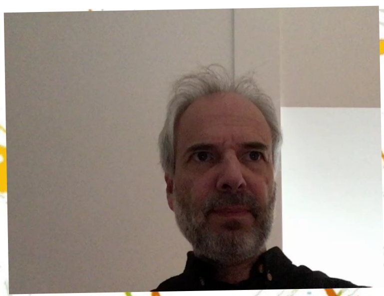


TIME SERIES ANALYSIS & RECURRENT NEURAL NETWORKS #3

- Statistical inference in ARMA models
- Granger causality

Main lecture: Daniel Durstewitz

Exercises: Leonard Bereska, Manuel Brenner, Daniel Kramer, Georgia Koppe

Heidelberg University

Parameter estim. in AR(p)

scalar TS $\{x_t\}$, $t = 1 \dots T$

$$x_t = a_0 + \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t \,, \quad \varepsilon_t \sim W(0, \sigma^2)$$

Solve for param. by LSE

$$LSE(\{a_{0:p}\}) = \sum_{t=p+1}^{T} (\underbrace{x_t}_{\text{truly observed!}} - \underbrace{\hat{x}_t}_{\text{prediction}})^2$$

$$\hat{x}_t = E\left[a_0 + \sum a_i x_{t-i} + \varepsilon_t\right]$$
$$= a_0 + \sum_{i=1}^{p} a_i x_{t-i}$$

$$\underbrace{\begin{pmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_T \end{pmatrix}}_{\substack{(T-p)\times 1 \\ \underline{x}_T}} = \underbrace{\begin{pmatrix} 1 & x_p & \cdots & x_1 \\ 1 & x_{p+1} & \cdots & x_2 \\ \vdots & \vdots & & \vdots \\ 1 & x_{T-1} & \cdots & x_{T-p} \end{pmatrix}}_{\substack{(T-p)\times(p+1) \\ \underline{X}_p}} \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}}_{\substack{(p+1)\times 1 \\ \underline{a}_p}} + \underbrace{\begin{pmatrix} \varepsilon_{p+1} \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{\substack{(T-p)\times 1 \\ \underline{\varepsilon}}}$$

$$\Rightarrow \underline{x}_T = \underline{X}_p \underline{a}_p + \underline{\varepsilon}$$

Petersen & Pedersen (2012) matrix cookbook

$$LSE(\{a_{a_p}\}) = \sum (x_t - \hat{x}_t)^2$$
$$= (\underline{x}_T - \underline{X}_p \underline{a}_p)^T (\underline{x}_T - \underline{X}_p \underline{a}_p)$$

$$\frac{\partial LSE(\underline{a}_p)}{\partial \underline{a}_p} = \cancel{-2} \underline{X}_p^T (\underline{x}_T - \underline{X}_p \underline{a}_p) \stackrel{!}{=} 0$$

$\underline{a}_p^T \underline{X}_p^T \underline{X}_p \underline{a}_p$

$(p+1)\times(p+1)$

$$\Rightarrow \underline{X}_p^T \underline{x}_T = (\underline{X}_p^T \underline{X}_p) \underline{a}_p$$
$$\Rightarrow \hat{\underline{a}}_p = (\underline{X}_p^T \underline{X}_p)^{-1} \underline{X}_p^T \underline{x}_T$$

- vector-valued TS

$\{\underline{x}_t\}$, $t = 1 \dots T$

$\underline{x}_t = (x_{1t} \dots x_{Kt})^T$

K: # of simult. obs. TS

- fMRI

- spatial struc., water temp. across diff. loc.

- shares from diff. companies

→ Vector AR (VAR) model

$$\underline{x}_t = \underbrace{\underline{a}_0}_{K\times 1 \text{ vector}} + \sum_{i=1}^{p} \underbrace{A_i}_{K\times K \text{ matrix}} \underline{x}_{t-i} + \underbrace{\underline{\varepsilon}_t}_{K\times 1} \,, \quad \underline{\varepsilon}_t \sim W(0, \Sigma)$$

$$\underline{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Kt} \end{pmatrix}$$

$$A_i = \begin{pmatrix} a_{11}^{(i)} & \cdots & a_{1K}^{(i)} \\ \vdots & \ddots & \vdots \\ a_{K1}^{(i)} & \cdots & a_{KK}^{(i)} \end{pmatrix}$$

capture cross-corr.

auto-corr.

