

Time series

# Preliminary definitions

# Time series definition [2]

## Informal definition

A time-series is a set of observation  $x_t$  each one being recorder at a specific time  $t$ .

## Formal definition

A time series model for the observed data  $x_t$  is a specification of the joint distribution (or possible only the means covariance) of a sequence of random variable  $X_t$  of which  $x_t$  is postulated to be a realization

## A binary process

Consider the sequence of iid random variables, with  $P[X_t = 1] = p$  and  $P[X_t = -1] = 1 - p$

## Random walk

The random walk is obtained by cumulatively summing iid random variables. Thus a random walk with zero mean is obtained by defining

# Stationarity, autocovariance and autocorrelation[2]

## Mean Function

Let  $X_t$  be a time series with  $E(x_t^2) < \infty$ . The mean function of  $X_t$  is  $\mu_X(t) = E(X_t)$ . The covariance function of  $X_t$  is  $\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$   $\forall r, s$

## Weakly stationary TS

$X_t$  is weakly stationary if i)  $\mu_X(t)$  is independent from time  $t$  and ii)  $\gamma_X(t+h)$  is independent of  $t \forall h$

## Autocovariance function

At lag  $h$  the auto-covariance function is defined as  $\gamma_X(h) = \text{Cov}(X_{t+h}, X_t)$

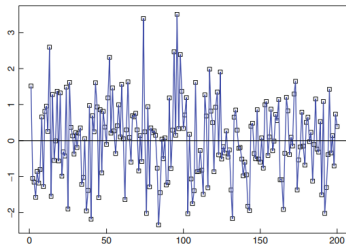
## Autocorrelation function

At lag  $h$  the autocorrelation function is defined as

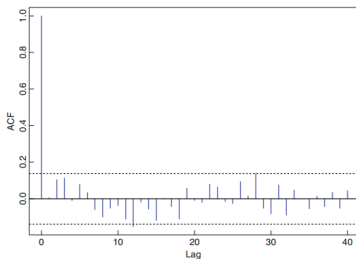
$$\rho(h)_X = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Cor}(X_{t+h}, X_t)$$

# Stationarity, autocovariance and autocorrelation [2]

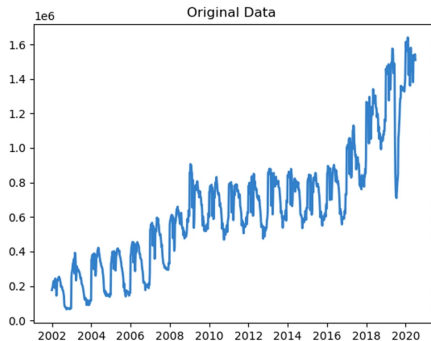
**Figure 1-12**  
200 simulated values of IID  
 $N(0,1)$  noise



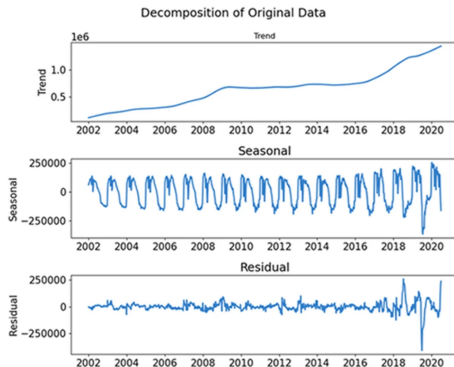
**Figure 1-13**  
The sample autocorrelation  
function for the data of  
Figure 1-12 showing the  
bounds  $\pm 1.96/\sqrt{n}$



# Stationarity, autocovariance and autocorrelation [4]



(a)



(b)

## Definition [2]

### Linear process

Time series  $X_t$  is a linear process if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

for all  $t$ , where  $Z_t \approx \text{WN}(0, \sigma^2)$  and  $\psi_j$  is a sequence of constants with  $\sum_{j=-\infty}^{\infty} \psi_j < \infty$ . In terms of backward shift operator  $B$   $BX_t = X_{t-1}$  we have  $X_t = \phi(B)Z_t$ . Therefore the previous definition can be rewritten as  $X_t = \phi(B)Z_t$  in which  $\phi(B)$  can be thought as a linear filter that when applied to the white noise input series  $Z_t$  produces the output  $X_t$ .

