Time series models

Oliver Snellman

March 2020

1 Time series processes

Time series is a sequentially organized collection of evenly spaced random variables $y_t, t = 1, \dots, T$. The aim of time series analysis is to build statistical models to represent the observed time series, their fluctuations and the dependence structure of consecutive observations.

In the heart of the model there is a probability distribution of the shock process ε_t, and the model describes a transformation of that shock process $y_t = f(\varepsilon_t)$, which results in a joint probability density function for y_t.

Typically instead of explicitly deriving the joint pdf for y_t, it is more convenient to just define the model in terms of the error-distribution, equations characterizing the dynamics of the system, and the parameter space.

The **lag operator** used in the following expressions is defined according to normal conventions as

$$
\begin{aligned}
By_t &= y_{t-1} \\
B^h y_t &= y_{t-h}
\end{aligned}
$$

with

$$
\begin{aligned}
1 - \phi_1 B - \dots - \phi_p B^p &=: \phi(B) \\
1 + \theta_1 B + \dots + \theta_q B^q &=: \theta(B)
\end{aligned}
$$

AR(p)

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \qquad\qquad \varepsilon_t \sim iid(0, \sigma^2) \\
&\iff \\
y_t &= (\phi_1 B + \dots + \phi_p B^p) y_t + \varepsilon_t \\
&\iff \\
(1 - \phi_1 B - \dots - \phi_p B^p) y_t &= \varepsilon_t \\
&\iff \\
\phi(B) y_t &= \varepsilon_t
\end{aligned}
$$

A sufficient condition for the **stationarity** of an AR(p) process is, that the roots of the polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad z \in C$$

lie outside of the unit circle in the complex plane, or equivalently

$$\phi(z) \neq 0 \quad |z| \leq 1$$

MA(q)

$$\begin{aligned} y_t &= \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma^2) \\ &\iff \\ y_t &= \theta(B)\varepsilon_t \end{aligned}$$

ARMA(p,q)

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ &\quad + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t & \varepsilon_t \sim iid(0, \sigma^2) \\ &\iff \\ \phi(B) y_t &= \theta(B)\varepsilon_t \\ &\iff \\ y_t &= \frac{\theta(B)}{\phi(B)}\varepsilon_t & \frac{\theta(B)}{\phi(B)} =: \pi(B) \\ &\iff \\ y_t &= \pi(B)\varepsilon_t \end{aligned}$$

ARIMA(p,d,q)

If the process y_t is nonstationary, but its d^{th} difference is stationary and invetrible ARMA(p,q) process, we call y_t an ARIMA(p,d,q) process.

The difference operator is defined as

$$\begin{aligned} \Delta &= 1 - B \\ &\iff \\ \Delta y_t &= y_t - y_{t-1} \end{aligned}$$

and

$$\begin{aligned} \Delta^d y_t &= \Delta(\Delta^{d-1} y_t) \\ &= (\Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}) - (\Delta^{d-1} y_{t-1} - \Delta^{d-1} y_{t-2}) \end{aligned}$$

Hence, if $\Delta^d y_t$ follows $ARMA(p,q)$, then y_t follows $ARIMA(p,d,q)$.

ARCH(s)

$$
\begin{aligned}
y_t &= h_t^{\frac{1}{2}} \varepsilon_t & \varepsilon_t &\sim iid(0,1) \\
h_t &= \omega + \alpha_1 y_{t-1}^2 + \dots + \alpha_s y_{t-s}^2 & \omega &> 0, \quad \alpha \geq 0
\end{aligned}
$$

GARCH(r,s)

$$
\begin{aligned}
y_t &= h_t^{\frac{1}{2}} \varepsilon_t & \varepsilon_t &\sim iid(0,1) \\
h_t &= \omega + \beta_1 h_{t-1} + \dots + \beta_r h_{t-r} \\
&\quad + \alpha_1 y_{t-1}^2 + \dots + \alpha_s y_{t-s}^2 & \omega &> 0, \quad \alpha, \beta \geq 0
\end{aligned}
$$

AR(p)-GARCH(r,s)