$E[\varepsilon_{it}\varepsilon_{jt'}] = 0$ for $t \neq t'$ $\forall i,j$

$E[\varepsilon_{it}\varepsilon_{jt}] = \sigma_{ij}^2$

Theorem: - any AR(p) can be reexpressed as a p-variate VAR(1)

- any VAR(p) can be reexpressed as a Kp-variate VAR(1)

proof by construc.

for instance AR(2): $x_t = a_0 + \sum_{i=1}^{2} a_i x_{t-i} + \varepsilon_t$

$$\underline{x}_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix}.$$

Conditions for stat. in a VAR(1)

- Note that A is square

  $AR(p) \to p \times p$

  $VAR(1) \to K \times K$ $\qquad \underline{x}_t = \underline{a}_0 + A \underline{x}_{t-1} + \underline{\varepsilon}_t$

  $VAR(p) \to pK \times pK$

Theorem: condition for stat. of VAR(1) is $\boxed{\max |eig(A)| < 1}$

"sketch of proof":

- ignore $\underline{a}_0$
- assume A is invertible

$$A = V \Lambda V^{-1}$$

matrix of eig. vec. (V) ; diag. matrix of eig. val. (Λ)

$$E[\underline{\varepsilon}_t] = 0$$

$$\underline{x}_T = A \underline{x}_{T-1} = (V \Lambda V^{-1}) \underline{x}_{T-1}$$

$$= (V \Lambda \cancel{V^{-1}})(\cancel{V} \Lambda V^{-1}) \underline{x}_{T-2}$$

$$= \ldots = (V \Lambda \cdot \Lambda \cdot \Lambda \cdots V^{-1}) \underline{x}_1$$

$$= V \Lambda^{T-1} V^{-1} \underline{x}_1 \quad , \underline{x}_1 \neq 0$$

→ if $\max |\lambda_i| > 1 \to$ diverge

$\max |\lambda_i| < 1 \to$ converge

$\max |\lambda_i| = 1 \to$ rand. walk "integrator"

Maximum Likelihood inference of param.

- consistent estimators
- efficient "

assume: AR(p)

Gaussian WN!

$$x_t = a_0 + \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

$$\underbrace{a_0 + \sum_{i=1}^{p} a_i x_{t-i}}_{\text{mean } E[x_t | x_{t-p} \dots x_{t-1}]} \qquad E[\varepsilon_t \varepsilon_{t'}] = 0 \text{ for } t \neq t'$$

"likelihood func." $p(X|\theta)$

$$\underbrace{L_X(\{a_i\}, \sigma^2)}_{\theta} = \overbrace{f(\{x_{p+1} \dots x_T\} | \{a_i\}, \cancel{p}, \sigma^2, \{x_1 \dots x_p\})}^{f_\theta(\{x_{p+1} \dots x_T\})}$$

$$= \prod_{t=p+1}^{T} f_\theta(x_t | x_1 \dots x_{t-1})$$ Bayes' rule recursively appl.

$p(A,B,C) = p(A|B,C)\,p(B|C)\,p(C)$

$$= \prod_{t=p+1}^{T} f_\theta(x_t | x_{t-p} \dots x_{t-1})$$ by model assumpt.

→ p^{th}-order Markov process

$$x_t | x_{t-p} \dots x_{t-1} \sim N\left[a_0 + \sum_{i=1}^{p} a_i x_{t-i}, \sigma^2\right]$$

$$\hat{x}_t = E[x_t | x_{t-p} \dots x_{t-1}]$$

$$\varepsilon_t = x_t - \hat{x}_t$$

$$\begin{pmatrix} \varepsilon_{p+1} \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

$$L(\theta) = \prod_{t=p+1}^{T} (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2}(x_t - \hat{x}_t)^2/\sigma^2}$$

$(T-p) \times (T-p)$ identity

$$= (2\pi)^{-(T-p)/2} |\sigma^2 I|^{-1/2} e^{-\frac{1}{2}\varepsilon^T \varepsilon / \sigma^2}$$

max. log L w.r.t. θ

$$\log L(\theta) = \ell(\theta) = \cancel{-\frac{T-p}{2}\log(2\pi)} \cancel{-\frac{T-p}{2}\log(\sigma^2)} - \frac{1}{2}\varepsilon^T\varepsilon\,\sigma^{-2}$$