## Definition [2]

### Linear process

The time series  $X_t$  is an ARMA(1,1) process if its stationary and satisfiy (for every t)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where  $Z_t \approx \text{WN}(0, \phi^2)$  and  $\phi + \theta \neq 0$  or in terms of filters  $\phi$  and  $\theta$

$$\phi(B)X_t = \theta(B)Z_t$$



## Remarks

- A stationary solution of the ARMA(1,1) equation exists if and only if  $\phi \neq \pm 1$
- If  $|\phi| < 1$ , then the unique stationary solution is given by  $X_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}$ . In this case we say that  $X_t$  is causal or a causal function of  $Z_t$  or a causal function of  $Z_t$  since  $X_t$  can be expressed in terms of the current and past values  $Z_s, s \leq t$
- If  $|\phi| > 1$ , then the unique stationary solution is given by  $X_t = -\theta\phi^{-1}Z_t - (\phi + \theta) \sum_{j=1}^{\infty} \phi^{-j-1}Z_{t-j}$ . The solution is non-causal, since  $X_t$  is then a functional of  $Z_s, s \geq t$

## Wold decomposition [2]

Prediction operator based on the infinite past  $X_t, -\infty < t < n$

$$\tilde{P}_n X_{n+h} = \lim_{m \rightarrow -\infty} P_{m,n} X_{n+h}$$

Wold decomposition  $X_t, -\infty < t < n$

$X_t$  is a non-deterministic stationary time series, then

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$$

where  $V_t$  is deterministic

## Definition [2]

### ARMA (p,q)

$X_t$  is an ARMA(p,q) process if  $X_t$  is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

where  $Z_t \approx WN(0, \sigma^2)$  and the polynomials  $(1 - \phi_1 z - \dots - \phi_p z^p)$  and  $(1 + \theta_1 z + \dots + \theta_q z^q)$  have no common factor.

## Definition [2]

### Causality

An ARMA(p,q) process  $X_t$  is causal, or a causal function of  $Z_t$  if there exist constants  $\phi_j$  such that  $\sum_{j=0}^{\infty} |\phi_j| < \infty$  and

$$X_t = \sum_{j=0}^{\infty} \sum \phi_j Z_{t-j} \quad \forall \quad t$$

this is equivalent to the condition

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \quad |z| < 1$$

# Spectral Analysis

# Spectral Analysis [2]

## Spectral density

Given a zero mean stationary time series  $X_t$  with autocovariance function  $\gamma(\cdot)$  satisfying  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ , the spectral density of  $X_t$  is the function defined by

$$f(\lambda) = \frac{1}{2\pi} \sum e^{-ih\lambda} \gamma(h) \quad -\infty < \lambda < \infty$$

with the condition that

$$f(\lambda) \geq 0 \quad \forall \lambda$$

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda \quad \forall h$$

# Spectral Analysis [2]

## Spectral Representation of The ACVF

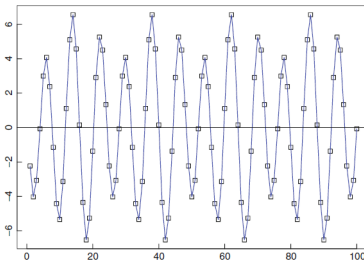
A function  $\gamma(\cdot)$  defined on the integers is the ACVF of a stationary time series if and only if there exists a right-continuous, non-decreasing, bounded function  $F$  on  $[-\pi, \pi]$  with  $F(-\pi) = 0$  such that:

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\lambda} dF(\lambda)$$

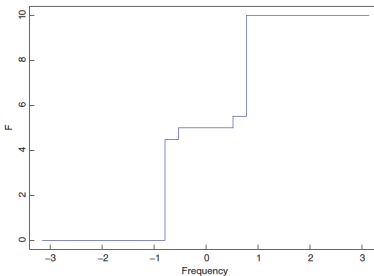
for all integers  $h$ .  $F$  is a generalized distribution function that is called the spectral distribution function of  $\gamma(\cdot)$

# Spectral Analysis [2]

**Figure 4-1**  
A sample path of size  
100 from the time series  
in Example 4.1.2



**Figure 4-2**  
The spectral distribution  
function  $F(\lambda)$ ,  $-\pi \leq \lambda \leq \pi$ ,  
of the time series  
in Example 4.1.2