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t \\
u_t &= h_t^{\frac{1}{2}} \varepsilon_t & \varepsilon_t &\sim iid(0,1) \\
h_t &= \omega + \beta_1 h_{t-1} + \dots + \beta_r h_{t-r} \\
&\quad + \alpha_1 u_{t-1}^2 + \dots + \alpha_s u_{t-s}^2 & \omega &> 0, \quad \alpha, \beta \geq 0 \\
&\iff \\
\phi(B) y_t &= u_t
\end{aligned}
$$

ARMA(p,q)-GARCH(r,s)

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} \\
&\quad + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + u_t \\
u_t &= h_t^{\frac{1}{2}} \varepsilon_t & \varepsilon_t &\sim iid(0,1) \\
h_t &= \omega + \beta_1 h_{t-1} + \dots + \beta_r h_{t-r} \\
&\quad + \alpha_1 u_{t-1}^2 + \dots + \alpha_s u_{t-s}^2 & \omega &> 0, \quad \alpha, \beta \geq 0 \\
&\iff \\
\phi(B) y_t &= \theta(B) \varepsilon_{t-1} + u_t
\end{aligned}
$$

2 Stationarity and invertibility

Weak stationarity

Process y_t is called weak (or covariance) stationary if the first two moments (expected value and autocovariance) are time-invariant.

$$
\begin{aligned}
(E) &= \mu \quad \text{for all} \quad t = 0, \pm 1, \pm 2, \dots \\
Cov(y_t, y_{t+h}) &= \gamma_{t,t+h} \quad \text{for all} \quad h, t = 0, \pm 1, \pm 2, \dots \\
&= \gamma_{0,h} \\
&= \gamma_h
\end{aligned}
$$

Weak stationarity does not imply strict stationarity in general, but it does with normally distributed random variables.

To check for weak stationarity, find out if $E(y_t), Var(y_t)$ and $Cov(y_t)$ indeed do not depend on time.

Strict stationarity

A process $\{y_t; t = 0, \pm 1, \pm 2, \dots\}$ is called strictly (or strongly) stationary if the whole distribution of the process is time-invariant, i.e. collections of random variables

$$(y_{t_1}, \dots, y_{t_m}) \quad \sim \quad (y_{t_1+h}, \dots, y_{t_m+h}) \qquad \forall\ t_1, \dots t_m,\ h \geq 0, \quad m > 0$$

have the same m-dimensional joint probability distribution for all integers $t_i,\ h$ and $m > 0$. Hence, the entire probability structure (distribution) of the process y_t is time invariant.

Then it also holds, that

- **SS1** The random variables y_t have the same distribution $\forall t$
- **SS2** The random vectors (y_t, y_{t+h}) and (y_s, y_{s+h}) have the same distribution for all t and s and all (fixed) h
- **SS3** y_t is weakly stationary if $E(y_t^2) < \infty$ holds.
- **SS4** Strict stationarity is conserved under (measurable) transformation g. If z_t is is strictly stationary, then so is also

$$y_t = g(z_{t+m}, \dots, z_{t-n}),\ m, n \geq 0$$

There are cases where strict stationarity does not imply weak stationarity, for example with Cauchy distributed random variables.

Weak white noise

A weakly stationary sequence of iid random variables, ε_t, is called weak white noise, i.e. the expected value and variance of this process are finite and time-invariant, and the covariance is zero.

$$\begin{aligned} \varepsilon_t &= \mu \\ Var(\varepsilon_t) &= \sigma^2 \\ Cov(\varepsilon_j, \varepsilon_s) &= 0 \qquad j \neq s \end{aligned}$$

Weak white noise processes are used as building blocks to construct more complicated processes.

Strong white noise

A strictly stationary sequence of iid random variables, ε_t, is called strong white noise. In other words, the cumulative distribution function F() of the random vectors $(\varepsilon_{t_1}, \dots, \varepsilon_{t_m})$ and $(\varepsilon_{t_1+h}, \dots, \varepsilon_{t_m+h})$, when evaluated at $(x_1, \dots, x_m)$, is the same $F(x_1)F(x_2)\dots F(x_m)$.

Invertibility

A time series process is invertible, if the innovations can be translated into a representation of past observations. An invertible process allows for the innovations, which drive the process on the background, to be identified from the past values of the that process.

More practically, invertibility refers to the possibility of transforming AR or ARMA processes into MA processes, and vice versa, MA and ARMA processes into AR representation. If an AR process is not invertible, it is not possible to identify the innovations driving the process, and hence it is impossible to forecast the series into the future.