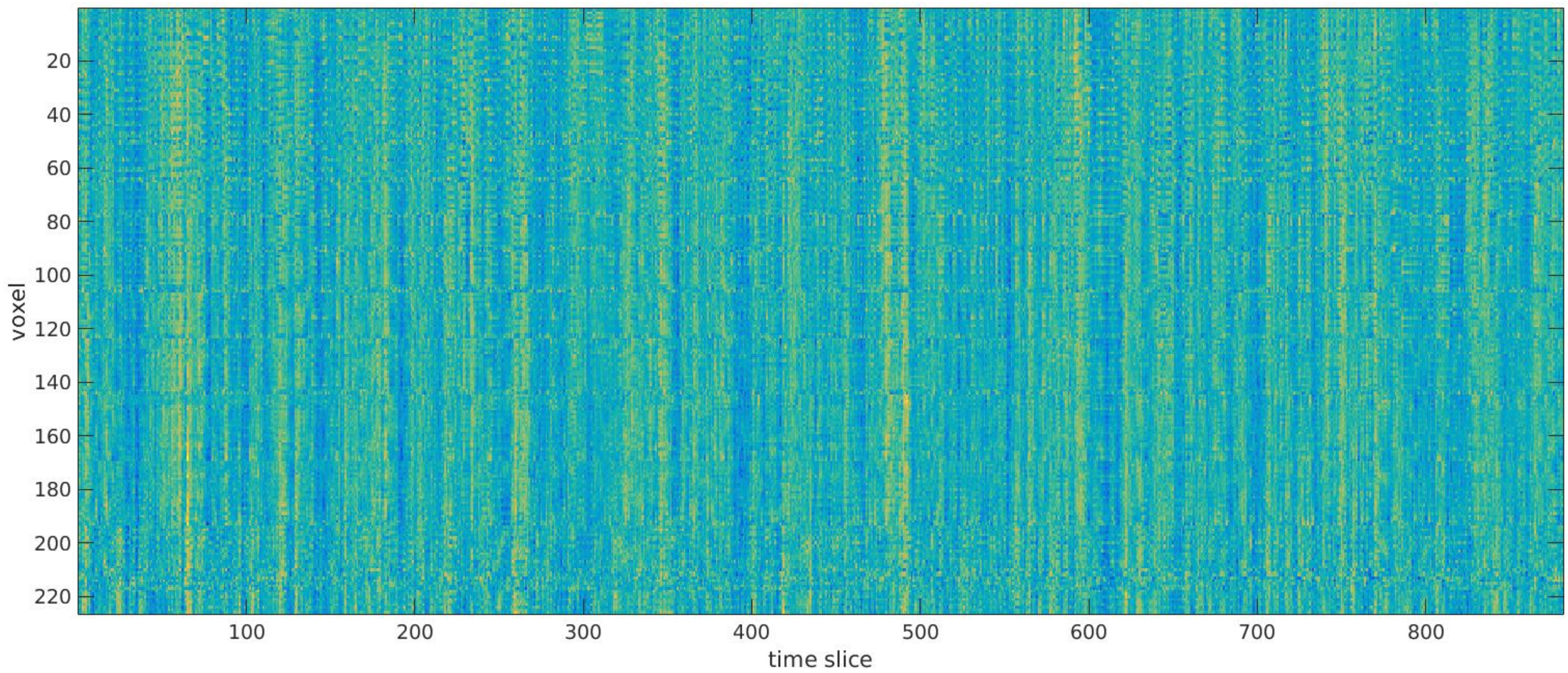
$$\Rightarrow \frac{\partial \ell(\theta)}{\partial \sigma} = -\frac{T-p}{\cancel{2}}\cancel{2}\left(\frac{\sigma}{\sigma^2}\right) - \frac{1}{\cancel{2}}\varepsilon^T\varepsilon(-\cancel{2}\sigma^{-3}) \stackrel{!}{=} 0 \quad | \cdot \sigma$$

