y_t\}_{t\in\mathbb{Z}}$, or equivalently that $\{y_t\}_{t\in\mathbb{Z}}$ is a causal autoregressive process.
- Is this the only (stationary) solution on $\mathbb{Z}$?
- Consider two solutions and their difference:

$$\begin{aligned} y_t &= \rho y_{t-1} + \varepsilon_t \\ y_t^* &= \rho y_{t-1}^* + \varepsilon_t \\ y_t - y_t^* &= \rho(y_{t-1} - y_{t-1}^*) \end{aligned}$$

- Thus, the difference between any two solutions is a solution to the homogenous equation: $y_t = \rho y_{t-1}$.
- The only stationary solution of the homogenous equation (on $\mathbb{Z}$) is $\{y_t\}_{t\in\mathbb{Z}} = 0$ (with probability 1).
- This implies that the above solution is the only stationary solution on $\mathbb{Z}$ for the equation $y_t = \rho y_{t-1} + \varepsilon_t$ with $|\rho| < 1$.
- Set, for brevity, $y_0 = 0$ and verify that the "backward" solution from above is not well-defined in the limit.
- The next question is: Are there also stationary solutions on $\mathbb{Z}$ in case that $|\rho| = 1$ or $|\rho| > 1$?

- Is there also a stationary solution on $\mathbb{Z}$ in case $|\rho| > 1$?

$$y_t = -\sum_{j=1}^{\infty} \left(\frac{1}{\rho}\right)^j \varepsilon_{t+j}, \quad t \in \mathbb{Z}$$

- For all t, y_t only depends on later dated ε_j's.
- In this case, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is not causal for $\{y_t\}_{t\in\mathbb{Z}}$.
- This solution can be found analogously to the stationary solution found in the stable case, but we can also resort to the so-called backward shift operator L.

The Backward Shift Operator

- Now, some for the moment "magical" calculations are performed.
- Considering a solution on $\mathbb{Z}$ allows writing:

$$
\begin{aligned}
y_t = \rho y_{t-1} + \varepsilon_t, &\quad t \in \mathbb{Z} \\
(1 - \rho L)\{y_t\} &= \{\varepsilon_t\},
\end{aligned}
$$

  where the second equation is short-hand notation.[1]

- Now, suppose that one can calculate with $(1 - \rho L)$ like with complex valued functions, e. g., $(1 - \rho z)$:

$$
(1 - \rho z)^{-1} = \frac{1}{1 - \rho z} = \sum_{j=0}^{\infty} \rho^j z^j, \quad \forall\, |z| < \frac{1}{\rho}
$$

[1] Proper notation is: $(1 - \rho L)\{y_t\}_{t \in \mathbb{Z}} = \{\varepsilon_t\}_{t \in \mathbb{Z}}$

- Consider now $(1 - \rho L)\{y_t\} = \{\varepsilon_t\}$ and use the above result to obtain formally:

$$\begin{aligned}\{y_t\} &= (1 - \rho L)^{-1}\{\varepsilon_t\}, \quad |\rho| < 1 \\ &= \left(\sum_{j=0}^{\infty} \rho^j L^j\right)\{\varepsilon_t\} \\ &= \sum_{j=0}^{\infty} \rho^j \{\varepsilon_{t-j}\},\end{aligned}$$

  i. e., these calculations lead to the unique stationary solution on $\mathbb{Z}$.

- This type of calculations is not "only magic", but has, of course, some proper mathematical background (see, e. g., Deistler, 1975 for one approach).