Wold decomposition

Every weakly stationary process has an MA(∞) representation.

Stationary AR(p) (in companion form) can be inverted to a MA form

$$y_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$$

Autocovariances

$$\begin{aligned} Cov(y_t, y_{t-j}) &= E[(y_t - E[y_t])(y_{t-j} - E[y_{t-j}])] \\ &= E[(y_t, y_{t-j})^2] - E(y_t)E(y_{t-j}) \end{aligned}$$

For an AR(p) process, autocovariances can be obtained alternatively by multiplying both sides of the expression by y_{t-h} and taking expectations to obtain the h^{th} autocovariance

$$\begin{aligned} \gamma_h &= \mathbf{E}[y_t, y_{t-h}] \\ &\Longleftrightarrow \\ \mathbf{E}[y_t y_{t-h}] &= \phi_1 \underbrace{\mathbf{E}[y_{t-1} y_{t-h}]}_{\gamma_{h-1}} + \cdots + \phi_p \underbrace{\mathbf{E}[y_{t-p} y_{t-h}]}_{\gamma_{h-p}} + \mathbf{E}[\varepsilon_t y_{t-h}] \\ &\Longleftrightarrow \end{aligned}$$

$$\gamma_h = \begin{cases} \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma^2 & when \quad h = 0 \\ \phi_1\gamma_{h-1} + \cdots + \phi_p\gamma_{h-p} & when \quad h > 0 \end{cases} \tag{1}$$

Autocorrelations

Divide the h^{th} order autocovariance γ_h by the variance $\gamma_0 = \mathbf{E}[y_t y_t]$ to get the h^{th} order autocorrelation.

For an AR(p) process this entails

$$
\begin{aligned}
\rho_h &= \frac{\gamma_h}{\gamma_0} \\
&\Longleftrightarrow \\
\frac{\mathbf{E}[y_t y_{t-h}]}{\mathbf{E}[y_t y_t]} &= \phi_1 \frac{\mathbf{E}[y_{t-1} y_{t-h}]}{\mathbf{E}[y_t y_t]} + \cdots + \phi_p \frac{\mathbf{E}[y_{t-p} y_{t-h}]}{\mathbf{E}[y_t y_t]} \\
&\Longleftrightarrow \\
\frac{\gamma_h}{\gamma_0} &= \phi_1 \underbrace{\frac{\gamma_{h-1}}{\gamma_0}}_{\rho_{h-1}} + \cdots + \phi_p \underbrace{\frac{\gamma_{h-p}}{\gamma_0}}_{\rho_{h-p}} \\
&\Longleftrightarrow \\
\rho_h &= \phi_1 \rho_{h-1} + \cdots + \phi_p \rho_{h-p}
\end{aligned}
\tag{2}
$$

Note, that the 0^{th} order autocorrelation is $\rho_{h-h} = \rho_0 = \frac{\gamma_0}{\gamma_0} = 1, \ \ \forall h$.

Partial autocorrelation function

Measures the correlation between y_t and y_{t-h}, when the linear effect of the random variables $y_{t-1}, \ldots, y_{t-h+1}$ has been first eliminated. For AR(p) the PACF is simply

$$
\begin{aligned}
\alpha_h &= \phi_h \quad for \ \ h \leq p \\
\alpha_h &= 0 \quad for \ \ h > p
\end{aligned}
$$

Yule Walker

Yule Walker equations allow unknown parameters of an AR(p) model to be estimated on the basis of empirical autocovariances and autocorrelations.

$$
\begin{aligned}
\phi &= \mathbf{\Gamma^{-1}} \gamma \\
\phi &= \mathbf{P^{-1}} \rho
\end{aligned}
$$

See github.com/OliverSnellman/Cool-econometrics/, Yule Walker.pdf for a more thorough walkthrough.

Some definitions

Non-causal: General case, where the process is determined by all of the past and forecoming shocks

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

Causal: Only past (realized) shocks have impact on the current value of the process

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

A process y_t is called **deterministic**, is the value of y_t can be perfectly linearly forecasted without any forecast error, for all t.

Processes which are not deterministic, are **non-deterministic**, i.e. they have shocks/errors included.