$$\Rightarrow -(T-p) + \varepsilon^T\varepsilon\,\sigma^{-2} \stackrel{!}{=} 0$$

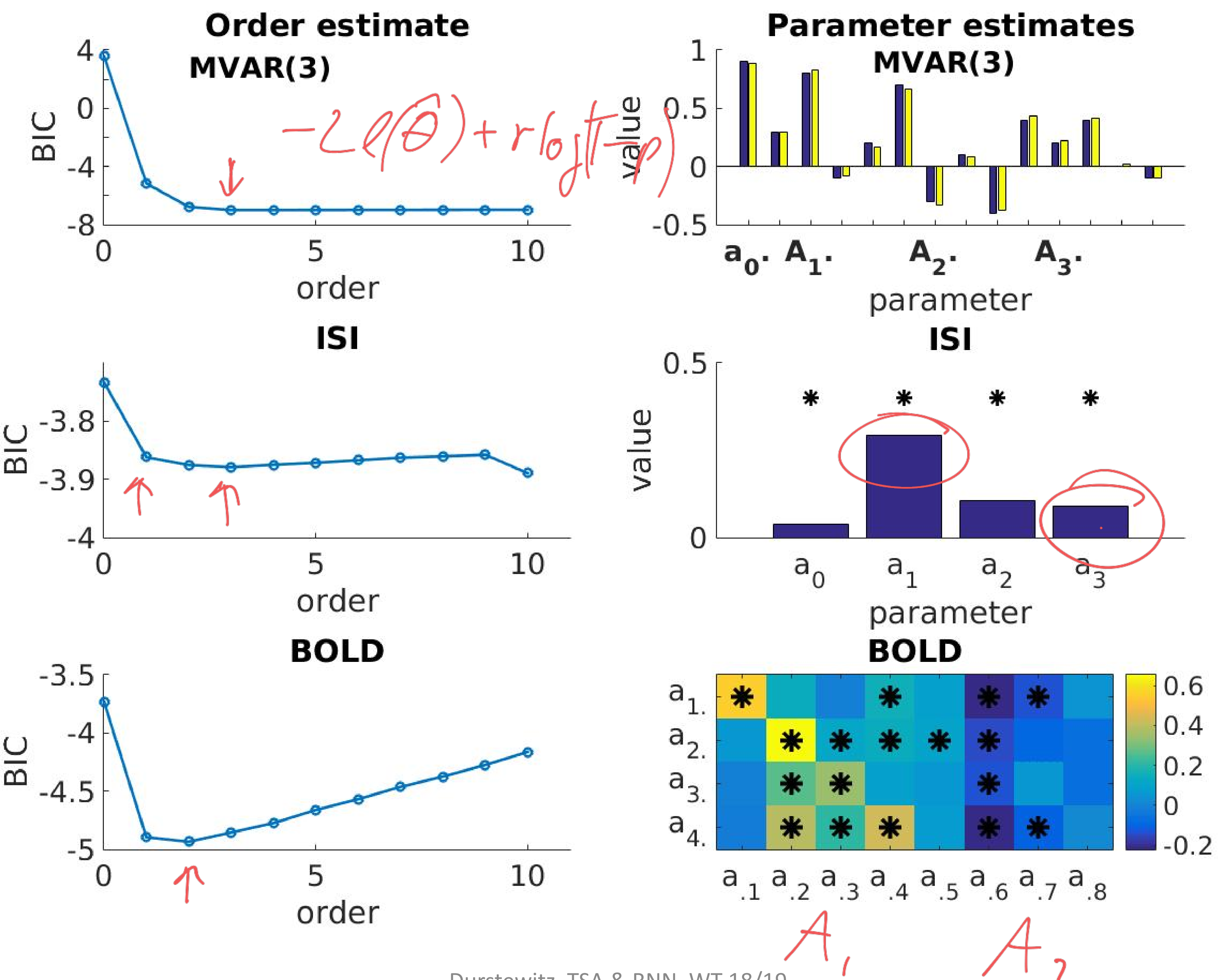
$$\Rightarrow \varepsilon^T\varepsilon\,\sigma^{-2} = T-p$$

$$\Rightarrow \hat{\sigma}^2 = \varepsilon^T\varepsilon/(T-p) = \frac{1}{T-p}\sum_{t=p+1}^{T}(x_t - \hat{x}_t)^2$$

$$\frac{\partial \ell(\theta)}{\partial \{a_i\}} = -\frac{1}{2}\sum_{t=p+1}^{T}(x_t - \hat{x}_t)^2/\cancel{\sigma^2} \stackrel{!}{=} 0$$



Source: data from Bähner et al. (2015) Neuropsychopharmacology



Source: Durstewitz (2017) *Advanced Data Analysis in Neuroscience*. Springer.