- L is a linear operator on the set of all (n-dimensional) stochastic processes with $\mathcal{T} = \mathbb{Z}$ with uniformly bounded second (non-central) moments.
- Note that L is bijective for $\mathcal{T} = \mathbb{Z}$, thus L^{-1} is well-defined.
- Define a set $\overline{T}$ of operators in L as:

$$\overline{T} = \left\{ \sum_{s=-\infty}^{\infty} B_s L^s \;\middle|\; \sum_{s=-\infty}^{\infty} \|B_s\| < \infty \right\}$$

- Furthermore, define a set of complex valued functions:

$$\overline{P} = \left\{ B(\lambda) = \sum_{s=-\infty}^{\infty} B_s e^{-i\lambda s} : [-\pi, \pi] \to \mathbb{C}^{n \times n} \;\middle|\; \sum_{s=-\infty}^{\infty} \|B_s\| < \infty \right\}$$

- Addition and multiplication are defined straightforwardly on both $\overline{T}$ and $\overline{P}$.
- Then $\mathcal{J} : \overline{T} \to \overline{P}$ with:

$$\mathcal{J}\left(\sum_s B_s L^s\right) = \sum_s B_s e^{-i\lambda s}$$

  defines an isomorphism between $\overline{T}$ and $\overline{P}$.
- On $\overline{P}$, calculations can be done "as usual", e. g., calculate the inverse and then use isomorphism.
- Inversion on $\overline{P}$ requires that the function $B(\lambda)$ is invertible over the interval $\lambda \in [-\pi, \pi]$.
- Thus, one needs to assume that $B(\lambda) = \sum_s B_s e^{-i\lambda s}$ has full rank for all $-\pi \leq \lambda \leq \pi$.
- This is equivalent to the assumption that:

$$\left|\sum_s B_s z^s\right| \neq 0 \;\; \forall \; z \text{ with } |z| = 1$$

- In the $\mathrm{AR}(1)$ case, it holds that $B(L) = 1 - \rho L$.
- The root of $B(z) = 1 - \rho z$ is at $z^* = \rho^{-1}$, which has absolute value not equal to one, in case $|\rho| \neq 1$.
- This implies that:
$$B^{-1}(L) = \sum_{s=-\infty}^{\infty} C_s L^s,$$
with $\sum_s |C_s| < \infty$ exists, i. e., is well-defined, in $\overline{T}$.
- The summability assumption above corresponds to the condition in the proposition about convergence in mean square above (i. e., to $\sum_j \psi_j \varepsilon_{t-j}$).

- Consider for $|\rho| > 1$:

$$(1 - \rho L)\{y_t\} = \{\varepsilon_t\}$$

- It holds that:

$$\begin{aligned}(1-\rho z)^{-1} &= \frac{1}{1-\rho z} = \frac{1}{\rho z(\frac{1}{\rho z} - 1)} \\ &= -\frac{1}{\rho z}\frac{1}{1-\frac{1}{\rho z}} = -\frac{1}{\rho z}\sum_{j=0}^{\infty}\left(\frac{1}{\rho}\right)^j z^{-j} \\ &= -\sum_{j=1}^{\infty}\left(\frac{1}{\rho}\right)^j z^{-j}\end{aligned}$$

- Thus, one can write:

$$\{y_t\} = (1-\rho L)^{-1}\{\varepsilon_t\} = -\sum_{j=1}^{\infty}\left(\frac{1}{\rho}\right)^j \{\varepsilon_{t+j}\}$$

- Now, consider the case $\rho = 1$:

$$y_t = y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}, \;\; t \geq 0$$

$$y_t = y_0 - \sum_{j=t}^{-1} \varepsilon_{t-j}, \;\; t < 0$$

- $\mathrm{Var}(y_t)$ grows in t, therefore this is not a stationary solution for any choice w.r.t. y_0.
- The solution does not converge in $\mathcal{L}^2$ for $t \to \infty$.
- This raises the question: Are there stationary solutions (on $\mathbb{N}$ or $\mathbb{Z}$)?
- Remember, the set of all solutions is given by a particular solution and the solution set of the homogenous equation.

- The homogenous equation is $y_t = y_{t-1}$, which has as solution set $y_t = y^*$ for some y^*.
- Thus, in case y^* is a random variable with finite variance, the solution to the homogenous equation is stationary.
- This implies that there is no stationary solution in the set of all solutions.
- The characteristic polynomial $a(z) = 1 - z$ has, in case $\rho = 1$, its root at the point $z = 1$, i. e., it has a unit root.
- Similar arguments apply for any $|\rho| = 1$ and $\rho \neq 1$.
- Thus, in case of a unit root, the $\mathrm{AR}(1)$ equation does not have a stationary solution (on either $\mathbb{N}$ or $\mathbb{Z}$).

