4. Model Parametrization

- model classification
- general model structure
- time series models
- state-space models
- uniqueness properties

Model Classification

- SISO/MIMO models
- linear/nonlinear models
- parametric/nonparametric models
- time-invariant/time-varying models
- time domain/frequency domain models
- lumped/distributed parameter models
- deterministic/stochastic models

General model structure

$$\mathcal{M}(\theta): \quad y(t) = G(q^{-1};\theta)u(t) + H(q^{-1};\theta)e(t)$$
$$\mathbf{E}\, e(t)e(s)^T = \Lambda(\theta)\delta_{t,s}$$

- $y(t)$ is ny-dimensional output
- $u(t)$ is nu-dimensional input
- $e(t)$ is an i.i.d. random variable with zero mean (white noise)
- q^{-1} is backward shift operator
- H, G, Λ are functions of the parameter vector θ
- this model is a genearal linear model in u and e

Feasible set of parameters

θ take the values such that

- H^{-1} and $H^{-1}G$ are asymptotically stable
- $G(0;\theta) = 0$ and $H(0;\theta) = I$
- $\Lambda(\theta) \succeq 0$

General SISO model structure

$$A(q^{-1})y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t), \quad \mathbf{E}[e(t)e(t)^T] = \lambda^2$$

where

$$\begin{aligned}
A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_p q^{-p} \\
B(q^{-1}) &= b_1 q^{-1} + b_2 q^{-2} + \cdots + b_n q^{-n} \\
C(q^{-1}) &= 1 + c_1 q^{-1} + \cdots + c_m q^{-m} \\
D(q^{-1}) &= 1 + d_1 q^{-1} + \cdots + d_s q^{-s} \\
F(q^{-1}) &= 1 + f_1 q^{-1} + \cdots + f_r q^{-r}
\end{aligned}$$

Special cases

output error structure

$$y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + e(t)$$

in this case $H(q^{-1};\theta) = 1$

the output error is the difference between the measurable output $y(t)$ and the model output $B(q^{-1})/F(q^{-1})u(t)$

if $A(q^{-1}) = 1$ in the general model structure

$$y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t)$$

- G and H have no common paramater
- possible to estimate G consistently even if choice of H is not appropriate

Time series models

stationary models

- ARMAX: AutoRegressive Moving Average model with Exogenous inputs
- ARMA: AutoRegressive Moving Average model
- ARX: AutoRegressive model with Exogenous inputs
- AR: AutoRegressive model
- MA: Moving Average model

non-stationary models

- ARIMA: AutoRegressive Integrated Moving Average model
- ARCH, GARCH (not discussed here)

ARMAX models

an autoregressive moving average model with an exogenous input:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t)$$

where

$$\begin{aligned} A(q^{-1}) &= I - (A_1 q^{-1} + \cdots + A_p q^{-p}) \\ B(q^{-1}) &= B_1 q^{-1} + B_2 q^{-2} + \cdots + B_m q^{-m} \\ C(q^{-1}) &= I + C_1 q^{-1} + \cdots + C_r q^{-r} \end{aligned}$$

and $e(t)$ is white noise with covariance Σ

the parameter vector is

$$\theta = (A_1, \ldots, A_p, B_1, \ldots, B_m, C_1, \ldots, C_r)$$

(the noise covariance could be a parameter to be estimated too)

Special cases of ARMAX models

- ARMA: $A(q^{-1})y(t) = C(q^{-1})e(t)$
- AR: $A(q^{-1})y(t) = e(t)$
- MA: $y(t) = C(q^{-1})e(t)$
- FIR: $y(t) = B(q^{-1})u(t) + e(t)$
- ARX: $A(q^{-1})y(t) = B(q^{-1})u(t) + e(t)$

applying the backward shift operator explicitly

$$\begin{aligned} y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) \\ + B_1 u(t-1) + \cdots + B_m u(t-m) \\ e(t) + C_1 e(t-1) + \cdots + C_r e(t-r) \end{aligned}$$

special cases:

- autoregressive moving average (ARMA) models

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + e(t) + C_1 e(t-1) + \cdots + C_r e(t-r)$$

- autoregressive (AR) models

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + e(t)$$

- moving average (MA) models

$$y(t) = e(t) + C_1 e(t-1) + \cdots + C_r e(t-r)$$

y consists of a finite sum of stationary white noise (e), so y is also stationary