# Time-Invariant Linear Filters [2]

## Linear process

The process  $Y_t$  is the output of a linear filter  $C = \{c_{t,k}, t, k = 0 \pm 1, \dots\}$  applied to an input process  $X_t$  if  $Y_t = \sum_{k=-\infty}^{\infty} c_{t,k} X_k \quad t = 0, \pm 1, \dots$

## Time invariant

The filter is said to be time-invariant if the weights  $c_{t,t-k}$  are independent of  $t$  e.g.  $c_{t,t-k} = \phi_k$

$$Y_t = \sum_{k=-\infty}^{\infty} \phi_{t,k} X_{t-k}$$

$$Y_{t-s} = \sum_{k=-\infty}^{\infty} \phi_{t,k} X_{t-s-k}$$

The TLF  $\phi$  is to be causal if  $\phi_j = 0$  for  $j < 0$

# Time-Invariant Linear Filters [2]

## Transfer function

Let  $X_t$  be a stationary time series with mean zero and spectral density  $f_X(\lambda)$ . Suppose that  $\Phi = \{\phi_j, j = 0, \pm 1, \dots\}$  is an absolutely summable TLF. Then the time series

$$Y_t = \sum_{j=-\infty}^{\infty} \phi_j X_{t-j}$$

is stationary with mean zero and spectral density

$$f_Y(\lambda) = |\Psi(e^{-i\lambda})|^2 f_X(\lambda) = \Psi(e^{-i\lambda}) \Psi(e^{i\lambda}) f_X(\lambda)$$

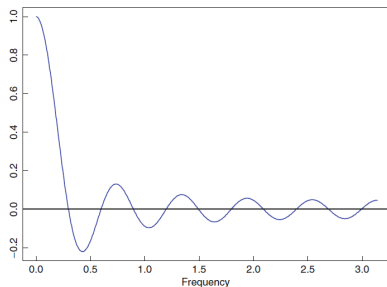
where  $\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \phi_j e^{-ij\lambda}$  where  $\Psi(e^{-i})$  is called the transfer function of the filter, and the squared modulus  $|\Psi(e^{-i})|$  is referred to as the power transfer function of the filter.

# Time-Invariant Linear Filters [2]

## Simple moving average

$$Y_t = \frac{1}{2q+1} \sum_{|j| \leq q} X_{t-j}$$

where  $\psi = (2q+1)^{-1}$ ,  $j = -q, \dots, q$  and  $\psi_j$

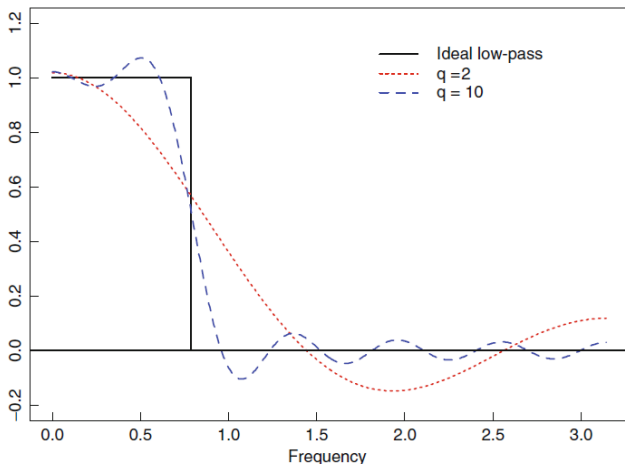


**Figure 4-12**  
The transfer function  $D_{10}(\lambda)$  for the simple moving-average filter

# Time-Invariant Linear Filters [2]

## Gibbs phenomenon

The poor approximation in the neighbourhood of cut-off frequency ( $\omega_c$ )



**Figure 4-13**

The transfer function for the ideal low-pass filter and truncated Fourier approximations  $\Psi^{(q)}$  for  $q = 2, 10$

# Estimation

## Estimation [2]

### Gaussian Likelihood

$$L(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma)^n r_0 \dots r_{n-1}}} \exp\left(-\frac{1}{2\sigma^2} \sum \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}\right)$$

### Maximum Likelihood Estimators

Differentiating  $\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$  partially with respect to  $\sigma^2$  and noting that  $\hat{X}_j$  and  $r_j$  are independent of  $\sigma$

$$\hat{\sigma}^2 = n^{-1} S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$$

$$S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1}$$

where  $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}$  are the values of  $\boldsymbol{\phi}, \boldsymbol{\theta}$  that minimize:

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln(n^{-1} S(\boldsymbol{\phi}, \boldsymbol{\theta})) + n^{-1} \sum_{j=1}^n \ln r_{j-1}$$

## Estimation [2]

### Akaike information criterion

$$AICC = -2 \ln L(\hat{\Phi}, \hat{\theta}, S((\phi_p, \theta_q)/n) + 2(p + q + 1)n/(n - p - q - 2)$$

### Residual

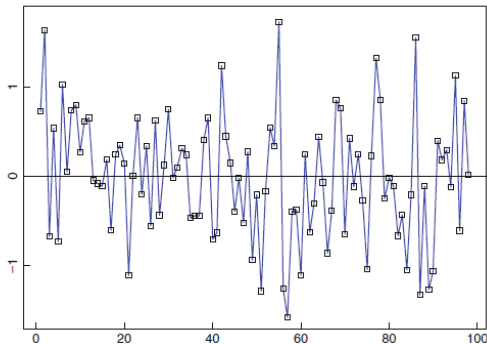
$$\hat{W}_t = (X_t - \hat{X}_t \phi, \theta) / (r_{t-1}(\hat{\Phi}, \hat{\theta}))^{1/2} \quad t = 1, \dots, n$$

This should have properties similar to the white noise sequence, thus one can define the rescaled residuals

$$\hat{R}_t = \hat{W}_t / \hat{\sigma}$$

$$\hat{\sigma} = \sqrt{\left( \sum_{t=1}^n \hat{W}_t^2 \right) / n}$$

# Residuals [2]

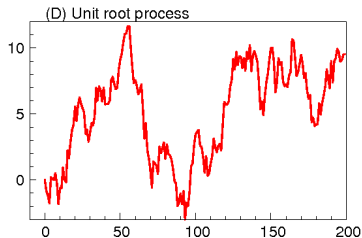
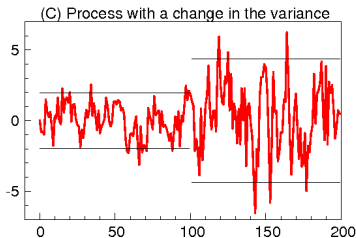
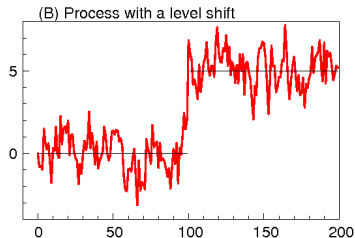
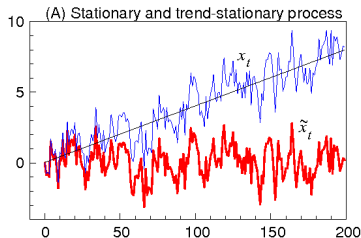


**Figure 5-5**  
The rescaled residuals after  
fitting the ARMA(1,1) model  
of Example 5.2.5 to the lake  
data



# Nonstationary time series

# Residuals [2, 1]



### Akaike information criterion

If  $d$  is a non-negative integer, then  $X_t$  is an ARIMA( $p, d, q$ ) process if

$$Y_t = (1 - B)^d X_t$$

is causal ARMA( $p, q$ ) process  $BX_t = X_{t-1}$ . Note that  $\nabla X_t = X_t - X_{t-1}$

# Multivariate time series

# Multivariate time series [2]

## Akaike information criterion

$\{\mathbf{X}_t\}$  is an ARMA (p,q) process if  $\{\mathbf{X}_t\}$  is stationary and if for every t,

$$\mathbf{X}_t - \phi_1 \mathbf{X}_{t-1} - \dots - \phi_p \mathbf{X}_{t-p} = Z_t + \Theta_1 Z_{t-1} + \dots + \Theta_q Z_{t-q}$$

where  $Z_t \approx \text{WN}(0, \Sigma)$

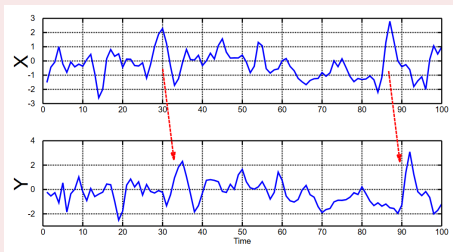
# Vector Autoregression [3]