- We now consider the autoregressive equation of order p:

$$a(L)\{y_t\} = \{\varepsilon_t\}$$
$$y_t = a_1 y_{t-1} + \ldots + a_p y_{t-p} + \varepsilon_t$$

- Existence and uniqueness of a stationary solution on $\mathbb{Z}$ depends upon the roots of the characteristic polynomial:

$$a(z) = 1 - a_1 z - \ldots - a_p z^p$$

- The fundamental theorem of algebra states that:

$$a(z) = -a_p \prod_{i=1}^{p}(z - z_i) = \prod_{i=1}^{p}(1 - \rho_i z),$$

with $0 \neq z_i = \rho_i^{-1}$ denoting the roots of $a(z)$.

- Similar calculations (combined for the various roots) as above lead to:

$$\begin{aligned} y_t &= \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \;\; |z_i| \neq 1, \; i = 1, \ldots, p \\ &= \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \;\; |z_i| > 1, \; i = 1, \ldots, p \end{aligned}$$

- In the second case, as before in the AR(1) case, it holds that $\{y_t\}_{t \in \mathbb{Z}}$ is a causal autoregressive process.
- The coefficients c_j can also be obtained by coefficient comparison.

- An $\mathrm{ARMA}(p,q)$ process is defined as a solution of:

$$y_t = a_1 y_{t-1} + \ldots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q},$$

  with $a_p \neq 0$, $b_q \neq 0$ and $\{\varepsilon_t\}_{t \in \mathcal{T}} \sim \mathrm{WN}(0, \sigma^2)$.

- The existence and uniqueness of a stationary solution depends – in case $a(z)$ and $b(z)$ have no common roots – only on $a(z)$.
- The equation:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t - \varepsilon_{t-1} \\ a(z) &= 1 - z \\ b(z) &= 1 - z \end{aligned}$$

  has the stationary solutions $y_t = \varepsilon_t + y^*$ for y^* with finite variance, despite the unit root of $a(z)$.

- Consider analogously the solution set to $y_t = \rho y_{t-1} + \varepsilon_t - \rho \varepsilon_{t-1}$ for $|\rho| \neq 1$.

Proposition

The ARMA equation $a(L)\{y_t\} = b(L)\{\varepsilon_t\}$ has a unique stationary solution on $\mathbb{Z}$ if $a(z) \neq 0$ for all $|z| = 1$. This solution is given by:

$$y_t = a(L)^{-1} b(L)\{\varepsilon_t\} = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}$$

If, in addition, $a(z) = 0$ implies $|z| > 1$, then $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is causal for $\{y_t\}_{t\in\mathbb{Z}}$, i. e.,:

$$y_t = a(L)^{-1} b(L)\{\varepsilon_t\} = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}$$

Remark: The presence of common roots of $a(z)$ and $b(z)$ that do not have modulus one does not change the above result, these common factors are cancelled out.

- If in the equation $a(L)y_t = b(L)\varepsilon_t$ all roots of $b(z)$ are outside the closed unit circle and $a(z)$ has no roots on the unit circle, then:[2]

$$\varepsilon_t = b^{-1}(L)a(L)\{y_t\} = \sum_{j=0}^{\infty} d_j y_{t-j}, \quad t \in \mathbb{Z}$$

  i. e., ε_t can be represented as an (infinite) weighted sum of $y_t, y_{t-1}, \ldots$.
- In this case, the stationary ARMA process is called invertible. [Like causality this is a relationship between two processes.]
- Causal and invertible:

$$y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

$$\varepsilon_t = \sum_{j=0}^{\infty} d_j y_{t-j}$$

[2] Consider all common factors to be cancelled.