- finite impulse response (FIR) models

$$y(t) = B_1 u(t-1) + \cdots + B_m u(t-m) + e(t)$$

- autoregressive with exogenous input (ARX) models

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + B_1 u(t-1) + \cdots + B_m u(t-m) + e(t)$$

Equivalent representation of AR(1)

write the first-order AR model recursively

$$
\begin{aligned}
y(t) &= Ay(t-1) + e(t) \\
&= A(Ay(t-2) + e(t-1)) + e(t) \\
&= A^2 y(t-2) + Ae(t-1) + e(t) \\
&= A^2(Ay(t-3) + e(t-2)) + Ae(t-1) + e(t) \\
&= A^3 y(t-3) + A^2 e(t-2) + Ae(t-1) + e(t) \\
&\;\;\vdots \\
&= \sum_{k=0}^{\infty} A^k e(t-k)
\end{aligned}
$$

- by assuming that i) t can be extended to negative index and ii) $|\rho(A)| < 1$
- y can be represented as *infinite moving average*

State-space form of AR models

define the state variable

$$x(t) = (y(t-1), y(t-2), \ldots, y(t-p))$$

the state-space form of AR model is

$$x(t+1) = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ I & 0 & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & I & 0 \end{bmatrix} x(t) + \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix} e(t)$$

- the characteristic polynomial of the dynamic matrix is

$$\det \tilde{A}(z) = \det(z^p - (A_1 z^{p-1} + A_2 z^{p-2} + \cdots + A_p))$$

- the AR process is stationary if its dynamic matrix $\mathcal{A}$ is stable

Non-uniqueness of MA models

consider examples of two MA models

$$\begin{aligned} y(t) &= e(t) + (1/5)e(t-1), \quad e(t) \sim \mathcal{N}(0, 25) \\ x(t) &= v(t) + 5v(t-1), \quad v(t) \sim \mathcal{N}(0, 1) \end{aligned}$$

that cannot be distinguished because of normality of the noise

- note that MA and AR processes are the inverse to each other (by swapping the role of y and e)

$$y(t) = -(1/5)y(t-1) + e(t), \quad x(t) = -5x(t-1) + v(t)$$

- an MA model is called **invertible** if it corresponds to a *causal* infinite AR representation – e.g., process with coefficient $1/5$

Properties of ARMA models

important properties of ARMA model:

$$A(q^{-1})y(t) = C(q^{-1})e(t)$$

- the process is **stationary** if the roots of the determinant of

$$A(z) = I - (A_1 z + A_2 z^2 + \cdots + A_p z^p)$$

  are outside the unit circle

- the process is said to be **causal** if it can be written as

$$y(t) = \sum_{k=0}^{\infty} \Psi(k) e(t-k), \quad \sum_{k=0}^{\infty} |\Psi(k)| \leq \infty$$

  (the process cannot depend on the future input)

- the process is **causal** if and only if the roots of the determinant of $A(z)$ lie outside the unit circle

- the process is **invertible** if the roots of the determinant of

$$C(z) = I + C_1 z + \cdots + C_r z^r$$

lie outside the unit circle

Non-stationary model

examples of non-stationarity and the use of differencing

- random walk: $x(t) = x(t-1) + w(t)$

$$z(t) \triangleq x(t) - x(t-1) = w(t)$$

  $z(t)$ is white noise which is stationary

- linear static trend: $x(t) = a + bt + w(t)$

$$z(t) \triangleq x(t) - x(t-1) = b + w(t) - w(t-1)$$

  $z(t)$ is a MA process

can we recover the original model from the fitted differenced series ?