## VAR(1)

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}$$

## Granger causality test (hamilton2020time)

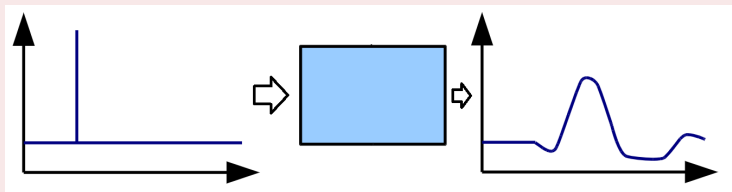
$y_2$  cause (in Granger sense)  $y_1$  if the coefficient  $a_{1,2}$  is significantly not equal to 0



# Vector Autoregression [3]

## Impulse response function

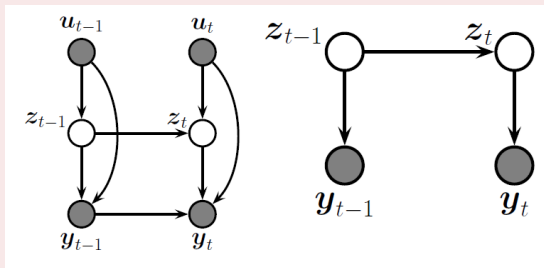
$y_2$  cause (in Granger sense)  $y_1$  if the coefficient  $a_{1,2}$  is significantly not equal to 0



# Space State Models [5]

## Definition

A state-space model (SSM) is a partially observed Markov model in which the hidden state  $z_t$  evolves over time according to a Markov process and each hidden state generates some observations  $y_t$  at each step. The main goal is to infer the hidden states given the observations (and also to predict the future observations)





# Space State Models [5]

## Non linear dynamical system

A state-space model (SSM) can be represented as a stochastic discrete time nonlinear dynamical system of the form

$$\begin{aligned}z &= f(\mathbf{z}_{t-1}, \mathbf{u}_t, \mathbf{q}_t) \\ y_t &= h(\mathbf{z}_t, \mathbf{u}_t, \mathbf{y}_{1:t-1}, \mathbf{r}_t)\end{aligned}$$

where  $\mathbf{z}_t \in \mathbb{R}^N$  are the hidden states,  $\mathbf{u}_t \in \mathbb{R}^N$  are the optional observed inputs,  $\mathbf{y}_t \in \mathbb{R}^N$  are observed output and  $\mathbf{f}$  is the transition function,  $\mathbf{q}_t$  is the process noise,  $\mathbf{h}$  is the observation function and  $\mathbf{r}_t$  is the observation noise. The transition model and the observational model are

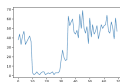
$$\begin{aligned}p(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{u}_t) &= p(\mathbf{z}_t | \mathbf{f}(\mathbf{z}_{t-1}, \mathbf{u}_t)) \\ p(\mathbf{y}_t | \mathbf{z}_t, \mathbf{u}_t, \mathbf{y}_{1:t-1}) &= p(\mathbf{y}_t | \mathbf{h}(\mathbf{z}_t, \mathbf{u}_t, \mathbf{y}_{1:t-1}))\end{aligned}$$

# Space State Models [5]

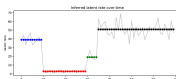
## Non linear dynamical system

Hidden Markov Model (HMM)  $\rightarrow$  a SSM in which the hidden states are discrete, thus  $z_t \in 1, \dots, K$

## Time series segmentation



(a)



(b)

Figure 29.6: (a) A sample time series dataset of counts. (b) A segmentation of this data using a 4 state HMM. Generated by `poisson_hmm_changepoint.py`.

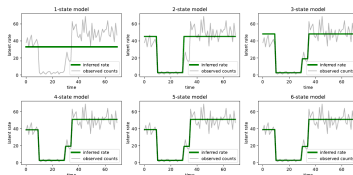


Figure 29.7: Segmentation of the time series using HMMs with 1-6 states. Generated by `poisson_hmm_changepoint.py`.