- Denote with H_t^y the Hilbert space (subspace of $\mathcal{L}^2$) spanned by $\{y_t, y_{t-1}, y_{t-2}, ...\}$.
- Denote with H_t^ε the Hilbert space (subspace of $\mathcal{L}^2$) spanned by $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.
- Then causality means that $y_t \in H_t^\varepsilon$.
- Invertibility means that $\varepsilon_t \in H_t^y$.
- Therefore, both together imply that $H_t^y = H_t^\varepsilon$.

MA Representation and Invertibility

- If $\{y_t\}_{t\in\mathbb{Z}}$ is the solution to an MA equation that is not invertible, e. g.,:

$$y_t = \varepsilon_t + 2\varepsilon_{t-1}, \;\; \varepsilon_t \sim \mathrm{WN}(0, \sigma^2),$$

  then there exists an $\{\varepsilon_t^*\}_{t\in\mathbb{Z}}$ such that:

$$y_t = \varepsilon_t^* + b_1\varepsilon_{t-1}^*,$$

  with $b^*(z) = 1 + b_1 z$ such that $b^*(z) \neq 0 \; \forall |z| \leq 1$.

- Thus, $\{y_t\}_{t\in\mathbb{Z}}$ has another MA representation with invertibility.
- In the discussion, we have excluded the case that $b(z)$ has roots with modulus one, but the result is also true in that case.

The Wold Representation

- Let $y_{t+\tau,t}$ denote the best linear predictor in mean square sense given $y_s,\ s \leq t$, i. e.,:

$$y_{t+\tau,t} = \underset{\tilde{y}}{\operatorname{argmin}}\ \mathbb{E}\left((y_{t+\tau} - \tilde{y})'(y_{t+\tau} - \tilde{y})\right),\ \ s.t.\ \tilde{y} \in H_t^y$$

- A (weakly) stationary process is called linearly regular, if:
  1. $\mathbb{E}(y_t) = 0$
  2. $\lim_{\tau\to\infty} y_{t+\tau,t} = 0$, or equivalently $S := \bigcap_{t\in\mathbb{Z}} H_t^y = 0$

- A (weakly) stationary process is called linearly singular, if:

$$y_{t+\tau,t} = y_{t+\tau}\ \ a.s.,$$

  or equivalently when $S = H^y = H_t^y$ for all $t \in \mathbb{Z}$.

- Here, H^y is the so-called time domain of $\{y_t\}_{t\in\mathbb{Z}}$, i. e., the Hilbert space spanned by $\{y_{i,t} | i = 1, \ldots, n,\ t \in \mathbb{Z}\}$.

The Wold Representation

Herman Wold (1938)

Theorem

Every weakly stationary process $\{y_t\}_{t\in\mathbb{Z}}$ can be represented in a unique way as:

$$y_t = u_t + v_t,$$

where $\{u_t\}_{t\in\mathbb{Z}}$ and $\{v_t\}_{t\in\mathbb{Z}}$ are obtainable via linear transformations of $\{y_t\}_{t\in\mathbb{Z}}$, with:

- $u_t \in H_t^y$
- $v_t \in H_t^y$
- $\mathbb{E}(u_t v_s') = 0, \ \ \forall s, t$

and where:

- $\{u_t\}_{t\in\mathbb{Z}}$ is linearly regular
- $\{v_t\}_{t\in\mathbb{Z}}$ is linearly singular

Theorem

Furthermore, every linearly regular process can be represented as:

$$u_t = \sum_{j=0}^{\infty} K_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \|K_j\|^2 < \infty,$$

where $H_t^u = H_t^\varepsilon$ and where $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$.

- In particular, one can choose $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ to be the one step prediction errors, $\varepsilon_t = u_t - u_{t,t-1}, \ t \in \mathbb{Z}$.
- Note that $\sum_{j=0}^{\infty} \|K_j\|^2 < \infty$ holds in this theorem.

- Consider a diagonal case with $p = 1$ and $n = 2$:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

  with $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ bivariate white noise with $\Sigma > 0$.