Integrated model

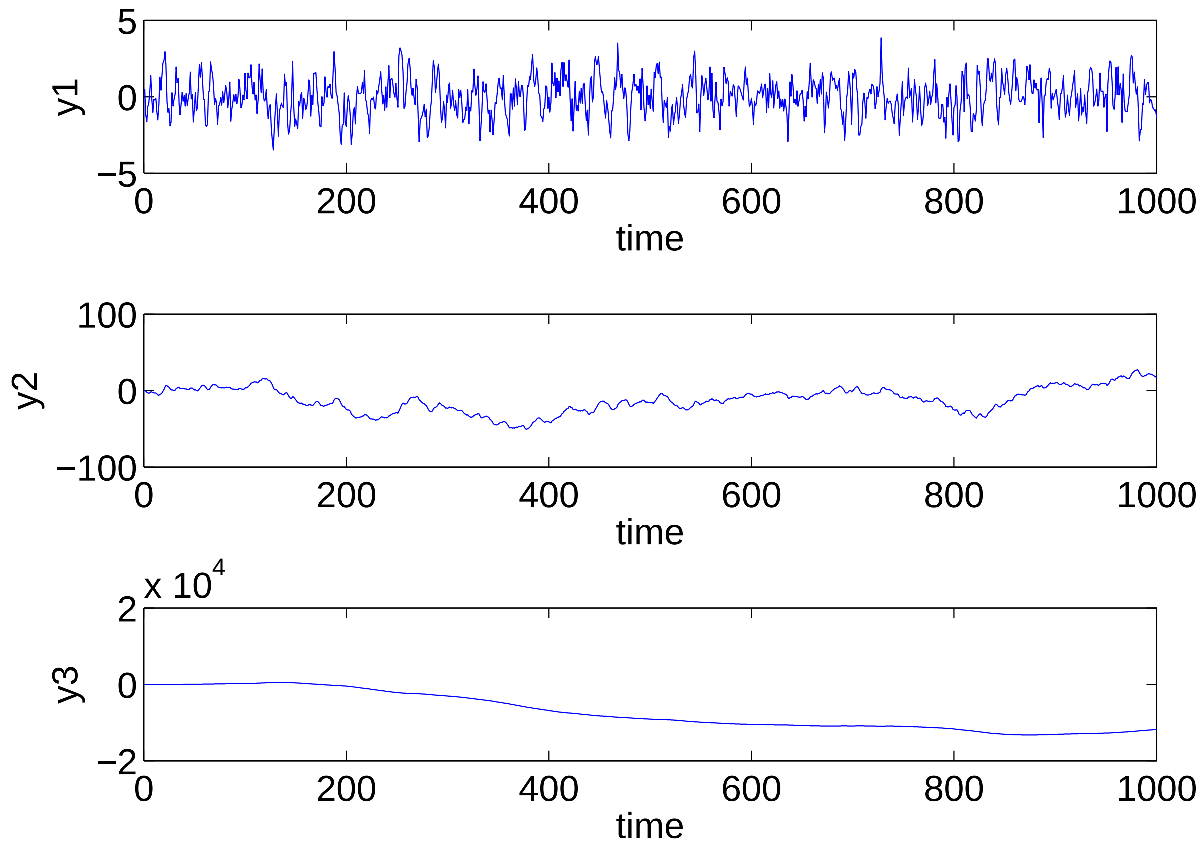
denote L a lag operator; a series $x(t)$ is **integrated** of order d if

$$(I - L)^d x(t)$$

is stationary (after dth difference)

- we use $I(d)$ to denote the integrated model of order d
- random walk is the first-order integrated model
- the lag of differencing is used to reduce a series with a trend
- for example, 12-lag of differencing removes additive seasonal effect

example: y_1 is a first-order AR process with coefficient 0.4 and is $I(0)$



- $y_2(t) = \sum_{k=0}^{t} y_1(k)$ (cumulative sum of y_1 is $I(1)$ – no exact reverting)
- $y_3(t) = \sum_{k=0}^{t} y_2(k)$ (cumulative sum of y_2 is $I(2)$ – momentum effect)

ARIMA models

$x(t)$ is an ARIMA process if the dth differences of $x(t)$ is an ARMA(p,r)

$$A(L)(I - L)^d x(t) = C(L)e(t)$$

and denoted by ARIMA(p, d, r)