# Space State Models [5]

## Time series segmentation

We want to segment a time series into different regimes, each of which correspond to a different statistical distribution. In particular we would like to segment this data stream in to  $K$  different regimes or states, each of which is associated with a Poisson observation model with rate  $\lambda_k$ :

$$p(y_t | z_t = k) = \text{Poi}(y_t | \lambda_k)$$

where a uniform prior over the initial states was considered. The transition matrix will be:

$$z_1 \approx \text{Categorical} \left( \left\{ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right\} \right)$$

$$z_t | z_{t-1} \approx \text{Categorical} \left( \left\{ \begin{pmatrix} p & z_t = z_{t-1} \\ \frac{1-p}{4-1} & \text{else} \end{pmatrix} \right\} \right)$$

$$p(y_{1:T} | K) \approx \max_{\lambda} \sum p(y_{1:T}, z_{1:T} | \lambda, K)$$

# Structural time series models [5]

## Structural time series (STS)

Defined in terms of linear-Gaussian SSMs. Differently from ARMA method, they have much more flexibility: one can create non-linear, non-Gaussian and even hierarchical extension. Represent the observed scalar time series as a sum of  $C$  individual components

$$f(t) = f_1(t) + f_2(t) + \dots + f_C(t) + \epsilon_t$$

Each single component (latent process)  $f_c(t)$  is modeled by a linear Gaussian state-space model which is also called dynamic linear model. Since these are linear, one can combine in to a single state-space model.

$$p(z_t|z_{t-1}, \theta) = \mathcal{N}(z_t|Fz_{t-1}, Q)$$

$$p(y_t|z_t, \theta) = \mathcal{N}(y_t|Hz_t + \beta^T u_t, \sigma_y^2)$$

where  $\mathbf{F}$  and  $\mathbf{Q}$  are block structure matrices, with one block per component. The vector  $\mathbf{H}$  then adds up all the relevant pieces from each component to generate the overall mean.

# Structural time series models [5]

## Local level model

The observation  $y_t \in \mathbb{R}^N$  are generated by a Gaussian with (latent) mean  $\mu_t$ , which evolves over over time according to a random walk

$$\begin{aligned}y_t &= \mu_t + \epsilon_{y,t} & \epsilon_{y,t} &\approx \mathcal{N}(0, \sigma_y^2) \\ \mu_t &= \mu_{t-1} + \epsilon_{t-1\mu,t} & \epsilon_{\mu,t} &\approx \mathcal{N}(0, \sigma_\mu^2)\end{aligned}$$

One can also assume that  $\mu_1 \in \mathcal{N}(0, \sigma_\mu^2)$ . Thus the latent mean at any future step has distribution  $\mu_t \in \mathcal{N}(0, t\sigma_\mu^2)$  so the variance grows with time. One can also use an autoregressive model in which  $\mu_t = \rho\mu_{t-1} + \epsilon_{\mu,t}$  where  $|\rho| < 1$ . In this case we have  $\mu_\infty \approx \mathcal{N}(0, \frac{\sigma_\mu^2}{1-\rho^2})$ : thus the uncertainty grows to finite asymptote instead of undoubtedly

# Structural time series models [5]

## Space representation of ARMA models

In this case we have

$$Z = \begin{pmatrix} Y_{t-p+1} \\ Y_{t-p+2} \\ \dots \\ Y_t \end{pmatrix} \quad (1)$$

And the observation equation is

$$Y_t = [0, 0, 0, \dots, 1]X_t \quad t = \pm 0, \pm 1, \dots \quad (2)$$

And the state equation is

$$X_{t+1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \dots & \phi_1 \end{pmatrix} X_t + \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 1 \end{pmatrix} \text{WN}(0, \sigma^2) \quad t = 0, \pm 1, \dots \quad (3)$$

# Bibliography I

- [1] <https://www.semanticscholar.org/paper/NON-STATIONARY-TIME-SERIES-AND-UNIT-ROOT-TESTING-Nielsen/2d0e62db75bdeafd2277fab1039f4866aef642b7>.
- [2] Peter J Brockwell e Richard A Davis. *Introduction to time series and forecasting*. Springer, 2002.
- [3] James D Hamilton. *Time series analysis*. Princeton university press, 2020.
- [4] Jianhua Hao e Fangai Liu. “Improving long-term multivariate time series forecasting with a seasonal-trend decomposition-based 2-dimensional temporal convolution dense network”. In: *Scientific Reports* 14.1 (2024), p. 1689.
- [5] Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL: <http://probml.github.io/book2>.