- In the above case, a stationary solution y_t exists, if both coordinates are stationary.
- For this, the roots of $(1 - \rho_1 z)(1 - \rho_2 z)$ matter, i. e., the roots of the determinant of $a(z)$.
- Exactly the same logic applies for the general case, compare also the discussion concerning solutions of multivariate difference equations.

- Consider the difference equation:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t,$$

with $A_i \in \mathbb{R}^{n \times n}$, $A_p \neq 0$ and $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$, $\Sigma > 0$.

- We use the notation $a(L) = I_n - A_1 L - \ldots - A_p L^p$.
- There exists a unique stationary solution on $\mathbb{Z}$ to the above equation if and only if $\det a(z) \neq 0$ for all $|z| = 1$.
- The process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is causal for the stationary solution $\{y_t\}_{t \in \mathbb{Z}}$ if and only if $\det a(z) \neq 0$ for all $|z| \leq 1$.

- Consider the difference equation

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t + B_1 \varepsilon_{t-1} + \ldots + B_q \varepsilon_{t-q},$$

with $A_i,\ B_i \in \mathbb{R}^{n\times n}$, $A_p \neq 0$, $B_q \neq 0$ and $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$, $\Sigma > 0$.

- In addition to the above conditions about the autoregressive polynomial, there is also a (multivariate) invertibility condition, given by $\det b(z) \neq 0$ for all $|z| \leq 1$.

Proposition

(1) The $\mathrm{ARMA}(p, q)$ equations have a unique stationary solution $\{y_t\}_{t\in\mathbb{Z}}$ on $\mathbb{Z}$, if $\det a(z) \neq 0$ for all $|z| = 1$.

(2) The solution is causal, i. e., $y_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$, if $\det a(z) \neq 0$ for all $|z| \leq 1$.

(3) The solution is invertible, i. e., $\varepsilon_t = \sum_{j=0}^{\infty} D_j y_{t-j}$, if $\det b(z) \neq 0$ for all $|z| \leq 1$.

The coefficients C_j are defined from $c(z) = a^{-1}(z)b(z)$ and the coefficients D_j from $d(z) = b^{-1}(z)a(z)$.

- Remember from above the case of common factors, e. g.,:

$$y_t = \rho y_{t-1} + \varepsilon_t - \rho \varepsilon_{t-1}$$

- In case $|\rho| \neq 1$, the unique stationary solution is given by $y_t = \varepsilon_t,\ t \in \mathbb{Z}$:

$$\begin{aligned} a(L) &= 1 - \rho L \\ b(L) &= 1 - \rho L \\ k(L) &= a^{-1}(L) b(L) = 1 \\ \{y_t\} &= k(L)\{\varepsilon_t\} = \{\varepsilon_t\} \end{aligned}$$

- The stationary solution does not depend upon $|\rho| \neq 1$.
- Thus, if one thinks of parameter estimation, the parameter ρ is not identified (think of Gaussian ML estimation).
- Now, what is the multivariate analogue of no common roots? Left Coprimeness.

Definition

A pair of matrix polynomials is called left coprime, if $(a(z), b(z))$ has no non-unimodular left divisors.

- A matrix polynomial $u(z)$ is a left divisor of $(a(z), b(z))$, if there exists a pair of polynomial matrices $(\overline{a}(z), \overline{b}(z))$ such that

$$(a(z), b(z)) = u(z)(\overline{a}(z), \overline{b}(z)).$$

- A matrix polynomial $u(z)$ is called unimodular, if $\det u(z) \equiv c \neq 0$, which is equivalent to $u^{-1}(z)$ also being a matrix polynomial.

- Going back to the example:

$$\begin{aligned} a(z) &= 1 - \rho z \\ b(z) &= 1 - \rho z \\ (1 - \rho z, 1 - \rho z) &= (1 - \rho z)(1, 1), \end{aligned}$$

  thus, there is a non-unimodular left divisor.

- In the scalar case only non-zero numbers are unimodular (matrix polynomials).