examples of scalar ARIMA models

- $x(t) = x(t-1) + e(t) + ce(t-1)$ can be arranged as

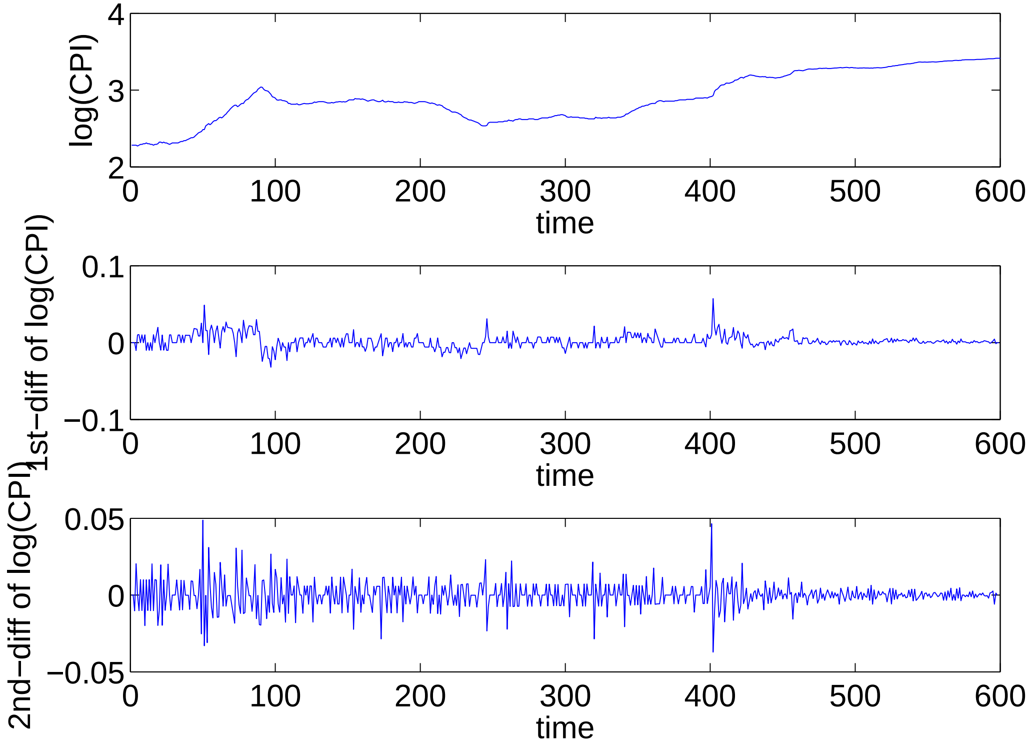
  $$(1 - L)x(t) = (1 + cL)e(t)$$

  which is ARIMA(0,1,1) or sometimes called *integrated moving average*
- $x(t) = ax(t-1) + x(t-1) - ax(t-2) + w(t)$ can be arranged as

  $$(1 - aL)(1 - L)x(t) = w(t)$$

  which is ARIMA(1,1,0)

example: $\log$ of CPI - consumer production index and its first, second differences



- $\log$ CPI shows the momentum type – characteristics of $I(2)$
- the first difference has no momentum but no mean-reverting
- the second difference seems to be mean-reverting and $I(0)$

State-space models

a linear stochastic model:

$$x(t+1) = A(\theta)x(t) + B(\theta)u(t) + \nu(t)$$
$$y(t) = C(\theta)x(t) + \eta(t)$$

$\nu(t)$ and $\eta(t)$ are white noise sequences with zero means and

$$\mathbf{E}\begin{bmatrix}\nu(t)\\ \eta(t)\end{bmatrix}\begin{bmatrix}\nu(s)\\ \eta(s)\end{bmatrix}^T = \begin{bmatrix}Q & S\\ S^T & R\end{bmatrix}\delta_{t,s}$$

- $\nu(t)$ is the *process noise*
- $\eta(t)$ is the *measurement noise*
- needs to transform to the so-called *innovation form* to compare with the standard model

Choosing a class of model structures

important factors:

- **Flexibility:** the model structure should describe most of the different system dynamics expected in the application
- **Parsimony:** the model should contain the smallest number of free parameters required to explain the data adequately
- **Algorithm complexity:** the form of model structure can considerably influence the computational cost
- **Properties of the criterion function:** for example, the asymptotic properties of prediction-error method depends crucially on the criterion function and the model structure

Uniqueness properties

question: can we describe a system *adequately* and *uniquely* ?

define $\mathcal{D}$ the set of θ for which

$$(\hat{G}, \hat{H}, \hat{\Lambda}) \text{ gives a } \textit{perfect description} \text{ of the true system}$$

three possibilities of this set can occur:

- the set $\mathcal{D}$ is empty or underparametrization
- the set $\mathcal{D}$ contains one point
- the set $\mathcal{D}$ consists of several points or overparametrization