Multivariate ARMA Processes

Transfer Function and Second-Order Properties

- In the stationary case, the transfer function $k(z) = a^{-1}(z)b(z)$ and Σ fully describe the second-order properties of $\{y_t\}_{t\in\mathbb{Z}}$, given as $y_t = \sum_{j=-\infty}^{\infty} K_j \varepsilon_{t-j}$.
- In the scalar case, in addition to $\mathbb{E}(y_t) = 0$, it directly follows that:

$$\mathbb{E}(y_t^2) = \mathbb{E}\left(\sum_{j=-\infty}^{\infty} k_j \varepsilon_{t-j}\right)^2 = \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} k_j^2$$

$$\mathbb{E}(y_t y_{t-h}) = \mathbb{E}\left(\sum_{j=-\infty}^{\infty} k_j \varepsilon_{t-j} \sum_{m=-\infty}^{\infty} k_m \varepsilon_{t-h-m}\right) = \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} k_j k_{j-h},$$

  i. e., the second-order properties depend upon $k(z)$ and Σ.
- Identification of ARMA systems for (given) $k(z)$ and Σ means to select a unique $(a(z), b(z))$ and Σ that corresponds to $k(z)$ and Σ.

Theorem

Let $T(p,q)$ denote the set of all ARMA systems (a,b) that fulfill:

- $a(0) = I_n, \ \det a(z) \neq 0 \ \forall \ |z| \leq 1$
- $b(0) = I_n, \ \det b(z) \neq 0 \ \forall \ |z| \leq 1$
- $\Sigma > 0$
- (a,b) are left coprime
- deg$(a(z)) \leq p$, deg$(b(z)) \leq q$
- (A_p, B_q) has full rank n

Then it holds that $T(p,q)$ is identifiable., i. e., (a,b,Σ) is unique for (k,Σ).

- Note that a VAR model is always identified (when $\Sigma > 0$).
- The current discussion is about reduced-form identification, not about "structural" identification (with potentially "over-identifying" restrictions stemming from economic theory).

Deterministic Components

- Up to now, the discussion assumed – unlikely, e. g., for many economic time series – that $\mathbb{E}(y_t) = 0$ for all t.
- In case the expected values are finite for all t, one can always consider:

$$y_t = \mathbb{E}(y_t) + (y_t - \mathbb{E}(y_t)) = T_t + y_t^0,$$

  where the first component – the trend – is often modeled by some deterministic function of time (constant, linear trend, seasonal dummies) and where for the mean-zero component y_t^0 a time series model is used.
- As shall be seen, the properties of the two components may impact the statistical analysis in many ways.
- To perform such a decomposition $\mathbb{E}(y_t)$, in general, has to be estimated.

- Consider the $\mathrm{AR}(1)$ case with intercept ν and commence again with recursive substitution:

$$\begin{aligned} y_t &= \nu + A_1 y_{t-1} + \varepsilon_t \\ &= (I_n + A_1 + \ldots + A_1^j)\nu + A_1^{j+1} y_{t-j-1} + \sum_{i=0}^{j} A_1^i \varepsilon_{t-i} \end{aligned}$$

- Now, if all eigenvalues of A_1 are smaller than one in absolute value, then:[3]
  - The sequence A_1^i is absolutely summable
  - $A_1^j \to 0$ for $j \to \infty$
  - $(I_n + A_1 + \ldots + A_1^j)\nu \to (I_n - A_1)^{-1}\nu$ for $j \to \infty$

[3] Relate this condition to the discussion on the roots of the determinant of $a(z)$.

- Thus, with "starting time infinite past", one arrives at the solution:

$$y_t = \mu + \sum_{i=0}^{\infty} A_1^i \varepsilon_{t-i}, \quad t \in \mathbb{Z},$$
$$\mu := (I_n - A_1)^{-1}\nu$$

- This shows that the solution process has expected value equal to μ (and not – as one might have guessed first – ν).
- Exercise: Verify that $\{y_t\}_{t\in\mathbb{Z}}$ as given above satisfies the $\mathrm{AR}(1)$ equations.
- Because of the assumption that all eigenvalues of A_1 are smaller than one in absolute value, one refers to this case also as the stable case or as stable autoregression.

- The ACF can easily be calculated for the $\mathrm{AR}(1)$ case (and coincides with the ACF of the case without intercept):

$$
\begin{aligned}
\Gamma_y(h) :&= \mathbb{E}(y_t - \mu)(y_{t-h} - \mu)' \\
&= \lim_{n\to\infty} \sum_{i=0}^{n} \sum_{j=0}^{n} A_1^i \mathbb{E}(\varepsilon_{t-i}\varepsilon'_{t-h-j})(A_1^j)' \\
&= \lim_{n\to\infty} \sum_{i=0}^{n} A_1^{h+i} \Sigma (A_1^i)' = \sum_{i=0}^{\infty} A_1^{h+i} \Sigma (A_1^i)'
\end{aligned}
$$

- This sum exists and is independent of t in case $|\lambda_{\max}(A_1)| < 1$.
- Later chapters discuss what happens when this constraint is not fulfilled (in particular the case of the eigenvalue 1 will be considered).

- It will be seen below that the question whether a stochastic process is difference stationary or trend stationary is an important question.[4]

- Consider again the stable $\mathrm{AR}(1)$ case, now with intercept and linear trend (with $\gamma \in \mathbb{R}^n$):

$$y_t = \nu + \gamma t + A_1 y_{t-1} + \varepsilon_t$$

- Given y_0, recursive substitution leads for $t > 0$ to:

$$y_t = (I_n + A_1 + \ldots + A_1^{t-1})\nu + \gamma \sum_{j=0}^{t-1} A_1^j (t-j) + A_1^t y_0 + \sum_{j=0}^{t-1} A_1^j \varepsilon_{t-j}$$

[4] Precise definitions will be given later.

- In this case there exists no y_0 that renders $\{y_t\}_{t\in\mathbb{N}}$ stationary.
- Considering again the starting value from the case without intercept for y_0, the solution is stationary up to the trend component, i. e., it is trend stationary.
- One can also consider the first difference of $\{y_t\}$, i. e., $\Delta y_t = y_t - y_{t-1},\ t \in \mathbb{N}$, which leads to:

$$\begin{aligned}\Delta y_t &= \gamma(I_n + A_1 + \ldots + A_1^{t-1}) + A_1^{t-1}\left(\nu - (I_n - A_1)y_0\right) + \varepsilon_t - \varepsilon_{t-1} \\ &= -\sum_{j=0}^{t-2} A_1^j (I_n - A_1)\varepsilon_{t-1-j}\end{aligned}$$

- Thus, considering first differences “removes” the “linear trend”, as can be guessed when remembering the discussion about the kernels of moving averages or filters.

- For any solution one can also "take the differences" at the equation level:

$$
\begin{aligned}
y_t &= \nu + \delta t + A_1 y_{t-1} + \varepsilon_t \\
y_{t-1} &= \nu + \delta(t-1) + A_1 y_{t-2} + \varepsilon_{t-1} \\
\Delta y_t &= \delta + A_1 \Delta y_{t-1} + \varepsilon_t - \varepsilon_{t-1}
\end{aligned}
$$

- This shows that Δy_t is the solution to an $\mathrm{ARMA}(1,1)$ equation with $b(z) = I_n - I_n z$, i. e., with unit roots in the MA polynomial.
- This property of the MA polynomial can, of course, also be directly verified for the solution given above.

- The above example already shows that the integration of deterministic components in difference equations is "a bit complicated".

- For this reason, the above mentioned decomposition is often preferred, e. g.,:

$$\begin{aligned} y_t &= \mu + \gamma t + y_t^* \\ a(L)\{y_t^*\} &= \{\varepsilon_t\}, \end{aligned}$$

  which directly implies that $\mathbb{E}(y_t) = \mu + \gamma t$.

- One situation in which this decomposition is not pursued is cointegration analysis in VAR models, where restricted trend coefficients play an important role.