Uniqueness properties for a scalar ARMA model

let the true ARMA model be given by

$$A(q^{-1})y(t) = C(q^{-1})e(t), \quad \mathbf{E}\, e(t)^2 = \lambda^2$$

$\mathcal{D}$ is the set of $\hat{A}, \hat{B}, \hat{C}, \hat{\lambda}$ for which

$$\frac{C(q^{-1})}{A(q^{-1})} = \frac{\hat{C}(q^{-1})}{\hat{A}(q^{-1})}, \quad \hat{\lambda^2} = \lambda^2$$

in order for these equalities to have a solution, we must have

$$\mathrm{deg}(\hat{A}) \geq \mathrm{deg}(A), \quad \mathrm{deg}(\hat{C}) \geq \mathrm{deg}(C)$$

or,

$$n^* \triangleq \min\left\{\mathrm{deg}(\hat{A}) - \mathrm{deg}(A), \mathrm{deg}(\hat{C}) - \mathrm{deg}(C)\right\} \geq 0$$

- A and C have no common factor
- $\frac{C(q^{-1})}{A(q^{-1})}$ and $\frac{\hat{C}(q^{-1})}{\hat{A}(q^{-1})}$ must have the same poles and zeros

these implies

$$\hat{A}(q^{-1}) = A(q^{-1})D(q^{-1}), \quad \hat{C}(q^{-1}) = C(q^{-1})D(q^{-1})$$

where $D(q^{-1})$ has arbitrary coefficients

$$\text{deg}(D) = \min\{\text{deg}(\hat{A}) - \text{deg}(A), \text{deg}(\hat{C}) - \text{deg}(C)\} = n^*$$

- $n^* > 0$: infinitely many solutions of $\hat{C}, \hat{A}, \hat{\lambda}$ (by varying D)
- $n^* = 0$: this gives $D(q^{-1}) = 1$, or at least one of $\hat{A}$ and $\hat{C}$ has the same degree as the true polynomial

Non-uniqueness of general state-space models

consider the multivariable model

$$\begin{aligned} x(t+1) &= A(\theta)x(t) + B(\theta)u(t) + \nu(t) \\ y(t) &= C(\theta)x(t) + \eta(t) \end{aligned}$$

where $\nu(t)$ and $\eta(t)$ are mutually independent white noise with zero means and covariance R_1, R_2 resp.

also consider a second model

$$\begin{aligned} z(t+1) &= \bar{A}(\theta)z(t) + \bar{B}(\theta)u(t) + \bar{\nu}(t) \\ y(t) &= \bar{C}(\theta)z(t) + \eta(t) \end{aligned}$$

where $\mathbf{E}\,\bar{\nu}(t)\bar{\nu}(s)^T = \bar{R}_1\delta_{t,s}$ and

$$\bar{A} = QAQ^{-1}, \quad \bar{B} = QB, \quad \bar{C} = CQ^{-1}, \quad \bar{R}_1 = QR_1Q^T$$

for some nonsingular matrix Q

the two models are equivalent:

- they have the same transfer function from u to y

$$G(q^{-1}) = \bar{C}(qI - A)^{-1}\bar{B} = CQ^{-1}(qI - QAQ^{-1})^{-1}QB = C(qI - A)^{-1}B$$

- the outputs y from the two models have the same second-order properties, $i.e.$, the spectral densities are the same

$$\begin{aligned} S_y(\omega) &= \bar{C}(e^{i\omega} - \bar{A})^{-1}\bar{R}_1(e^{i\omega} - \bar{A})^{-*}\bar{C}^* + R_2 \\ &= CQ^{-1}(e^{i\omega} - \bar{A})^{-1}QR_1Q^*(e^{i\omega} - \bar{A})^{-*}Q^{-*}C^* + R_2 \\ &= C[Q^{-1}(e^{i\omega} - \bar{A})Q]^{-1}R_1[Q^*(e^{i\omega} - \bar{A})^*Q^{-*}]^{-1}C^* + R_2 \\ &= C(e^{i\omega} - A)^{-1}R_1(e^{i\omega} - A)^{-*}C^* + R_2 \end{aligned}$$

the model is not unique since Q can be chosen arbitrarily

References

Chapter 6 in
T. Söderström and P. Stoica, *System Identification*, Prentice Hall, 1989

P.S.P. Cowpertwait and A.V. Metcalfe, *Introductory Time Series with R*, Springer, 2009

R.H. Shumway and D.S. Stoffer, *Time Series Analysis and Its Applications: with R Examples*, 3rd edition, Springer, 2011

Chapter 12-13 in
D. Ruppert and D.S. Matteson, *Statistics and Data Analysis for Financial Engineering*, 2nd edition, Springer, 2015